Guido Visconti

# Fundamentals of Physics and Chemistry of the Atmosphere

*Second Edition*

EXTRAS ONLINE

Springer

# Fundamentals of Physics and Chemistry of the Atmosphere

Guido Visconti

# Fundamentals of Physics and Chemistry of the Atmosphere

Second Edition

EXTRAS ONLINE

Springer

Guido Visconti
Dipartimento di Scienze Fisiche e Chimiche
Università dell'Aquila
Coppito, L'Aquila, Italy

ISBN 978-3-319-29447-6 ISBN 978-3-319-29449-0 (eBook)
DOI 10.1007/978-3-319-29449-0

Library of Congress Control Number: 2016932892



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

*This book is dedicated to my family and to the late sisters, Anna and Rita, and brother, Sante.*

*No weather will be found in this book*

Mark Twain, *The American Claimant*

# Preface

This is the second edition of a book published about 15 years ago. Someone says that is better to write a new book rather than work on a second edition especially after such long time. Part of the problem is that the contract for a second edition was signed just before my hometown was hit by a 6.3 earthquake in 2009. In any case I think an honorable compromise was reached considering that the book is largely rewritten.

The first five chapters are an introduction to the general topics of atmospheric physics, and they deal with thermodynamics, radiation, dynamics with applications, and chemistry. Then the sixth chapter introduces to remote sensing. However, each one of these chapters contains one of the main novelties of this book, and that is the so-called examples. These either show some applications of the matter introduced in the chapter or represent a much more detailed explanation of the same topic. Sometimes the examples contain simple programs (MATLAB or FORTRAN) to solve problems.

The chapter on the origin and evolution of the atmosphere has been canceled because this theme has advanced so much (especially in connection with the exoplanets research) that it would require a textbook of its own.

Starting with Chap. 7 the book looks very similar to the previous edition but contains much more material. This chapter has a quite detailed treatment of the vorticity and its properties. Chapter 8 gives more details on the oceanic boundary layer and some introduction to the classical concepts of turbulence. Chapter 9 contains a complete new paragraph on clouds in planetary atmospheres. Atmospheric waves are treated in Chap. 10, and the examples contain a rather complete exposition about mountain waves including simple programs. Chapter 11 is very similar to the previous book with some additional information about the wave contribution to the general circulation. Chapter 12 is about theories on general circulation. Here we have rewritten the section on the Hadley circulation with an explicit calculation based on the work of Sobel and Schneider. Also in the examples the same problem is solved in the shallow water approximation.

Chapter 13 gives more detailed information about radiative transfer calculations that are necessary for the introduction to simple climate models. These are treated

in Chap. 14 and 15. In particular Chap. 14 ends with examples that introduce the entropy approach to energy balance climate models, and as a preparatory step the same method is used to calculate the temperature profile of an elementary atmosphere. In Chap. 15 the section on the performance of GCM has been very much expanded. It includes the most recent development about metrics and the Bayesian point of view. This requires an elementary introduction to Bayesian statistics. Statistics also enter in some of the examples about the evaluation of the effect of uncertainty in model parameters.

The chemistry of the troposphere is the topic of Chap. 16. This is another chapter largely rewritten and expanded. The simple models for tropospheric ozone have been rewritten and used to evaluate the gas isopleth in an urban atmosphere. Chapter 17 about circulation of the middle atmosphere has not changed very much except for some additional examples on equatorial waves and the Holton model on quasi biennial oscillation. The same is true for the following chapter about stratospheric ozone chemistry. In this case the examples are about the calculation of loss rate of polar ozone and the explicit calculation of the effects of the catalytic cycle, a solved exercise proposed in the book by Andrews.

Another major difference with the first edition is Chap. 19 that has been extended to deal not only with chaos but also with nonlinear phenomena. As examples of nonlinear phenomena, the Stommel model for the thermohaline circulation is discussed in connection with climate theory, and there is an extensive treatment of the difference equations made so popular by Edward Lorenz. Then the delayed differential equations are discussed in connection with the ENSO and the aerosol-cloud problem seen as predator–prey problem as developed by Koren and Feingold. The interesting parts of this chapter are the examples with programs that solve most of the topics described in the chapter.

Finally a new chapter was added on the controversial theme of geoengineering. This is a huge field now, and we just discussed a few options on carbon capture and sequestration using what we had learned in Chap. 16 about the carbon cycle. Then we had an excuse to reintroduce the energy balance model as described by Kleidon and Renner. Also some additional requirements are treated in the examples concerning the radiative effects and the role of the aerosol in the cloud albedo (Twomey effect).

All these took a considerable effort (all the figures were drawn or redrawn by the author), and I asked many times during these years if the game was worth the candle especially when there is so much material around. After 14 years the field of atmospheric physics and chemistry has expanded impressively, and there are many excellent books published not considering the amount of material available on the Internet. Our intent is to take the reader through an approach that deepens some of the most unknown aspect of the field that many times are buried in “historical” forgotten paper that remains very instructive. A classical example is the problem of the breezes that most of the books limit to a couple of pages but contain many useful insights. Similar examples could be found in the tropospheric chemistry and nonlinear problems. Our ideas were to use very simple arguments wherever possible that could be translated in simple and comprehensible calculations.

Other important reasons for a second edition are errors. Reading back the first edition, I found many mistakes, some of them the results of simple distraction but others had some deep and wrong roots. Most of them have been corrected, but nothing guarantees that we have not introduced new ones. On the other hand the gigantic amount of material referred before contains often errors that went beyond even the referees (this is the perverse side of the peer review).

As before the contribution of students and friends has been fundamental. Frank Marzano (who should have been a coauthor) suggested many features of this edition. Former students, now in the professional lineage, have very much contributed, in particular Gabriele Curci and Paolo Ruggieri. I have to give them all the credits they deserve to help out especially with the MATLAB scripts.

The support of my family must be acknowledged considering that we all went through the stress of a destructive earthquake from which neither the city nor the university have recovered yet. Their encouragement has been constant, and this is one of the reasons this work is dedicated to them.

L'Aquila, Italy Guido Visconti

# Contents


**1 Fundamentals: Thermodynamics of the Atmosphere** ..... 1
1.1 Simple Laws ..... 1
1.1.1 The Scale Height ..... 4
1.1.2 The Potential Temperature ..... 5
1.1.3 Static Stability ..... 7
1.2 The Thermodynamics of Water Vapor ..... 8
1.2.1 The Equation of Clausius–Clapeyron ..... 8
1.2.2 About Eutectics ..... 10
1.3 Some Effects of Water Vapor ..... 13
1.3.1 The Tephigram or Thermodynamic Diagram ..... 16
1.3.2 The Skew T–Log P Diagram (Emagram) ..... 18
1.3.3 The Conditional Convective Instability ..... 20
1.4 The Distribution of Water Vapor in the Atmosphere ..... 22
E.1 Examples ..... 24
E.1.1 Was the Atmosphere Drier During the Ice Age? ..... 24
E.1.2 More on the Clausius–Clapeyron (C–C) Equation ..... 25
E.1.3 The Equivalent Potential Temperature ..... 26
E.1.4 The Saturated Adiabat ..... 27
E.1.5 Constructing an Emagram ..... 29
E.1.6 The Equal-Area Requirement ..... 31
E.1.7 The Virtual Temperature ..... 32
E.1.8 Using Diagrams in Forecasting ..... 33
References ..... 35

**2 Fundamentals: Radiation in the Atmosphere** ..... 37
2.1 The Definition of Radiometric Variables ..... 37
2.2 The Solar Radiation ..... 39
2.3 Scattering and Absorption of Solar Radiation ..... 43
2.3.1 Rayleigh Scattering ..... 44
2.3.2 The Absorption of Solar Radiation ..... 48

2.4 Infrared Radiation ..... 52
2.4.1 The Equation of Radiative Transfer ..... 53
2.4.2 The Radiative–Convective Atmosphere ..... 57
2.4.3 The Runaway Greenhouse ..... 59
E.2 Examples ..... 62
E.2.1 Rayleigh Scattering from Natural Light (Sunlight) ..... 62
E.2.2 A Simple Way to Evaluate Ozone Absorption ..... 64
E.2.3 The Radiative Time Constant ..... 66
E.2.4 A Simple Model for the Greenhouse Effect ..... 67
References ..... 69

**3 The First Laws of Motion** ..... 71
3.1 Scales and Orders of Magnitude ..... 72
3.2 The Basic Equations ..... 73
3.2.1 The Total Derivative ..... 74
3.2.2 The Continuity Equation ..... 75
3.2.3 Pressure Forces ..... 77
3.2.4 Friction Forces ..... 78
3.2.5 The Equations of Motion in an Inertial System ..... 80
3.3 Vorticity and Circulation ..... 83
3.3.1 Some Properties of Vorticity and Circulation ..... 85
3.3.2 The Vorticity Equation ..... 88
E.3 Examples ..... 94
E.3.1 The Coriolis Acceleration ..... 94
E.3.2 The Inertial Oscillation ..... 96
E.3.3 The Rossby Adjustment Problem (Nonrotating) ..... 98
E.3.4 The Rossby Adjustment Problem (Rotating Case) ..... 99
E.3.5 Energetics of the Adjustment ..... 101
References ..... 103

**4 Dynamics: Few Simple Applications** ..... 105
4.1 The Geostrophic Motion ..... 105
4.1.1 The Geostrophic Streamfunction ..... 109
4.1.2 The Quasi-geostrophy: The Isallobaric Wind ..... 111
4.2 The Thermal Wind ..... 112
4.2.1 Thermal Wind in the Atmosphere ..... 116
4.3 More About Geostrophic Wind ..... 118
4.3.1 Margules Formula ..... 118
4.3.2 Inertial Instability ..... 120
4.4 The Natural Coordinate System ..... 122
4.5 Some Application of Circulation and Vorticity ..... 126
4.5.1 The Sea Breeze ..... 126
4.5.2 Some Other Local Winds ..... 127
4.5.3 The Rossby Waves ..... 130

E.4 Examples ..... 135
E.4.1 The Sea Breeze Circulation ..... 135
E.4.2 The Circulation Around Lows and Highs ..... 136
E.4.3 Effects on the Propagation of Long Waves ..... 138
References ..... 140

**5 Atmospheric Chemistry** ..... 141
5.1 Characteristics of the Atmospheres ..... 141
5.2 Atmospheric Composition and Chemistry ..... 146
5.3 Chemical Kinetics ..... 148
5.4 Chemistry and Transport ..... 151
E.5 Examples ..... 154
E.5.1 Units for Chemical Abundance ..... 154
E.5.2 The Chapman Model for Atmospheric Ozone ..... 155
E.5.3 Calculation of Photolysis Rate ..... 157
E.5.4 Photodissociation and Vertical Transport ..... 158
E.5.5 A Time-Dependent Case ..... 159
References ..... 160

**6 Introduction to Remote Sensing** ..... 161
6.1 Observations of the Atmosphere ..... 161
6.2 Thermal Emission Measurements ..... 163
6.3 Ozone Measurements from Satellite ..... 164
6.4 Atmospheric Properties from Radio Occultation (RO) ..... 168
6.5 A Few Things About Radar ..... 171
6.6 Lidar Measurements ..... 175
E.6 Examples ..... 177
E.6.1 Refractive Index of Air ..... 177
E.6.2 The Abel Transform ..... 178
References ..... 180

**7 The Atmospheric Motions** ..... 181
7.1 The Thermodynamic Equation ..... 181
7.2 The Isentropic Coordinate System ..... 184
7.2.1 The Vorticity Equation in Isentropic Coordinates ..... 186
7.3 The Ertel Potential Vorticity ..... 188
7.3.1 The Application of the Potential Vorticity ..... 190
7.3.2 Ozone and Vorticity ..... 192
7.3.3 More on Rossby Waves ..... 194
7.4 The Non-stationary Solutions ..... 196
7.4.1 Numerical Solutions of a Flow Above an Obstacle: The Stationary Case ..... 198
7.4.2 Numerical Solutions of a Flow Above an Obstacle: The Non-stationary Case ..... 199

7.5 Quasi-Geostrophic Vorticity ..... 202
7.5.1 The Equation of Quasi-Geostrophic Potential Vorticity .. 204
7.6 Potential Vorticity Inversion ..... 206
7.6.1 A Periodic Potential Vorticity Anomaly ..... 208
7.6.2 Rossby Waves and Vorticity Inversion..... 209
7.7 Scaling of the Shallow Water Equations ..... 211
7.7.1 Scaling of the Equations of Motion..... 211
7.7.2 Scaling of the Vorticity and Divergence Equations ..... 213
E.7 Examples..... 216
E.7.1 Ertel Potential Vorticity in a Barotropic Fluid..... 216
E.7.2 Conservation of Potential Vorticity ..... 216
E.7.3 Scaling and Vorticity Inversion ..... 218
E.7.4 Rossby Waves in Shallow Water..... 219
E.7.5 Flow Over an Obstacle: The Numerical Solution ..... 222
References..... 223

**8 The Planetary Boundary Layer**..... 225
8.1 Turbulence and Diffusion ..... 225
8.2 Turbulent Friction ..... 229
8.2.1 The Mixing Length ..... 231
8.3 The Surface Layer..... 233
8.4 The Ekman Layer ..... 237
8.5 The Secondary Circulation..... 242
8.5.1 Spin-Down in a Teacup..... 244
8.6 Turbulent Diffusion from Discrete Sources ..... 246
8.6.1 The Characteristics of Smoke Plumes ..... 247
8.6.2 The Gaussian Plume..... 249
E.8 Examples..... 252
E.8.1 Boundary Layer in the Ocean ..... 252
E.8.2 The Transfer of Sensible and Latent Heat ..... 252
E.8.3 The Fluxes in the Presence of Vegetation ..... 255
E.8.4 The Kolmogorov Spectrum ..... 258
References..... 260

**9 Aerosols and Clouds**..... 261
9.1 Sources of Atmospheric Aerosols ..... 261
9.2 The Size Distribution of Atmospheric Aerosols ..... 263
9.3 Nucleation and Growth ..... 266
9.3.1 Nucleation from Water Vapor Condensation ..... 267
9.3.2 The Growth by Condensation ..... 270
9.3.3 Droplet Growth by Collision and Coalescence..... 271
9.3.4 The Statistical Growth..... 275
9.4 Formation and Growth of Ice Crystals ..... 276
9.5 Stratospheric Aerosols ..... 281
9.5.1 The Sulfate Aerosol Layer ..... 282
9.5.2 Polar Stratospheric Clouds ..... 284

9.6 Clouds in Planetary Atmospheres ..... 286
E.9 Examples ..... 290
E.9.1 The Lognormal Size Distribution ..... 290
E.9.2 A Few Things More About the Köhler Curve ..... 292
E.9.3 Sedimentation of Particles ..... 293
References ..... 294

**10 Waves in the Atmosphere** ..... 297
10.1 Some Properties of the Waves ..... 297
10.2 Gravity Waves in Shallow Water ..... 300
10.3 Orographic Waves ..... 302
10.4 Internal Gravity Waves ..... 304
10.5 Three-Dimensional Rossby Waves ..... 307
10.6 The Physics of Gravity Waves ..... 311
10.6.1 The Equation of Quasi-geostrophic Potential Vorticity ... 311
10.6.2 The Eliassen–Palm Flux ..... 312
10.6.3 Energetics of Gravity Waves ..... 315
10.7 Breaking, Saturation, and Turbulence in the Upper Atmosphere .. 317
E.10 Examples ..... 322
E.10.1 Is the Phase Velocity a Vector? ..... 322
E.10.2 The Quasi-geostrophic Potential Vorticity in Log P Coordinates ..... 325
E.10.2 The Eliassen–Palm Flux Terms ..... 326
E.10.3 Energy and EP Flux ..... 326
E.10.4 The WKB Approximation ..... 328
E.10.5 The Numerical Solution to the Wave Equation ..... 329
E.10.6 A Few More Things About Mountain Waves ..... 330
E.10.7 Waves Forced by Sinusoidal Ridges ..... 331
References ..... 336

**11 The Data on the Atmospheric Circulation** ..... 339
11.1 The General Features ..... 339
11.2 The Energy Budget of the Atmosphere ..... 342
11.2.1 Forms of Energy ..... 346
11.2.2 Decomposition of Transport ..... 348
11.2.3 The Details of the Energy Budget ..... 350
11.3 The Mean Zonal Circulation ..... 352
E.11 Examples ..... 357
E.11.1 Waves and Momentum Flux ..... 357
E.11.2 Waves and Vorticity Flux ..... 359
E.11.3 More on Pseudomomentum ..... 361
References ..... 362

**12 Theories on the General Circulation of the Atmosphere** ..... 365
12.1 The Equatorial Circulation ..... 365
12.1.1 Gill's Symmetric Circulation ..... 366

12.1.2 The Nonlinear Symmetric Circulation ..... 370
12.1.3 The Vorticity Equation and Viscosity ..... 376
12.2 The Middle Latitude Circulation ..... 378
12.2.1 The Baroclinic Instability: Qualitative Treatment ..... 380
12.2.2 The Baroclinic Instability: The Eady Problem ..... 383
12.2.3 The Baroclinic Instability: The Charney Problem ..... 387
12.2.4 The Baroclinic Instability: Two-Level Model ..... 389
12.3 Energetics of the Baroclinic Waves ..... 395
12.3.1 Energy in the Two-Level Model ..... 398
12.3.2 The Parameterization of Transport ..... 400
12.4 The General Circulation: A Reductionist Approach ..... 403
12.4.1 The Inertial Instability ..... 405
12.4.2 A Comparison Among the Planets ..... 406
E.12 Examples ..... 408
E.12.1 The Thermodynamic Equation ..... 408
E.12.2 The Hadley Circulation as a Shallow Water Case ..... 409
E.12.3 The Hadley Circulation: Numerical Solution ..... 410
E.12.4 The Hadley Circulation on Slow-Rotating Planet? ..... 411
E.12.4 Transport by Eddies ..... 413
References ..... 415

**13 Radiation for Different Uses** ..... 417
13.1 Parameterization of Gaseous Absorption ..... 417
13.1.1 The Ozone Absorption ..... 419
13.1.2 The Water Vapor Absorption ..... 421
13.2 The Interaction of Solar Radiation with Particulates in the Atmosphere ..... 424
13.2.1 Optical Properties of the Particles ..... 425
13.2.2 Phase Functions and Mie Scattering ..... 432
13.3 Radiative Transfer in the Presence of Scattering ..... 433
13.3.1 Few Simple Applications of the δ-Eddington Approximation ..... 438
13.4 The Transfer of Infrared Radiation ..... 441
13.4.1 The Formal Solution ..... 441
13.5 Molecular Spectra ..... 443
13.5.1 Spectral Line Shape ..... 444
13.6 Models for the Line Absorption ..... 446
13.6.1 A Formulation of the Infrared Flux ..... 448
13.6.2 The Band Absorptivities According to Cess and Ramanathan ..... 450
13.7 δ-Eddington in the Infrared ..... 451
E.13 Examples ..... 452
E.13.1 Color for Nonabsorbing Spheres ..... 452
E.13.2 A Simple Model for Scattering ..... 452

E.13.3 Reflectivity and Transmission from Nonconservative Scattering ..... 455
E.13.4 A MATLAB Program for the Delta-Eddington ..... 456
E.13.5 Infrared Flux from Methane ..... 458
References ..... 459

**14 Simple Climate Models** ..... 461
14.1 Energy Budget ..... 461
14.2 Zero-Dimensional Models and Feedback ..... 462
14.3 One-Dimensional Energy Balance Climate Models ..... 468
14.3.1 North's Model ..... 469
14.3.2 The Stability of the One-Dimensional Model ..... 473
14.3.3 The Sellers Model ..... 476
14.3.4 The Time Dependence of EBM ..... 479
14.4 The Radiative–Convective Models ..... 483
14.4.1 The Radiative–Convective Models and the Greenhouse Effect ..... 487
14.4.2 Can We Put Together the Radiative–Convective and Energy Balance Climate Models? ..... 490
E.14 Examples ..... 491
E.14.1 Stability of North's Model ..... 491
E.14.2 Time-Dependent Solution of North's Model ..... 492
E.14.3 Temperature Profile from Maximum Entropy Principle ..... 495
E.14.4 Entropy Production and Energy Balance Models ..... 497
References ..... 502

**15 The Application of Climate Models** ..... 503
15.1 The Climate System ..... 503
15.2 The Solar Radiation and the Orbital Parameters ..... 508
15.3 Some Experimental Data on the Ice Ages ..... 511
15.4 The 100 Kyear Cycle and the Lithosphere–Atmosphere Coupling ..... 513
15.5 Stochastic Resonance ..... 519
15.6 The Global Warming: A Simple Exercise ..... 523
15.6.1 The Near Future Climate of the Earth as a Problem of Electrical Engineering ..... 524
15.7 The General Circulation Models ..... 531
15.7.1 The Model Equations ..... 533
15.8 The Performances of GCMs ..... 535
15.8.1 The Taylor Diagram ..... 535
15.8.2 The Feedback Factor ..... 539
15.8.3 The Bayesian Point of View ..... 542
15.8.4 The Bayesian Evaluation of Models: Part 1 ..... 544
15.8.5 The Bayesian Evaluation of Models: Part 2 ..... 546

E.15 Examples ..... 549
E.15.1 100 Kyear Glacial Cycle: Details ..... 549
E.15.2 A Multi-state Climate Model for the Timing of Glaciations ..... 550
E.15.3 The Wigley – Schlesinger Model ..... 555
E.15.4 A Model to Explore Climate Sensitivity ..... 557
E.15.5 Properties of Two-Dimensional Gaussian Distribution ..... 562
E.15.6 A Simple Example ..... 565
References ..... 567

**16 Chemistry of the Troposphere** ..... 569
16.1 Introduction ..... 569
16.2 The Minor Gas Inventory ..... 570
16.2.1 Methane ..... 572
16.2.2 Nitrous Oxide ..... 573
16.2.3 Atmospheric Chlorine ..... 574
16.3 The Biogeochemical Cycle for Carbon ..... 576
16.3.1 Carbonate/$CO_2$ System: A Bit of Marine Chemistry ..... 578
16.3.2 How Long Will the Biosphere Survive? ..... 584
16.4 Chemistry of the Troposphere ..... 589
16.4.1 Methane Oxidation ..... 590
16.4.2 The Chemistry of Urban Air ..... 593
16.4.3 Can We Control Air Quality? ..... 595
16.4.4 The Atmospheric Sulfur Cycle ..... 597
16.5 Modes of a Chemical System ..... 600
E.16 Examples ..... 604
E.16.1 The Simple Carbon Cycle ..... 604
E.16.2 The Carbon Cycle with the Ocean ..... 605
E.16.3 The Oxygen Cycle Is Connected with the Carbon Cycle ..... 606
E.16.4 The Simple Polluted Atmosphere ..... 607
E.16.5 The Isopleth Diagram for Ozone ..... 608
E.16.6 The Lifespan of the Biosphere ..... 609
E.16.7 An Example on Chemical Modes ..... 611
References ..... 612

**17 Dynamics of the Middle Atmosphere** ..... 615
17.1 Thermal Structure of the Stratosphere ..... 616
17.2 The Eulerian Mean Circulation ..... 618
17.2.1 The Transformed Eulerian Mean ..... 620
17.2.2 An Attempt to Understand the Origin of the Residual Circulation ..... 622
17.2.3 The Sudden Stratospheric Warming ..... 623
17.3 Tracers Transport in the Stratosphere ..... 626
17.3.1 The Two-Dimensional Diffusion Coefficients ..... 627

17.3.2 Self Consistent Transport in Two Dimensions ..... 630
17.3.3 Eddies and the Troposphere–Stratosphere Flux ..... 633
17.4 Transport in Isentropic Coordinates ..... 638
17.4.1 Stratospheric Dynamics and Ertel Potential Vorticity .... 640
17.4.2 The Slope of the Tracers ..... 642
17.4.3 The Tracer Correlation: Age of Air and Transport ..... 646
17.4.4 The Conservative Coordinates ..... 651
E.17 Examples ..... 657
E.17.1 Troposphere–Stratosphere Exchange ..... 657
E.17.2 Equatorial Waves ..... 659
E.17.3 The Simplest Theory on Quasi-Biennial Oscillation ..... 663
References ..... 668

**18 Stratospheric Chemistry** ..... 671
18.1 The Ozone Distribution ..... 672
18.2 The Ozone Homogeneous Chemistry ..... 674
18.2.1 The Catalytic Cycles in the Gaseous Phase ..... 676
18.2.2 The Odd Hydrogen Catalytic Cycle ..... 677
18.2.3 The Odd Nitrogen Catalytic Cycle ..... 679
18.2.4 The Bromine and Chlorine Catalytic Cycles ..... 681
18.2.5 The Effects of the Catalytic Cycles ..... 684
18.3 Heterogeneous Chemistry ..... 686
18.4 The Perturbations to the Ozone Layer ..... 689
18.4.1 The Global Ozone Trend ..... 691
18.4.2 Natural and Anthropic Perturbations: Volcanic Eruptions ..... 694
18.4.3 Natural and Anthropic Perturbations: The Effect of Aviation ..... 701
18.5 Polar Ozone ..... 705
18.5.1 The Theory on the Polar Ozone ..... 706
E.18 Examples ..... 709
E.18.1 The Equivalent Effective Stratospheric Chlorine (EESC) ..... 709
E.18.2 Few More Things About Polar Stratospheric Clouds ..... 711
E.18.3 How to Calculate the Loss Rate of Ozone Over Antarctica ..... 712
References ..... 713

**19 Chaos and Nonlinearities** ..... 715
19.1 Simple Examples from the Theory of Dynamic Systems ..... 716
19.1.1 The Poincarè Section ..... 717
19.1.2 Fractal Dimension ..... 719
19.2 The Climate ..... 720

19.3 Is El Niño Chaotic? ... 726
19.4 Dimensions of Weather and Climate Attractors ... 728
19.5 A Bridge to Nonlinearities: The Loop Oscillator ... 733
19.6 The Thermohaline Circulation According to Stommel ... 736
19.6.1 The Model ... 736
19.6.2 Stability of the Solutions ... 740
19.7 The Difference Equations ... 741
19.7.1 Examples for Transitive and Intransitive System ... 743
19.8 Nonlinearity and Delayed Differential Equations ... 745
19.8.1 ENSO as a Delay Oscillator ... 747
19.8.2 Aerosol–Cloud–Precipitation as the Predator–Prey Problem ... 750
E.19 Examples ... 754
E.19.1 The Lorenz System: The Mother of All Chaotic Systems ... 754
E.19.2 The Logistic Map as an Example of Difference Equation ... 758
E.19.3 The Lyapunov Exponent ... 759
E.19.4 MATLAB Program for El Niño Delayed Oscillator ... 762
E.19.5 MATLAB Program for the Predator–Prey Problem ... 762
References ... 763

**20 Geoengineering** ... 765
20.1 A Short Inventory of Geoengineering Technologies ... 766
20.2 Carbon Sequestration and Storage ... 767
20.3 What Geoengineering Can Do ... 771
20.4 Shortwave Options ... 774
20.4.1 Increase Albedo ... 774
20.4.2 Stratospheric Aerosol or How to Create a Volcanic Eruption ... 776
20.5 Space Shields ... 779
20.6 Can Solar Radiation Management Work? ... 781
20.7 A Cure for the Ozone Hole with Geoengineering ... 784
E.20 Examples ... 787
E.20.1 Back to Radiative Transfer ... 787
E.20.2 The Twomey Effect ... 788
E.20.3 Energy Balance Model ... 790
References ... 792

**Erratum** ... E1

**Index** ... 795

# List of Symbols

| | |
|---|---|
| $\alpha$ | Albedo, specific volume, dielectric polarizability |
| $A$ | Wave action |
| $A_{oz}$, $A_{wv}$ | Fraction of solar radiation absorbed by ozone and water vapor |
| $A(u)$ | Absorptivity |
| $B_\nu$ | Planck function |
| $\beta$ | Meridional gradient of Coriolis parameter, Bowen ratio, linewidth parameter |
| $C$ | Circulation |
| $\chi$ | Mass mixing ratio, size parameter |
| $C_p$, $C_v$ | Specific heat per unit mass at constant pressure and constant volume |
| $c$, $c_p$, $c_g$ | Speed of light, phase velocity, group velocity |
| $C_D$ | Drag coefficient |
| $D$ | Molecular diffusion coefficient |
| $D_e$ | Thickness of the Ekman layer |
| $\delta$ | Declination of the sun |
| $\varepsilon$, $\varepsilon_0$ | Emissivity, dielectric constant |
| $e$, $e_s$ | Water vapor pressure and saturation pressure |
| $E$,$\mathbf{E}$ | Electric field, Eliassen and Palm flux, energy density, collision efficiency |
| $E_I$, $E_k$, $E_p$ | Internal, kinetic, and potential energy of the atmosphere |
| $f$ | Coriolis parameter, volume mixing ratio |
| $\phi$ | Latitude |
| $F_A$, $F_{AO}$ | Atmosphere and atmosphere plus ocean heat flux |
| $F_\nu$ | Flux of radiation |
| $\phi$ | Phase of the wave |
| $f(v)$ | Maxwell velocity distribution |
| $\Phi$, $\Phi_e$, $\Phi_l$ | Geopotential, escape flux, limiting flux |
| $F^+$, $F^-$, $F^\uparrow$, $F^\downarrow$ | Upward and downward fluxes |
| $F[y(x)]$ | Functional |
| $g$ | Gravitational acceleration, asymmetry factor |

| | |
|---|---|
| $G$ | Gravitational constant, gain of the antenna, Green function |
| $G_F$ | Feedback of climatic system |
| $\Gamma$, $\Gamma_d$, $\Gamma_s$ | Lapse rate, dry and saturated |
| $G(y)$, $H(z)$ | Form factor for smoke plumes |
| $g_s$ | Conductance |
| $H$ | Scale height, hour angle, sensible heat flux |
| $h$ | Enthalpy, hour angle |
| $I$ | Infrared radiation |
| $I_\nu$ | Intensity of radiation |
| $k$ | Boltzmann constant |
| $K$ | Eddy diffusion coefficient, kinetic energy |
| $K(R, r)$ | Coagulation coefficient |
| $k$, $l$ | Wave numbers in the x and y directions |
| $J$ | Photodissociation coefficient |
| $L$ | Latent heat of condensation |
| $L_r$ | Rossby deformation radius |
| $L_H$ | Latent heat flux |
| $\lambda$ | Wavelength, escape parameter, climate sensitivity |
| $m$ | Complex refractive index |
| $m$, $m_v$, $m_d$ | Mass per unit surface, mass, water vapor mass, mass of dry air |
| $m\mu$ | Cosine of zenith angle, dynamic viscosity |
| $M$ | Angular momentum |
| $N$ | Buoyancy frequency, density of dipoles |
| $n$ | Refractive index |
| $\nu$ | Frequency, viscosity |
| $n(r)$ | Size distribution for aerosols |
| $N(z)$ | Columnar density |
| $n(z)$ | Number density |
| $\Omega$ | Angular velocity of the Earth, solid angle |
| $\boldsymbol{\omega}$ | Vorticity (vector) |
| $\omega$ | Vertical velocity on pressure coordinates, frequency, single scattering albedo |
| $p$ | Pressure |
| $P$ | Potential vorticity, irradiated power, potential energy |
| $\Psi_N$, $\Psi_D$ | Angular momentum for the nebula and the disk |
| $\psi$ | Streamfunction |
| $\Psi$ | Montgomery streamfunction |
| $\Pi$ | Ertel potential vorticity |
| $P(\theta)$ | Phase function |
| $P_0$ | Dipole moment |
| $P_t$, $P_r$ | Transmitted and received power by a radar |
| $q$ | Heat, humidity, quasi-geostrophic potential vorticity |
| $Q_{\text{ext}}$, $Q_{\text{abs}}$, $\text{Q}_{\text{sca}}$ | Extinction, absorption, and scattering efficiency |
| $\dot{Q}$ | Heating rate |

| | |
|---|---|
| $\rho$ | Gas density |
| $R$, $R_v$ | Radius of the Earth or planet, gas constant, reflectivity, gas constant for water vapor |
| $r_a$ | Aerodynamic resistance |
| $R_a$ | Reflectivity of the atmosphere |
| $R_e$ | Reynolds number |
| $R_i$ | Richardson number |
| $S$ | Poynting vector (module), band strength |
| $\mathbf{s}$ | Vector in the natural coordinate system |
| $s$ | Entropy per unit mass |
| $S_0$ | Solar constant |
| $\sigma$ | Stefan Boltzmann constant, surface density, static stability parameter |
| $\sigma_{x,y,z}$ | Semidispersion for smoke plume |
| $\sigma_m(\theta)$ | Cross section for molecular scattering |
| $T$ | Temperature |
| $\tau$ | Time constant, optical thickness |
| $\Theta$ | Temperature |
| $\tau_{ij}$ | Viscous stresses |
| $\theta$, $\theta_e$, $\theta_w$ | Potential temperature and angle, equivalent potential temperature, wet bulb potential temperature |
| $u$ | Internal energy |
| $U$ | Most probable velocity, optical path |
| $\bar{u}$ | Average zonal wind |
| $u_*$ | Friction velocity |
| $u_g$, $v_g$ | Geostrophic wind component |
| $u_E$ | Radiative zonal wind velocity |
| $v$, $V$ | Velocity, gas volume |
| $\mathbf{V}_\mathrm{g}$, $\mathbf{V}_\mathrm{a}$ | Geostrophic wind (vector), ageostrophic wind |
| $w$, $w_s$ | Mass mixing ratio, saturation mass mixing ratio |
| $W$ | Water vapor amount |
| $x$, $y$, $z$ | Coordinates |
| $X$ | Mass streamfunction |
| $Z$ | Geopotential height, radar reflectivity |
| $\zeta$ | Vertical component of relative vorticity |
| $z^*$ | Log–pressure vertical coordinate |

The original version of this book was revised. An erratum to this book can be found at http://dx.doi.org/10.1007/978-3-319-29449-0_21

# Chapter 1
# Fundamentals: Thermodynamics of the Atmosphere

In order to introduce even the most simple questions about atmospheric physics, we need to refresh some basic physics concept. We will start with thermodynamics and continue with radiation (Chap. 2) and very essential fluid dynamics (Chap. 3). This scheme will give us the possibility to compare some characteristics of the planetary atmospheres. We all have studied thermodynamics as a part of general physics and we may have wondered about the purpose of all those theorems and demonstrations. Are they of any utility, for example, for changing a tire on our car or talking to the plumber? Actually one of the most enlightening applications of thermodynamics is to study the atmosphere or, in general, complex systems. Of course we need to study more deeply real gases like water vapor because, in a sense, it is the fuel of the atmosphere. Also the atmosphere is actually a mix of different gases and one should know under which conditions this mix may be treated as a perfect or real gas. We will start from the most elementary concepts and then, as always, we will be very careful about the jargon.

## 1.1 Simple Laws

The applications we have in mind for those things learned in the early years are not many. It will be useful to introduce definitions more typical of meteorologists; for example, we start from the equation of a perfect gas

$$pV = \left(\frac{m_g}{M}\right) RT \tag{1.1}$$

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_1) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_1

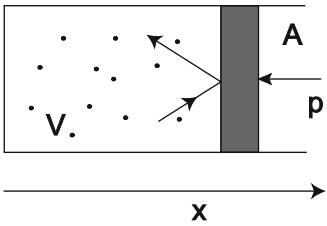



**Fig. 1.1** The gas contained in a cylindrical box closed with a piston

where $p$ is the pressure, $V$ the volume, and $m_g$ and $M$ the mass of the gas and its molecular mass, respectively. $R$ is the gas constant and $T$ the temperature. Usually, Eq. (1.1) is simplified introducing the *specific volume* $\alpha$ and a new gas constant defined as $R' = R/M$. In the case of air, we consider a molecular mass obtained as a weight average of the two major gases composing the atmosphere (nitrogen and oxygen). We get $M = 0.78\ 28 + 0.21\ 32 = 28.9$ and the value of the new constant $R'$ becomes $8314/28.9 = 286.7$ J kg$^{-1}$ K$^{-1}$. Equation (1.1) becomes

$$p\alpha = RT \tag{1.2}$$

where $R$ has been substituted for $R'$.

It is quite interesting to obtain the same law from the kinetic theory of gases that applies simple mechanics laws to a gas contained for simplicity in a box (Fig. 1.1).

Each molecule that strikes the piston will give up part of its momentum and will exert a force on the surface that will result in the pressure $p$. If the component of the velocity is $v_x$, then it will give up a momentum $2mv_x$. The number of molecules per second striking the surface A is the one contained in a cylinder with volume and is equal to $nAv_x$ where $n$ is the number density. The force is then

$$F = nAmv_x 2mv_x \tag{1.3}$$

The pressure is obtained by dividing by the surface A. We need also to consider that only half of the molecules are directed toward the piston and also we need to consider the average quadratic value for $v_x$. We then get

$$p = nm\left\langle v_x^2 \right\rangle \tag{1.4}$$

There is nothing special about the $x$ direction so that we can easily assume that

$$\left\langle v_x^2 \right\rangle = \frac{1}{3}\left(\left\langle v_x^2 \right\rangle + \left\langle v_y^2 \right\rangle + \left\langle v_z^2 \right\rangle\right) = \frac{\left\langle v^2 \right\rangle}{3} \tag{1.5}$$

so that the pressure is simply

$$p = \left(\frac{2}{3}\right) n \left\langle \frac{mv^2}{2} \right\rangle \tag{1.6}$$

If we multiply this equation by the volume $V$, we get

$$pV = \left(\frac{2}{3}\right) N \left\langle \frac{mv^2}{2} \right\rangle \tag{1.7}$$

where $N$ is the total number of molecules in the volume. The temptation is too strong to compare (1.7) with (1.1) and get

$$\left(\frac{m_g}{M}\right) RT = \left(\frac{2}{3}\right) N \left\langle \frac{mv^2}{2} \right\rangle$$

And we can define the kinetic energy as

$$\left\langle \frac{mv^2}{2} \right\rangle = \frac{3}{2} kT \tag{1.8}$$

with k the Boltzmann constant $k = R/N_a$, where $N_a$ is the Avogadro number and R the gas constant.

A very insightful application of (1.8) can be made to find how the pressure changes with altitude. First of all we write (1.2) in a different way introducing the Boltzmann constant $k$ defined as $k = R/N_a$. It is easy to show that

$$p = nkT \tag{1.9}$$

where $n$ is the number of molecules for unit volume

We consider two parallel planes in the atmosphere separated by a distance $dz$, and then the force on each molecule times their density must be balanced by the change in pressure

$$Fndz = dp = kTdn$$

that is equivalent to

$$F = kT \frac{1}{n} \frac{dn}{dz} = kT \frac{d}{dz} [\ln(n)] = -\frac{d}{dz}(PE) \tag{1.10}$$

where we have assumed that the force derived from some potential, that is, the difference in potential energy ($PE$). Then

$$d(\ln \ n) = -d(PE)/kT$$

that can be integrated to give

$$n = \cos t \ \mathrm{e}^{-PE/kT} \tag{1.11}$$

This equation gives the number density (pressure) as a function of the potential energy of a molecule, that is, $mgz/kT$, but it also constitutes the Boltzmann law, that is, the probability of finding a molecule in a determined energy state is proportional to the exponential of that energy divided by $kT$.

The pressure change with altitude is also an application of Equation (1.2). We start from the equation of hydrostatic equilibrium. The pressure change across a layer of thickness $dz$ must be equal to the weight for unit surface of the atmospheric column of the same thickness:

$$dp = -\rho g dz \tag{1.12}$$

The $z$ coordinate is oriented upward from the ground. Equation (2.3) can be easily integrated to give

$$p(z) = \int_z^\infty g\rho dz' \tag{1.13}$$

From this equation we see that the pressure as a function of altitude must depend on the density and then on the temperature.

### *1.1.1 The Scale Height*

A very important connection can be made between thermodynamic quantities and a typical meteorological variable like the *geopotential*. This actually coincides with the gravitational potential when we take as reference the sea level. We can write

$$d\Phi = gdz = -\alpha dp \tag{1.14}$$

where $\Phi$ is the geopotential whose units are $\mathrm{m}^{-2}\mathrm{s}^{-2}$. A quantity which is often used in meteorology is the *geopotential height* $Z$ defined as

$$Z = \frac{\Phi}{g_0} = \frac{1}{g_0}\int_z^\infty g dz' \tag{1.15}$$

The geopotential height differs from the geometric altitude insofar as the acceleration of gravity decreases with altitude. However, in order to find appreciable differences, we need to consider altitudes of several tens of kilometers. In any case the introduction of the geopotential gives some useful relations. Starting from the gas equation and using the hydrostatic equilibrium we get

$$\frac{dp}{dz} = -\frac{pg}{RT} \tag{1.16}$$

from which we get

$$d\Phi = -RT\frac{dp}{p} \tag{1.17}$$

and this can be integrated between pressure levels $p_1, p_2$

$$\Phi_1 - \Phi_2 = R\int_{p_1}^{p_2} T\frac{dp}{p} \tag{1.18}$$

If we divide both sides by $g_0$, we obtain

$$Z_2 - Z_1 = \frac{R}{g_0}\int_{p_1}^{p_2} T\frac{dp}{p} \tag{1.19}$$

The integral can be easily solved for an isothermal atmosphere to give

$$Z_2 - Z_1 = H\ln\left(\frac{p_1}{p_2}\right) \tag{1.20}$$

which is the same as

$$p_2 = p_1 \exp\left[-\left(Z_2 - Z_1\right)/H\right]$$

where $H = RT/g_0$ is the scale height. $H$ gives an indication how fast atmospheric density and pressure decrease with altitude and has been defined only for an isothermal atmosphere. For example, for a temperature of 290 K, $H$ is about 8 km. If we consider a layer defined by the pressures $p_1$ and $p_2$ to which correspond the geopotential heights $Z_1$ and $Z_2$, then we can define an average scale height $\overline{H}$ through Eq. (2.11)

$$Z_2 - Z_1 = -\overline{H}\ln\left(\frac{p_2}{p_1}\right) \tag{1.21}$$

where $\overline{H}$ is the scale height calculated for $\overline{T}$.

### *1.1.2 The Potential Temperature*

Application of the first law of thermodynamics gives the relation between the change in internal energy $du$, heat provided to the system $\delta q$, and work done $dw$

$$\delta q - dw = du \tag{1.22}$$

Equation (2.13) has some interesting consequences. Using the specific volume, it can be rewritten as $\delta q = pd\alpha + C_v dT$, where $C_p$ is the specific heat per unit mass at constant pressure. We notice that $pd\alpha = -\alpha dp$ and we get an expression equivalent to the first law:

$$\delta q = C_v dT + RdT - \alpha dp = C_p dT - \alpha dp \tag{1.23}$$

It is convenient at this point to define the *enthalpy* $h$

$$h = u + p\alpha \tag{1.24}$$

and we find immediately $dh = C_p dT$. From the definition of geopotential, it follows another form of the first law of thermodynamics:

$$dq = d\left(h + \Phi\right) = d\left(C_p T + \Phi\right) \tag{1.25}$$

This very simple equation implies that if the motion of an air parcel is adiabatic ($dq = 0$), then the sum of the enthalpy and geopotential is conserved. Actually, we can get a very useful relation just putting $dq = 0$ in Eq. (1.25):

$$-\frac{dT}{dz} = \frac{g}{C_p} = \Gamma_d \tag{1.26}$$

This represents the *dry adiabatic lapse rate*, that is, the rate of change of temperature with altitude when the motion of the air parcel can be considered adiabatic. If we substitute $g = 9.81$ ms$^{-2}$ and $C_p = 1004$ j kg$^{-1}$ K$^{-1}$, we get $\Gamma_d = 9.8$ K km$^{-1}$. Actually, the average gradient observed in the atmosphere is lower than this value, being about 6.5 K km$^{-1}$ and will be denoted by $\Gamma = -dT/dz$. Once we have introduced the adiabatic gradient, it is very simple to define the potential temperature $\theta$. This would be the temperature assumed by an air parcel initially at pressure $p$ brought adiabatically at pressure $p_0$ assumed as reference at 1000 hPa. Using Eq. (2.14) for an adiabatic and eliminating the specific volume through the gas equation, we have

$$\frac{C_p}{R}\frac{dT}{T} - \frac{dp}{p} = 0$$

which can be integrated from the reference pressure $p_0$ where the temperature is the potential temperature $\theta$ to the level $p$ where the temperature is $T$. We get

$$\frac{C_p}{R}\ln\frac{T}{\theta} = \ln\frac{p}{P_0}$$

and then for the potential temperature

$$\theta = T\left(\frac{p_0}{p}\right)^{R/C_p} \tag{1.27}$$

The exponent $R/C_p$ is about 0.286. The importance of the potential temperature is that it is a conservative quantity as we will see in the course of the book.

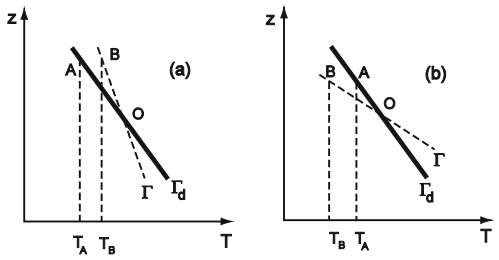



**Fig. 1.2** The static stability: on the *left* negative stability; on the *right* positive stability

### *1.1.3 Static Stability*

Consider an air parcel at some altitude where the measured temperature gradient is $\Gamma$, with the dry adiabatic gradient being $\Gamma_d$. The air parcel does not contain water vapor and it is lifted from point O to A (Fig. 1.2a). Its temperature will change according to the dry adiabat and will assume the value $T_A$. This value is lower than the surrounding atmospheric temperature $T_B$, and at this point the density of the air parcel is greater than that of the surrounding air and will tend to return to the initial position. On the other hand, if $\Gamma_d$ is less than $\Gamma$, the opposite will happen, and, as shown in Fig. 1.2b, the air parcel will move away from the initial point. The first case corresponds to a situation of positive static stability, while the second case is an example of negative static stability. It is interesting to relate the stability to the rate of change of the potential temperature with altitude. Suppose the air parcel of the previous examples has a volume $dV$, density $\rho'$, and temperature $T'$. The surrounding atmosphere, in hydrostatic equilibrium, satisfies the equation $dp/dz = -\rho g$. The air parcel will be subjected to a force per unit volume given by $-(\rho' - \rho)\, g$ and to acceleration $-(\rho' - \rho)\, g/\rho'$. As a function of temperature, we will get

$$\frac{d^2\delta z}{dt^2} = -g\frac{\Gamma_d - \Gamma}{T}\delta z \tag{1.28}$$

In writing Eq. (1.28), we have considered that the pressure inside and outside the air parcel is the same. Also $T \cong T'$ and $T - T' = (\Gamma_d - \Gamma)\,\delta z$ and $\delta z$ is the vertical displacement of the air parcel.

The right-hand side of Eq. (2.18) has the dimension of frequency squared and it can be related to the potential temperature. Using the logarithmic derivative of the potential temperature

$$\frac{1}{\theta}\frac{\partial\theta}{\partial z} = \frac{1}{T}\frac{dT}{dz} + \frac{R}{C_p}\frac{\rho g}{p} = \frac{\Gamma_d - \Gamma}{T} \tag{1.29}$$

and comparing Eqs. (2.18) and (2.19), we get

$$N^2 = \frac{g}{\theta}\frac{d\theta}{dz} \tag{1.30}$$

$N$ is called the *Brunt–Väisälä frequency* and is a measure of the static stability. From Eq. (2.18), we have $\delta z = A\exp(iNt)$, so that for $N^2 > 0$ the air parcel will oscillate around its equilibrium position (positive stability) and if $N^2 < 0$ the air parcel will move away from the equilibrium position (negative stability). These two conditions correspond to

$$N^2 > 0 \rightarrow \frac{d\theta}{dz} > 0 \quad N^2 < 0 \rightarrow \frac{d\theta}{dz} < 0$$

In the case of positive stability, the potential temperature will increase with altitude and the opposite will happen for negative stability.

## 1.2 The Thermodynamics of Water Vapor

A topic in thermodynamics that is often neglected in general physics courses is that of the properties of condensable gases. In studying atmospheric physics, the obvious condensable gas is water vapor, at least for the atmosphere of the Earth. The pressure and atmospheric conditions on our planet are such as to allow the existence of water in its three main phases (gas, liquid, and solid). This peculiarity has a fundamental influence on the weather and the climate of our planet.

We will again start with a brief review of elementary thermodynamics. A good starting point is the introduction of the equation of Clausius–Clapeyron, which gives the saturation pressure of water vapor as a function of temperature.

### *1.2.1 The Equation of Clausius–Clapeyron*

Water (just one chemical component) can appear on three different phases, that is, liquid, gas, and ice (solid). Once we have specified the number of component $\chi$ and the number of phases, $\phi$, the Gibbs law gives the number of independent variable $f$ needed to specify the system:

$$f = \chi - \phi + 2 \tag{1.31}$$

In the particular case of water, $\chi$ =1, so that $f = 3 - \phi$. This means that if the phase is just one (like gas), we need two variables to specify the state like pressure and temperature. If the phases are two like liquid and vapor, then we need to specify only one variable like temperature $T$. If all three phases are in equilibrium with $f = 0$, that means there is only one value for the variables. This is illustrated in Fig. 1.3 where we have traced the boundaries between the phase in a $p$, $T$ plane.

We now consider a liquid in equilibrium with its vapor phase and we consider two isotherms as shown in Fig. 1.4a. When there is only the gas phase, the pressure

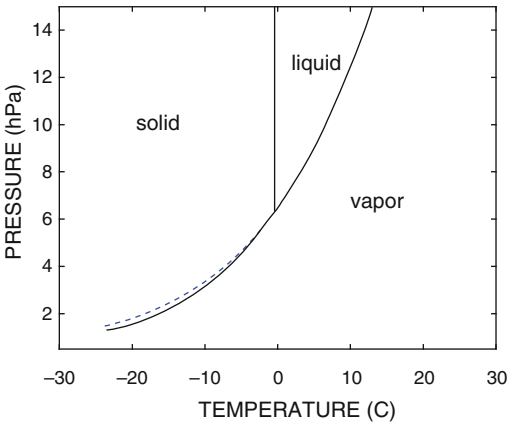



**Fig. 1.3** The phase diagram for water in the *p, T* plane. The three curves indicate those points for which two phases coexist at equilibrium

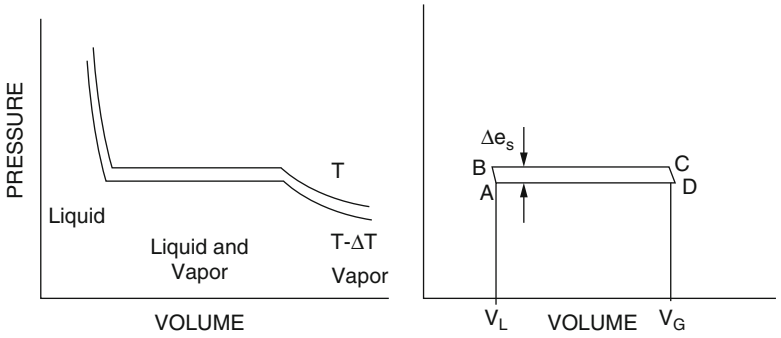



**Fig. 1.4** Isothermal lines for water vapor (**a**, on the *left*) and the elementary Carnot cycle obtained by connecting the two isothermal at constant pressure. The liquid volume ($V_L$) and gas volume ($V_G$) are also indicated

increases as the volume decreases. When the vapor starts to condense and the liquid and gas phase coexist, then the pressure remains constant up to the point where all the gas has condensed. After that the pressure increases rapidly. We consider two isotherms, one at $T$ and the other at $T - \Delta T$, and connect them with two adiabats to constitute an elementary Carnot cycle (see Fig. 1.4b). Starting from point A at pressure $e_s - de_s$ (with $e_s$ being the saturation pressure) and temperature $T - \Delta T$, we execute a slight compression to pressure $e_s$ to which corresponds a volume $V_L$ (point B). In order to increase the saturation pressure, the temperature will increase to $T$. At constant temperature the gas expands to volume $V_G$ (point C); from this point a new expansion is carried out to pressure $e_s - de_s$ and temperature $T - \Delta T$ (point D). Finally through adiabatic compression, the gas goes back to the initial condition. The isothermal expansion from B to C requires that some heat be provided: this is the *latent heat of condensation* $L$ ($2.5\ 10^6$ J kg$^{-1}$). In the same expansion, the work done is $e_s\,(V_G - V_L)$. On the other isotherm (D-A), heat must be subtracted and the work done will be $(e_s - de_s)\,(V_G - V_L)$. If

the work done along the adiabatic paths is neglected, the total work done will be $de_s\,(V_G - V_L)$, and because the efficiency of the Carnot cycle is $dT/T$, we will have $de_s\,(V_G - V_L) = LdT/T$ and then the Clausius–Clapeyron equation

$$\frac{de_s}{dT} = \frac{L}{T}\,(V_G - V_L) \tag{1.32}$$

From this equation, we can get an expression which gives the saturation pressure as a function of temperature. We notice that $V_G$ is always larger than $V_L$ and in particular $V_G$ is the vapor specific volume, that is, $V_G = R_v T/e_s$

$$\frac{de_s}{dT} = \frac{Le_s}{R_v T^2} \tag{1.33}$$

where $R_v$ is the gas constant for water vapor (461 J kg$^{-1}$ K$^{-1}$). Integrating Eq. (1.33) we obtain

$$\ln \frac{e_s}{e_0} = \frac{L}{R_v}\left(\frac{1}{T_0} - \frac{1}{T}\right) \tag{1.34}$$

where the constants $e_0$ and $T_0$ can be obtained from the experimental data of the phase transition for water vapor. We have $e_0 = 611$ hPa when $T_0 = 273$ K. The same relation can be obtained for the saturation pressure with respect to ice. The only thing that changes is the heat of sublimation (2.834 $10^6$ jKg$^{-1}$) so the saturation pressure with respect to ice is lower than respect to water as shown in Fig. 1.3. Notice that integration of (1.33) is possible with the assumption that the heat of condensation or sublimation is independent of temperature. Otherwise, the equation is a bit more complex (see problem).

### *1.2.2 About Eutectics*

For what we plan to do in the book, we need to talk about thermodynamics of solution. In the previous pages, we have concentrated on one chemical component system (i.e., water vapor). Now we consider the simplest example, a solution of salt (NaCl) that could be used later when we will study cloud formation. For a two-component system, $\chi = 2$. Equation (1.31) gives $f = \chi - \varphi + 2 = 4 - \varphi$. A complete representation of a two-component system requires three dimensions.

For simplicity, we choose two like temperature and concentration expressed in practical salinity unit (psu). One psu corresponds to 1 g of salt in 1 kg of water.

This phase diagram is shown in Fig. 1.5. When considering that the variables are reduced to temperature and concentration, the available phases reduce to $f = 3 - \varphi$. In this system, three different phases are possible: the pure water ice, the solution saltwater that we call *brine*, and pure solid salt written as salt hydrate NaCl $2H_2O$. The solid lines separate the three phases according to the phase rule. If we cool

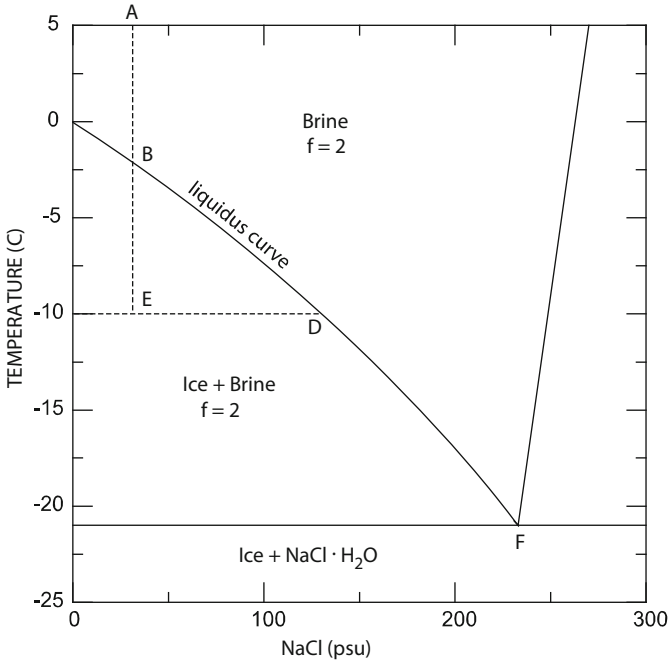



**Fig. 1.5** The water–salt phase diagram

the solution starting from point A (35 psu and a temperature of 5C), we reach point B where the freezing of the solution starts and we have the coexistence of the solid phase (ice) with brine. The amount of ice is proportional to the segment DE, while the amount of liquid (brine) is given by CE. Along the liquidus, both phases (brine and ice) are in equilibrium so that $f = 1$. This means we can move along this line by simply changing temperature with concentration following the rule or vice versa. The point in which all three phases coexist (F) is called eutectic point. The temperature for this solution is $-21.2$ C and the concentration 233 psu. This point is somewhat similar to the triple point for water when gas, liquid, and ice coexist. Sea salt (NaCl) is one of main constituents of the condensation nuclei in clouds, and we will resume the topic later on. However, this introduction is important because it gives the opportunity to introduce another point related to the formation of polar stratospheric clouds (PSC) and the stratospheric aerosol that forms after catastrophic eruptions. Again the simplest case is to deal with a solution of sulfuric acid ($H_2SO_4$) and water. In this case, the phase diagram is more complex because different hydrates are present beside the simplest one $H_2SO_4$ $H_2O$. Figure 1.6a shows the phase diagram for a sulfuric acid solution in water. On the same figure, we report the composition of a condensed particle as a function of temperature for an ambient partial pressure of water vapor of $5 \times 10^{-4}$ hPa (dashed line) and $6 \times 10^{-4}$ hPa (6 ppm) (solid line). The line indicated as ice gives the region where simple

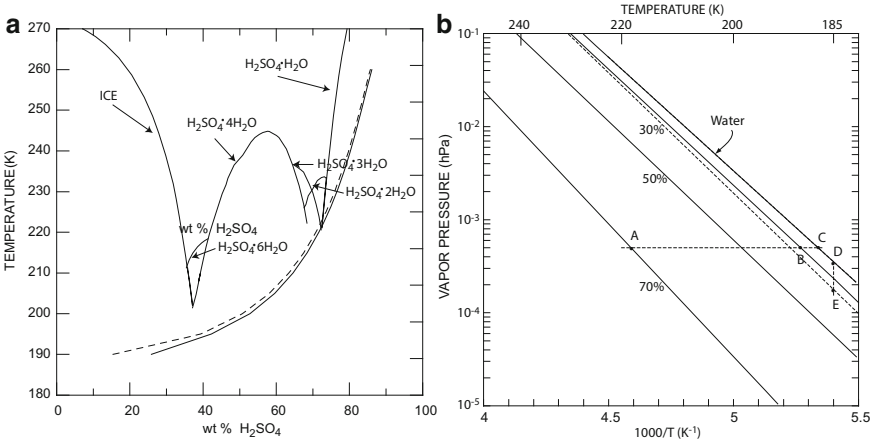



**Fig. 1.6** (**a**) The phase diagram for the solution of sulfuric acid in water is shown. The *dotted line* shows temperature as a function of composition for a mixing ratio of 5 ppm and the *solid line* for 6 ppm. (**b**) For the same solution, the water vapor pressure is shown as a function of $T^{-1}$. The *dotted line* refers to pressure over pure ice

water ice would form. For this region, it is possible to evaluate the water vapor pressure as a function of weight percentage of sulfuric acid. In principle a Clausius–Clapeyron equation can be written also for these solutions as

$$\log P_w = a_1 + a_2 w + a_3 w^2 + a_4 w^3 \tag{1.35}$$

where $P_w$ is the water vapor pressure and $w$ is the weight fraction of sulfuric acid.

The coefficients $a_i$ ($i = 1, 4$) are given in the form $a_i = A + B/T$ where $A$ and $B$ are assigned from experimental measurements. Figure 1.6b (on the right) shows the results of calculated water vapor pressure as a function of temperature and weight percentage. Actually, the temperature scale is shown as $1/T$, and so the pressures are linear with this variable. This diagram is rather interesting. For example, if we start from point A at a pressure of 5 $10^{-4}$ hPa and cool the solution down to point B, the composition of the condensate will change from 70 to 30 %. Also if the temperature falls (point D) to 185 K, the ambient pressure will change from 5 to about 3.1 $10^{-4}$ hPa, and the difference in the amount of water will be absorbed by the condensate that will lower further the sulfuric acid content. On the other hand, if water will freeze out (point B) from this point on, further cooling will force the condensate to move along the dotted line (BE).

Some of the concept we have just illustrated will be used when we will deal with cloud formation but also when we will talk about heterogeneous chemistry.

## 1.3 Some Effects of Water Vapor

Water vapor is a condensable gas (actually it can also sublimate going from gas to ice) under the conditions usually found in the troposphere and the stratosphere. Our conclusions about the thermodynamics of dry air must be modified to take into account water vapor and its important properties.

The first thing to be defined is the mass mixing ratio $w$ as the ratio between the mass of water vapor $m_v$ and the mass of dry air $m_d$ in some volume:

$$w = \frac{m_v}{m_d} \tag{1.36}$$

In particular, if we refer to a unit volume, then the mass mixing ratio coincides with the ratio of densities. If $p$ is the total pressure of the mixture of air and water vapor, then the density ratio can be expressed through the gas equation

$$w = \frac{e/R_v T}{(p-e)/R_d T} \tag{1.37}$$

where $e$ is the partial vapor pressure and $R_d$ and $R_v$ the constants for dry air and water vapor, respectively. We can also write

$$w = 0.622 \frac{e}{p-e} \approx 0.622 \frac{e}{p} \tag{1.38}$$

because the partial pressure is always much smaller than the total pressure.

The presence of water vapor also changes all the previous conclusions about the vertical temperature gradient. If we consider an air parcel that moves upward containing water vapor, its temperature will decrease until the physical conditions are such that condensation occurs. This will happen when the mixing ratio reaches the saturation value. The elementary heat quantity $dq$, in the first law, will be substituted by $-Ldw_s$, where $w_s$ is the saturation ratio and $L$ the latent heat of condensation. The minus sign simply means that when water vapor condenses, $w_s$ decreases ($dw_s$ negative) and heat is provided to the parcel ($dq$ positive). The first law then becomes

$$-Ldw_s = C_p dT + gdz \tag{1.39}$$

Dividing through by $dz$ and writing

$$\frac{dw_s}{dz} = \frac{dw_s}{dT}\frac{dT}{dz} \tag{1.40}$$

we obtain for the *saturated adiabatic lapse rate*

$$\Gamma_s = -\frac{dT}{dz} = \frac{\Gamma_d}{1 + \left(L/C_p\right)\left(dw_s/dT\right)} \tag{1.41}$$

We see that the saturated lapse rate is lower than the dry adiabatic lapse rate. The reason is very simple: when water vapor starts to condense in the parcel as it moves upward, heat is liberated in the condensation process. This heat warms up the parcel. We will see later that water vapor is more or less present in all the terrestrial planets so that using Eq. (1.41) we calculate the dry and the saturation lapse rates. It will be convenient to calculate $dw_s/dT$ directly from the Clausius–Clapeyron equation and from Eq. (1.39). We get easily

$$\frac{dw_s}{dT} = \frac{Lw_s}{R_v T^2}$$

From this relation, we see that the value of the saturation mixing ratio is critical and its evaluation is not straightforward. With the introduction of latent heat, it is possible to extend the concept of potential temperature to those processes which include condensation. The starting point is always the first law written in somewhat different form:

$$\frac{dq}{T} = C_p \frac{dT}{T} - R\frac{dp}{p} \tag{1.42}$$

From the logarithmic derivative of the potential temperature, we obtain

$$C_p \frac{d\theta}{\theta} = C_p \frac{dT}{T} - R\frac{dp}{p}$$

Combining this equation with Eq. (1.42), with $dq = -Ldw_s$, we get

$$-\frac{L}{C_p T} dw_s = \frac{d\theta}{\theta}$$

Considering that, as a first approximation, $dw_s/T$ does not depend on temperature (see example E.1.2):

$$-\frac{Lw_s}{C_p T} = \ln\theta + \text{cost} \tag{1.43}$$

The constant can be obtained with the conditions $w_s \to 0, \theta \to \theta_e$, so that the *equivalent potential temperature* $\theta_e$

$$\theta_e = \theta \exp\left(\frac{Lw_s}{C_p T}\right) \tag{1.44}$$

Consider now the following process. We start with an air parcel containing water vapor. This air parcel will behave as dry air until the pressure and temperature conditions are such that condensation starts. From this point the temperature of the air parcel will change in such a way to maintain the saturation lapse rate. This expansion is also called pseudoadiabatic because the heat source is still within the air parcel. When all the water vapor has condensed out and the water has precipitated, the parcel will again behave as dry air. We can then define the equivalent potential temperature as the temperature that would be obtained by expanding the air mass in a pseudoadiabatic way until all the water vapor is condensed and then compressing the air parcel following the dry adiabatic lapse rate to the reference pressure of 1000 hPa. Similar to the potential temperature, the equivalent temperature is also conserved in the pseudoadiabatic processes. This is a quantity that is used frequently in meteorology and contains an element of complexity. In addition to the potential temperature, we can characterize further the air parcel containing water vapor. The parcel can be brought to saturation conditions in two ways: (a) cooling the air at constant pressure without changing the water vapor content until the saturation temperature is reached and (b) adding or subtracting water vapor to the parcel without any contribution from latent heat. In the first case, the temperature at which the condensation starts coincides with *the dew point temperature*, $T_{\mathrm{d}}$, which can be easily evaluated using Eq. (1.38). In the second case, we will reach the *wet bulb temperature* $T_{\mathrm{w}}$, which is less than the temperature of the air parcel but higher than the dew point temperature. The experimental procedure consists of covering the bulb with a cloth wick wet with clean water. As the water evaporates, the temperature decreases to $T_{\mathrm{w}}$. To understand a little better the meaning of the wet bulb temperature, we can write Eq. (1.39) at constant pressure

$$-Ldw_s = C_p dT \tag{1.45}$$

Consider that we start from an initial temperature $T$ and a mixing ratio $w$ and we calculate the temperature $T_{\mathrm{w}}$, that is, the saturation temperature for a mixing ratio $w'$. Integrating Eq. (1.45) we obtain

$$C_p\left(T - T_w\right) = -L\left(w - w'\right) = L\left(w' - w\right)$$

which is a simple heat balance. We can then calculate the wet bulb temperature:

$$T_w = T - L\left(w' - w\right)/C_p \tag{1.46}$$

Obviously because $w'$ is a function of $T_{\mathrm{w}}$, Eq. (1.46) can be solved, for example, graphically (as we shall see). Different meanings can be attributed to $T_{\mathrm{w}}$. As shown above, when the bulb is at this temperature, the latent heat deposited by condensation is exactly equal to the heat subtracted by evaporation. We can see also that the drier the initial air parcel, the larger must be the temperature decrease.

For saturated air there is no net evaporation and the wet bulb temperature is the same as that of the dry bulb.

All these processes can be easily grasped if we use graphical methods. The reason is obviously a historical one because today it is possible to make all these calculations on a pocket calculator. The calculation in any case is not so easy because the relation that gives the saturation mixing ratio as a function of the temperature is an exponential one. We will describe here one of these diagrams, while in the Appendix we will give details about its construction.

### *1.3.1 The Tephigram or Thermodynamic Diagram*

The first diagram we will illustrate has typically been used by meteorologists working on daily weather forecasting. This diagram is also known as a *tephigram* and it is very instructive because it permits us to visualize concepts we will use often when dealing with thermodynamic transformations in the presence of water vapor.

The state of an air mass can be defined in terms of temperature and pressure or alternatively by temperature and potential temperature. The original tephigram was invented by the English meteorologist Napier Shaw and utilized the potential temperature as ordinate and the temperature as abscissa. Because the potential temperature is related to the entropy (which Shaw denoted by $\phi$), it was natural for him to call the diagram $T\phi$, *tephi*. The version we will illustrate first is the one that uses the pressure as a vertical coordinate (or a function of it), and a general sketch is given in Fig. 1.7. This diagram is known as the diagram of Stüve or *emagram*. We will return later to the Shaw diagram because it offers many

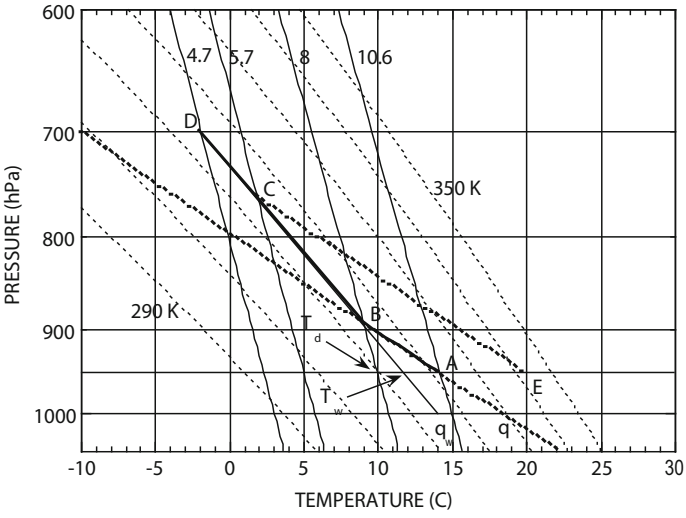



**Fig. 1.7** The tephigram with the history of an air parcel, as described in the text. The *thin dotted lines* are saturated adiabats with equivalent potential temperatures in the range 290–350 K. The *thick dotted lines* are the dry adiabats. The *solid lines* are the constant mixing ratio curves

advantages with respect to the emagram, although the latter is more didactic. In Fig. 1.7, as ordinate we show the pressure in hPa and as the abscissa the temperature in degrees centigrade. With respect to these axes, it is possible to draw the curves at constant potential temperature, which are actually straight lines because the pressure scale is logarithmic. These are the dry adiabats and in the figure they are represented by thick dotted lines. We can also draw curves at constant equivalent potential temperature which are the saturated adiabats. In Fig. 2.3 these are represented by the thin dotted lines which go from 350 K to 290 K spaced by 10 K. Based on the definition of saturation mixing ratio, the relative curves can be traced. In the figure these are the solid lines with smaller slopes and they are indicated by numbers like 10.6, 8.0, 5.7, and 4.7, which are the mixing ratios in units of grams per kilogram.

To show the use of this diagram, we try to solve one of the problems proposed in Chap. 2 of Wallace and Hobbs. An air parcel with a mixing ratio of 8 g $kg^{-1}$ has an initial temperature of 14 ° C at a pressure of 950 hPa. On the diagram, this condition is indicated with the point A, and it is possible to see that the saturation mixing ratio at this point is 10.6 g $kg^{-1}$. The water vapor then is not in saturated condition. At this point the parcel is lifted up to a pressure of 700 hPa and moves initially along the dry adiabat (line AB) up to the point where the mixing ratio curve at 8 g $kg^{-1}$ is intercepted (point B). This happens at a pressure of 890 hPa and from this point on the parcel will cool following the saturated adiabats with an equivalent potential temperature of 315 K (line BD). At the pressure of 700 hPa, the total mass of water vapor condensed is given by 8–4.7 = 3.3 g $kg^{-1}$. If, for example, 70 % of this water is lost as rain at the 700 hPa level, we have 1 g $kg^{-1}$ of liquid water. If we assume that the maximum altitude corresponds to the top of a hill, we can imagine that after that the parcel will start to go down to the other side. When the parcel descends to higher pressure, it heats up and the liquid water evaporates so that the mixing ratio reaches the value of 5.7 g $kg^{-1}$. This happens at a level of 760 hPa and a temperature of 1.8 °C (point C). If the air parcel reaches 950 hPa, it will follow the dry adiabat down to point E where its temperature will be 20 ° C. At this point, we state some definitions: the altitude at which the saturation is reached (point B) is called the *lifting condensation level* (LCL). If starting from this point the parcel is brought to the pressure of 1000 hPa along the saturated adiabat, its temperature would be $\theta_w$. This is called the *wet bulb potential temperature.* The potential temperature $\theta$ at the same altitude is given by continuing along the dry adiabat. On the same diagram, the dew point temperature is easily determined because from point A we follow the isobar until we cross the line of the constant mixing ratio, 8 g $kg^{-1}$. The wet bulb temperature is obtained from the intersection of the saturated adiabat through B and the isobar at 950 hPa. If, from the same point B, the air parcel should have followed a constant mixing ratio line, it would have reached at 950 hPa, the temperature corresponding to the dew point, $T_d$. The wet bulb potential temperature is a conservative quantity during the dry and wet adiabatic processes. The final result of the transformations the air parcel has experienced along the paths A, B, C, D, and E is a heating from the initial temperature of 14 °C to the final temperature of about 20 °C. From a simple examination of a diagram of this kind, it is then possible to make a forecast for the air parcel and to evaluate, for example, how much liquid water is available for precipitation.

### 1.3.2 The Skew T–Log P Diagram (Emagram)

This diagram is used professionally by all meteorologists and is directed to make the angle between isotherms and adiabats more nearly 90°. In principle, this can be done using a coordinate transformation from the $p$, $T$ (or $\alpha$) space to a generic $B$ and $A$ space where $B$ and $A$ are new variables related to the usual thermodynamic variables. It is shown in one of the exercises that with the requirements of equal area in both spaces implies a condition for A:

$$\left(\frac{\partial A}{\partial \alpha}\right)_{\ln p} = -\frac{1}{R}\left(\frac{\partial p}{\partial \ln p}\right)_{\alpha} = -\frac{p}{R} \tag{1.47}$$

This relation can be integrated by multiplying both sides by $d\alpha$ to obtain at constant pressure

$$A = -\frac{p\alpha}{R} + F(\ln p) = -T + F(\ln p) \tag{1.48}$$

The function $F(\ln p)$ can be chosen in such way that $F(\ln p) = -K \ln p$ so that the new coordinate system will be

$$\begin{aligned} A &= T + K \ln p \\ B &= -R \ln p \end{aligned} \tag{1.49}$$

The angle between isotherms and adiabats will depend on the arbitrary constant $K$. The isotherm will have an equation with $T_i$ generic temperature:

$$A = T_i + K \ln p = T_i - \frac{K}{R}B \Rightarrow B = -\frac{R}{K}A + \frac{R}{K}T_i \tag{1.50}$$

Isotherms will be straight parallel lines whose inclination depends on K and will be sloping on the right for decreasing pressure.

To obtain adiabats, we simply use the definition to get

$$\ln T = \frac{R}{C_p}\ln p + \left(\ln \theta_i - \frac{R}{C_p}\ln p_o\right) \tag{1.51}$$

where the term in the parenthesis is a constant for chosen $\theta_i$. Because $T$ is not a coordinate curve with constant potential, temperature will not be straight lines but rather will be concave upward. While the isotherms will be straight lines running from lower left to upper right (notice the negative coefficient *K/B*), the adiabats will be roughly at right angle. We have constructed an emagram as an example and this is shown in Fig. 1.8a. Detail of this construction can be found in one of the problems at the end of this chapter. We can see that the ordinate scale is again given by the log pressure, while the value of the abscissa is simply indicating the intersection

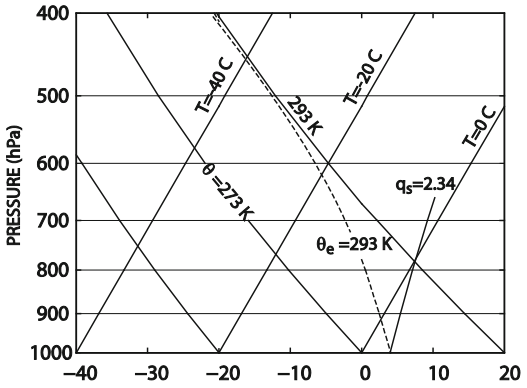



**Fig. 1.8a** The skew T–log P diagram. The vertical coordinate is still log p, while the abscissa is related to both pressure and temperature. The isotherm at 0, −20, and −40 °C is shown together with dry adiabats at 292 K and 273 K. Also shown is the wet adiabat at 293 K and the saturated mixing ratio of 2.34 g/kg

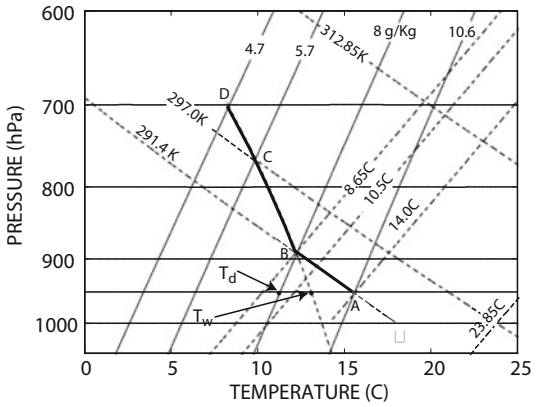



**Fig. 1.8b** The same history shown in Fig. 1.7 is here traced on the skew T diagram

of isotherms with the reference pressure of 1000 hPa. The adiabats are now well separated from the isotherms. The pseudoadiabat at 293 K is also shown and as expected tends to coincide with the dry adiabat at the same temperature for very small water mixing ratio. Saturated mixing ratio line at 2.34 g/kg is also shown. Lower value for $q_s$ will be on the left of this line.

Figure 1.8b shows the same parcel history of Fig. 1.7. The particle starts at point A at a pressure of 950 hPa and 14 C temperature, and it is raised along the dry adiabat up to 890 hPa where now the temperature has reached 8.65 C. Point B is where the condensation starts because it is the intersection with the 8 g/kg mixing ratio. From this point on, the parcel moves along the wet adiabat to reach the 700 hPa pressure level. At this point the parcel has a water content of 4.7 g/Kg and has lost 8–4.7 = 3.3 g of water. If the parcel should descend from point C, the temperature at 1000 hPa would be 23.85 C almost 10 C warmer than the initial value.

The wet adiabat corresponds to an equivalent potential temperature of 312.85 K, and for reference the corresponding dry adiabat is shown.

On the same diagram, we also show how to obtain the initial potential temperature $\theta$ and the dew point temperature $T_d$ and the wet bulb temperature $T_w$. The latter

is obtained by pronging the wet adiabat to the 950 hPa pressure level. $T_d$ is obtained by intersecting the initial mixing ratio at constant pressure level.

The advantages of the skew T–log P diagram are due to the fact that the chart then makes more evident the temperature discontinuities as a function of pressure. In this diagram, the pressure is on a logarithm scale, and the isotherms and adiabats are almost normal. As an application of these diagrams, we can discuss a number of interesting effects that the presence of water vapor have on the atmospheric stability. In practice, we can repeat the same arguments used to establish the stability of a dry atmosphere, but we should include the possibility of phase change for water.

### *1.3.3 The Conditional Convective Instability*

We have seen that the static stability in the case of dry air depends on the difference between the adiabatic lapse rate and the actual lapse rate. In the real atmosphere, water vapor is always present, so the observed lapse rate will probably between the dry and wet one given by Eq. (1.41). This calls for reexamination of the discussion about the stability criteria. We can refer to Fig. 1.9 assuming that the observed lapse rate ($\Gamma$) is between the saturated lapse rate $\Gamma_s$ and the dry one $\Gamma_d$. The air parcel that moves from its equilibrium position at O cools along the dry adiabat until the water vapor it contains starts to condense. This is called the *condensation level*. At this point the air parcel is colder than the surrounding atmosphere so that, in the absence of external forces acting on the mass, it will return to its initial position. If there are forces acting and these are such to take the parcel beyond A, the temperature will change according to the saturated adiabat.

Once the parcel is beyond point B, its temperature will be higher than the surrounding atmosphere, and it will have positive buoyancy and move away freely. Point B corresponds to the *level of free convection*. The important thing to notice at this point is that a temperature profile can be stable with respect to the dry adiabat ($\Gamma < \Gamma_d$) but unstable with respect to a saturated adiabat ($\Gamma_s < \Gamma$). If atmospheric motions are able to force the air parcel beyond point A, then we have a free motion

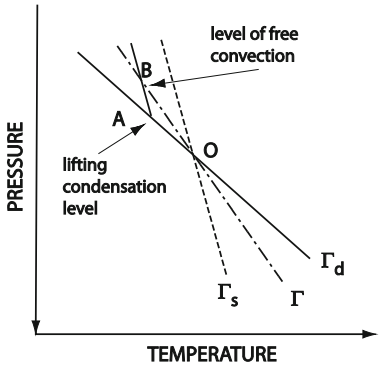



**Fig. 1.9** Scheme for conditional instability. See text for discussion

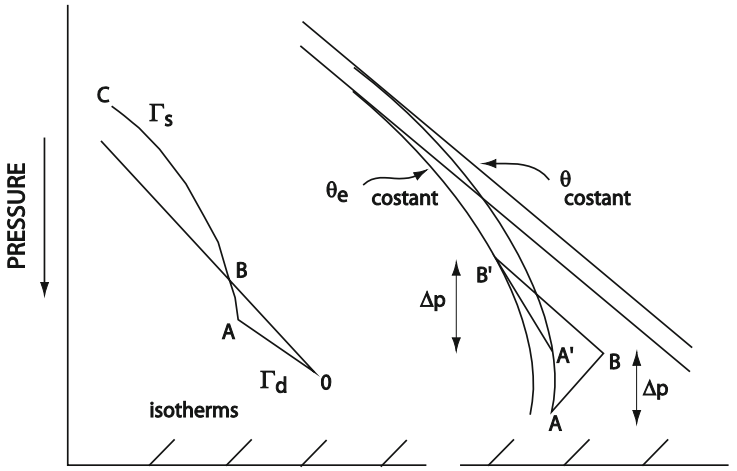



**Fig. 1.10** Two examples of application of a tephigram. On the *left* is shown the same case as in Fig. 1.9. The *sloping lines* on the bottom are isotherms

that corresponds to a *convective instability*. This is known as *conditional instability.* The same concepts can be illustrated using a tephigram, as illustrated in Fig. 1.10. Consider an air parcel initially at point O (left part of the figure). At first the parcel will move along the dry adiabat until the condensation level is reached at point A. Beyond A, the parcel will follow the saturated adiabat AC. Along the path OAB, the temperature of the air parcel is lower than the temperature of the surrounding air, represented by OB, so that the parcel has negative buoyancy and has to be forced. Beyond B the temperature of the air parcel is higher than the surroundings and the mass is in a free convection regime.

On the right side of the same figure, we consider now a layer confined between points A and B and of thickness $\Delta p$. Initially, this layer is stable because the temperature in B is higher than in A, which is in a condition of saturation. If this layer moves upward, point A will remain on the saturated adiabat, while B will move along the dry adiabat until it reaches the condensation level in B′. At this point, the temperature lapse rate in the layer is larger than the saturation lapse rate, and the layer is convectively unstable.

Application of these concepts in a more sophisticated form is useful especially for meteorological phenomena near the equator. Here, the convergence of the air masses coming from opposite hemispheres, coupled with the large amount of water vapor contained in them, is an important energy source for the atmospheric circulation. In general, however, convective instability can develop in any region where the low levels of the atmosphere are occupied by humid and warm air, while the upper layers contain drier air masses. Under conditions in which these masses are lifted, very violent convective activity (severe thunderstorms) may result. This is a quite common situation that develops in summer in the region of the USA known as the Great Plains.

## 1.4 The Distribution of Water Vapor in the Atmosphere

The circumstance that water vapor is so abundant on the Earth, and subject to condensation, suggests a very simple way to treat its distribution with altitude. Air masses containing water vapor are lifted in convective motion; during ascent they cool up to the saturation point. As a first approximation, we could imagine that the water vapor mixing ratio corresponds to the saturated mixing ratio. Actually, when looking at real data, we note that saturated conditions are met very rarely. An empirical, global averaged relation gives the relative humidity

$$\frac{e}{e_s} \cong 0,8 \left( \frac{p}{p_0} \right) \tag{1.52}$$

where $p_o$ is the ground pressure. An intuitive explanation of this behavior is that, when saturated conditions are reached, the condensed water precipitates as ice or liquid, depending on a rather complicated mechanism. The upper layers are depleted of water vapor with respect to the saturated value. Starting from Eq. (1.34), it is possible to find an almost analytical solution to the water vapor distribution. Note that Eq. (1.52) is valid only in the troposphere because in the stratosphere the mixing ratio is very small, although it may increase a little with altitude. We start by rewriting (1.34) with the pressure expressed in Pascals

$$e_s = 2,5049 \cdot 10^{11} \exp\left( -\frac{5417}{T} \right) \tag{1.53}$$

where $T$ is the temperature at altitude $z$. The atmosphere is in hydrostatic equilibrium so pressure will change as $p/p_0 = \exp(-gz/RT)$. This can be substituted in Eq. (1.52) to give

$$e = e_g \exp(-z/H_w) \tag{1.54}$$

where $e_g$ is the vapor pressure at the ground obtained by combining Eqs. (1.52) and (1.53), while $H_w$ is given by

$$\frac{1}{H_w} = \frac{5417}{z} \left( \frac{1}{T} - \frac{1}{T_g} \right) + \frac{g}{RT} \tag{1.55}$$

Assuming the temperature lapse rate $\Gamma = -dT/dz$ is constant, we obtain

$$H_w = \frac{RT/g}{1 + 5417R\Gamma/gT_g} \tag{1.56}$$

and considering that the scale height is given by $H = RT/g$, we have

$$\frac{H}{H_w} = 1 + 5417R\Gamma/gT_g \tag{1.57}$$

This relation shows that the ratio of the scale heights is independent of the local $T$.

Assuming $\Gamma = 6.5\ \text{km}^{-1}$ and $T_g = 288$ K, we have $H_w/H \approx 0.22$ and for water vapor a scale height around 2–3 km. This means that the mixing ratio decreases rapidly with altitude. We need to make a few points about (1.57). We have used without restraint the Clausius–Clapeyron equation. Actually, when the temperature falls below zero degrees, the saturation pressure should be evaluated not with respect to the liquid phase but with respect to ice. However this would also be an approximation considering the supercooling of liquid droplets. When the temperature falls below zero at some altitude, the water could be in the liquid state (droplets) or present as vapor. In these conditions there is not a spontaneous passage to the solid phase unless there are condensation nuclei (impurities of very small size or aerosols). On the other hand, it is very difficult to measure the vapor pressure with respect to ice because pressures are very small. Recently this problem has received renewed interest because of the problem of polar stratospheric clouds in connection with the ozone hole. As a matter of fact, there are now new measurements available to express the vapor pressure with respect to ice, according to the relation

$$\log e = -2667.4/T + 10.553 \tag{1.58}$$

This must be compared with the Clausius–Clapeyron equation. Both expressions include the triple point of water (where all three phases coexist) $e = 611.65$ Pa, $T = 273.15$ K.

The difference between these two relations is shown in Fig. 1.3 where we notice that above 0 °C the difference is small, while for low temperatures (which we often find in the upper troposphere or lower stratosphere), the difference may be important. In the same figure, the triple point is well evidenced, and also we note how the melting temperature of the ice is almost independent of pressure. We will discuss at length this topic when dealing with the heterogeneous chemistry in the stratosphere. For the time being, we can conclude this part by reporting a calculation based on the previous equations about the water vapor content of the lower atmosphere of the Earth. In Fig. 1.11, we show two profiles for the water vapor volume mixing ratio calculated at saturation or taking into account the observed data, so that the relative humidity is proportional to the local pressure according to Eq. (1.52). The mixing ratio by volume is easily converted into mixing ratio by mass, so that the values range from a few grams per kilogram to a few milligrams per kilogram going from the surface to the tropopause. The mixing ratio at the tropopause presents a discontinuity because the temperature in this region is almost constant or tends to increase slightly. The value of the mixing ratio in this region is a function of temperature. The main source of water vapor is at the surface, and as a consequence the tropopause acts as a "cold finger" for water vapor as in the vacuum apparatus. In practice, what determines the water vapor mixing ratio is the coldest temperature found in the tropopause. In this sense, the cold finger controlling the humidity of the stratosphere is actually the tropical tropopause, which is one of the coldest regions in the Earth's lower atmosphere. The mixing ratio observed in the stratosphere is then a little lower than the one given in Fig. 1.11, that is, only a few

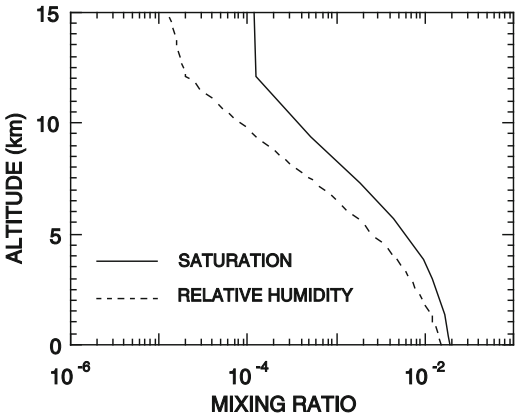



**Fig. 1.11** The water vapor mixing ratio in the atmosphere of the Earth calculated at saturation (*solid line*) and according to Eq. (2.35) (*dashed line*)

parts per million (ppm or $10^{-6}$). In principle, clouds will form every time the vapor pressure exceeds the saturation pressure locally. However, as we will see later, the problem of cloud formation is much more complex.

It is rather important to note that elementary thermodynamics has taken us far along the road to study the atmospheres of the planets. It is hoped that this will be an encouragement not only to students but mainly to teachers.

## E.1 Examples

### *E.1.1 Was the Atmosphere Drier During the Ice Age?*

We can carry out now a very simple exercise, based on the work of Wallace Broecker, a well-known oceanographer and climatologist. Through a number of experimental methods, it is possible to show that the *snow line* during the last ice age was lower by roughly 800 m. The snow line is the altitude where the temperature falls below zero degrees. Broecker has shown with very simple thermodynamic arguments that this implies a drier atmosphere, that is, an atmosphere with less water vapor. In principle, this is obvious because in a colder atmosphere, the saturation pressure would be reduced and excess water vapor would condense and rain out. However, the originality of Broecker's method is that he used simple thermodynamics to estimate even the temperature structure of the atmosphere during the last ice age. The known data are the pressure and altitude where the snow line is located. In the present epoch, an average value for the tropics is assumed to be 5.3 km at a pressure of 500 hPa, while this value would have been lowered to 4.5 km at 570 hPa, during the ice age. For the present epoch, we assume a relative humidity of 65 % and a surface temperature of 26 °C. The observed temperature lapse rate is then $\Gamma = 26/5.3 = 4.9$ K km$^{-1}$. If the dry adiabatic lapse rate determines the temperature, then at 5.3 km we would have a temperature

of $26 - 9.8 \cdot 5.3 = -25.9$ °C. In radiative equilibrium we need heating of the same amount. This heating can be provided by water vapor condensing out. We easily find that, for each gram mol of water (18 g) that condenses, a mole of air is heated by 1.55 °C. A total heating of 25.9 °C then requires the condensation of 16.7 g mol of water vapor per mole of air. The assumption of a 65 % relative humidity implies that at the snow line the saturation pressure is 611 Pa and consequently the partial pressure is $0.65 \cdot 611 = 397$ Pa. We can then assume that at the ground the total mixing ratio would be $16.7 + 7.94 = 24.64$ g mol per mole of air, which implies a relative humidity at the ground of 72 %, not too far from the 65 % value we assumed for the entire range of altitude. Exactly the same calculation can be repeated now for ice age condition assuming that the surface temperature might have been lower. For a surface temperature colder by 1 °C, we would find a surface mixing ratio of water of 19.9 g mol, while for – 5 °C we would find 22 g mol. We then observe a reduction of the water vapor at the ground ranging from 81 to 92 %.

Later in this book we will discuss the ice age at length and also global warming, and we need to keep in mind these very simple arguments. Less water vapor in the atmosphere means less heat absorbed by the atmosphere and this could be a way to accelerate the cooling. On the other hand, if the Earth warms, the opposite reasoning implies that it should have more water vapor in the atmosphere, and this again could accelerate warming. But things are much more complex than that because the balance between radiative forcing and condensation may depend on the water vapor distribution with altitude. We need to slow down a bit and continue on the learning curve.

### *E.1.2 More on the Clausius–Clapeyron (C–C) Equation*

We obtained a simple way to calculate the saturation pressure of water vapor integrating Eq. (1.33). The integration assumed that the latent heat is constant. The most simple dependence of the latent heat on temperature is linear so that we could write

$$L_v = L_{v0} + \left(C_{pv} - C_{pw}\right)\left(T - T_0\right) \tag{E.1.1}$$

where $L_{v0}$ is a reference value for $L_v$ and $C_{pv}$ and $C_{pw}$ refer to the specific heat for vapor and liquid water, respectively. Substituting Eq. (E.1.1) in Eq. (1.33), we get

$$\frac{1}{e_{sw}}\frac{de_{sw}}{dT} = \frac{L_{v0} + \left(C_{pw} - C_{pv}\right)T_0}{R_v T^2} - \frac{C_{pw} - C_{pv}}{R_v T}$$

This equation can be integrated by parts to give

$$\ln\frac{e_{sw}}{e_{s0}} = \frac{L_{v0} + \left(C_{pw} - C_{pv}\right)T_0}{R_v}\left(\frac{1}{T_0} - \frac{1}{T}\right) - \frac{C_{pw} - C_{pv}}{R_v}\ln\frac{T}{T_0} \tag{E.1.2}$$

Considering that $T_0 = 273.15$ K, $e_{s0} = 6.11$ hPa, $L_{v0} = 2.5 \times 10^6$ Jkg$^{-1}$, $C_{pv} = 1850$ Jkg$^{-1}$K$^{-1}$, and $C_{pw} = 4218$ Jkg$^{-1}$K$^{-1}$, Eq. (E.1.2) reduces to

$$e_{sw} = 6.11 \exp\left(53.49 - \frac{6808}{T} - 5.09 \ln T\right) \tag{E.1.3}$$

While in case $L_v$ is assumed to be independent from temperature, the saturation pressure is given by Eq. (1.34). The same procedure can be followed for ice and in this case for latent heat depending on the temperature the saturation pressure over ice is given by

$$e_{si} = 6.11 \exp\left(26.16 - \frac{6293}{T} - 0.555 \ln T\right) \tag{E.1.4}$$

For latent heat independent from temperature, we have Eq. (1.58).

### *E.1.3 The Equivalent Potential Temperature*

The equivalent potential temperature defined by Eq. (1.44) has been obtained with the approximation that $dw_s/T$ does not depend on temperature. To show that this is an acceptable approximation, consider

$$d\left(\frac{w_s}{T}\right) = \frac{w_s}{T}\left(\frac{dw_s}{w_s} - \frac{dT}{T}\right) \tag{E.1.5}$$

We shall show that the first term in the parenthesis on the right is much larger than the second. Their ratio is given by

$$\frac{T}{w_s}\frac{dw_s}{dT}$$

And considering the definition of $w_s$, we get

$$\frac{T}{w_s}\frac{dw_s}{dT} = \frac{T}{e_s}\frac{de_s}{dT} - \frac{T}{p}\frac{dp}{dT}$$

The first term on the right can be expressed with the Clausius–Clapeyron equation:

$$\frac{T}{w_s}\frac{dw_s}{dT} = \frac{L}{R_v T} - \frac{T}{p}\frac{dp}{dT}$$

The order of magnitude of both terms on the right along a saturated adiabat gives around 0 °C:

$$\frac{T}{w_s}\frac{dw_s}{dT} \approx 20 - 5 = 15$$

So that we can neglect the second term in (E.1.5) and obtain

$$d\left(\frac{w_s}{T}\right) = \frac{dw_s}{T}$$

So that for the equivalent potential temperature, we get

$$-\frac{L}{C_p T}dw_s = -\frac{L}{C_p}d\left(\frac{w_s}{T}\right) = \frac{d\theta}{\theta}$$

### *E.1.4 The Saturated Adiabat*

We can start from Eq. (1.39) and express $dw_s$ as

$$\frac{dw_s}{w_s} = \frac{de_s}{e_s} - \frac{dp}{p}$$

So that Eq. (1.39) becomes

$$-Lw_s\left(\frac{de_s}{e_s} - \frac{dp}{p}\right) = C_p dT + g dz$$

And dividing by $dz$ and rearranging, we get

$$-\frac{dT}{dz}\left(1 + \frac{Lw_s}{C_p e_s}\frac{de_s}{dT}\right) = \frac{g}{C_p} + \frac{Lw_s}{C_p}\frac{\rho g}{p}$$

Considering the C–C equation, we have $\frac{1}{e_s}\frac{de_s}{dT} = \frac{L}{R_v T^2}$

and the saturated lapse rate

$$-\frac{dT}{dz} = \Gamma_s = \Gamma_d \frac{1 + \frac{Lw_s}{RT}}{1 + \frac{\varepsilon L^2}{C_p R}\frac{w_s}{T^2}} \tag{E.1.6}$$

This equation can be integrated numerically to get a saturated adiabat. When the saturation mixing ratio goes to 0, the saturated lapse rate coincides with the dry lapse rate. It is rather interesting to compare this expression with the definition of equivalent potential temperature given by Eq. (1.44). In this case, when $\theta_e$ has been fixed, the derivation of (1.44) with respect to altitude must be zero:

$$\frac{1}{\theta}\frac{d\theta}{dz} + \frac{L}{C_p}\left(\frac{1}{T}\frac{dw_s}{dz} - \frac{w_s}{T^2}\frac{dT}{dz}\right) = 0 \tag{E.1.7}$$

After some algebra, we reach a conclusion very similar to (E.1.6), that is,

$$-\frac{dT}{dz} = \Gamma_s = \Gamma_d \frac{1 + \frac{Lw_s}{RT}}{1 + \frac{\varepsilon L^2}{C_p R}\frac{w_s}{T^2} - \frac{w_s L}{C_p T}} \tag{E.1.8}$$

We see that an additional term in the denominator and the ratio between the two terms is given by

$$\frac{\varepsilon L}{RT} \approx \frac{0.622 \cdot 2.5 \times 10^6}{287 \cdot 300} \approx 18$$

Again we see that the approximation used earlier is reasonable. These results are shown in Fig. E.1.1. The moist adiabat obtained by using the condition $\theta_e = \cos t$ is not distinguishable from the one obtained by solving Eq. E.1.6. Although the construction of these curves is simple, we give below a MATLAB program to integrate the simple equation $dT/dp = \Gamma_s/\rho g$.

```
[z,y]=ode45('gradfun',[4e4 1e5],[210.21 210.21]);
semilogx(z,y(:,1), z,y(:,2),'r'),xlabel('p'),
  axis([4e4 1e5 210 280])
function ydot=gradfun(z,y)
ydot=zeros(2,1);
rd=287;
cp=1004;
gammad=-rd/cp;
lat=2.5e6;
esat=611*exp(19.83-5417/y(1))
epsi=0.622;
ws=epsi*esat/z;
num=1+lat*ws/(rd*y(1));
den=1+epsi*lat*lat*ws/(cp*rd*y(1)*y(1));
aux=num/den;
ydot(1)=-aux*gammad*y(1)/z;
ydot(2)=-gammad*y(2)/z;
```

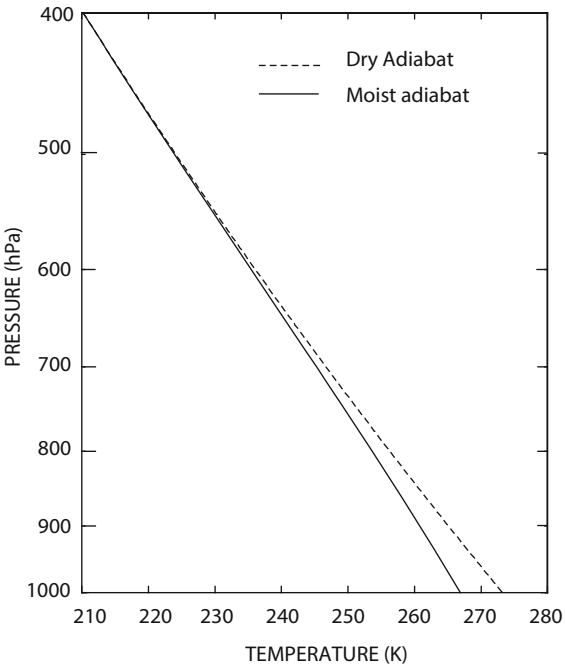



**Fig. E.1.1** The dry and moist adiabat obtained with the program given below. The potential temperature is 273.15 K

## *E.1.5 Constructing an Emagram*

In Sect. 1.3.1, we have shown what the emagram is. Here, we want to show how to construct it. The MATLAB program has been inspired by the one presented in Kerry Emanuel's page http://ocw.mit.edu/OcwWeb/Earth--Atmospheric--and-Planetary-Sciences/12-811Spring-2005/Tools/index.htm. It is very simple and is reported below.

```
function success=tskew(pz,tz,rhz)
%
% pz = pressure (hPa), tz = temperature (C),
   rhz = relative humidity (0-1)
%
pz=1000.;
tz=273.;
rhz=0.6;
p=[1050:-50:600];
pplot=transpose(p);
t0=[0:1:25];
[ps1,ps2]=size(p)
ps=max(ps1,ps2);
[ts1,ts2]=size(t0);
ts=max(ts1,ts2);
```

```
for i=1:ts,
   for j=1:ps,
      tem(i,j)=t0(i)+30.*log(0.001.*p(j))
      thet(i,j)=(273.15+tem(i,j)).*(1000./p(j)).^.287
      es=6.112.*exp(19.83.*tem(i,j)./(273.15+tem(i,j)));
      q(i,j)=622.*es./(p(j)-es)
      thetaea(i,j)=thet(i,j).*exp(2.5.*q(i,j)
        ./(tem(i,j)+273.15));
   end
end
p=transpose(p);
t0=transpose(t0);
temp=transpose(tem);
theta=transpose(thet);
thetae=transpose(thetaea);
qs=transpose(q);
h=contour(t0,pplot,temp,[8.65,10.5,14,20],'k');
hold on
set(gca,'ytick',[1000:100:600])
set(gca,'yscale','log','ydir','reverse')
set(gca,'fontweight','bold')
set(gca,'ytick',[600:100:1000])
set(gca,'ygrid','on')
hold on
h=contour(t0,pplot,theta,[291.4,297, 312.85],'b');
h=contour(t0,pplot,qs,[4.6,5.7,8,10.7],'g');
h=contour(t0,pplot,thetae,[312.85],'r');
hold off
xlabel('Temperature (C)','fontweight','bold')
ylabel('Pressure (mb)','fontweight','bold')
```

The core of the program is the following instructions:

```
for i=1:ts,
   for j=1:ps,
      tem(i,j)=t0(i)+30.*log(0.001.*p(j))
```

Here, he builds a matrix for the temperature (tem) expressed in degrees centigrade. This expression reproduces the first transformation in (1.49). The reader may change the coefficient (30) to obtain different inclination for the curves.

```
      thet(i,j)=(273.15+tem(i,j)).*(1000./p(j)).^.287
      es=6.112.*exp(19.83.*tem(i,j)./(273.15+tem(i,j)));
      q(i,j)=622.*es./(p(j)-es)
```

These instructions simply build matrix for the potential temperature and the saturated mixing ratio.

```
      thetaea(i,j)=thet(i,j).*exp(2.5.*q(i,j)
        ./(tem(i,j)+273.15));
```

This instruction simply gives the equivalent potential temperature as a matrix. Notice that the factor 2.5 before q(i,j) is simply an approximation to $L/C_p$

```
      end
 end
```

To draw the lines at constant temperature, mixing ratio, etc., the instruction contour is used in which the specific values to be drawn are specified.

### *E.1.6 The Equal-Area Requirement*

We would like to justify the Eq. (1.47). We start from a transformation from the system $p, \alpha$ to the system A, B. As shown before, A and B are functions of the initial thermodynamic variable so that each point p, −a corresponds to a point in A, B. Any closed cycle in one diagram is a closed cycle on the other.

We shall require that the area enclosed in the two diagrams is the same to insure that the plot is a thermodynamic diagram. Thus

$$\oint -pd\alpha = \oint AdB$$

We have

$$\oint (pd\alpha + AdB) = 0$$

This implies that the integrand must be an exact differential like $ds$

$$\oint (pd\alpha + AdB) = ds$$

where $ds$ could be a function of $\alpha$ and B

$$ds\,(\alpha, B) = \left(\frac{\partial s}{\partial \alpha}\right)_B d\alpha + \left(\frac{\partial s}{\partial B}\right)_\alpha dB$$

Comparing with the previous expression

$$p = \left(\frac{\partial s}{\partial \alpha}\right)_B \quad \text{and} \quad \mathrm{A} = \left(\frac{\partial s}{\partial B}\right)_\alpha$$

If we derive the first term with respect to B and the second with respect to $\alpha$

$$\left(\frac{\partial p}{\partial B}\right)_\alpha = \frac{\partial^2 s}{\partial \alpha \partial B} \quad \text{and} \quad \left(\frac{\partial A}{\partial \alpha}\right)_B = \frac{\partial^2 s}{\partial \alpha \partial B}$$

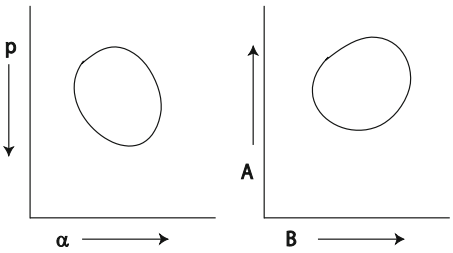



**Fig. E.1.2** Representation of an equal-area transformation from a .−$p$, α space to a A, B space

So that

$$\left(\frac{\partial p}{\partial B}\right)_{\alpha} = \left(\frac{\partial A}{\partial \alpha}\right)_{B}$$

When $T$ is substituted for B, we get (1.47) (Fig. E.1.2).

## *E.1.7 The Virtual Temperature*

Virtual temperature is introduced in connection with saturated air. In a mixture of dry air and vapor, the equation of state for the mixture reads

$$(p_d + e)\, V = RT\left(\frac{m_d}{M_d} + \frac{m_v}{M_v}\right) = T\left(R_d m_d + R_v m_v\right)$$

where "d" refers to dry air and "v" to vapor. We can eliminate the volume through the definition of specific volume

$$\alpha = \frac{V}{m_d + m_v}$$

to get

$$p\alpha == T\left(\frac{R_d m_d + R_v m_v}{m_d + m_v}\right) = qT\left(\frac{m_d}{m_v} R_d + R_v\right)$$

where we have introduced the specific humidity

$$q = \frac{m_v}{m_v + m_d}$$

In practice, the mixture can be treated as a perfect gas with a constant R given by

$$R = (1 - q)\, R_d + R_v q = R_d\left[1 + q\left(\frac{1}{\epsilon} - 1\right)\right] = R_d\,(1 + 0.608q)$$

Finally, the equation of state can be written as

$$p\alpha = R_d\,(1 + 0.608q)\,T$$

or

$$p\alpha = R_d T_v$$

where

$$T_v = (1 + 0.608q)\,T \tag{E.1.9}$$

is the virtual temperature. Considering the values for $q$ that does not exceed values of the order of 0.02, the virtual temperature may differ by 2–3 °C from the usual temperature. This difference although small may have a large influence on convection.

### *E.1.8 Using Diagrams in Forecasting*

Suppose now that we have made a sounding whose results can be plotted on a skew T–ln P diagram like in Fig. E.1.3. This is a common emagram in which we have the isotherms that slope at right going upward, the dry adiabat sloping to the left and constant mixing ratio given by the almost straight lines. The sounding is the heavy line that delimitates the gray area. The other border of the area gives a particular moist adiabat. Starting from point A, the parcel will move along a dry adiabat up to intersect the moist adiabat that corresponds to its initial water vapor content (3.4 g/kg). The intersection point is the so-called lifting condensation level (LCL). During this ascent, the parcel temperature is lower than the environment so that the particle must be forced to rise. If this forcing continues, the parcel will reach the so-called *level of free convection* (LFC) because after that the temperature of the parcel will be higher than the environment, and the parcel will move freely. The parcel starts from 1000 hPa at about 287 K with a content of 10 g/kg of water vapor and moves along the dry adiabat at 299 K to intersect the moist adiabat with an equivalent potential temperature of 330 K. We may calculate the energy necessary to push the parcel from the initial point up to the lifting condensation level. Considering the density difference $\Delta\rho$ between the environment and the parcel, we get as energy per unit mass

$$\int_0^z \frac{\Delta\rho}{\rho} g dz = \int_{p_{surf}}^{p} \frac{\Delta T}{T}\frac{dp}{\rho} = R\int_{p_{surf}}^{p} \left(T' - T\right) d\ln p$$

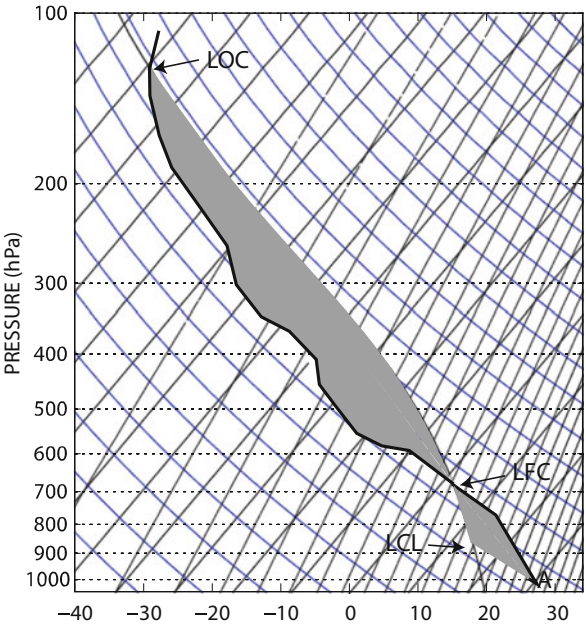



**Fig. E.1.3** A sounding plotted on the emagram. The heavy line is the sounding. The segment from A to LCL is a dry adiabat, while the curve from LFC to LOC is the moist adiabat

This energy is called convective inhibition energy (CINE). As a matter of fact, the temperature of the parcel is lower in the environment, and consequently the parcel must be pushed upward. Considering the example, the temperature difference between $p_{\text{surf}} = 1000$ hPa and $p = 700$ hPa is on the average 5 °C so that the integral has the value 500 J kg$^{-1}$. We can relate this energy to the velocity that the parcel should have initially using the simple relation

$$\text{CINE} = \int_{p_{\text{surf}}}^{p} \frac{dv}{dt} v dt = \frac{1}{2}\left(v^2 - v_{\text{surf}}^2\right)$$

Assuming $v = 0$, we get

$$v_{\text{surf}} = (2\text{CINE})^{1/2} \approx 33\,\text{ms}^{-1}$$

This is the rather large value for the initial value of the parcel velocity in order to reach the LCL. Between this level and the limit of convection (LOC), the parcel will accelerate because its temperature is higher than the surrounding. The corresponding energy is called convective available potential energy CAPE) and can be evaluated in the same way as CINE:

$$\text{CAPE} = -R\int_{\text{LCL}}^{\text{LOC}} \left(T' - T\right) d\ln p$$

CAPE is the energy that could be released during the convection. CINE and CAPE are very important numbers for the evaluation of the occurrence and intensity of thunderstorm. If CINE is positive, it means energy must be spent to raise the parcels (e.g., by orographic forcing). Once the parcels are beyond LCL, they can release CAPE energy. The other quantity that can be estimated from a sounding is the amount of precipitation using the decrease in water mixing ratio with altitude.

## References[1]

### *Books*

Bohren CF, Albrecht BA (1998) Atmospheric thermodynamics. Oxford
Goody R (1995) Principles of atmospheric physics and chemistry. Oxford
Hess SL (1979) Introduction to theoretical meteorology. Constable
North GR, Erukhimova TL (2009) Atmospheric thermodynamics: elementary physics and chemistry. Cambridge
Wallace JM, Hobbs PV (1977) Atmospheric science: an introductory survey. Academic
Tsonis AA. An introduction to atmospheric thermodynamics. Cambridge

### *Articles*

Broecker WS (1997) Mountain glaciers: recorders of atmospheric water vapor content. Global Biogeochem Cycles 11:589
Zhang R, Wooldridge PJ, Abbatt JPD, Molina MJ (1993) Physical chemistry of the $H_2SO_4/H_2O$ binary system at low temperatures: stratospheric implications. J Phys Chem 97:7351

[1]This chapter, as the following six, has been changed completely from the previous edition. It contains the basic elements of thermodynamics applied to the atmosphere. The textbook by Wallace and Hobbs is essential for all the introductory parts. More advanced texts are those of Tsonis, North and Erukhimova, and Bohren. Not surprisingly, a real treasure is the book by Hess, now very rare.

# Chapter 2
# Fundamentals: Radiation in the Atmosphere

We assume that the most elementary notions about blackbody radiation are already known. We will use those concepts to find, for example, the average temperature of a planet. Unfortunately, also in the case of radiation, in general physics courses, the rule is to go straight from Planck's law to the second quantization. In this way, it is simply ignored that Planck's law has so many applications. The average physics student confuses "radiative" and "radioactive." The problem of radiative transfer is completely neglected to the point that even the simplest notions about absorption are not mentioned.

At this point, however, we cannot proceed any further without a "baby step" in treating the radiative processes. A preliminary point is to examine the radiative field that is surrounding us due to solar radiation and planetary radiation, that is, the infrared radiation coming from the atmosphere and the Earth's surface.

## 2.1 The Definition of Radiometric Variables

Based on our experience, we know that students learn everything about the lens equation, but very few of them are able to evaluate how much energy impinges on the film (or chip) of their camera. Some basic definition is then necessary for the variables related to the energy flux of the electromagnetic waves. Nowadays, students can quite easily visualize a radiation beam, for example, a laser beam. In the old days, the nearest thing to a radiation beam was a light of an electric torch. Both examples have their advantages. A laser beam is characterized by a direction of propagation, but the divergence is so small that one is not aware of the beam divergence. A light torch on the other hand makes clear that radiation is emitted within a finite solid angle. If we refer to Fig. 2.1, we see that the *intensity of radiation* is a vector whose direction and orientation are determined by the propagation, with the magnitude given by



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_2

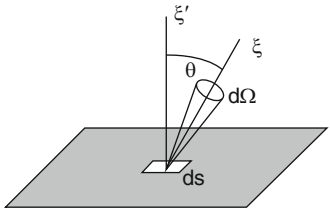



**Fig. 2.1** The elements to define the intensity of radiation

$$I_\nu = \frac{\text{power}}{\text{unit of surface, unit solid angle, unit of frequency}} \tag{2.1}$$

The intensity is then measured in W $m^{-2}$ $sr^{-1}$. The flux of radiation in the $\xi'$ direction is given by

$$F_\nu\left(\xi'\right) = \int I_\nu\left(\xi\right)\cos\ \theta d\Omega \tag{2.2}$$

The flux calculation requires projection of the radiation intensity on the direction $\xi'$ and integration over the entire solid angle. The solid angle is justified by the torch example, but it is also obvious when we consider a receiving system or a detector. In this case, the solid angle corresponds to the acceptance angle (again, think about your camera).

A classic example of radiation intensity is given by the Planck function, that is, the power emitted per unit area, unit frequency, and unit of solid angle by a blackbody at temperature $T$:

$$B_\nu(T) = \frac{2h\nu^3}{c^2}\left(e^{h\nu/kT} - 1\right)^{-1} \tag{2.3}$$

where $h$ is the Planck constant, $c$ the speed of light, and $k$ the Boltzmann constant. Equation (2.3) can also be expressed as power per unit surface, unit frequency, and unit of wavelength:

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5}\left(e^{hc/\lambda kT} - 1\right) \tag{2.3a}$$

The blackbody radiation is isotropic so that if we consider a surface at temperature $T$, we can easily evaluate the flux normal to the surface. $I_\nu(\xi)$, now, is constant and is the Planck function $B_\nu$ so that the integral of Eq. (2.2) reduces to

$$F_\nu = 2\pi B_\nu \int_0^{\pi/2} \sin\theta\cos\theta\ d\theta = \pi B_\nu \tag{2.4}$$

If this relation is integrated across the frequency, we get $B = \sigma T^4/\pi$. The intensity of radiation sometimes is indicated as *radiance*.

We have defined all the relevant quantities about radiation in connection with Planck's law simply because it is much easier.

## 2.2 The Solar Radiation

The energy that the Earth and the planets receive from the sun determines the physical characteristics of the planetary atmospheres together with the emitted planetary (infrared) radiation. We have seen in Chap. 1 that the power per unit area at the top of the atmosphere of the Earth, that is, the solar constant, has a value around 1380 W $m^{-2}$. The value of this constant is measured today directly outside the atmosphere from orbiting satellites.

The sun is a star that has a radius of 695,000 km and a surface temperature of about 5770 K, while the temperature of the nucleus reaches probably 15 million degrees. The sun can be considered with good approximation as a blackbody, with a maximum emission occurring at about 0.5 μm (see exercises). This can be seen from Fig. 2.2 which compares the solar spectrum taken outside the atmosphere with a spectrum taken at the surface at solar zenith angle of 60°. We notice that even in the most external part of the sun, there are absorption phenomena

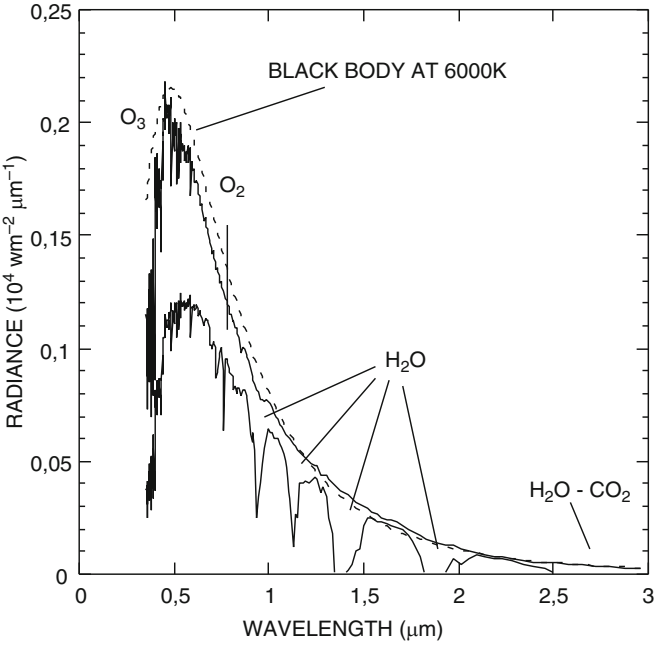



**Fig. 2.2** The spectrum of the solar radiation outside and below the atmosphere of the Earth. The spectrum is referred to a 60° zenith angle. The gases responsible for absorption are indicated. The *dotted line* is a blackbody emission at 6000 K

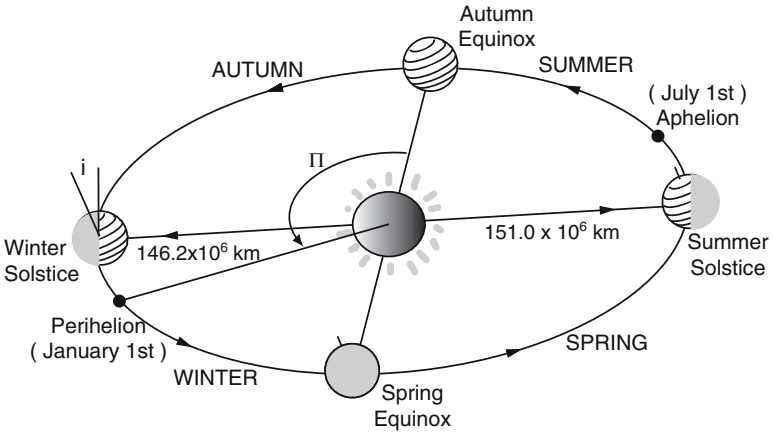



**Fig. 2.3** The geometry of the revolution of the Earth around the sun during the seasons. At the solstices the rotation axis is contained in the plane normal to the ecliptic passing through the Earth and the sun. At the equinoxes the rotation axis and the normal to ecliptic are contained in a plane perpendicular to the sun–Earth line. In the figure, the sun–Earth distance at the aphelion and perihelion is also reported, as is the length of the perihelion, $\Pi$

that change considerably the shape of the emission. The Earth's atmosphere also reduces the power reaching the surface due to the absorption by ozone and other atmospheric gases that are indicated in the figure. Particularly important is the ozone absorption that blocks the radiation below 300 nm. We will see later that atmospheric absorption is responsible for the attenuation of about 15 % of the power reaching the surface. Besides the actual power that reaches the surface, we need to take into account how this power is distributed during the year, as a function of latitude. It is something we all learn from physical geography textbooks most of which contain a figure very similar to Fig. 2.3, where the Earth is shown in its orbit around the sun. As we have seen before, the rotation axis of the Earth makes an angle of about 23° with the normal to the plane of the ecliptic. This implies that the different regions of the Earth are exposed differently to the solar radiation.

At the winter solstice, the rotation axis at the north pole points in a direction away from the sun and during the winter the regions of the polar arctic circle are always in the dark. The winter solstice happens between December 22 and 23, and the Earth reaches the minimum distance from the sun (perihelion) about 10 days later. The opposite happens at the summer solstice (June 21–22) when the rotation axis points toward the sun: in this case, the high latitude regions are always illuminated. At the equinox, the rotation axis and the normal to the ecliptic are in a plane perpendicular to the sun–Earth line. At that moment, there is no longer asymmetry between the two hemispheres, and the night has exactly the same duration as the day (March 20–21, September 22–23). The modulation introduced in the power received at the different latitudes causes the seasons, which are symmetrical in the two hemispheres. The length of the perihelion shown in the figure is the angle between the autumn equinox and the winter perihelion. It is to note that the Earth is nearer to the sun in the winter

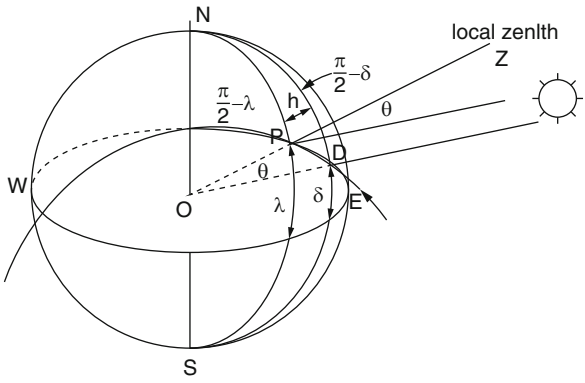



**Fig. 2.4** The geometry of solar illumination on the Earth. The thick line indicates the path of the sun on the Earth. $\delta$ is the declination with respect the equatorial plane. $\phi$ is the latitude, h the hour angle, and $\theta$ the zenith angle (Adapted from Sellers 1965)

than in the summer. However, this does not really influence the heat received in the season because the dominating effect is the average height of the sun above the horizon. We will talk at length about this when treating the theory of the ice ages.

To arrive at a more quantitative treatment of distribution of the solar radiation, we can refer to the scheme given in Fig. 2.4. To understand this figure, the observer must imagine himself to be at the point P at latitude $\lambda$. In this case, the local vertical is given by the direction $OZ$, and then the zenith angle of the sun (the angular distance between the vertical and the direction of the sun) is $\theta$.

As we have seen already, the incident flux is given by

$$F = F_0 \cos\theta \tag{2.5}$$

where with $F_0$ we have indicated the instantaneous flux when the earth is at distance $d$ from the sun, which is at an average distance $d_m$. Considering that the flux goes as the inverse of the square of the distance, we can write

$$F = S_0 \left(\frac{d_m}{d}\right)^2 \cos\theta \tag{2.6}$$

What is of interest to us is the average value of the flux $\langle F \rangle$ in some region at latitude $\theta$, so that

$$\langle F \rangle = S_0 \left(\frac{d_m}{d}\right)^2 \int_{t\ \text{sunrise}}^{t\ \text{sunset}} F(t)dt/ \quad \text{length} \quad \text{of the day} \tag{2.7}$$

which represents the average value of the insulation during the daytime. From Fig. 2.4, we should find a relationship between the solar zenith angle and the hour angle $h$, that is, the angle at which the Earth should rotate so that the meridian at $P$ is just below the sun. From the spherical triangle $PDN$, we have

$$\cos\theta = \operatorname{sen}\phi\ \operatorname{sen}\delta + \cos\phi\ \cos\delta\ \cos\ h$$

where $\delta$ is the solar declination, that is, the angular distance of the sun with respect to the equatorial plane, and $\varphi$ is the latitude.

At this point, it is simple to find a relationship between the hour angle and the time because if $\Omega$ is the angular velocity of the Earth, then $dh = \Omega dt$ and substituting in Eq. (2.7) we obtain

$$\langle F \rangle = \frac{S_0}{2\pi}\left(\frac{d_m}{d}\right)^2 \int_{-H}^{H} (\text{sen}\ \phi\ \text{sen}\ \delta + \cos\ \phi\ \cos\ \delta \cos\ h)\, dh \tag{2.8}$$

where $H$ indicates hour angle for sunrise and sunset for which the solar zenith is $\pi/2$. Equation (2.8) can be easily integrated to give

$$\langle F \rangle = \frac{S_0}{\pi}\left(\frac{d_m}{d}\right)^2 (H\ \text{sen}\ \phi\ \text{sen}\ \delta + \cos\ \phi\ \cos\ \delta\ \text{sen}\ H) \tag{2.9}$$

where $H$ is a function of latitude and is obtained with $\theta = 0$ in the equation that gives the zenith angle

$$\cos\ H = -\tan\ \phi\ \tan\ \delta \tag{2.10}$$

that must be substituted in the (2.9) for each latitude. With this relation, it is possible to evaluate the insulation outside the atmosphere of the Earth, as it is shown in Fig. 2.5. Actually, this figure shows the W m$^{-2}$ accumulated in 1 day. We notice an almost perfect symmetry with respect to the summer solstice (SS). The symmetry

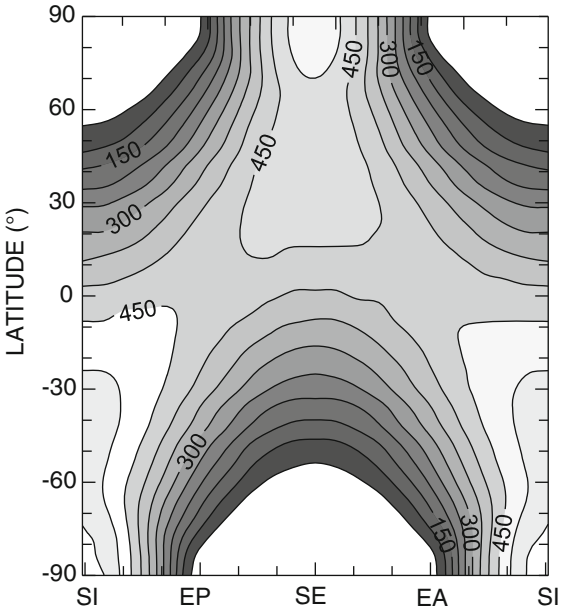



**Fig. 2.5** The average power per unit surface (in W m$^{-2}$) incident at the top of the Earth's atmosphere as a function of the season and latitude. WS, SE, SS, and AE are in order the winter solstice, the summer equinox, the summer solstice, and the autumn equinox. The average is carried out on 1 day. The *dark* zones correspond to the polar night

is not complete, however, with respect to the equinoxes, because in the winter the Earth is nearer to the sun. This fact implies a slight asymmetry especially for the highest values.

It is quite surprising that the regions with maximum in the insulation are those at high latitude in the summer. This happens because the long period of insulation compensates for the very high zenith angle of the incident radiation. The darkest zones are those not illuminated during the polar night. It should be noted, however, that the power indicated refers to the radiation incident at the top of the atmosphere. The radiation reaching the surface is quite different from that shown in Fig. 2.5. This will depend on the distribution of the cloudiness and the surface albedo, both functions of the season and the latitude. Actually, the use of satellite data has made it possible to obtain average values of the albedo and its distribution. For the clouds, only recently have statistics been developed that give indications on their distribution or probability. However, by their intrinsic nature, clouds are a very complex subject.

Let us return for a moment to the practical evaluation of the hour angle and the declination of the sun. A good approximation is to refer to a circular orbit of the Earth around the sun. In this case the declination can be written as

$$\delta = 23.45^\circ \cos\left[\frac{2\pi\ (d - d_r)}{365}\right]$$

where $d$ is the day of the year and $d_\mathrm{r}$ is the day of the summer solstice. For example, for March 3 and non-leap years, $d = 31 + 28 + 3 = 62$ and $d_\mathrm{r} = 173$ because the summer solstice is on June 22. In this case, the declination is simply 12.54°. If for the same day we want to calculate the solar zenith angle at 3 in the afternoon at a latitude of 43° north, we need to evaluate the hour angle. This is defined as a function of the *Coordinated Universal Time* (UTC) or GMT, the Greenwich Mean Time. If we know the longitude of the place $\lambda_e$, the hour angle is given by

$$h = 2\pi\left(\frac{UTC - 12}{t_d}\right) - \lambda_e.$$

In this case, $t_d$ is the length of the day (24 h), and the longitude is positive if west and negative if east. If we take my place in Italy, situated at 13°E, then the simple formula above gives for $h$ 43°. Substituting in the expression for the zenith angle, we get $\theta = 47.62^\circ$.

## 2.3 Scattering and Absorption of Solar Radiation

As we have already seen, the solar radiation entering the atmosphere of the Earth is absorbed by its gaseous components. At relatively low altitudes, another important process is the diffusion. In this first part, we will sketch the processes of diffusion

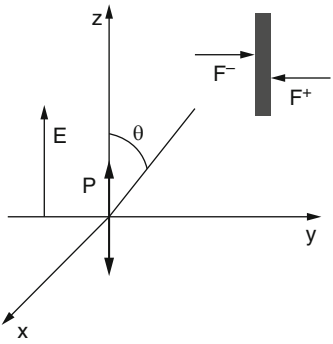



**Fig. 2.6** The electric field $E$ of an electromagnetic wave linearly polarized incident on a charge $e$ at the origin of the system of axis. $\theta$ is the direction considered for the emission of radiation. A layer with diffuse fluxes is also shown

(scattering) and absorption only due to the gases. The same processes can take place in the presence of particulates (i.e., small particles or aerosol). However, their complexity is such to suggest postponing their treatment for a few chapters down the line. As usual, we will start from things that should be known from general physics: we will only apply them to some specific problem.

### *2.3.1 Rayleigh Scattering*

The simplest way to deal with Rayleigh scattering is to refer to the model given in Fig. 2.6. In it, the electric field of a linearly polarized electromagnetic wave is incident on a bound electron. Under the action of the varying electric field, the charge starts to oscillate around the origin along the field direction. If the incident field is written in the form $E = E_0 \text{ sen } \omega t$, the equation of motion for the charge is

$$\frac{d^2z}{dt^2} + \omega_0^2 z = \frac{e\,E_0}{m}\text{sen } \omega t \tag{2.11}$$

where $\omega_0$ is a typical frequency of oscillation. The solution for Eq. (2.11) is

$$z = \frac{e\,E_0 \text{ sen } \omega t}{m\left(\omega_0^2 - \omega^2\right)} \tag{2.12}$$

and the dipole moment is given by (see for example Bekefi and Barrett 1977)

$$p = ez = \frac{e^2\,E_0 \text{ sen } \omega t}{m\left(\omega_0^2 - \omega^2\right)} \tag{2.13}$$

If we use the formula that gives the energy radiated by an oscillating dipole, we get for the electric field

$$E = \frac{\omega^2 p_0\, sen\,\theta}{4\pi\varepsilon_0 c^2 r} sen\ \omega\,(t - r/c) \tag{2.14}$$

where $r$ is the distance from the center of the dipole, $p_0 = e^2 E_0/m\left(\omega_0^2 - \omega^2\right)$, and the other symbols have the usual meanings. From the definition of energy flux (Poynting vector), we have, after averaging over the time,

$$S = \frac{\omega^4 p_0^2 \,\mathrm{sen}^2\,\theta}{32\,\pi^2 \varepsilon_0 c^3 r^2} \tag{2.15}$$

and substituting for the value of $p_0$, we have

$$S = \frac{\omega^4 \mathrm{sen}^2\,\theta}{32\,\pi^2 \varepsilon_0 c^3 r^2}\left(\frac{e^2 E_0}{m\left(\omega_0^2 - \omega^2\right)}\right)^2 \tag{2.16}$$

This relation shows a number of things. First of all because $\omega_0 \gg \omega$, the scattered power is actually proportional to the fourth power of the frequency, that is, to the inverse of the fourth power of the wavelength. This means that the long wavelengths (red) are scattered less than the short ones (blue); this may explain the blue color of the sky. Besides, for a polarized wave, the scattered radiation is not isotropic and is even zero in the dipole direction and has a maximum in the direction of incident radiation. At this point, we could find the irradiated power for unit of solid angle, $dP/d\Omega$, noting that $S = dP/\mathrm{d}A = dP/(r^2 d\Omega)$ so that

$$\frac{dP}{d\Omega} = \frac{\omega^4 \mathrm{sen}^2\,\theta}{32\,\pi^2 \varepsilon_0 c^3}\left(\frac{e^2\,E_0}{m\left(\omega_0^2 - \omega^2\right)}\right)^2 \tag{2.17}$$

To use this equation, it is necessary to put in relation microscopic quantities appearing in Eq. (2.17) with macroscopic quantities like the index of refraction. In practice, it is possible to show (see, e.g., Liou) that

$$\frac{e^2}{m\left(\omega_0^2 - \omega^2\right)} = \frac{3\varepsilon_0}{N}\frac{n^2 - 1}{n^2 + 2} \tag{2.18}$$

where $n$ is the index of refraction and $N$ the density of the dipoles (i.e., of the molecules). Substituting in (2.17), we get

$$\frac{dP}{d\Omega} = \frac{9\pi^2 \varepsilon_0 c\,\sin^2\,\theta}{2N^2\lambda^4}\left(\frac{n^2 - 1}{n^2 + 2}\right)^2 E_0^2 \tag{2.19}$$

where we have explicated the dependence over the wavelength. The equation we have obtained is valid for a very particular situation because it gives the power scattered from a polarized wave. The natural light is the superposition of a wave

polarized in random directions so that we could find a cross section for this process that should then be generalized. The Poynting vector $S = c\varepsilon_0 E_0^2/2$ is related to the cross section $\sigma_m$ by the equation

$$\frac{dP}{d\Omega} = \sigma_m(\theta)\, S \tag{2.20}$$

This shows how the cross section is the fraction of the flux subtracted from the incident radiation and irradiated in the $\theta$ direction for unit solid angle. From the definition of the Poynting vector, we find immediately

$$\sigma_m(\theta) = \frac{9\pi^2}{N^2\lambda^4}\left(\frac{n^2-1}{n^2+2}\right)^2 \text{sen}^2\theta \tag{2.21}$$

From this cross section, it is possible to obtain the total one integrating over the solid angle, that is,

$$\sigma_m = \int_0^{4\pi} \sigma_m(\theta)\, d\Omega = 2\pi \int_0^{\pi} \frac{9\pi^2}{N^2\lambda^4}\left(\frac{n^2-1}{n^2+2}\right)^2 sen^3\theta d\theta$$

which gives

$$\sigma_m = 8\pi^3\left(n^2-1\right)^2/3N^2\lambda^4 \tag{2.22}$$

where we have used the approximation that $n^2+2 \approx 3$. The cross section defined in this way refers to a single polarization direction and to a single molecule and so must be measured in $m^2$. More often, it is preferred to use a volume scattering coefficient that is obtained by multiplying Eq. (2.22) by $N$. The value of the refraction index of air for standard conditions of pressure and temperature depends weakly on the wavelength and ranges from 1.00029668 at 270 nm up to 1.00027269 a 4000 nm. The corresponding value of the Rayleigh cross section at 270 nm is $8.960°10^{-30}$ $m^2$ which can be easily scaled with the wavelength. When we consider natural light, the cross section does not change, but it changes the angular dependence of the diffuse radiation, and Eq. (2.21) must be substituted by (see Examples)

$$\sigma_m(\theta) = \frac{\pi^2\left(n^2-1\right)^2}{2N^2\lambda^4}\left(\frac{6+3\delta}{6-7\delta}\right)\left(1+\cos^2\theta\right) \tag{2.23}$$

In this equation, the quantity appearing in parentheses is the depolarization factor $(6+3\delta)/(6+7\delta) = 1.06$. This factor arises from the fact that the orientation of the molecular dipoles is not necessarily the same as the incident field. Equation (2.23) shows that diffuse radiation is emitted in the case of natural light in all directions. We can introduce a *phase function* as the ratio of the energy scattered in the direction $\theta$ for a unit of solid angle to the average radiation scattered over the

entire solid angle. This definition implies that the phase function be normalized to unity and in the particular case of Eq. (2.23) is given by

$$\int_0^{4\pi} P(\theta)\, d\Omega = 1 \Rightarrow P(\theta) = \frac{3\left(1+\cos^2\theta\right)}{4} \tag{2.24}$$

The Rayleigh scattering simply redistributes the radiation and does not imply any absorption. The redistribution however gives rise to an attenuation of the radiation along the propagation direction. It is possible to establish a relation between the loss of intensity $dI$ of the radiation passing through a layer of thickness $dz$

$$dI = -\sigma_m N I\, dz \tag{2.25}$$

Integrating this equation, we have the result

$$I = I_0 \exp\left(-\int_0^{\infty} \sigma_m N dz\right) = I_0 e^{-\tau} \tag{2.26}$$

The quantity $\tau = \int \sigma_m N dz$ is called the optical thickness or optical depth and for the atmosphere at a wavelength of 500 nm (in the visible) has a value of about 0.1.

Now we can ask what happens to the solar flux in the presence of Rayleigh scattering. If we look at the phase function of Eq. (2.24), we can see that most of the flux is scattered either in the forward or in the backward direction. Essentially, we may assume that the diffuse flux is reduced to a stream traveling in the forward direction and another in the opposite direction. If we refer to Fig. 2.6, we can indicate as $F^-$ the flux in the positive $y$ direction and $F^+$ the flux in the opposite direction. To be closer to reality, we will assume that the optical thickness increases with increasing $y$. After traversing a layer of thickness $d\tau$, the flux $F^-$ will be decreased by the amount $F^-/2$ and increased by $F^+/2$. The first is due to "absorption," while the second is due to scattering of the flux traveling in the negative $y$ direction (following Stephens 2004):

$$\frac{dF^-}{d\tau} = 0.5\left(F^+ - F^-\right) \tag{2.27a}$$

For the other flux, we will have exactly the same expression

$$\frac{dF^+}{d\tau} = 0.5\left(F^+ - F^-\right) \tag{2.27b}$$

One very simple reason is that the net flux $F^+ - F^-$ must not change across $d\tau$ because this is conservative scattering and energy must not be lost. Then if we sum and subtract Eqs. (2.27a and 2.27b), we get

$$\frac{d\left(F^{+}-F^{-}\right)}{d\tau}=0 \tag{2.28a}$$

$$\frac{d\left(F^{+}+F^{-}\right)}{d\tau}=\left(F^{+}-F^{-}\right) \tag{2.28b}$$

The solution to this system is very easy to find, with the fluxes given by

$$\begin{aligned} F^{+} &= D + C\,(1+\tau) \\ F^{-} &= D - C\,(1-\tau) \end{aligned} \tag{2.29}$$

The integration constants $C$ and $D$ can be found from the boundary conditions. For simplicity, we will assume that the incident flux $F_0$ coincides with $F^{-}(0)$. We will also make the case that the medium has a total optical thickness $\tau^{*}$ (e.g., at the origin), and the reflectivity at this point is zero so that $F^{+}(\tau^{*})=0$. From these two boundary conditions, we get for the constants

$$C=\frac{-F_0}{(2+\tau^{*})} \qquad D=\frac{F_0\,(1+\tau^{*})}{(2+\tau^{*})}$$

An interesting quantity is the reflectivity, defined as $R=F^{+}(0)/F_0$ and which can be easily found to be

$$R=\frac{\tau^{*}}{(2+\tau^{*})} \tag{2.30}$$

We see that the reflectivity increases with increasing optical thickness. However, a much more interesting property is to derive Eq. (2.30) with respect to wavelength. The result is that for large values of optical thickness, the reflectivity is independent of the wavelength, that is, the medium looks white.

Finally, we have found that Rayleigh scattering makes the sky look blue but also makes the sky look white when we look at the horizon because then the optical thickness is large. A very easy extension of these concepts may be made for the clouds, but to be correct we have to wait a few more chapters.

### *2.3.2 The Absorption of Solar Radiation*

In this first approach to absorption, we will take into account only the gases in the atmosphere and will not consider the absorption due to particles.

Absorption takes place every time radiation impinging on a molecule or an atom of a gas composing the atmosphere causes a transition, which implies a loss of energy. We will look more closely at absorption when dealing with photodissociation. For the moment it is enough to say that absorption takes place

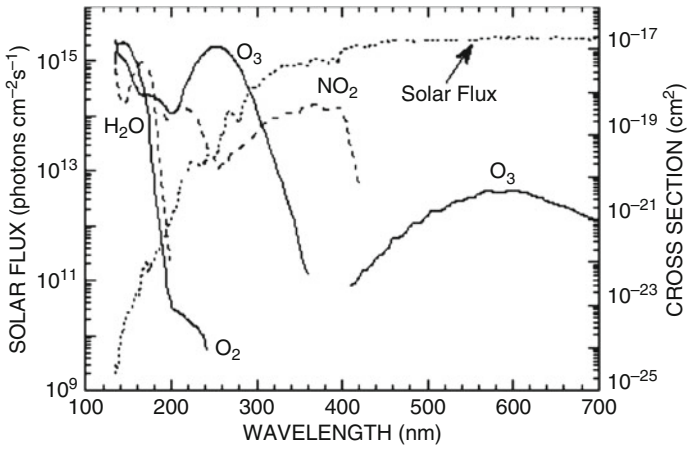



**Fig. 2.7** The flux of solar radiation in units of photons cm$^{-2}$ s$^{-1}$ for an interval of 5 nm, and the absorption cross section in cm$^2$ for a few gases in the atmosphere of the Earth

for wavelengths below 1000 nm. This spectral region can be divided roughly in the interval of the extreme ultra violet (EUV) up to 120 nm, in the range between 120 and 300 nm, and finally from 300 to 1000 nm. The EUV is absorbed in the high atmosphere above 100 km from gases like $O_2$ and $N_2$ that have absorption bands in this spectral region. The intermediate wavelengths are absorbed in the mesosphere and stratosphere, and the longest are absorbed in the lower stratosphere and troposphere. In Fig. 2.7, we show the most important absorption cross sections in the spectral region between 130 and 700 nm. It can be noted that in the ultraviolet, the most important absorbing gases are molecular oxygen, water vapor, and ozone. Water vapor however has important absorption bands in the visible region and in the near infrared that are not shown in this figure. Oxygen has an absorption region corresponding to the so-called Schumann–Runge bands, between 175 and 200 nm, that have a very complex structure and are not reported here.

The absorption by a gas can be easily evaluated when the cross section is known. If at wavelength $\lambda$ the cross section of $i$-species is $\sigma_i(\lambda)$, then the optical thickness from the top of the atmosphere to the height $z$ is given by

$$\tau_\lambda(z) = \sum_i \sigma_i(\lambda) N_i(z) \tag{2.31}$$

where $N_i(z)$ is the so-called columnar density of the $i$-species, measured in molecules or atoms per unit surface

$$N_i(z) = \int_z^\infty n_i(z) dz \tag{2.32}$$

where $n_i(z)$ is the number or atoms per unit volume. To understand better, we can make a simple example. In the actual practice, the integral of Eq. (2.32) should be calculated numerically because the density cannot be expressed analytically. We can refer to the case of molecular oxygen, which has a constant volume mixing ratio of about 0.21. Using Eq. (2.32) for oxygen, we find

$$n_i(z) = f_i n_m = \frac{f_i p(z)}{kT} = \frac{f_i \rho(z) R}{k}$$

and as a consequence the integral

$$N_i(z) = \frac{f_i R}{k} \int_z^\infty \rho(z) dz = -\frac{f_i R}{kg} \int_z^\infty dp = \frac{f_i R}{kg} p \tag{2.33}$$

where $R$ is the gas constant, $k$ is the Boltzmann constant, and $f_i = 0.2096$ is the mixing ratio for oxygen. This very elementary exercise is also useful to refresh simple relationships between moles, Avogadro number, and so on.

When the optical thickness is known, we can proceed to the calculation of the solar flux as a function of the altitude. We can use what we learned with Rayleigh scattering and the attenuation of the solar flux in that case.

We introduce here a notation that is common in radiative transfer theory. If the solar zenith angle is $\theta$, then a more convenient quantity is $\mu_0 = \cos\theta$, and the monochromatic flux at height $z$ will be

$$F_\lambda(z) = F_\lambda(\infty) \exp\left(\frac{-\tau_\lambda(z)}{\mu_0}\right) \tag{2.34}$$

where $F_\lambda(\infty)$ is the monochromatic flux at the top of the atmosphere. This flux is expressed in W m$^{-2}$ over a specified range of wavelengths and is represented in Fig. 2.8 in slightly different units. To convert the photons into watts, it is necessary to multiply by the energy of a single photon $hv$. From the figure, we see that the UV flux is rapidly absorbed in the upper atmosphere first by oxygen and then by ozone. The absorption of the flux implies that some energy is deposited in the atmosphere. To evaluate this energy, we simply note that for the $j$-species with cross section $\sigma_j(\lambda)$, the power absorbed by a single molecule is $\sigma_i(\lambda) F_\lambda(z)$, so that the total absorption

$$A = \sum_\lambda n_i(z) \sigma_i(\lambda) F_\lambda(z) \tag{2.35}$$

This relation, however, except for the sign, is simply the derivative with respect to altitude of Eq. (2.34). The sign comes about because to get the absorbed flux, we need to subtract Eq. (2.34) from $F_\lambda(\infty)$ so that the absorption is correctly given by

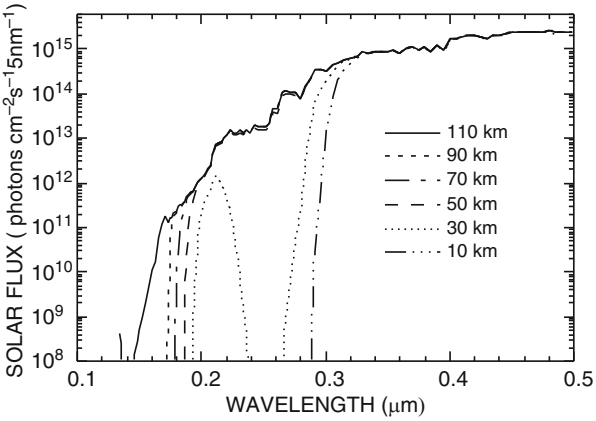



**Fig. 2.8** The solar flux, in the same units as in Fig. 2.7, as a function of wavelength, at different heights. Notice the rapid attenuation due to oxygen absorption and the cut below 300 nm due to ozone absorption

$$F(z) = \sum_{\lambda} F_{\lambda}(\infty)\left[1 - \exp\left(\frac{-\tau_{\lambda}}{\mu_0}\right)\right] \tag{2.36}$$

and the absorbed flux increases with decreasing altitude. The power absorbed per unit volume will then be given by the derivative with respect to height of Eq. (2.36). If we know the absorbed power per unit volume, it is possible to evaluate the temperature change in the atmospheric layer of $dz$ simply by noting that the heat capacity per unit volume is $\rho C_p$. The rate of change of temperature in the layer will be

$$\frac{\partial T}{\partial t} = -\frac{1}{\rho C_p}\frac{\partial F}{\partial z} \tag{2.37}$$

This equation expresses the conservation of energy because it states that the rate of change with altitude of the power density is responsible for heating or cooling the layer. The importance of the heat capacity in determining the heating in a given volume can be appreciated from Fig. 2.9. The left part of the figure shows the power absorbed per unit volume by ozone and molecular oxygen, while the right part shows the heating rate as a function of altitude. It is quite clear that where the heat capacity is small (in the upper atmosphere), the heating rate due to oxygen reaches several tens of degrees per day even if the absorbed power is only a few erg cm$^{-2.}$ It is interesting to compare this figure with Fig. 1.1 for the Earth. We can see that we are on the right track to explain the thermal structure of the middle atmosphere. We still need to know what is the process that balances heating produced by the absorption of solar radiation, and we may suspect that it could be the long wavelength or infrared emission.

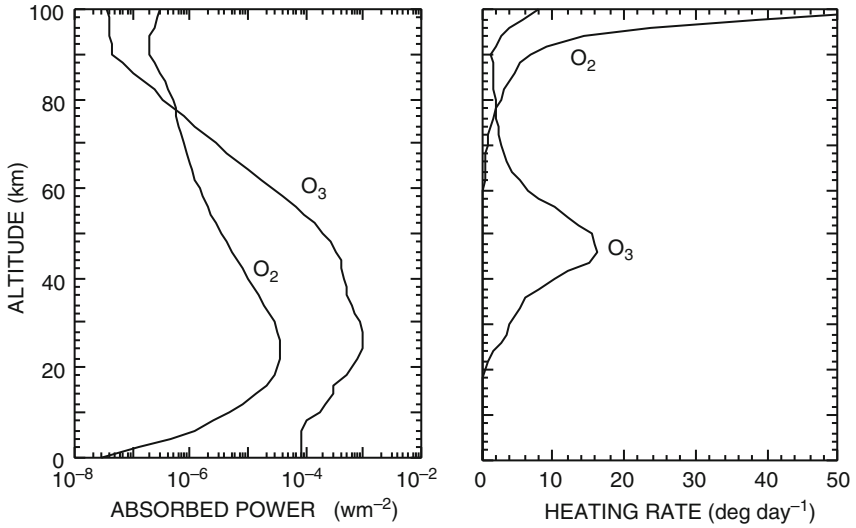



**Fig. 2.9** The absorbed power per unit volume (*left*) and the heating rate (*right*) as a function of altitude. The increasing specific heat as the height decreases makes oxygen mainly responsible for heating in the thermosphere and ozone responsible in the stratosphere

## 2.4 Infrared Radiation

For infrared radiation, as a first approximation, the atmosphere can be considered as a blackbody emitting according to Planck's law. On the other hand, we have seen that the Earth–atmosphere system is not exactly a blackbody because its emissivity is slightly less than one.

When for the first time a meteorological satellite measured the spectrum of the infrared radiation emitted by the Earth, we had almost didactic proof that the Earth–atmosphere system is not a blackbody. Figure 2.10 shows one of these spectra measured at middle latitudes. Actually, the function shown is not exactly that of Planck, rather the quantity $\lambda B_\lambda$. The reason is that this function is proportional to the fraction of the total emitted power up to a certain wavelength. In the figure the spectra are shown as a function of both the wavelength and the wave number, that is, the inverse of the wavelength expressed in cm. If we compare the measured spectra with that of a blackbody at 300 K, we notice that the former has a series of "holes" due to the gas absorption. It is interesting to note that the base of the absorption features corresponds to a temperature that is roughly the same as the effective level of the emission. We notice, at this point, that this level is higher for the ozone bands (temperatures characteristic of the stratosphere) than for the carbon dioxide or water vapor bands (temperatures of the tropopause). We can observe a spectral region free of absorption features between 8 and 12 μm that is called the *atmospheric window*. The most interesting thing about this figure is however a "visualization" of the greenhouse effect. The area of the spectrum represents the total power per unit

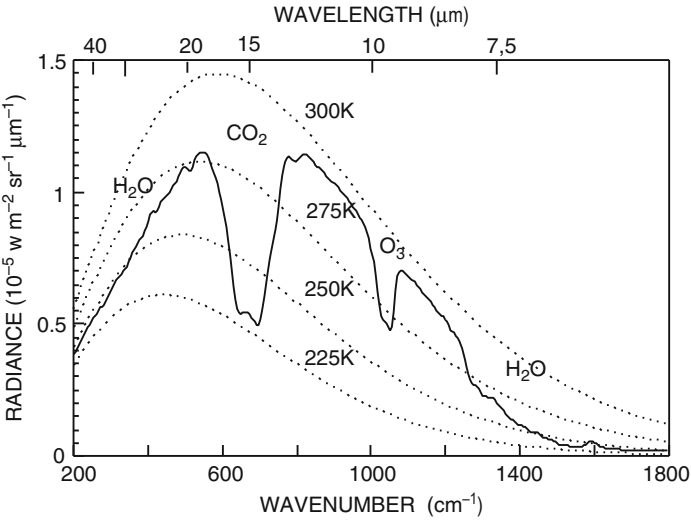



**Fig. 2.10** The function $\lambda B_\lambda$, measured for the Earth (*solid line*) and calculated for a blackbody at different temperatures (*dotted lines*). The gases responsible for the main absorption features are indicated

surface emitted by the Earth. If the concentration of one of the gases responsible for the absorption increases, the area of the corresponding hole also increases, and the total emitted power decreases. In order to maintain equilibrium between power absorbed from the sun and power emitted from the planet, the surface temperature must increase. This obviously is a qualitative explanation for the greenhouse effect, and we will show how this problem may be treated at different levels of complexity.

### *2.4.1 The Equation of Radiative Transfer*

If we consider a medium (e.g., a gas of density $\rho$) traversed by radiation in the $z$ direction, then in a layer of thickness $dz$, the radiation absorbed $dI$ will be given by

$$dI = -Ik\rho dz \tag{2.38}$$

where $k$ is the absorption coefficient that is characteristic of a specific gas. The integration of Eq. (2.38) gives the familiar Beer–Lambert law, $I = I_0 \exp(-\tau/\mu_0)$, already seen when writing Eq. (2.26), where $I_0$ is the intensity of radiation for $z = 0$, while the quantity $d\tau = -\rho k dz$ is again the optical thickness or optical depth that we have already seen for the solar radiation. However, in the case of thermal radiation, beside absorption we need to consider also the emission, which can be characterized by $B$ (the Planck function) if we assume thermal equilibrium. In the layer of thickness $dz$ we have, combining emission and absorption,

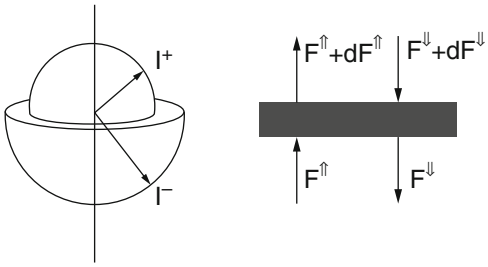



**Fig. 2.11** The two-stream approximation assumes that the intensities $I^+$ and $I^-$ are different in two hemispheres but isotropic in each of them. On the *right* the flux changes within an atmospheric layer are shown

$$dI = -Ik\rho dz + Bk\rho dz$$

$$\frac{dI}{d\tau} = B - I \tag{2.39}$$

It is to note that in Eq. (2.39), we have specialized the definition for $\tau$ in the sense that we have established that the optical thickness increases with decreasing altitude $z$, so that $d\tau = -\rho k dz$. In general, the intensity depends, other than $\tau$, also on the zenith angle $\theta$ and the azimuthal angle $\phi$. A drastic simplification in the resolution of Eq. (2.39) is obtained by assuming that the intensity is only a function of the zenith angle. In this case, if we set $\mu = \cos\theta$, we can have a more general expression for Eq. (2.39), referred to a direction $s$, which makes an angle $\theta$ with the $z$ direction

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - B \tag{2.39a}$$

This represents a simplified form of the equation of the radiative transfer. This equation can be written for the two hemisphere separated by the horizontal plane at altitude $z$, as is shown in Fig. 2.11 where with $I^+$ and $I^-$, we have indicated $I(\tau, \mu)$ and $I(\tau, -\mu)$, respectively. At this point, Eq. (2.39a) can be written for the two hemispheres (see for example Houghton 1977)

$$\mu \frac{dI(\tau, +\mu)}{d\tau} = I(\tau, \mu) - B$$

$$-\mu \frac{dI(\tau, -\mu)}{d\tau} = I(\tau, -\mu) - B \tag{2.40a,b}$$

The flux in the two hemispheres will be reminiscent of the definition of flux Eq. (2.4), $F^{\uparrow} = \pi I^+$ and $F^{\downarrow} = \pi I^-$, in the upper and lower hemispheres, respectively. The two Eq. (2.40a,b) can be written as a function of the fluxes multiplying through by $2\pi\mu$ and integrating with respect to $\mu$ between 0 and 1 so that we obtain

$$(2/3)\frac{dF^{\uparrow}}{d\tau} = F^{\uparrow} - \pi B$$

$$-(2/3)\frac{dF^{\downarrow}}{d\tau} = F^{\downarrow} - \pi B \tag{2.41a,b}$$

The solution to this system of equations can be found very easily in the approximation of *radiative equilibrium*. In practice, as is outlined on the left of Fig. 2.11, the net flux absorbed by the layer is given by $dF^{\downarrow} - dF^{\uparrow}$, so that the divergence is given by $d\left(F^{\downarrow} - F^{\uparrow}\right)/dz$. We have seen previously that this corresponds to a power density, which is the absorbed power per unit volume. In the layer we will then have a heating rate $dT/dt$

$$\frac{dT}{dt} = -\frac{1}{\rho C_p}\frac{d}{dz}\left(F^{\uparrow} - F^{\downarrow}\right) \tag{2.42}$$

where $\rho$ denotes the atmospheric density. The negative sign can be justified by the fact that we have heating when the net absorbed flux decreases with altitude and vice versa. When the temperature of the layer does not change with time, then the net flux is constant with altitude or with the optical thickness. This corresponds to the approximation of radiative equilibrium and can be used to solve Eq. (2.41). If we put for simplicity $\tau = 3\tau/2$ and sum Eqs. (2.41a) and (2.41b)

$$\frac{d}{d\tau}\left(F^{\uparrow} - F^{\downarrow}\right) = \left(F^{\uparrow} + F^{\downarrow}\right) - 2\pi B = 0; \quad B = \frac{1}{2\pi}\left(F^{\uparrow} + F^{\downarrow}\right) \tag{2.43a}$$

Subtracting, we have another relation

$$\frac{d}{d\tau}\left(F^{\uparrow} + F^{\downarrow}\right) = \left(F^{\uparrow} - F^{\downarrow}\right) = C_1 \tag{2.43b}$$

that can be integrated easily

$$F^{\uparrow} + F^{\downarrow} = C_1\tau + C_2 \tag{2.44}$$

The integration constants can be obtained considering that at the top of the atmosphere ($t = 0$), the flux $F^{\downarrow} = 0$ so that $C_2 = F^{\uparrow}(0)$. With the same reasoning from (2.42), we have also $C_1 = F^{\downarrow}(0)$ so that the values of the average flux $F^{\uparrow} + F^{\downarrow}$ and the net flux $F^{\uparrow} - F^{\downarrow}$ can be determined:

$$F^{\uparrow} + F^{\downarrow} = F^{\uparrow}(0)\,(\tau + 1)\,; \quad F^{\uparrow} - F^{\downarrow} = F^{\uparrow}(0) \tag{2.45}$$

Subtracting and summing, we obtain

$$F^{\uparrow} = \frac{1}{2}F^{\uparrow}(0)\,(\tau + 2)\,; \quad F^{\downarrow} = \frac{1}{2}F^{\uparrow}(0)\tau \tag{2.46}$$

Using Eq. (1.4), we can write the upward flux at the top as a function of the effective temperature

$$\sigma T_e^4 = F^{\uparrow}(0)$$

and noting that $B = \sigma T^4/\pi$, the temperature as a function of $\tau$ is given by

$$T^4 = \frac{1}{2}T_e^4\left(\frac{3}{2}\tau + 1\right) \tag{2.47}$$

where $\tau$ has been replaced by its effective value. An interesting consequence of this solution is to discover what happens at the surface, where the optical thickness is $\tau*$ and the temperature is $T_s$; we have

$$F^{\uparrow}\left(\tau^*\right) = \sigma T_s^4 = \frac{1}{2}\sigma T_e^4\left(\frac{3}{2}\tau^* + 2\right) \tag{2.48}$$

The temperature at the bottom of the atmosphere at $\tau^*$ is given by Eq. (2.47), so that we have a discontinuity between the air temperature and that of the surface

$$T_s^4 - T\left(\tau^*\right) = \frac{1}{2}T_e^4 \tag{2.49}$$

In the same way for $\tau = 0$, we have

$$T(0) = T_0 = 2^{-1/4}T_e$$

This result is similar to the one obtained for an atmosphere having emissivity $\varepsilon$, as in Eq. (E.2.17). In Fig. 2.12, the temperature as a function of $\tau$ is shown for values of the optical thickness that are typical of the atmosphere of the Earth. In this figure, we do not see any discontinuity at the ground, where the temperature is around 385 K. The temperature $T_0$ in this case is about 212.8 K, and this solution corresponds to radiative equilibrium. Near the ground however, the lapse rate arising from radiative equilibrium is not consistent with the convective stability criteria we

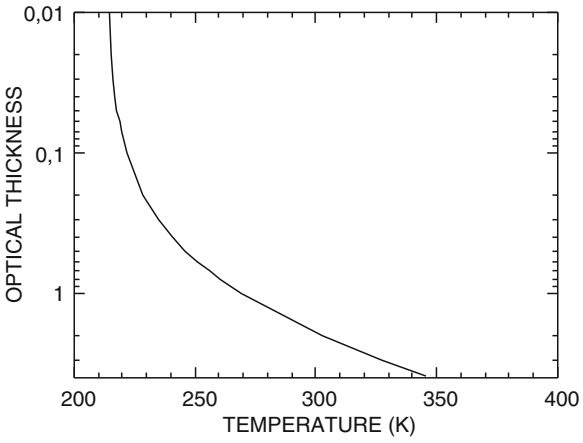



**Fig. 2.12** The radiative equilibrium temperature as a function of the optical thickness for a temperature $T_e = 254.4°\text{K}$

studied in Chap. 2 when dealing with thermodynamics. At this point, something must be introduced to maintain a realistic solution.

### 2.4.2 The Radiative–Convective Atmosphere

We have seen already that the distribution of water vapor in the atmosphere of the Earth can be approximated with an exponential function that has a scale height of roughly 2–3 km. On the other hand, the optical thickness in the infrared largely depends on the water vapor, so that with a good approximation, we can assume

$$\tau = \tau_0 \exp\left(\frac{-z}{H_w}\right) \tag{2.50}$$

where $\tau_0$ is the maximum optical thickness and $H_w$ the scale height of water vapor, which we assume to be 2 km. In this way, we can establish a correspondence between optical thickness and altitude so that we can evaluate the lapse rate as a function of the optical thickness (Fig. 2.13).

The result of this simple calculation is shown in Fig. 2.12. What we notice immediately is that the lapse rate around 10 km is much larger than the limit value of 9.8 K km$^{-1}$ (i.e., the adiabatic lapse rate). Beyond this point, the atmosphere is convectively unstable. This means that the radiative equilibrium solution is not representative of reality below a certain altitude, where the lapse rate should be maintained lower than the adiabatic lapse rate by convective processes. This

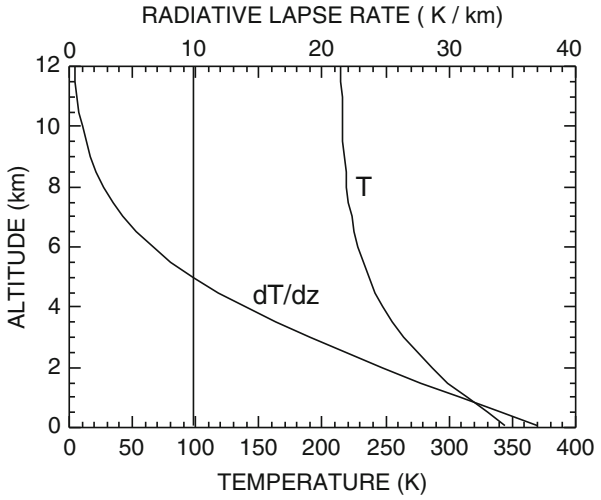



**Fig. 2.13** The temperature and the lapse rate as functions of optical thickness. The value of the adiabatic lapse rate is indicated by the *vertical line*

conclusion means that the radiative fluxes calculated for the radiative equilibrium are no longer valid in an atmosphere where convection is present in the layers near the surface. We can elaborate on the physics by considering that there is a temperature discontinuity at the surface. This temperature difference establishes a convective heat flux between the atmosphere and surface that acts to reduce the discontinuity. We will call this *convective* flux, and as a consequence of it, the difference between the upward and downward radiative fluxes will no longer be constant as in Eq. (2.43b), and the heating rate instead of as in Eq. (2.42) will be

$$\frac{dT}{dt} = -\frac{1}{\rho C_p}\frac{d}{dz}\left(F^{\uparrow} - F^{\downarrow} + F_c\right) \tag{2.51}$$

where we have indicated with $F_c$ the convective flux. To understand a little better what is happening with the fluxes, we can refer to Fig. 2.14, where the radiative fluxes are calculated with the previous equations. In the case of an atmosphere in radiative equilibrium, the difference between the upward flux and downward flux is constant. At the surface the upward flux equals exactly the sum of the solar absorbed flux and infrared downward flux. At the top of the atmosphere, the upward flux must be equal to the net solar flux (i.e., incident less the reflected one). With a 30 % albedo and an incident average flux of 340 W m$^{-2}$, the net solar flux amounts to 238 W m$^{-2}$. In the case of a radiative convective atmosphere, the solution can

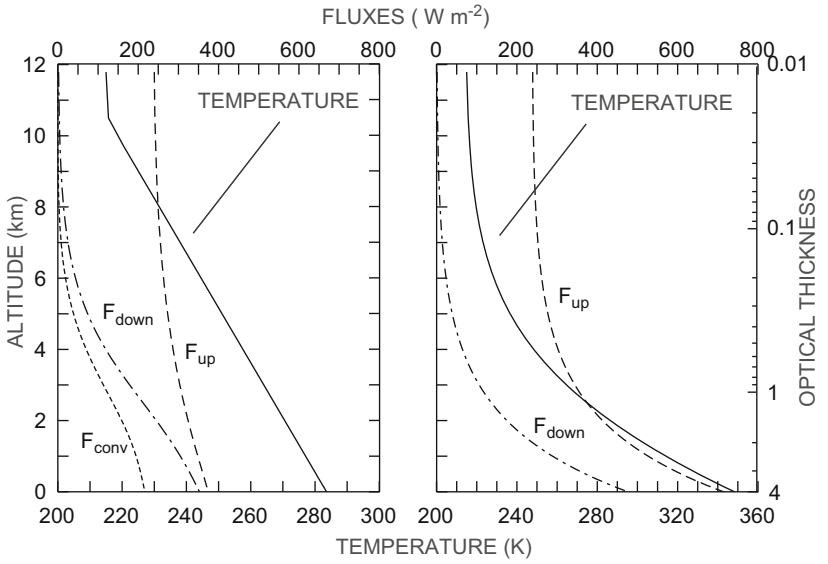



**Fig. 2.14** Temperature and radiative fluxes for an atmosphere in radiative equilibrium (*left*). On the *right* the same quantities are drawn for an atmosphere that includes a convective layer that simulates the troposphere. Notice that the equilibrium temperature at the ground for the radiative equilibrium is about 373 K while for the radiative convective equilibrium it is only 285 K

be found integrating Eq. (2.43a, b). In this case, the Planck function depends only on the optical thickness once the temperature is assigned. The integration, which could be done numerically, gives at the ground an upward flux of 373 $\mathrm{Wm}^{-2}$ and a downward flux of 352 $\mathrm{Wm}^{-2}$ that are not in equilibrium with the solar flux. The constraint on this integration is the upward radiative flux at the top being in equilibrium with the net solar flux. We must then require the existence of a convective flux of $352 + 238 - 373 = 217$ $\mathrm{Wm}^{-2}$. This quantity includes actually also what is called *latent heat flux*. This flux is generated when water in the soil or at the surface is evaporated, requiring heat. The conclusion is then that the radiative equilibrium solution at some point is convectively unstable and an adiabatic profile is established that implies a convective flux that decreases the surface temperature of about 90 K. Although it may look a little strange, we have learned enough to reach some important conclusions. We have introduced the radiative equilibrium and the stability criteria and examined if these are consistent. We have also found that the gas that contributes mainly to the greenhouse effect is water vapor. The abundance of this gas depends on the temperature at the least in the troposphere. This corresponds to the so-called water vapor feedback, that is, the higher the temperature, the greater the water vapor content, the more efficient the greenhouse effect. But when will this game end? It is rather intuitive to think that some cold region of the atmosphere controls the water vapor content, and in our case this region could be the tropopause. With increasing water vapor, the temperature of the tropopause could be too warm to maintain a low value for the mixing ratio of water vapor. At this point, without any "cold trap," water could evaporate from the oceans and reduce the Earth to somewhat like Venus. This is called the *runaway greenhouse*.

### 2.4.3 The Runaway Greenhouse

We need to point out that the topic we are going to illustrate is actually much more complex. Strangely enough, however, when Andrew Ingersoll introduced the runaway greenhouse, in a paper published in 1969 in the *Journal of the Atmospheric Sciences*, he used very simple arguments, similar to those we will use here. The starting point is to relate the optical thickness of the region where the radiative equilibrium dominates (the lower stratosphere) to the mass absorption coefficient $K$ of water vapor according to the relation

$$\tau = \int_0^\infty K \rho_v dz \tag{2.52}$$

In the stratosphere we may assume that the mass mixing ratio $\rho_v/\rho$ is constant with altitude, so that Eq. (2.48) can be easily rewritten with the help of the hydrostatic equilibrium

$$\int_z^\infty K\frac{em_v}{pm}\rho dz = K\frac{em_v}{pmg}\int_z^\infty dp$$

where with $e$ we have indicated the water vapor pressure and with $m_v$ and $m$ the molecular mass of water vapor and air. This equation can be easily integrated:

$$\tau = \frac{Kp_v m_v}{mg} \tag{2.53}$$

At the tropopause, if the relative humidity is $r$, we can put $e = re_s$, where with $e_s$ we have indicated the saturation pressure. Combining Eq. (2.53) with Eq. (2.47) and considering that the downward flux is equal to the net solar flux, we obtain an explicit equation between solar flux and temperature at the tropopause

$$(1-\alpha)\frac{Q}{2} = \sigma T^4\left[1+\frac{e_s(T)}{p_c}\right]^{-1} \tag{2.54}$$

where $p_c = 2mg/3rkm_v$, which according to Eq. (2.52) is the value of $e$ corresponding to the optical thickness equal to 1. This condition for $p_c$ is consistent for the terrestrial situation for which the optical thickness is about 4. If we assume a value for the relative humidity of 100 % (i.e., saturation) and an average absorption coefficient of 0.13 $m^2$ $kg^{-1}$, $p_c$ is about 8 mb and becomes 6 mb for a relative humidity of 80 %. Once $Q$ is assigned, there are only certain values of $T$ that can be solutions of Eq. (2.50).

This is shown in Fig. 2.15, where we represent the solutions for the radiative equilibrium at three different solar fluxes corresponding to the temperatures $T_e = 206$ K (Mars), 254.5 K (Earth), and 299.3 K (Venus). At the same time, if we assume that

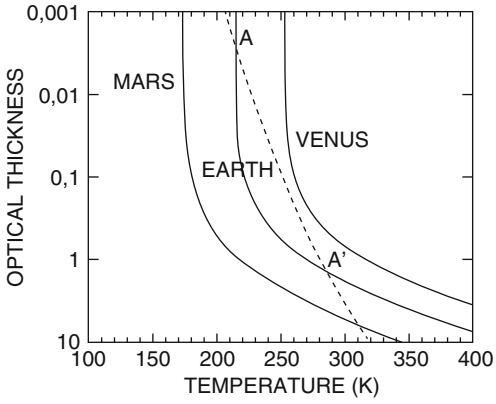



**Fig. 2.15** The radiative equilibrium temperature is shown for three different values of the solar constant. These values correspond to the orbits of Venus, Earth, and Mars, with the value of the albedo constant and equal to 0.2. *Dotted line* corresponds to the temperature for which saturation is reached according to Eq. (2.50)

at the tropopause the pressure of water vapor is the saturation pressure given by Eq. (2.53), we can obtain the corresponding temperature as a function of the optical thickness. The right-hand side of Eq. (2.52) for temperatures low enough tends to $\sigma T^4(0)$, while for high temperatures, it becomes

$$\frac{\sigma T^4 p_c}{e_s(T)}$$

Between these extremes there is a maximum value of the temperature corresponding to the maximum infrared flux in equilibrium with the incident solar flux. The maximum is due to the fact that the saturation pressure grows faster with temperature than the fourth power of the same quantity appearing in Eq. (2.54). In the case of Mars, the maximum flux is around 372.8 $\mathrm{Wm}^{-2}$, for the Earth it is 391 $\mathrm{Wm}^{-2}$, and for Venus it is 390.5 $\mathrm{Wm}^{-2}$. These values must be compared with the values of the solar incident flux for the planets, which are, respectively, 102.2, 238, and 455 $\mathrm{Wm}^{-2}$. It is obvious that only for the Earth and Mars can Eq. (2.52) be satisfied and equilibrium reached. In the case of Venus, the emitted flux is always lower than the one absorbed so that the temperature of the planet would increase forever.

This effect can also be seen from Fig. 2.15, where the radiative solution is shown compared with the solution obtained from Eq. (2.47), which is the optical thickness at saturation. Between the points $A$ and $A$', the atmosphere is supersaturated, and it is reasonable to expect a radiative equilibrium with a saturated lapse rate. The same curve does not intercept the radiative solution for Venus, and this means that saturation is never reached. In principle, in these conditions all the water present on the planet could evaporate: this is the runaway greenhouse. There are several limitations to this reasoning, but the principle is robust. What should really happen is that in the region of radiative equilibrium, the outgoing radiation must be equal to the incident radiation. The lower atmosphere should then be able to provide the upward net infrared flux. We have already seen that in the lower atmosphere, the radiative solution is unstable, so that it must be substituted with a convective atmosphere that is colder and thus gives a lower infrared flux. A convective troposphere produces a lower infrared flux, because it transports more water vapor upward, increasing the total optical thickness of the atmosphere. We have neglected for obvious reasons all the mechanisms that associate clouds with condensation. The clouds can change drastically all the results of the infrared fluxes. We will show in the next chapter, which deals with the evolution of the atmosphere, and in all the chapters dealing with climate, that things are much more complex. We already know, for example, that convective processes dominate the troposphere and that moist adiabats should be used to describe the temperature changes with altitude. However, it is noteworthy that, already in Chap. 3, we can already discuss the most interesting problems of the planetary atmospheres.

## E.2 Examples

### *E.2.1 Rayleigh Scattering from Natural Light (Sunlight)*

Consider a small homogeneous spherical particle with radius much smaller than the wavelength of incident radiation. This is assumed to be an electric field $\mathbf{E}_0$ that produces impinging on the particle a dipole moment $\mathbf{P}_0$ such that

$$\mathbf{p}_0 = \alpha \mathbf{E}_0 \tag{E.2.1}$$

where $\alpha$ is the polarizability of the mean. Because the impinging field oscillates, the dipole also will oscillate according to the relation

$$\mathbf{p} = \mathbf{p}_0 \exp\left[ik\left(r - ct\right)\right] \tag{E.2.2}$$

and will emit a scattered radiation whose field is given by at distance r and in the direction $\gamma$ (see Fig. E.2.1)

$$\mathbf{E} = \frac{1}{4\pi\varepsilon_0 c^2 r}\frac{\partial^2 \mathbf{p}}{\partial t^2}\sin\gamma \tag{E.2.3}$$

When (E.2.2) is substituted in (E.2.3), we get

$$\mathbf{E} = -\mathbf{E}_0 \frac{\exp\left[ik\left(r - ct\right)\right]}{r} k^2 \alpha \ \sin\ \gamma \tag{E.2.4}$$

We now consider the scattering of unpolarized sunlight by air molecules. The unpolarized light can be decomposed in two components parallel to the horizontal plane, $\mathbf{E}_{0l}$ and perpendicular $\mathbf{E}_{0r}$. These two components have a random phase relationship, and we can consider separately the scattering of the two components.

Then according to $t$ (E.2.4), we have

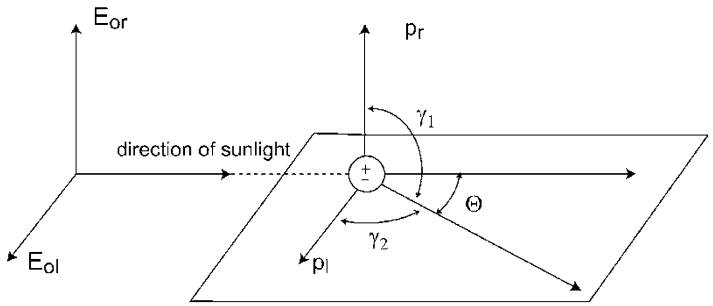



**Fig. E.2.1** Scattering by a dipole

$$E_r = -E_{0r}\frac{\exp\left[ik\left(r-ct\right)\right]}{r}k^2\alpha\,\sin\,\gamma_1$$
$$E_l = -E_{0l}\frac{\exp\left[ik\left(r-ct\right)\right]}{r}k^2\alpha\,\sin\,\gamma_2 \tag{E.2.5}$$

The corresponding intensities are proportional to the square roots of the electric field. The figure gives also $\sin\gamma_1 = 1$ and $\sin\gamma_2 = \sin\left(\pi/2-\Theta\right) = \cos\Theta$. We have then for the intensities

$$\begin{aligned} I_r &= I_{0r}k^4a^2/r^2 \\ I_l &= I_{0l}k^4a^2\cos^2Q/r^2 \end{aligned} \tag{E.2.6}$$

The total intensity is the sum

$$I = I_r + I_l = \left(I_{0r} + I_{0l}\cos^2Q\right)k^4a^2/r^2$$

For unpolarized light $I_{0r} = I_{0l} = I_0/2$ so that

$$I = \frac{I_0}{r^2}\alpha^2\left(\frac{2\pi}{\lambda}\right)^4\frac{1+\cos^2\Theta}{2} \tag{E.2.7}$$

This relation derived for the first time by Rayleigh gives the intensity of scattered unpolarized light as proportional to the incident radiation and inversely proportional to the fourth power of the wavelength. The intensity decreases with distance and depends strongly on the scattering angle $\Theta$ (Fig. E.2.2).

An interesting consequence of this approach is that natural light has a degree of polarization that depends on the direction with respect to direction of incident light.

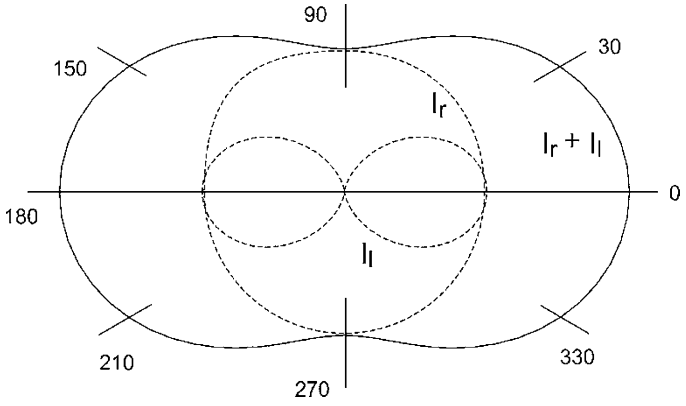



**Fig. E.2.2** The polar diagram of the scattered radiation by a Rayleigh molecule. The *dashed lines* illustrate the distribution of the perpendicular component (*circle*) and the parallel component (lobed figure). The *solid line* is the total intensity

The degree of polarization of natural scattered light can be defined as

$$LP(\Theta) = \frac{I_r - I_l}{I_r + I_l} = \frac{\cos^2\Theta - 1}{\cos^2\Theta + 1} = -\frac{\sin^2\Theta}{\cos^2\Theta + 1} \tag{E.2.8}$$

We see that the degree of polarization has a maximum (1) for $\Theta = \pi/2$ that is in the direction perpendicular to the direction of sunlight. Actually in the real case, there is no point in which polarization is 100 % due to the presence of multiple scattering (this example is taken from Liou 1980).

## *E.2.2 A Simple Way to Evaluate Ozone Absorption*

The method presented here has been invented by Richard Lindzen and Douglas Will and published in 1973 paper on *Journal of the Atmospheric Sciences*. As we have illustrated, the intensity as a function of wavelength is given by

$$I_\nu = I_{\nu\infty} \exp(\sigma_\nu u)$$

where $u = \int_\infty^z n dz / \cos\theta$ is the optical thickness, $n$ is the ozone density, $\theta$ is the zenith angle, and $I_{\nu\infty}$ is the intensity outside the atmosphere.

And the absorbed power per unit volume with $\sigma_\nu$ the cross section for absorption

$$Q = \int_\nu \sigma_\nu I_\nu n d\lambda$$

The assumption here is that for ozone, most of the absorption depends only on the same gas. This also means that the quantity $\eta = Q/n$ depends only on $u$. The other assumption can be summarized as follows:

1. Above 45 km where $u < 5.37 \cdot 10^{16} \text{cm}^{-2}$, the atmosphere is almost transparent to all ozone bands. The only absorption may be due to Huggins and Hartley band.
2. Below 30 km where $4 \cdot 10^{18} < u < 8 \cdot 10^{18} \text{cm}^{-2}$, most of the radiation in the Hartley and Huggins bands has been absorbed, while the atmosphere is almost transparent to the Chappuis band.

As shown in Fig. E.2.3, the Hartley and Chappuis band can be approximated simply with the constant absorption cross section and constant flux so that $\eta$ can written as

$$\eta = I_H k_H \Delta\lambda_H \exp(-k_H u) + I_C k_C \Delta\lambda_C \exp(-k_C u) \tag{E.2.9}$$

For the Huggins band, the solar flux is constant, while the absorption cross section can be approximated by an exponential function (Fig. E.2.3b)

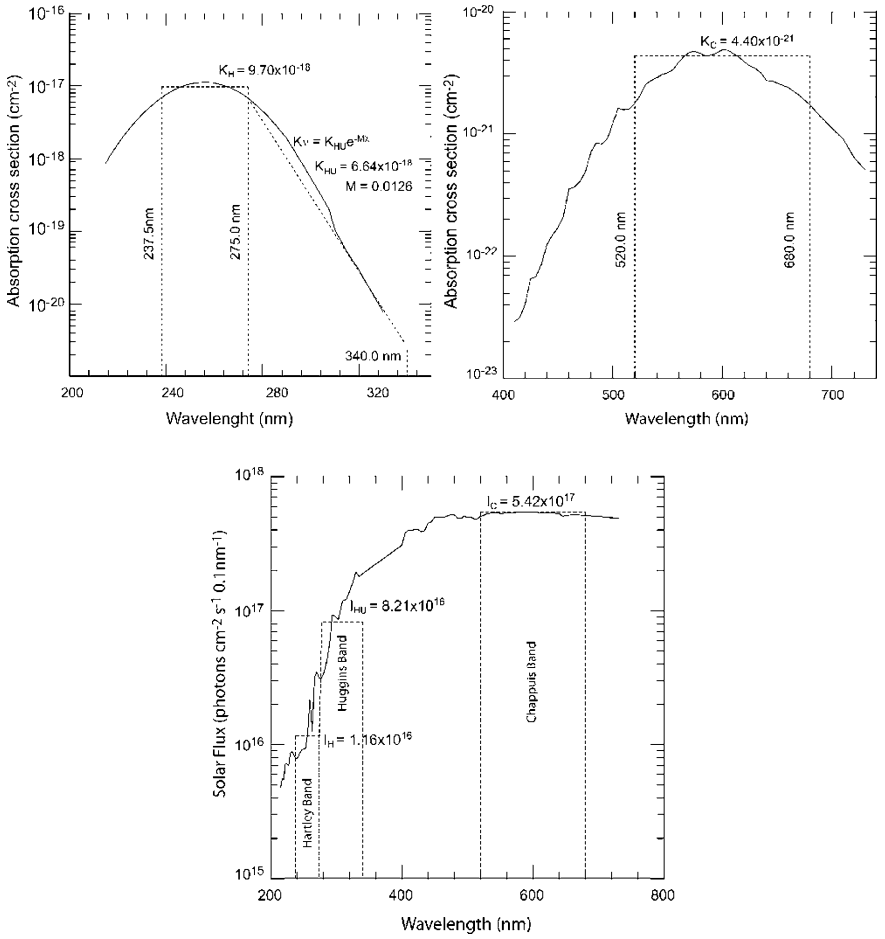



**Fig. E.2.3** The Hartley and Huggins bands

$$k_\nu = k_{Hu} \exp(-M\lambda)$$

The integration over $\lambda$ can be performed to get

$$\eta_{Hu} = \int_{\lambda_{\text{short}}}^{\lambda_{\text{long}}} k_{Hu} \exp(-M\lambda) I_{Hu} \exp[-k_{Hu} \exp(-M\lambda)]$$

to get

$$\eta_{Hu} = \frac{I_{Hu}}{Mu} \left\{ \exp\left[-uk_{Hu} \exp\left(-M\lambda_{\text{long}}\right)\right] - \exp[-uk_{Hu} \exp(-M\lambda_{\text{short}})] \right\} \quad \text{(E.2.10)}$$

The total heating will be then the sum of (E.2.9) and (E.2.10). The relevant parameters appearing in these relations are reported as follows: notice that the absorption coefficient is expressed in $cm^2$, and the ozone column is in molecules $cm^{-2}$:

$$I_H = 9 \cdot 10^{-2} w \text{ cm}^{-2} \text{ nm}^{-1}$$

$$k_H = 9.67 \cdot 10^{-18} \text{ cm}^2$$

$$\Delta\lambda_H = 37.5 \text{ nm}$$

$$I_C = 1.8w \text{ cm}^{-2}\text{nm}^{-1} \quad k_C = 4.40 \cdot 10^{-21} \text{ cm}^2 \quad \Delta\lambda_C = 165 \text{ nm} \quad M = 0.0126$$

$$I_{Hu} = 0.53w \text{ cm}^{-2}\text{nm}^{-1} \quad k_{Hu} = 1.99 \cdot 10^{-17}\text{cm}^2 \quad \lambda_{\text{short}} = 275\text{nm} \quad \lambda_{\text{long}} = 275\text{nm}$$

By fine-tuning some of the parameters, it is possible to obtain an overall error of less than 5 % with respect to a detailed calculation.

### *E.2.3 The Radiative Time Constant*

The Earth can be considered a black (or gray body) subject to solar forcing. However, this may change (e.g., the energy absorbed in some place) during the day. If we denote by $m$ the mass of the atmosphere for unit surface, and by $C_p$ the specific heat at constant pressure, then for small perturbations of the temperature $T'$, due to changes $Q'$ of the solar constant $Q$, we have

$$mCp\frac{d\left(T+T'\right)}{dt} = \left(Q+Q'\right)(1-\alpha) - \sigma\left(T+T'\right)^4 \tag{E.2.11}$$

At the equilibrium, $dT/dt = 0$ and the perturbation $T'$

$$mCp\frac{dT'}{dt} = Q'(1-\alpha) - 4\sigma T^4\left(\frac{T'}{T}\right) \tag{E.2.12}$$

The solution of this differential equation in $T'$ can be found is given by

$$T'(t) = \int_0^t \frac{Q'(1-\alpha)\,\mathrm{e}^{(t-t'/\tau)}}{mC_p}dt' \tag{E.2.13}$$

where

$$\tau = \frac{mCp}{4\sigma T^3} \tag{E.2.14}$$

**Table E.2.1** The radiative time constant (in terrestrial days for three planets)

| Planet | Atmospheric mass (kg) | Surface ($m^2$) | Mass/unit surface | Surface temp. (K) | Constant $\tau$ (days) | Day length |
|---|---|---|---|---|---|---|
| Venus | $4.8 \times 10^{20}$ | $4.54 \times 10^{14}$ | $1.04 \times 10^{6}$ | 735 | 133 | 233 |
| Earth | $5.2 \times 10^{18}$ | $5.1 \times 10^{14}$ | $1.04 \times 10^{4}$ | 288 | 21 | 1 |
| Mars | $2.5 \times 10^{16}$ | $1.44 \times 10^{14}$ | $1.73 \times 10^{2}$ | 210 | 0.9 | 1 |

The time constant given by Eq. (E.2.14) is the time it takes for the atmosphere to react to changes in the radiative forcing; that is, if the sun were to be turned off, the atmosphere could feel the effects in a time of the order of $\tau$.

It is interesting to compare the radiative time constant with the length of the day for the terrestrial planets with an atmosphere (Venus, Earth, and Mars). This comparison gives us an idea of what kind of diurnal variation in temperature we could expect for the planets. Table E.2.1 summarizes some of the planets' characteristics.

A constant value for the specific heat for three planets (1000 $\mathrm{jkg^{-1}K^{-1}}$) has been used. We can see from this table that the ratio between the length of the day and the time constant is short only for Mars. For this planet, we could expect pronounced diurnal effect that is negligible for the other. For the Earth, the integral (E.2.13) can be solved to give approximately

$$\Delta T = \frac{Q'\,(1-\alpha)\,t_{\mathrm{day}}}{mC_p} \approx 2K$$

For Venus, because the atmosphere is optically thick, this simple approximation may not be correct. Also for the Earth, we know that nights are colder than days, but the temperature difference is not only due to radiation but mainly to the interaction with the atmospheric boundary layer and the ground. We will return to this problem later on.

### *E.2.4 A Simple Model for the Greenhouse Effect*

We have then already isolated a few topics of study regarding the interaction of solar radiation with the clouds and the gases in the atmosphere and the transfer of infrared radiation. We have also implicitly assumed that an equivalent temperature $T_e$ can be obtained considering the planet as a blackbody and equating the absorbed solar power to the emitted power

$$\pi a^2 Q\,(1-\alpha) = 4\pi a^2 \sigma T_e^4 \Rightarrow Te = \left[\frac{Q\,(1-\alpha)}{4\sigma}\right]^{1/4} \qquad \text{(E.2.15)}$$

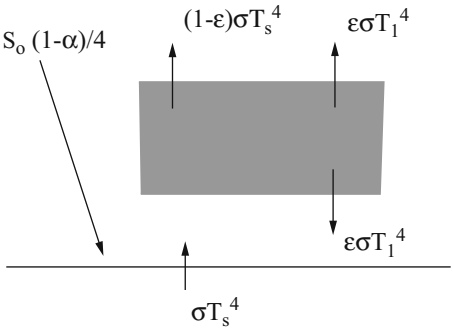



**Fig. E.2.4** A simple model of the greenhouse effect. The atmosphere is approximated as a medium at temperature $T_1$ and emissivity $\varepsilon$. The solar radiation is not absorbed

A very simple treatment Petty (2006), which should give a useful indication of the physics of the problem, can be developed if we consider the atmosphere to be made up of a single layer of temperature $T_1$ and emissivity $\varepsilon$, as shown in Fig. E.2.4. If the surface temperature is $T_s$, assuming that the transmissivity (i.e., the fraction of radiation transmitted by the atmosphere) is $(1-\varepsilon)$, we can write down the thermal equilibrium for the atmosphere and the surface–atmosphere system. We set the energy absorbed by the surface (solar radiation and infrared radiation coming from the atmosphere) equal to the energy emitted. The same condition is imposed for the atmosphere. We get then

$$2\varepsilon\sigma T_1^4 = \varepsilon\sigma T_s^4; \quad S_0\,(1-\alpha)\,/4 + \varepsilon\sigma T_1^4 = \sigma T_s^4 \tag{E.2.16}$$

Equation (E.2.16) describes a very simple system, which can be solved for the atmospheric temperature $T_1$ and the surface $T_s$:

$$\begin{aligned} T_1 &= T_s/2^{1/4} < [S_0\,(1-\alpha)\,/4\sigma]^{1/2} \\ \sigma T_s^4 &= S_0\,(1-\alpha)\,/\,[4\,(1-\varepsilon/2)] \end{aligned} \tag{E.2.17}$$

The temperature of the atmosphere is then lower than the surface temperature, while the latter is much higher than that which would result from Eq. (E.2.15): in other words, the addition of an atmospheric layer warms up the surface. Note that we have implicitly assumed that the atmosphere does not absorb solar radiation, and we have also found the existence of a discontinuity between the atmosphere and the surface. The emissivity, which in this very simple scheme is just the fraction of infrared radiation absorbed by the atmosphere, is mainly determined by the presence of water vapor in the troposphere. This means that if the surface temperature changes, then the atmospheric temperature will also change and, as a consequence, so will its water vapor content. A higher temperature implies more evaporation and this will result in a change of emissivity. This mechanism is also known as *water vapor feedback*, and it is very efficient in amplifying any initial perturbation to the climatic system.

# References[1]

## *Books*

Bekefi G, Barrett AH (1977) Electromagnetic vibrations, waves and radiation. MIT Press, Cambridge, MA
Houghton JT (1977) The physics of the atmospheres. Cambridge University Press, Cambridge
Liou KN (1980) An introduction to atmospheric radiation. Academic, Amsterdam/Boston
Petty GW (2006) A first course in atmospheric radiation. Sundog Publishing, Madison
Sellers WD (1965) Physical climatology. The University of Chicago Press, Chicago
Stephens GL (2004) Remote sensing of the lower atmosphere: an introduction. Oxford University Press, Oxford

## *Articles*

Held M, Suarez M (1979) The sensitivity of an energy balance climate model to the variation in the orbital parameters. J Geophys Res 20:4825
Ingersoll AP (1969) The runaway greenhouse: a history of water on Venus. J Atmos Sci 26:1191
Lindzen RS, Will DI (1973) An analytic formula for heating due to Ozone absorption. J Atmos Sci 30:513
Robinson DT, Catling DC (2012). An analytic radiative–convective model for planetary atmospheres. Astrophys J 757:104

[1]For the basic concepts of electromagnetism, the chapter relies on the Bekefi and Barrett textbook. A very useful introductory text is by Petty. A few figures are taken from Sellers book, and the two-stream approximation is basically that given in the Houghton textbook. Some very nice application can be found in the Stephens book.

# Chapter 3
# The First Laws of Motion

In planning this book, we thought that, after a short introduction, we would start to look at the properties of the atmosphere layer by layer going from the lowest up. Then we realized that such an approach would not justify the title of the book.

A physicist tends to obtain the most general laws, which are then applied to explain certain data or processes. It seems more logical that, after several introductions and diversions, we start to treat that part of physics that is regularly neglected by general physics courses. This part for us is fundamental and is of course fluid dynamics. This does not mean that we know everything about thermodynamics or radiative transfer; it is only that, without learning something about dynamics, we cannot proceed any further in the study of the atmosphere. We will see that this will be enough to allow us to study the troposphere. On the other hand, to proceed to the study of the stratosphere, we absolutely need to know something about chemistry and more about radiation.

The importance of the dynamics is quite obvious even when we watch the weatherman on TV. He often talks about air masses, fronts, and low-pressure or high-pressure areas. It seems so easy that sometimes the forecast looks like a simple extrapolation of the current situation. Actually, this is still the way meteorologists make the forecast. Whatever the power of the computer they have, you still need someone with good experience and who sometimes just neglects equations and dynamics. This is true especially on a relatively small scale, like the Mediterranean area, and for relatively short times. If you want to forecast the evolution of the atmospheric circulation on the planetary scale and as far as you may go in time (we will see about this), you then need dynamics and large computers. In this case you may be able to know more precisely (i.e., quantitatively) where and when and how much it will rain, for example. Dynamics is not only important for weather forecasting, but it is even more so for other regions of the atmosphere. For example, the well-known "ozone hole" and the fact that you find it only over the south pole is due to a combination of radiative, chemical, and dynamic effects. A large part of



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_3

the uncertainties related to global warming is due to the fact that the atmospheric circulation changes with the thermal structure and with the circulation are associated precipitation, deserts, and so on.

## 3.1 Scales and Orders of Magnitude

Our study must keep in mind that the theories on atmospheric dynamics must be compared with the events that happen in the only laboratories we have at hand. These are the atmospheres of Venus, Earth, Mars, Jupiter, Saturn, and all those planets and satellites which have an atmosphere that can be observed. The temporal and spatial scales on which the atmospheric motions develop are wide. They range from the breezes, which have a diurnal period and are of interest for a few kilometers, to the atmospheric tides which still have a diurnal period but are of interest on a planetary scale. With longer time scales, we described the atmospheric circulation and that of the ocean. On a global scale, on an even longer time scale, we described some of the ocean–atmosphere interactions.

To talk about "scales" in such simple terms may be misleading. For example, a large spatial scale may not be simply related to the geographical extent. Actually, the large-scale motions are those influenced by planetary rotation. If we take a typical length $L$ of an atmospheric perturbation and a typical velocity $U$ of the corresponding air mass, then the typical time scale of this air mass is $L/U$. With $L = 1000$ km e $U = 10$ ms$^{-1}$, we get a characteristic time of about 1 day. This is a time comparable to the period of rotation for the Earth, so that we can argue that the motion will be influenced by the rotation itself. We can then use a condition such that, if $\Omega$ is the angular velocity of the Earth or the planet, then the rotation will influence the motion if we have

$$L/U \geq \Omega^{-1} \tag{3.1}$$

which is the same as

$$R_0 = U/(2\Omega L) \leq 1 \tag{3.2}$$

This nondimensional parameter is known as the *Rossby number*, and we have mentioned, maybe in a presumptuous way, it in the first chapter. At this point we will define a large-scale motion such as the one for which the Rossby number is of the order of 1 or less. We will see later that the meaning of $R_0$ is much more profound, but in any case, Eq. (3.2) can be misleading because it gives the impression that we have small Rossby numbers associated with fast-rotating planets. For the Earth and the data specified above, we have $R_0 \approx 0.07$. Also, from Eq. (3.2) we cannot see that what determines the local Rossby number is the local vertical component of the angular velocity. This means an underestimation of $R_0$ at low latitudes. However, if the wind velocity (or that of the oceanic current) is small, the motion can still be

classified as large scale. For example, the characteristic velocity of the Gulf current is about 0.1 $ms^{-1}$, so even if the spatial scale is only 100 m, the Rossby number is still very small. Actually, the circulation in oceans of the Earth is much more similar to the circulation in the atmosphere of Jupiter, which is a very fast-rotating planet.

The considerations made so far are about kinematics, but the dynamic consequences of a small Rossby number are even more important. A small value for $R_0$ means that the fluid tends to rotate with the solid part of the planet, but for us the deviations from this rotation are important. To this end we will introduce a reference system that rotates with the planet in order to make more evident such deviations. This reference system is not inertial and we expect to find a few apparent forces, like the Coriolis force. We can see easily that, if the Rossby number is small, then the Coriolis force is important in determining the motion. Acceleration is of the order of $U/T = U^2/L$, while the Coriolis acceleration is of the order of

$$2\Omega \times U = 2\Omega U \sin\phi$$

where φ is the latitude. The motion is steady if the acceleration is small with respect to the Coriolis acceleration that can then equilibrate all the external "forces" acting on the fluid element. We will have then

$$R_0 = U/fL \leq 1 \tag{3.3}$$

where we define the *Coriolis parameter* as

$$f = 2\Omega \sin\phi \tag{3.4}$$

Equation (3.4) takes now into account correctly the latitude and shows that the Coriolis acceleration is zero at the equator.

Another characteristic of the large-scale circulation is that the ratio between the vertical and horizontal scale is quite small. The average depth of the ocean is about 4 km and of the same order is the thickness of the atmosphere where we have atmospheric motion. On the other hand, the horizontal scales are hundreds or thousands of kilometers. The motions then take place in a very thin layer of fluids and as consequence can be treated as quasi-horizontal.

## 3.2 The Basic Equations

When we start a new topic, we are always more enthusiastic for the novelty, probably because the previous one was becoming a bore. With fluid dynamics we must learn to see things a little bit differently, starting from the concept of derivative with respect to time. Actually, the rules we are going to give next may have been studied in calculus courses. However, we are going to give to those rules a very important physical meaning and show that they are very intuitive.

### 3.2.1 The Total Derivative

Think about a little balloon moving in the air and of which we want to study how its temperature $T$ changes. That temperature will be a function of position and time, $T(x, y, z, t)$. If the balloon moves to the point with coordinates $x + \Delta x$, $y + \Delta y$, and $z + \Delta z$ in the time $\Delta t$, its temperature will change by an amount

$$\Delta T = \left(\frac{\partial T}{\partial x}\right)\Delta x + \left(\frac{\partial T}{\partial y}\right)\Delta y + \left(\frac{\partial T}{\partial z}\right)\Delta z + \left(\frac{\partial T}{\partial t}\right)\Delta t \tag{3.5}$$

Dividing by $\Delta t$ we get for the derivative

$$\frac{DT}{Dt} = \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt} + \frac{\partial T}{\partial z}\frac{dz}{dt} + \frac{\partial T}{\partial t} \tag{3.6}$$

In a Cartesian system $x$, $y$, and $z$ (we will see how it is oriented), we establish once and for all that the velocity components along the axis are

$$u = \frac{dx}{dt}; \qquad v = \frac{dy}{dt}; \qquad w = \frac{dz}{dt} \tag{3.7}$$

So that Eq. (3.6) becomes, keeping in mind the definition of gradient and scalar product,

$$\frac{DT}{Dt} = \frac{\partial T}{\partial t} + \mathbf{V} \cdot \nabla T \tag{3.8}$$

The total derivative operator $D/Dt$ is distinct from the usual derivative because the total derivative makes sense only if the quantity to be derived is an exact differential. For example, temperature, pressure, and velocity can be expressed as exact differentials but not heat. The total derivative expresses the rate of change of temperature in a coordinate system that moves with the air parcel (in our case the balloon). The air parcel is a material element, like the mass in mechanics, and contains always the same mass of air. The total derivative is made up of two parts, the first expresses the *local derivative* $\partial T/\partial t$ and $\mathbf{V}\cdot\nabla$ expresses the *convective derivative*. If the system used to make the measurements is fixed with respect to the surface, we call it an *Eulerian reference frame*. If the system moves with the air parcel, we call it *Lagrangian*. Normally, we use the first one, but as we will see, the Lagrangian reference has introduced an appreciable simplification into the treatment of the stratospheric motions. In the Eulerian reference frame, the change in temperature will be

$$\frac{\partial T}{\partial t} = \frac{DT}{Dt} - \mathbf{V} \cdot \nabla T \tag{3.9}$$

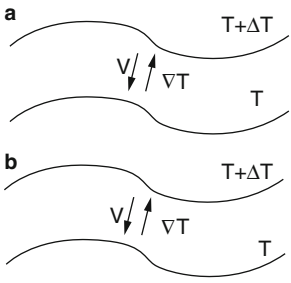



**Fig. 3.1** Isotherms and temperature advection. In case **(a)** the temperature gradient has the opposite sign of the velocity (positive advection). In case **(b)** we have the opposite situation

The second term on the right is called *advective* and is one of the most important in meteorology. To understand how this term works, we can refer to Fig. 3.1. Curves labeled $T$ and $T + \Delta T$ are isotherms on the plane of the figure, so that the arrows represent the generic direction of the gradient. For the sake of simplicity, we assume that the component of the wind on that plane is parallel to the gradient. Thus, we can have two cases: the first one is when the direction of the wind is opposite to the gradient (case a), and the second one when the wind has the same direction (case b). In the first case, the advection (i.e., the term $-\mathbf{V} \cdot \nabla T$) is positive, and, as a matter of fact, the downwind regions heat up. In case b, the advection is negative and the same regions cool. An example may help us to clarify further the concept. We assume that a cold front has passed over a meteorological station where the temperature is 10 °C and falls at a rate of 3 °C per hour. The wind arrives from the north at 40 km $\text{hr}^{-1}$ and the vertical component is zero. At another station 100 km to the north, the temperature is −2 °C.

We should be able now to evaluate what is the rate of change of the temperature for the masses following the cold front. The first thing that interests us is the advection, which can be easily calculated starting from the temperature gradient – (2 + 10)/100 km directed from south to north. The wind velocity is directed from north to south and is then negative. The rate of change of temperature is given according to Eq. (3.8)

$$\frac{DT}{Dt} = \frac{-3\ ^\circ\text{C}}{\text{hr}} + \frac{-40\ \text{km}}{\text{hr}} \times \frac{-(10+2)\ ^\circ\text{C}}{100\ \text{km}} = 1.8\ ^\circ\text{C}\ \text{hr}^{-1}$$

which represents heating. In practice, what is happening is that the advection warms the air mass by about 4.8 °C $\text{hr}^{-1}$, and this heating largely compensates for the local cooling at the station.

### *3.2.2 The Continuity Equation*

As a first application of the total derivative, we can evaluate the Eulerian change of the air density $\rho$; thus, we obtain

$$\frac{D\rho}{Dt} = \frac{\partial \rho}{\partial t} + \mathbf{V} \cdot \nabla \rho \tag{3.10}$$

This equation assumes a particular form when mass is conserved. We consider an elementary mass contained in the volume $\Delta V$, as in Fig. 3.2, with density $\rho$. From that volume we consider a net outgoing mass flux $\rho \mathbf{V} \cdot \mathbf{n} dS$, where $\mathbf{n}$ is the normal to the elementary surface and $\mathbf{V}$ is the local velocity. The net change of mass inside the volume is given by the integral of the flux

$$\frac{\partial}{\partial t} \int_V \rho \, dV = -\int_S \rho \mathbf{V} \cdot \mathbf{n} \, dS = -\int_V \nabla \cdot (\rho \mathbf{V}) \, dV \tag{3.11}$$

and we get another form of the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \ \mathbf{V}) = 0 \tag{3.12}$$

Noting that $\nabla \cdot (\rho \mathbf{V}) = \rho \nabla \cdot \mathbf{V} + (\mathbf{V} \cdot \nabla) \rho$, we have

$$\frac{1}{\rho} \frac{D\rho}{Dt} = -\nabla \cdot \mathbf{V}$$

It is interesting to give an interpretation of the velocity divergence. Consider an elementary mass $\delta M = \rho \delta V = \rho \delta x \delta y \delta z$, as shown in Fig. 3.3. The mass is conserved in the system that moves with the air parcel, so we can write

$$\frac{1}{\delta M} \frac{D}{Dt} (\delta M) = \frac{1}{\rho \delta V} \frac{D}{Dt} (\rho \delta V) = \frac{1}{\rho} \frac{D\rho}{Dt} + \frac{1}{\delta V} \frac{D}{Dt} (\delta V) = 0 \tag{3.13}$$

where

$$\frac{1}{\delta V} \frac{D}{Dt} (\delta V) = \frac{1}{\delta x} \frac{D}{Dt} (\delta x) + \frac{1}{\delta y} \frac{D}{Dt} (\delta y) + \frac{1}{\delta z} \frac{D}{Dt} (\delta z)$$

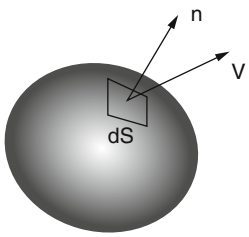



**Fig. 3.2** The mass flux from the interior of the elementary volume $dV$

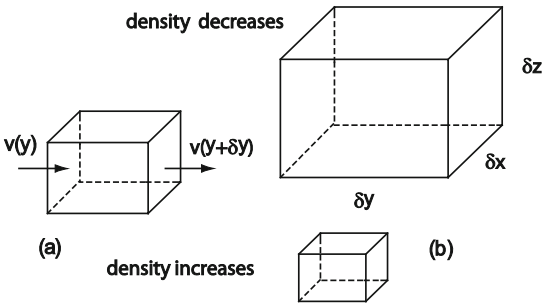



**Fig. 3.3** The elementary volume and the faces normal to the y axis that moves at different velocities **(a)**. In **(b)** the divergence is associated with the change of volume: positive when the volume increases, negative when it decreases

To understand better the terms on the right, we can refer to Fig. 3.3. The faces normal to the $y$ axis at distance $\delta y$ have slightly different velocities and the difference is given by

$\delta v = v\,(y + \delta y) - v\,(y)$ where $\delta v = D\,(y + \delta y)\,/Dt - Dy/Dt$

The same is true for the other components: the final result is

$$\lim_{\delta x,\delta y,\delta z}\left[\frac{1}{\delta V}\frac{D}{Dt}\,(\delta V)\right] = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \tag{3.14}$$

This equation gives a simple meaning to the velocity divergence. Positive divergence means that the volume of the parcel increases, and vice versa. This, according to the continuity equation, has an effect on the density because of the conservation of mass. If the divergence is positive, the density decreases, while if the divergence is negative, the density increases. If the density is constant, the divergence is zero and the fluid is called incompressible.

### *3.2.3 Pressure Forces*

We refer again to our elementary volume, as in Fig. 3.4. The net force along the $y$ axis is given by $F_x(y) - F_x\,(y + \delta y)$. The pressure in the $y$ direction

$$p(y) = p - \frac{\partial p}{\partial y}\frac{\delta y}{2} \quad p\,(y + \delta y) = p + \frac{\partial p}{\partial y}\frac{\delta y}{2}$$

The net force is given by

$$\left(p - \frac{\partial p}{\partial y}\frac{\delta y}{2} - p - \frac{\partial p}{\partial y}\frac{\delta y}{2}\right)\delta x\;\delta z = -\frac{\partial p}{\partial y}\delta x\delta y\;\delta z$$

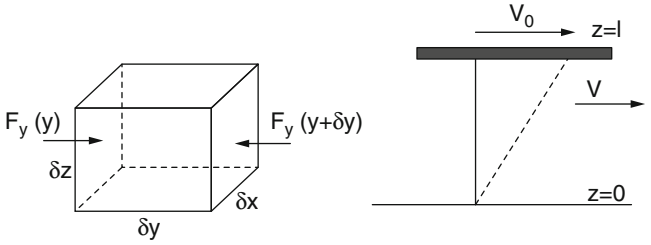



**Fig. 3.4** Pressure and viscous forces. On the *left* the forces on an elementary fluid portion are shown. On the *right* the drag force is illustrated

The mass of the element is $\delta m = \rho \delta x \delta y \delta z$ and its acceleration

$$\frac{F_x}{\delta m} = -\frac{1}{\rho}\frac{\partial p}{\partial x} \qquad \frac{F_y}{\delta m} = -\frac{1}{\rho}\frac{\partial p}{\partial y} \qquad \frac{F_z}{\delta m} = -\frac{1}{\rho}\frac{\partial p}{\partial z} \tag{3.15}$$

and in vectorial terms

$$\frac{\mathbf{F}}{\delta m} = -\frac{1}{\rho}\nabla p \tag{3.16}$$

The acceleration on the air mass has the opposite sign of the gradient. This conclusion may be obvious because the gas expands from the high to the low pressure.

### *3.2.4 Friction Forces*

The first contact of the student with the viscous force is a figure like Fig. 3.4. Here, a raft has to be maintained in uniform motion with velocity $v_0$. Then the force that must be used on the raft is given by $F = \mu A v_0 / l$, where $A$ is the surface of the raft, $\mu$ is the coefficient of dynamic viscosity, and $l$ is the thickness of the fluid. At the base of the fluid, the velocity is zero. The mechanism through which the raft drags the rest of the fluid is due to the friction between the surface $A$ and the fluid. The same friction exists between two adjacent layers of the fluid separated by a distance $dz$. The viscous stress, that is, the ratio $F/A$, is given by

$$\tau_{zy} = \mu \frac{\partial v}{\partial z}$$

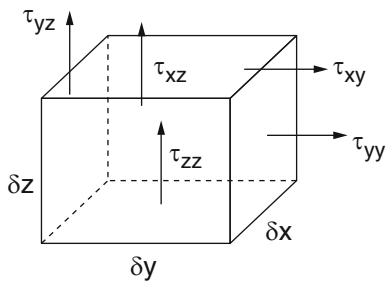



**Fig. 3.5** Some components of the stress tensor

where the subscripts on $\tau_{zy}$ indicate the stress component in the direction $y$ due to the change in velocity in the direction $z$. If we think about it a little more, we see that actually the viscosity transports momentum in the direction $z$. The momentum transported for unit time and unit surface has the dimension of pressure and it is the stress.

At this point, as already done for the pressure, we can find the net force exerted on an elementary mass by the viscous stress. In Fig. 3.5 we report some of the components of what is more properly defined as a *tensor*. Besides the component we already defined, we can define normal stresses to the faces

$$\tau_{xx},\ \tau_{yy},\ \tau_{zz}$$

To better understand the definition of the subscripts, we have to remember that the first subscript indicates the direction of the normal to the face on which the stress is exerted. The second subscript indicates the direction of the stress itself. If we refer to $\tau_{zy}$, its change with the $z$ direction is

$$\tau_{zy} + \frac{\partial \tau_{zy}}{\partial z}\frac{\delta z}{2} - \left(\tau_{zy} - \frac{\partial \tau_{zy}}{\partial z}\frac{\delta z}{2}\right) = \frac{\partial \tau_{zy}}{\partial z}\delta z$$

The total force on the $\delta x \delta y$ will be obtained by multiplying the net pressure for the surface. The acceleration in the $y$ direction will be

$$\frac{1}{\rho}\frac{\partial \tau_{zy}}{\partial z} = \frac{1}{\rho}\frac{\partial}{\partial z}\left(\mu \frac{\partial v}{\partial z}\right)$$

The total force per unit mass in the $x$ direction will be the sum of the stresses that act on the six faces of the volume element

$$F_x = \frac{\mu}{\rho}\left(\frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{yx}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z}\right) = \frac{\mu}{\rho}\nabla^2 u \tag{3.17a}$$

and similarly

$$F_y = \frac{\mu}{\rho}\nabla^2 v \qquad F_z = \frac{\mu}{\rho}\nabla^2 w \tag{3.17b}$$

The quantity $\nu = \mu/\rho$ is called *kinematic viscosity*. Notice that the scalar differential operator acts on the components of the velocity. To get a rough idea of this term, we can use the value of the air cinematic viscosity at sea level that is roughly

$\nu = 1.5\ 10^{-5}\ \mathrm{m^2\ s^{-1}}$. Consider then a wind that changes from 0 to 20 $\mathrm{ms^{-1}}$ for 1 to 2 km and then increases to 30 $\mathrm{ms^{-1}}$ at 3 km. We have

$$\frac{\partial^2 u}{\partial z^2} \approx \left(\frac{(30-20)}{10^3} - \frac{(20-0)}{10^3}\right)/10^3 = -10^{-5}\mathrm{m^{-1}s^{-1}}$$

Then the acceleration is of the order of $1.5\ 10^{-10}\ \mathrm{ms^{-2}}$. This acceleration term, we will see in a while, is completely negligible with respect to the Coriolis acceleration. Actually, we have seen in the previous chapter that the molecular viscosity is important in the upper atmosphere, but in the troposphere, it is important only for very thin layers. Then why all this fooling around with complicated equations? The reason is that down the line we will find a viscosity that is different from the molecular one (we will call it eddy), and, not knowing how to treat it, we will take as analogy the treatment given for the molecular viscosity. A last point, the way we have evaluated the second derivative, take it for granted. Somewhere in the next few chapters, we will talk about numerical methods, and you will see that, in order to evaluate second derivatives, you need at the least three points.

### *3.2.5 The Equations of Motion in an Inertial System*

Taking into account all the forces acting on a parcel of air (pressure, viscous, and gravitational), in an inertial reference frame, the second law of the dynamic is written as

$$\frac{D_a \mathbf{V}_a}{Dt} = \sum \mathbf{F} \tag{3.18}$$

where $\mathbf{V}_a$ is the velocity in the absolute reference frame and the derivative is made in the same system. The forces appearing on the right are all the external forces and do not include apparent forces like Coriolis and centrifugal. We, observers, are in a non-inertial system, and we need to find a relation between our relative velocity (we call it $\mathbf{V}$) and the absolute velocity. In principle this is a simple thing, as shown in Fig. 3.6. We have a vector $\mathbf{A}$ of constant modulus that rotates with

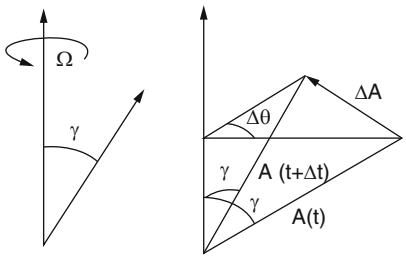



**Fig. 3.6** Vector of constant module A rotates with angular velocity Ω with angle $\gamma$ with respect to the rotation axis

angular velocity Ω and makes an angle $\gamma$ with the axis of rotation. It is very easy to show that $\Delta A = A\text{sen}\gamma\Delta\theta$ and so

$$\lim_{\Delta t\to 0}\frac{\Delta \mathbf{A}}{\Delta t} = \frac{d\mathbf{A}}{dt} = |\mathbf{A}|\sin\gamma\frac{d\theta}{dt}\frac{\Omega\times\mathbf{A}}{|\Omega\times\mathbf{A}|}$$

where $|\boldsymbol{\Omega}\times\mathbf{A}| = \Omega A\sin\gamma$,

$$\frac{D\mathbf{A}}{Dt} = \boldsymbol{\Omega}\times\mathbf{A} \tag{3.19}$$

The conclusion is that an observer rotating with the reference system does not see any change in **A**. An observer that is fixed with the inertial reference frame sees a change different from the one given by Eq. (3.19) for the observer in the inertial system. Both observers, however, see the same module that remains constant

$$\frac{D|\mathbf{A}|^2}{Dt} = 2\mathbf{A}\cdot(\boldsymbol{\Omega}\times\mathbf{A}) = 0 \tag{3.20}$$

Equation (3.19) can now be applied to any vector, and we consider in particular the distance **r** between our parcel and the origin of the rotating system with angular velocity **Ω**. In this system we can write

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

where **i**, **j**, and **k** are the unit vectors of the axis and $x$, $y$, and $z$ are the vector components. In the rotating system, the rate of change of vector **r** is given by

$$\left(\frac{D\mathbf{r}}{Dt}\right)_R = \frac{Dx}{Dt}\mathbf{i} + \frac{Dy}{Dt}\mathbf{j} + \frac{Dz}{Dt}\mathbf{k} \tag{3.21}$$

The inertial observer will see the entire reference frame rotating, with the unit vectors changing direction with time. At each of them we will apply the derivation rule of Eq. (3.19), so we get

$$\left(\frac{D\mathbf{r}}{Dt}\right)_I = \frac{Dx}{Dt}\mathbf{i} + \frac{Dy}{Dt}\mathbf{j} + \frac{Dz}{Dt}\mathbf{k} + x\frac{D\mathbf{i}}{Dt} + y\frac{D\mathbf{j}}{Dt} + z\frac{D\mathbf{k}}{Dt}$$

where

$$\frac{Dx}{Dt}\mathbf{i}+\frac{Dy}{Dt}\mathbf{j}+\frac{Dz}{Dt}\mathbf{k}=x\boldsymbol{\Omega}\times\mathbf{i}+y\boldsymbol{\Omega}\times\mathbf{j}+z\boldsymbol{\Omega}\times\mathbf{k}=\boldsymbol{\Omega}\times\mathbf{r}$$

In the inertial system, the rate of change of $\mathbf{r}$ will be given by

$$\left(\frac{D\mathbf{r}}{Dt}\right)_I=\left(\frac{D\mathbf{r}}{Dt}\right)_R+\boldsymbol{\Omega}\times\mathbf{r} \tag{3.22}$$

which is equivalent to a relation between velocities

$$\mathbf{v}_I=\mathbf{v}_R+\boldsymbol{\Omega}\times\mathbf{r} \tag{3.23}$$

where with $\mathbf{v}_I$ and $\mathbf{v}_R$, we have indicated the velocity in the inertial system and the relative velocity, respectively. Again, repeating the derivation rule on the inertial velocity, we get

$$\left(\frac{D\mathbf{v}_I}{Dt}\right)_I=\left(\frac{D\mathbf{v}_I}{Dt}\right)_R+\boldsymbol{\Omega}\times\mathbf{v}_I$$

Substituting from Eq. (3.23), we obtain the inertial acceleration as a function of measurable quantities in the rotating system.

$$\begin{aligned}\left(\frac{D\mathbf{v}_I}{Dt}\right)_I&=\left(\frac{D\mathbf{v}_R}{Dt}\right)_R+\boldsymbol{\Omega}\times\left(\frac{Dr}{Dt}\right)_R+\frac{D\boldsymbol{\Omega}}{Dt}\times\mathbf{r}+\boldsymbol{\Omega}\times(\mathbf{v}_R+\boldsymbol{\Omega}\times\mathbf{r})\\&=\left(\frac{D\mathbf{v}_R}{Dt}\right)_R+2\boldsymbol{\Omega}\times\mathbf{v}_R+\boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{r})+\frac{D\boldsymbol{\Omega}}{Dt}\times\mathbf{r}\end{aligned} \tag{3.24}$$

In this expression we know all the terms because, except for the relative acceleration, the second term on the right is simply the Coriolis acceleration, the third is the centrifugal acceleration, and the last one is zero if the angular velocity is constant. In order to use Eq. (3.24), for the equation of motion, we remember that in Eq. (3.18) the absolute derivative of the velocity is now the left-hand side of Eq. (3.24). The centrifugal force can be expressed as a potential, $\Phi$, that includes also the gravitational contribution. If we call $\mathbf{V}$ the velocity with respect to the rotating system, we have

$$\rho\left(\frac{D\mathbf{V}}{Dt}+2\boldsymbol{\Omega}\times\mathbf{V}\right)=-\nabla p+\rho\nabla\Phi+\mathbf{F}_r \tag{3.25}$$

where $\mathbf{F}_r$ are the friction forces.

Now that we have found an equation for the motion, we are not in a hurry to apply it. As a matter of fact, we still need to learn a few more things before going to the dynamics application. For one thing in order to apply Eq. (3.25), we have to specify

the reference frame, and in atmospheric physics, we have many reference systems mostly inherited from the meteorologists. We prefer then to wait and enter into another sector we may call "kinematics," as a further preparation to the maelstrom of the fluid dynamics.

## 3.3 Vorticity and Circulation

We will learn that sometimes it is not possible to solve Eq. (3.25) directly, so as often happens in physics; we will be looking for more general properties of the fluid in order to simplify the problem of solving the equation of motion. One of such quantities is known as *vorticity* and it is defined as the rotor of the velocity.

$$\boldsymbol{\omega} = \nabla \times \mathbf{V} \tag{3.26}$$

The vorticity will be relative if referred to the relative velocity $\mathbf{V}_r$ or absolute if referred to the absolute velocity $\mathbf{V}_a$. The first application of Eq. (3.26) can be made by referring to the inertial system so that, applying the rotor operation, we have

$$\boldsymbol{\omega} = \nabla \times (\mathbf{V}_r + \boldsymbol{\Omega} \times \mathbf{r}) = \boldsymbol{\omega}_r + 2\boldsymbol{\Omega} \tag{3.27}$$

The difference between the relative and absolute vorticity is the *planetary vorticity*, which is twice the angular velocity. In most of the cases, we are interested in the vertical component of the relative vorticity given by

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{3.28}$$

while the absolute vorticity is given simply $\eta = \zeta + 2\Omega \sin\phi$. From the rules of the vector calculus given a surface $A$ delimited by a curve $\Gamma$, we have

$$\int \boldsymbol{\omega} \cdot \mathbf{n} dA = \int_A (\nabla \times \mathbf{V}) \cdot \mathbf{n} dA = \oint_\Gamma \mathbf{V} \cdot \mathbf{dl}$$

with the last integral defining the circulation $C$. If the elementary $\delta A$ is perpendicular to the vertical component of the relative vorticity (i.e., an element of the Earth's surface), then we have the simple formula

$$\delta C / \delta A = \zeta$$

The vorticity can then be defined as the circulation for unit surface. This, among other things, makes clear why the planetary vorticity is just twice the angular velocity. If we imagine the surface of disk of radius $r$, which rotates with angular velocity $\Omega$, then the circulation is simply $2\pi r \Omega r$, and dividing by the area of the

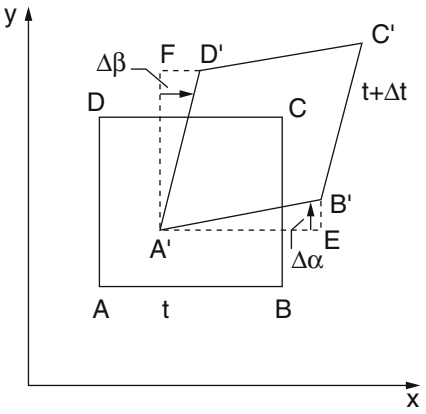



**Fig. 3.7** The distortion of a fluid element for the definition of vorticity

disk we obtain the desired result. Actually, this result can be extended to any fluid that usually has a differential rotation. This means that the angular velocity is a function of the radial distance of the axis of rotation. We can refer to Fig. 3.7, where a generic two-dimensional fluid element is shown on the plane $x, y$. The element is depicted at two successive instants $t$ and $t + \Delta t$. At the initial time, the element has the shape of rectangle $ABCD$ with dimension $\Delta x, \Delta y$, and after a time $\Delta t$, it changes into $A'B'C'D'$ in a nonrigid motion. The angular velocity around an axis parallel to $z$ is

$$\left(\frac{d\theta}{dt}\right)_z = \lim_{\Delta t \to 0} \frac{1}{2}\left(\frac{\Delta\alpha}{\Delta t} - \frac{\Delta\beta}{\Delta t}\right) = \frac{1}{2}\left(\frac{d\alpha}{dt} - \frac{d\beta}{dt}\right)$$

where $\Delta\alpha = EB'/A'E$ and $\Delta\beta = FD'/A'F$. We have also

$$EB' = \left[v_0 + \left(\frac{\partial v}{\partial x}\right)\Delta x\right]\Delta t - v_0\Delta t;\ A'E = \Delta x + \left[u_0 + \left(\frac{\partial u}{\partial x}\right)\Delta x\right]\Delta t - u_0\Delta t$$

From these two equations, we obtain then $d\alpha/dt = dv/dx$, and in the same way for $\beta$, we obtain $d\beta/dt = du/dy$. The equivalent angular velocity is

$$\left(\frac{d\theta}{dt}\right)_z = \frac{1}{2}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \tag{3.29}$$

On the right-hand side, we recognize the $z$ component of the vorticity already defined. In the case of a rigid body, the rotations of the angles are equal with opposite sign $\Delta\alpha = -\Delta\beta$ so that

$$\frac{d}{dt}(\Delta\alpha + \Delta\alpha) = 2\left(\frac{\partial\theta}{\partial t}\right)$$

and in this case the vorticity is twice the angular velocity.

### 3.3.1 Some Properties of Vorticity and Circulation

At this point, before going further, it is a good idea to define the reference system we will be using most of the time. This is a Cartesian system, as is shown in Fig. 3.8, with the $x$ axis tangent to a parallel and oriented toward east, the $y$ axis tangent to the meridian oriented toward north, and the $z$ axis coincident with the local vertical and oriented upward. The plane $xy$ is then tangent to the surface of the Earth at the point where the axis originates.

The definition we have given for the vorticity can be very easily interpreted in physical terms. To establish if the circulation is different from zero, we can introduce in the fluid a paddle wheel, as shown in Fig. 3.9. This is something like the "rotometer" introduced in electromagnetism. Here, we represent the horizontal flow in a channel. The case to the left could be the one in which we have friction on the walls. If the rotor is zero, the vorticity is zero and then the paddle wheel will not turn. This happens when the velocity field is uniform (i.e., the case on the right). If, as for the case on the right, the velocity changes, that is, a *shear* is present (as the meteorologists say), then vorticity is different from zero, and the wheel will turn because there is a net torque acting on it.

From the definition of the circulation, we can find a property of its rate of change with time. We can write

$$\frac{DC}{Dt} = \frac{D}{Dt}\oint \mathbf{V}\cdot\mathbf{dl} = \oint \frac{D\mathbf{V}}{Dt}\cdot\mathbf{dl} + \oint \mathbf{V}\cdot\frac{D}{Dt}\left(\mathbf{dl}\right) \tag{3.30}$$

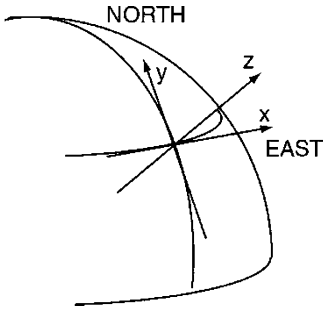



**Fig. 3.8** Illustration of the reference system

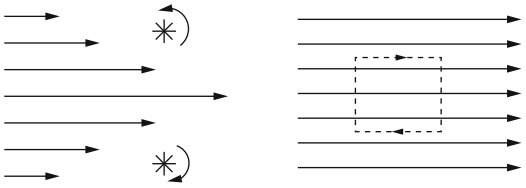


**Fig. 3.9** The vorticity for two different velocity fields of the fluid. To the *left*, the velocity shear introduces a counterclockwise vorticity (positive) in the upper part and clockwise vorticity in the lower part (negative). To the *right*, the integration along the path (and then the vorticity) is zero

The last term is the change with time of the infinitesimal vector that connects adjacent elements of the fluid. As shown in Fig. 3.12, if $\Delta\mathbf{r}$ is such distance then the infinitesimal change is given simply by $\Delta\mathbf{r}\,(t+\Delta t)-\Delta\mathbf{r}(t)=\Delta\mathbf{V}\Delta t$ so that at the limit $\Delta t\to 0$, we have $D\,(D\mathbf{l})\,/Dt=d\mathbf{V}$. This means that $\Gamma$ being a material contour, the infinitesimal change will depend only on the velocity of the fluid.

We get then for the rate of change of the circulation

$$\frac{DC}{Dt}=\oint_{\Gamma}\frac{D\mathbf{V}}{Dt}\cdot d\mathbf{l}+\oint_{\Gamma}\mathbf{V}\cdot d\mathbf{V}=\oint_{\Gamma}\frac{D\mathbf{V}}{Dt}\cdot d\mathbf{l}+\oint_{\Gamma}d|V|^2=\oint_{\Gamma}\frac{D\mathbf{V}}{Dt}\cdot d\mathbf{l} \tag{3.31}$$

Substituting for $D\mathbf{V}/Dt$ and neglecting the friction we have for the rate of change of the circulation

$$\frac{DC}{Dt}=-\oint_{\Gamma}(2\boldsymbol{\Omega}\times\mathbf{V})\cdot d\mathbf{l}-\oint_{\Gamma}\frac{\nabla p}{\rho}\cdot d\mathbf{l} \tag{3.32}$$

This equation simply tells us that there are two ways to change the relative circulation, the first through the Coriolis acceleration and the second because of the acceleration produced by the pressure gradients. The effect of the Coriolis acceleration is interesting because the first term on the right-hand side of Eq. (3.32) is the change of circulation due to the rotation of the planet. This term can be written following a vector rule as

$$-(2\boldsymbol{\Omega}\times\mathbf{V})\cdot d\mathbf{l}=-2\boldsymbol{\Omega}\cdot(\mathbf{V}\times d\mathbf{l})$$

The terms $\mathbf{V}\times\mathbf{dl}$ can be easily interpreted, as shown in Fig. 3.10, as the elementary area swept by the infinitesimal element $d\mathbf{l}$ in the time interval $dt$. The vector has the direction normal to the area itself. Then the scalar product of such a vector for the angular velocity can be written as

$$-\oint_{\Gamma}(2\Omega\times\mathbf{V})\cdot d\mathbf{l}=-2\Omega\frac{D\Sigma_n}{Dt}$$

where $\Sigma_n$ is the area included in the curve $\Gamma$ projected on a plane normal to $\boldsymbol{\Omega}$. We can easily see that in case the curve is drawn on the surface of the Earth, as in

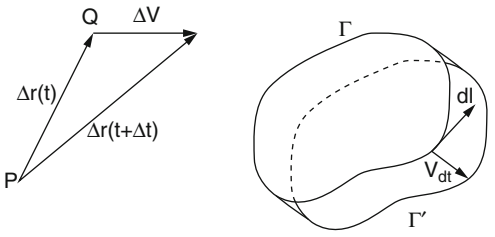



**Fig. 3.10** The infinitesimal change of vector $\Delta\mathbf{r}$ (*left*). The contour $\Gamma$ goes in the contour $\Gamma'$ after the time $\Delta t$ (*right*)

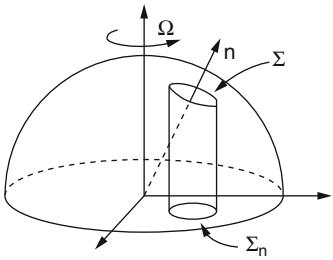



**Fig. 3.11** Illustration of the Bjerknes theorem

Fig. 3.11, and the area is $\Sigma$, then what counts is the projection on the equatorial plane that has the value $\Sigma_n = \Sigma \sin\phi$. The change of the relative vorticity is

$$\frac{DC}{Dt} = -2\Omega \frac{D\Sigma_n}{Dt} \tag{3.33}$$

This equation expresses *the Bjerknes circulation theorem*. If Eq. (3.36) is integrated from an initial state 1 to a final state 2 to which corresponds the areas $\Sigma_1$ and $\Sigma_2$ at the latitudes $\phi_1$ and $\phi_2$, respectively, we obtain

$$C_2 - C_1 = -2\Omega \left(\Sigma_2 \sin\phi_2 - \Sigma_1 \sin\phi_1\right)$$

which relates the change in the circulation to changes in latitude and area. Actually, Eq. (3.36) can be read in a different way. Keeping in mind Eq. (3.24), we can easily find a relation between relative circulation and absolute circulation

$$\frac{DC}{Dt} = \frac{DC_a}{Dt} - 2\Omega \frac{D\Sigma_n}{Dt}$$

that is essentially Eq. (3.33).

At this point we can discuss the other term appearing in Eq. (3.34), related to the pressure gradient. We use the Stokes theorem

$$-\int_\Gamma \frac{\nabla p}{\rho} \cdot d\mathbf{l} = -\int_\Gamma \left(\nabla \times \left(\frac{\nabla p}{\rho}\right)\right) \cdot \mathbf{n} dS$$

where

$$\left(\nabla \times \left(\frac{\nabla p}{\rho}\right)\right) = \frac{1}{\rho} \nabla \times \nabla p - \nabla p \times \nabla \left(\frac{1}{\rho}\right)$$

The first term on the right-hand side is zero so that the circulation of the pressure term becomes

$$-\int_\Gamma \frac{\nabla p}{\rho} \cdot d\mathbf{l} = -\int_\Gamma \frac{\nabla p \times \nabla \rho}{\rho^2} \cdot \mathbf{n} dS \tag{3.34}$$

This equation can be simplified when a relation can be established between $p$ and $\rho$, like $\rho = \rho(p)$. In this case the fluid is called *barotropic*, and a typical example is a liquid inside a container, such as water in the bathtub. At each level at depth $h$ with respect to the surface, we can write $p = \rho gh$. However, also for an isothermal atmosphere, pressure is simply proportional to density. In this case the integral appearing on the left of Eq. (3.37) is zero because

$$\oint_\Gamma \frac{\nabla p}{\rho} \cdot d\mathbf{l} = \oint_\Gamma \frac{dp}{\rho(p)} = 0$$

In general, isobars (constant pressure surfaces) are not parallel to constant density surfaces (isopycnics). In this case the fluid is called *baroclinic* and the vector $\nabla p \times \nabla \rho$ is not zero and so neither is the integral.

### 3.3.2 The Vorticity Equation

We can define a *current cylinder* as the surface formed by the streamlines that intersect with a closed curve normal to the streamlines, as shown in Fig. 3.12. In a similar manner, it is possible to define a *vortex line* or *filament* as the curve tangent to the vorticity vector at each point. The corresponding cylinder is a *vortex tube*. These elements have some properties that derive from the rules of vector analysis. For example, if we want to evaluate the flux of the velocity vector, we solve the integral

$$\int \mathbf{V} \cdot \mathbf{n} dA = \int \nabla \cdot \mathbf{V} dV$$

For zero divergence, this equation gives a constant flux along the tube defined by the velocity vector: then, where the section increases, the velocity decreases, and vice versa.

In the case of a vortex tube, we can evaluate in a similar way the vorticity flux to obtain

$$\boldsymbol{\omega_1} \cdot \mathbf{n} dA_1 = \boldsymbol{\omega_2} \cdot \mathbf{n} dA_2 \tag{3.35}$$

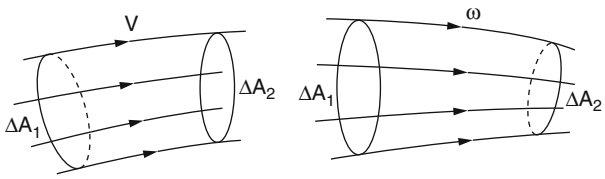



**Fig. 3.12** A current tube (*left*) and a vortex tube (*right*)

If the mass contained in the tube is $dM$, the conservation of the circulation that is implicit in Eq. (3.35) means that the ratio $C/dM$ is constant, so that if $\zeta$ is the vorticity component normal to the elementary area, we must have

$$C/dM = \zeta dS/\rho dl dS = \zeta/\rho dl \tag{3.36}$$

This means that the length of the tube will increase if the vorticity increases, and vice versa. The consequence of this property is evident if we recur to the equation of motion written with all its terms

$$\left(\frac{\partial \mathbf{V}}{\partial t}\right) + \mathbf{V}\cdot\nabla\mathbf{V} + 2\boldsymbol{\Omega}\times\mathbf{V} = -\frac{\nabla p}{\rho} + \nabla\Phi + \frac{\mathbf{F}_r}{\rho} \tag{3.37}$$

We notice that the quantity

$$(\nabla\times\mathbf{V})\times\mathbf{V} = (\mathbf{V}\cdot\nabla)\,\mathbf{V} - (1/2)\,\nabla\left(|\mathbf{V}|^2\right)$$

and the advection term in Eq. (3.37) can be substituted by defining the vorticity vector as $\boldsymbol{\omega} = \nabla\times\mathbf{V}$

$$\left(\frac{\partial \mathbf{V}}{\partial t}\right) + (2\boldsymbol{\Omega}+\boldsymbol{\omega})\times\mathbf{V} = -\frac{\nabla p}{\rho} + \nabla\left(\Phi - \frac{1}{2}|\mathbf{V}|^2\right) + \frac{\mathbf{F}_r}{\rho}$$

To find the vorticity tendency, we need only to calculate the curl of the above equation

$$\nabla\times\left(\frac{\partial \mathbf{V}}{\partial t}\right) + \nabla\times((2\boldsymbol{\Omega}+\boldsymbol{\omega})\times\mathbf{V}) = -\nabla\times\left(\frac{\nabla p}{\rho}\right) + \nabla\times\nabla\left(\Phi - \frac{1}{2}|\mathbf{V}|^2\right) + \nabla\times\left(\frac{\mathbf{F}_r}{\rho}\right)$$

The first term on the left-hand side is simply the local derivative of the vorticity, while the second can be written as

$$\nabla\times((2\boldsymbol{\Omega}+\boldsymbol{\omega})\times\mathbf{V}) = (2\boldsymbol{\Omega}+\boldsymbol{\omega})\,\nabla\cdot\mathbf{V} + (\mathbf{V}\cdot\nabla)\,(2\boldsymbol{\Omega}+\boldsymbol{\omega}) - (2\boldsymbol{\Omega}+\boldsymbol{\omega})\cdot\nabla\mathbf{V}$$

where we have taken into account that $\nabla\cdot(2\boldsymbol{\Omega}+\boldsymbol{\omega}) = 0$. We have then

$$\frac{\partial}{\partial t}(2\boldsymbol{\Omega}+\boldsymbol{\omega}) + (\mathbf{V}\cdot\nabla)\,(2\boldsymbol{\Omega}+\boldsymbol{\omega}) + (2\boldsymbol{\Omega}+\boldsymbol{\omega})\,\nabla\cdot\mathbf{V} - (2\boldsymbol{\Omega}+\boldsymbol{\omega})\cdot\nabla\mathbf{V} = \frac{\nabla\rho\times\nabla p}{\rho^2} + \nabla\times\left(\frac{\mathbf{F_r}}{\rho}\right) \tag{3.38}$$

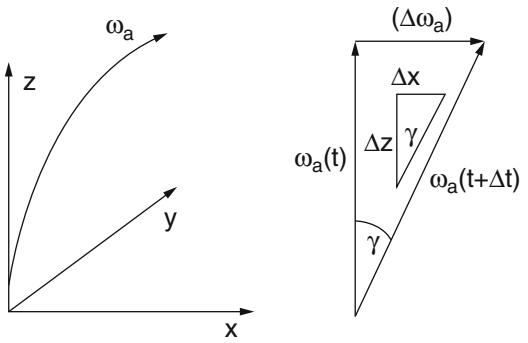



**Fig. 3.13** The coordinate system to illustrate the changes of vorticity (*left*). On the *right* the change in direction is shown

The quantity $2\boldsymbol{\Omega} + \boldsymbol{\omega} = \boldsymbol{\omega}_a$ is the absolute vorticity so that the second term on the left represents now the advection of the absolute vorticity. The vorticity equation becomes then

$$\frac{D\boldsymbol{\omega}_a}{Dt} + \boldsymbol{\omega}_a \nabla \cdot \mathbf{V} - \boldsymbol{\omega}_a \cdot \nabla \mathbf{V} = \frac{\nabla\rho \times \nabla p}{\rho^2} + \nabla \times \left(\frac{\mathbf{F_r}}{\rho}\right) \tag{3.39}$$

This equation tells us that the vorticity is generated by the baroclinic term (the non-parallelism between the pressure and density gradients), but it is also influenced by the two terms on the left-hand side.

To understand how these terms work, we can choose a very peculiar reference system, such that the $z$ axis is tangent to a vorticity filament (see Fig. 3.13). In this case the absolute vorticity can be written $\boldsymbol{\omega}_\mathrm{a} = \mathbf{k}\,\omega_\mathrm{a}$. The second and third term in Eq. (3.39) become

$$\begin{aligned} &\boldsymbol{\omega}_a \cdot \nabla \mathbf{V} - \boldsymbol{\omega}_a \nabla \cdot \mathbf{V} \\ &= \omega_\mathbf{a} \frac{\partial}{\partial z}(u\mathbf{i} + v\mathbf{j} + w\mathbf{k}) - \omega_a \mathbf{k}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) \\ &= \mathbf{i}\omega_a \frac{\partial u}{\partial z} + \mathbf{j}\omega_a \frac{\partial v}{\partial z} - \omega_a \mathbf{k}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \end{aligned} \tag{3.40}$$

The interpretation of these two terms is rather interesting. Starting from the second and being interested only in the changes of the $z$ component, we obtain

$$\frac{D(\zeta + f)}{Dt} = -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{3.41}$$

This equation indicates that when the divergence is negative, the vorticity increases, and vice versa. This is the same as saying that if the cross section of a vortex tube decreases, its rate of rotation increases.

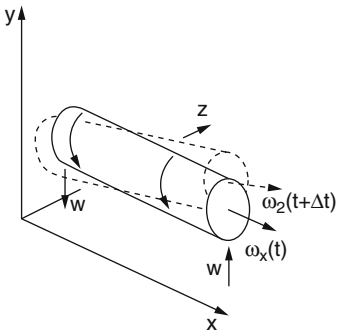



**Fig. 3.14** The change in inclination of a vortex tube caused by the horizontal component of the vorticity in the presence of a gradient of the vertical wind

The first term in Eq. (3.40) gives rise to another interesting effect on one of the components normal to the $z$ axis, for example, $\omega_x$. In this case we have for the change of the component

$$D(\omega_a)_x/Dt = \omega_a \partial u/\partial z$$

and in the time $\Delta t$ the fractional change will be

$$\frac{(\Delta\omega_a)_x}{\omega_a} = \frac{\partial u}{\partial z}\Delta t = \frac{\Delta x}{\Delta z} \tag{3.42}$$

As shown in Fig. 3.14, this corresponds to a rotation of the vortex tubes through an angle whose tangent is $\Delta x/\Delta z$.

The conclusion at this point is that the net production of vorticity is due to the baroclinic term, but its dissipation is due to the effects of stretching and rotation of the vortex tubes rather than the friction forces.

In the application of Eq. (3.38) to the atmosphere, we only consider the $z$ component so that it can be rewritten for this component. A complicated term to write may be

$$\boldsymbol{\omega}_a \cdot \nabla \mathbf{V} - \boldsymbol{\omega}_a \nabla \cdot \mathbf{V}$$

because we need to consider also the horizontal components of the vorticity, that is,

$$\omega_x = \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}; \qquad \omega_y = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}$$

so that the term becomes

$$(\omega_a \cdot \nabla \mathbf{V} - \omega_a \nabla \cdot \mathbf{V})_z = -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right)\frac{\partial w}{\partial x} + \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right)\frac{\partial w}{\partial y}$$

As we have noted earlier, the ratio $w/u << 1$, so that all the products involving $w$ can be neglected, with the result

$$(\boldsymbol{\omega}_a \cdot \nabla \mathbf{V})_z = -(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \left(\frac{\partial u}{\partial z}\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\frac{\partial w}{\partial x}\right)$$

In particular, the second term can be written as

$$\mathbf{k} \cdot \left(\frac{\partial \mathbf{V}}{\partial z} \times \nabla w\right)$$

Keeping all the pieces together, we have for the rate of change of the absolute vorticity $\eta = \zeta + f$

$$\frac{D\eta}{Dt} = -\eta\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \mathbf{k} \cdot \left(\frac{d\mathbf{V}}{dz} \times \nabla w\right) + \mathbf{k} \cdot \left(\frac{\nabla \rho \times \nabla p}{\rho^2}\right) + \mathbf{k} \cdot \left(\frac{\mathbf{F}_r}{\rho}\right) \tag{3.43}$$

This equation, even if written only for $z$, contains as a particular case the conservation of the absolute vorticity. This happens when the divergence is zero (first term on the right), when the motion is horizontal (second term zero), when the fluid is barotropic (third term zero), and finally when there is no friction (last term).

Equation (3.43) contains a stretching term (first term on the right-hand side) and rotation of the vortex line (second term). In Fig. 7.6 again the change in direction of the axis of a vortex tube is shown and is due to a shear of the vertical wind component. In this case the change in the relative vorticity due only to this term is given by

$$\frac{D\zeta}{Dt} = \omega_x \frac{\partial w}{\partial x}$$

As a consequence there is a vertical component of the vorticity given by

$$\Delta\zeta = \omega_x \frac{\partial w}{\partial x}\Delta t = \omega_x \frac{\Delta z}{\Delta x}$$

It is interesting at this point to proceed with a scale analysis of Eq. (3.43). Again, if we indicate with $L$, $H$, and $U$ as the horizontal scale, the vertical scale, and the horizontal velocity, respectively, we have

$$\frac{\partial \zeta}{\partial t} \approx \frac{U}{L}/\frac{L}{U} \approx \left(\frac{U}{L}\right)^2 \tag{3.44}$$

The divergence term needs a careful examination. First of all, we notice that Il $\zeta \approx U/L$ so that

$$\zeta/f \approx U/fL \approx Ro << 1 \tag{3.45}$$

In the divergence term, we can then neglect the relative vorticity with respect to the planetary vorticity, and because the left- and right-hand sides must be of the same order, we have

$$f\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)\approx\left(\frac{U}{L}\right)^2\Rightarrow\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)\approx Ro\zeta \tag{3.46}$$

This means that the ratio between divergence and vorticity is of the same order as the Rossby number. Notice that the equilibrium between the rate of change of vorticity and the divergence term is easily reached, because from the observations we have $U/L\approx 10^{-5}$, so that the divergence must be small. We can also estimate the order of magnitude of the vertical velocity. We have

$$\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)\approx-\frac{\partial w}{\partial z}\Rightarrow w\approx\frac{H}{f}\left(\frac{U}{L}\right)^2$$

This is because from Eq. (3.46) the horizontal divergence is of the order of $(U/L)^2/f$. We have then an interesting relation between Rossby number and vertical velocity

$$w/U\approx RoH/L \tag{3.47}$$

that is more restrictive than the one we found for a thin fluid. We can now proceed to the evaluation of the second term on the right-hand side of Eq. (3.43), simply noting that

$$\nabla w\approx\frac{w}{L}\;;\qquad\nabla w\frac{\partial v}{\partial z}\approx\frac{HU^2}{fL^3}\frac{U}{H}\approx\frac{1}{f}\left(\frac{U}{L}\right)^3 \tag{3.48}$$

Finally, we treat the baroclinic term, which needs a little elaboration. From geostrophic equilibrium and simple thermodynamics, we get

$$\nabla p/\rho\approx fU\;;\qquad\Delta\rho/\rho\approx-\Delta T/T;\quad\nabla\rho/\rho\approx fU/(gH)$$

the last from thermal wind. Finally,

$$\frac{\nabla\rho\times\nabla p}{\rho^2}\approx\frac{f^2U^2}{gH} \tag{3.49}$$

If these terms are compared with the divergence term or simply to $(U/L)^2$, we see that in Eq. (3.43) are at the least one order of magnitude smaller than the divergence term. As a first approximation, we can assume that the absolute vorticity is balanced by the stretching of the vortex tubes, so that Eq. (3.43) can be written as

$$\frac{D\,(\zeta+f)}{Dt}\approx-(\zeta+f)\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right) \tag{3.50}$$

Actually, the total derivative contains also the vertical advection term that we neglect, so that Eq. (3.50) is even simpler, and with the notation

$$d_h/dt = \partial/\partial t + u\partial/\partial x + v\partial/\partial y$$

we have

$$\frac{d_h\,(\zeta + f)}{dt} = -\,(\zeta + f)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{3.51}$$

If the total divergence is zero, this becomes

$$\frac{d_h\,(\zeta + f)}{dt} = -\,(\zeta + f)\,\frac{\partial w}{\partial z} \tag{3.52}$$

We have now laid all the basic stuff about dynamics, and the next chapter will show some application of the knowledge we gained.

## E.3 Examples

### *E.3.1 The Coriolis Acceleration*

The Coriolis acceleration arises when dealing with non-inertial systems (the Earth is one of them) and can be introduced in different ways.Suppose we set in motion a point at the pole toward the equator. An observer at rest with the "fixed stars" looking down at the pole will see a trajectory like the one shown in Fig. E.3.1a. For an observer at rest on the Earth, the trajectory will be like the one in Fig. E.3.1b. For an inertial observer there are no forces because the trajectory is a straight line, while the non-inertial observer will interpret the curved trajectory as produced by an acceleration to which he can assign a precise value. Based on the deviation, we see the Coriolis acceleration as due to a centrifugal force associated with the rotation of the Earth augmented by the rate $u/R$, where $u$ is the velocity component along the parallel and $R$ the distance between the point and axis of rotation. If we

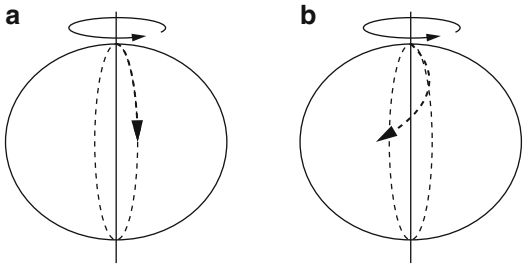



**Fig. E.3.1** The trajectory (represented with the *heavy line*) of a point moving from the pole to the equator. (**a**) The trajectory for an inertial system. (**b**) The trajectory for a system that is fixed with the Earth

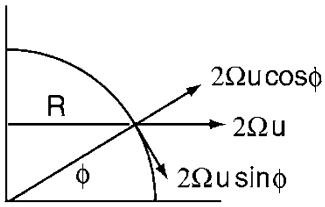



**Fig. E.3.2** The component of the Coriolis acceleration for motion along a parallel

refer to Fig. E.3.2 and considering carefully the different directions, we find for the centrifugal force

$$\left(\Omega + \frac{u}{R}\right)^2 \mathbf{R} = \Omega^2 \mathbf{R} + \frac{2\Omega u \mathbf{R}}{R} + \left(\frac{u}{R}\right)^2 \mathbf{R} \tag{E.3.1}$$

The terms on the right can be interpreted (they are all in the same directions and are oriented as **R**). The first is the acceleration that results from the rotation of the Earth, while the third is the same acceleration due to the additional rotation $u/R$. Usually for the atmospheric motions $|u| << \Omega R$ and so this term is negligible. The term at the center, on the other hand, is the Coriolis acceleration. From Fig. E.3.2 we see it has two components, one along the local vertical and one the along the meridional direction, and this has the value (defined positive toward the north)

$$\left(\frac{dv}{dt}\right)_{Co} = -2\Omega u \sin\phi \quad ; \qquad \left(\frac{dw}{dt}\right)_{Co} = 2\Omega u \cos\phi \tag{E.3.2}$$

This means that an air parcel that moves along the parallel from west to east undergoes acceleration toward the south, and vice versa. If the motion is along the meridian, to obtain the Coriolis acceleration we can invoke the conservation of angular momentum. We assume that the air parcel moves from radius $R$ and latitude $\phi_0$ to the radius $R + \delta R$ and latitude $\phi_0 + \delta\phi$, toward the equator. In this case the air parcel moves toward zones with high angular momentum so that its velocity will change by $\delta u$ in order to conserve the angular momentum. We have

$$\Omega R^2 = \left(\Omega + \frac{\delta u}{R + \delta R}\right)(R + \delta R)^2$$

We write the square as $(1 + 2\delta R/R)$, and neglecting higher-order terms, we obtain $\delta u = -2\Omega\delta R$ where $\delta R = -a\delta\phi sen\phi_0$ with $a$ as the radius of the Earth. We have then for the acceleration

$$\left(\frac{du}{dt}\right)_{Co} = 2\Omega a \frac{d\phi}{dt} \sin\phi_0 = 2\Omega v \sin\phi \tag{E.3.3}$$

where $v = ad\phi/dt$ is the velocity component in the north direction.

When a parcel moves along the meridian, acceleration is generated that acts along a parallel. The horizontal accelerations are proportional to the quantity $2\Omega \sin\phi$, which is called the *Coriolis parameter*, denoted by the symbol $f$. There is a more elegant way to obtain the Coriolis acceleration using a coordinate transformation.

### *E.3.2 The Inertial Oscillation*

It is rather interesting to study the motion of a parcel in the atmosphere subjected only to the Coriolis acceleration. Equation (3.25) in this case simplifies as

$$\begin{aligned} \frac{Du}{Dt} - fv &= 0 \\ \frac{Dv}{Dt} + fu &= 0 \end{aligned} \tag{E.3.4}$$

The first equation can be derived with respect to time, while the second can be used to eliminate $Dv/Dt$ so to obtain

$$\frac{D^2u}{Dt^2} + f^2u = 0$$

which has the solution

$$u(t) = u_0 \cos ft; \quad v(t) = v_0 \sin ft$$

with initial conditions $u(0) = u_0; \quad v(0) = 0$. If the air parcel is located at the origin that follows the inertial circle trajectory

$$x(t) = (u_0/f)\cos ft; \quad y(t) = (u_0/f)(\cos ft - 1) \tag{E.3.5}$$

The trajectory is a circle of radius $u_0/f$ and the particles move anticlockwise. It is interesting to note that the main role in producing the inertial oscillation is due to gravity. This can be understood by referring to Fig. E.3.3. The composition of the true gravity ($\mathbf{g}$) with the centrifugal force ($\boldsymbol{\Omega} \times \boldsymbol{\Omega} \times \mathbf{R}$) gives rise to an apparent gravity ($\mathbf{g_a}$). The normal at each latitude to the apparent gravity defines a new oblate equipotential surface called geoid. On the geoid the equator is "uphill" and the mass will tend to fall over the pole. The balance between the components of the centrifugal force and the true gravity keeps the mass in place.

It can be easily shown that from the right of Fig. E.3.3, the component of the true gravity tangent to the geoid is simply

$$g \sin\delta \approx g\delta = \Omega^2 a \sin\varphi \cos\varphi \tag{E.3.6}$$

which has a maximum at 45° where the angle $\delta$ is only 6 min or arc. As a consequence the tangent component is only 0.017 m s$^{-2}$.

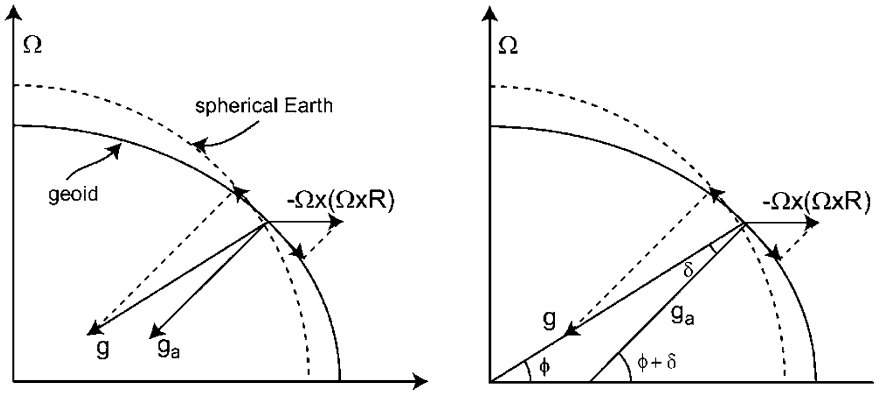



**Fig. E.3.3** *The* figure of the terrestrial geoid (*left*). The *solid line* represents the surface normal at each point to the apparent acceleration of gravity. The geometry of the two gravity accelerations

It is rather interesting to consider now the trajectory in a nonrotating reference frame. We use the plane tangent to the north pole and write the absolute acceleration in the fixed reference frame

$$\frac{d\mathbf{V}_f}{dt} = \frac{d\mathbf{V}_r}{dt} + 2\boldsymbol{\Omega} \times \mathbf{V}_r + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}) \tag{E.3.7}$$

In the rotating system, the Coriolis acceleration equilibrates the relative accelerations so that the first two terms on the right-hand side are zero and we get $\boldsymbol{\Omega}$

$$\frac{d\mathbf{V}_f}{dt} = \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}) \tag{E.3.8}$$

The term on the right cannot be an apparent force because we are in the inertial frame and actually is the component of the centrifugal force normal to the apparent gravity shown in Fig. E.3.3. Equation (E.3.8) can be projected on the two horizontal axis x and y to obtain

$$\frac{d^2x}{dt^2} + \Omega^2 x = 0; \quad \frac{d^2y}{dt^2} + \Omega^2 y = 0 \tag{E.3.9}$$

The fixed coordinate trajectory for a parcel leaving the north pole with initial velocities $u = u_0; v = 0$ is

$$x = \frac{2u_0}{f} \sin\left(\frac{ft}{2}\right); \quad y = 0 \tag{E.3.10}$$

where we have used the condition $\Omega = f/2$. The trajectory described by (E.3.10) is a straight line along which the parcel oscillates with a period $4\pi f^{-1}$. On the other hand the trajectory in the rotating coordinate system is described by (E.3.5).

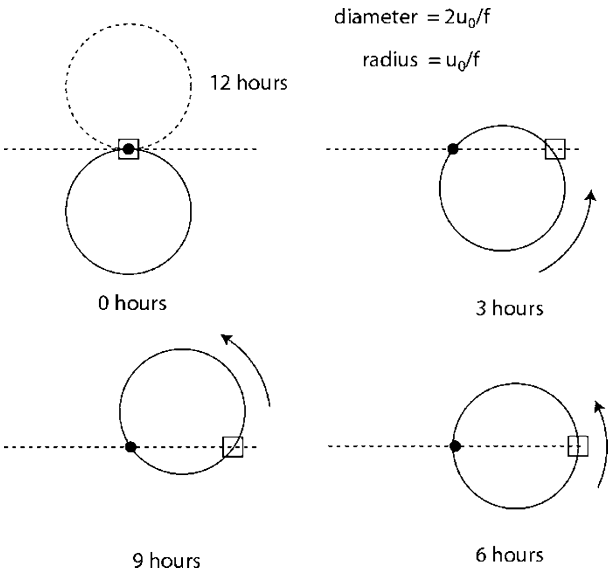



**Fig. E.3.4** Relative position of the fixed and rotating coordinate trajectories. The parcel position is shown by the *open square*. The north pole position by the *heavy dot* (Adapted from Durran 1993)

Figure E.3.4 shows the two trajectories. The linear oscillation happens along the dotted line with the parcel starting from the north pole (heavy dot). The rotating coordinate trajectory is given by the solid circle which moves in counterclockwise movement around the north pole.

### *E.3.3 The Rossby Adjustment Problem (Nonrotating)*

We consider a fluid of constant density that presents some discontinuity at the surface. Our intention is to show how those discontinuities are smoothed out. The first step is to consider a nonrotating fluid so that the acceleration is determined only by the pressure gradient. We have then

$$\begin{aligned} \partial u/\partial t &= -(1/\rho)\,\partial p/\partial x \\ \partial v/\partial t &= -(1/\rho)\,\partial p/\partial y \end{aligned} \tag{E.3.11}$$

To simplify things we assume that the total height of the fluid is $H + \eta$ with $\eta << H$ where H is a constant. Equation (E.3.11) becomes

$$\begin{aligned} \partial u/\partial t &= -g\partial\eta/\partial x \\ \partial v/\partial t &= -g\partial\eta/\partial y \end{aligned} \tag{E.3.12}$$

where we have assumed $p = \rho g \eta$. Equation (E.3.12) shows that currents are independent of depth so that the continuity equation can be integrated with respect to $z$ to obtain

$$\int_0^{H+\eta} (\partial u/\partial x + \partial v/\partial y + \partial w/\partial z)\, dz = \partial \eta/\partial t + H\,(\partial u/\partial x + \partial v/\partial y) = 0$$

Deriving the first of Eq. (E.3.11) with respect to $x$ and the second with respect to $y$, we obtain

$$\partial^2 \eta/\partial t^2 = gH\left(\partial^2 \eta/\partial x^2 + \partial^2 \eta/\partial y^2\right) = c^2 \nabla^2 \eta \tag{E.3.13}$$

where $c^2 = gH$ is the square of the phase velocity. The solution of this equation when there is no $y$ dependence is given by

$$\eta = \tfrac{1}{2}\left[G\,(x - ct) + G\,(x + ct)\right]$$

that is, a form of progressive and retrograde wave. The velocity can be obtained using the first of (E.3.12). Then we have

$$\partial u/\partial t = -g\tfrac{1}{2}\partial/\partial x\left[G\,(x + ct) + G\,(x - ct)\right]$$

With a simple integration with respect to time, we get the result

$$u = -\tfrac{1}{2}c^{-1} g\left[G\,(x + ct) - G\,(x - ct)\right] \tag{E.3.14}$$

This can be easily understood if we assume $G\,(x \pm ct) = \exp\,(x \pm ct)$. A very interesting initial condition can be obtained using

$$\eta = \eta_0\, \mathrm{sgn}(x) \quad \text{where} \quad \mathrm{sgn}(x) = \begin{cases} 1 \ \text{ for } x > 0 \\ -1 \text{ for } x < 0 \end{cases} \tag{E.3.15}$$

The solution is illustrated in Fig. E.3.5 where the initial condition is shown together with the time evolution. The initial discontinuity propagates in opposite direction with velocity $c$ leaving the fluid surface unperturbed but with a steady motion from left to right with velocity $g\eta_0/c$.

### *E.3.4 The Rossby Adjustment Problem (Rotating Case)*

In the rotating case, we should consider the conservation of potential vorticity, so we will start with a simplified version of Eq. (3.41) (Gill 1982)

$$\partial \zeta/\partial t + f\,(\partial u/\partial x + \partial v/\partial y) = 0 \tag{E.3.16}$$

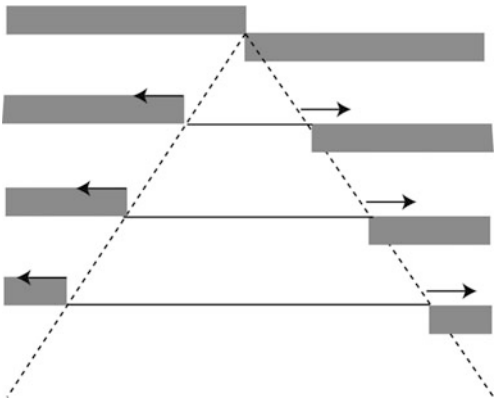


**Fig. E.3.5** The evolution of the initial perturbation of the surface of a nonrotating fluid

The horizontal divergence is eliminated with the continuity equation, so we get

$$\partial u/\partial x + \partial v/\partial y = -(1/H)\,\partial \eta/\partial t$$

$$\frac{\partial}{\partial t}\left(\frac{\zeta}{f} - \frac{\eta}{H}\right) = 0 \tag{E.3.17}$$

This equation expresses a particular case of the conservation of potential vorticity, and the quantity

$$Q' = \zeta/H - f\eta/H^2 \tag{E.3.18}$$

may be called perturbation potential vorticity that remains constant for every instant of time that is

$$Q'(x, y, t) = Q'(x, y, 0) \tag{E.3.19}$$

And in particular for initial conditions $u = v = 0$, we have

$$\zeta/f - \eta/H = (\eta_0/H)\,\mathrm{sgn}(x) \tag{E.3.20}$$

At this point the same as Eq. (E.3.13) can be obtained starting from

$$\begin{aligned} \partial u/\partial t - fv &= -g\partial \eta/\partial x \\ \partial v/\partial t + fu &= -g\partial \eta/\partial y \end{aligned} \tag{E.3.21}$$

We obtain substituting the horizontal divergence for the rotating case

$$\partial^2 \eta/\partial t^2 - c^2\left(\partial^2 \eta/\partial x^2 + \partial^2 \eta/\partial y^2\right) + fH\zeta = 0 \tag{E.3.22}$$

And again, substituting the potential vorticity from (E.3.20), we obtain

$$\partial^2\eta/\partial t^2 - c^2\left(\partial^2\eta/\partial x^2 + \partial^2\eta/\partial y^2\right) + f^2\eta = -f^2\eta_0\,\mathrm{sgn}(x) \tag{E.3.23}$$

We can work out a steady-state solution of this equation by rewriting it:

$$-c^2\left(\partial^2\eta/\partial x^2 + \partial^2\eta/\partial y^2\right) + f^2\eta = -f^2\eta_0\,\mathrm{sgn}(x)$$

And in case there is no dependence on $y$, we get

$$-c^2\partial^2\eta/\partial x^2 + f^2\eta = -f^2\eta_0\,\mathrm{sgn}(x) \tag{E.3.24}$$

A solution for this equation can be found for the two half space

$$\frac{\eta}{\eta_0} = \begin{cases} -1 + \exp\left(-x/L_\rho\right) & for\ \ x > 0 \\ 1 - \exp\left(x/L_\rho\right) & for\ \ x < 0 \end{cases}$$

where

$$L_\rho = c/\left|f\right| = (gH)^{1/2}/f \tag{E.3.25}$$

is the Rossby radius of deformation. It is quite interesting to find the solution for the velocity that can be done considering that zero acceleration in (E.3.21) is obtained when the Coriolis balances the pressure gradient. It is found that the velocity along $y$ is given by

$$v = (g/f)\,\partial\eta/\partial x = -\left(g\eta_0/fL_\rho\right)\exp\left(-\left|x\right|/L_\rho\right) \tag{E.3.26}$$

It is to note that the motion of the fluid is not along the direction of the pressure gradient but is perpendicular to it. As we will see this is a characteristic of the geostrophic motion the simplest approximation when the pressure gradient is balanced by the Coriolis acceleration. The consequence is a strong jet perpendicular and out of the page. The initial discontinuity is smoothed out with a characteristic length given by the Rossby radius of deformation (Fig. E.3.6).

### *E.3.5 Energetics of the Adjustment*

In the nonrotating case, the initial perturbation potential energy is given by

$$\int_0^{\eta_0} \rho g z dz = \tfrac{1}{2}\rho g\eta_0^2 \tag{E.3.27}$$

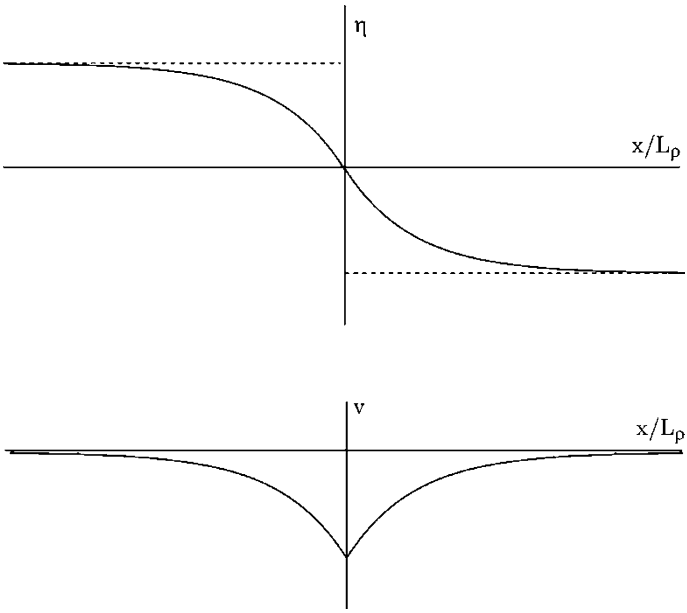



**Fig. E.3.6** The stationary solution of the adjustment for a rotating fluid. The initial condition is shown by *dashed line*. The *upper figure* represents the elevation, while the *bottom* shows the velocity normal to x

After the wave has passed, the kinetic energy has reached a value

$$\tfrac{1}{2}H\rho u^2 = \tfrac{1}{2}H\rho(g\eta_0/c)^2 = \tfrac{1}{2}\rho g\eta_0^2 \tag{E.3.28}$$

All the potential energy has been converted in kinetic energy. In the rotating case, however, the potential energy released per unit length is given by

$$P.E. = 2\tfrac{1}{2}\rho g\eta_0^2 \int_0^\infty \left[1 - \left(1 - \mathrm{e}^{-x/L_\rho}\right)^2\right] dx = \tfrac{3}{2}\rho g\eta_0^2 L_\rho \tag{E.3.29}$$

On the other hand, the kinetic energy per unit length in the equilibrium solution is given by

$$K.E. = 2\tfrac{1}{2}\rho H\left(g\eta_0/fL_\rho\right)^2 \int_0^\infty \mathrm{e}^{-2x/L_\rho} dx = \tfrac{1}{2}\rho g\eta_0^2 L_\rho \tag{E.3.30}$$

This represents only one third of the potential energy. The remaining energy could be taken into account only if the transient problem is worked out that is how the steady state is reached.

# References[1]

## *Books*

Cushman-Roisin B, Beckers JM (2011) Geophysical fluid dynamics. Academic, Waltham
Dutton JA (1976) The ceaseless wind: an introduction to the theory of atmospheric motion. McGraw Hill, New York
Gill AE (1982) Atmosphere-ocean dynamics. Academic, New York
Holton JR (1992) An introduction to dynamic meteorology. Academic, New York
Houghton JT (1977) The physics of the atmospheres. Cambridge University Press, Cambridge/New York
Pedlosky J (1987) Geophysical fluid dynamics. Springer, New York
Stull RB (2000) Meteorology for scientists and engineers. Brooks/Cole, Pacific Grove

## *Articles*

Durran DR (1993) Is the coriolis force really responsible for the inertial oscillation. Bull Am Meteor Soc 74:2179

[1]This introductory chapter to dynamics follows Holton and Pedlosky books. Some new examples have been added that are taken from Gill. Papers buried in the literature are very useful.

# Chapter 4
# Dynamics: Few Simple Applications

The plan for this book is to make a very rapid introduction to the very basic topic (thermodynamics, radiation, etc.) so that we could start as soon as possible to get serious. On the other hand, Chap. 3 was already too heavy with dynamics, and we thought that something in between very serious applications and simple exercises would be a little bit relaxing. We then start with a scale analysis of the equation of motions and then proceed to the most simple application of what we learned in the previous chapter.

## 4.1 The Geostrophic Motion

The simplest approximation to solve the equation of motion, Eq. (3.25), is known as *geostrophic*. This means in practice that the balance occurs between the pressure and the Coriolis forces. The reason for that is again based on the scale analysis, as we have seen before. We refer to a characteristic length $L$ that sets the horizontal scale of the motion and to a dimension $D$ that sets the depth of the fluid. We can find the order of magnitude of the acceleration with respect to the Coriolis term:

$$\frac{|D\mathbf{V}/dt|}{|2\boldsymbol{\Omega} \times \mathbf{V}|} = O\left[\frac{U^2/L}{\Omega U}, \frac{U}{\Omega L}\right] \tag{4.1}$$

This means that the ratio between acceleration and Coriolis term is of the same order of magnitude of the Rossby number and can be neglected as a first approximation. With the same argument, we can establish that the accelerations produced by the friction forces are negligible. Referring to Eq. (3.17), we have

$$\frac{F_r}{\rho} = \left[v\frac{U}{L^2}\right]$$



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_4

The ratio between the acceleration due to the friction and the Coriolis term is known as the *Ekman number E*

$$E - \frac{vU/L^2}{\Omega U} = \frac{v}{\Omega L^2} \tag{4.2}$$

We have seen that viscosity is of the order of $10^{-5}$ m$^2$ s$^{-1}$; thus with $L = 10^3$ km, we have $E = 10^{-13}$, that is, a very small number indeed. Carrying out further the scale analysis, it is clear that the ratio between vertical velocity $w$ and the horizontal components $u$, $v$ is of the order

$$\frac{w}{u} \approx O\left[\frac{D}{H}\right] \ll 1 \tag{4.3}$$

At this point, Eq. (3.25) can be simplified drastically and reduces to

$$\rho 2\boldsymbol{\Omega} \times \mathbf{V} = -\nabla p + \nabla \Phi \tag{4.4}$$

If this equation is applied to a thin fluid of thickness $D$ on the surface of the Earth, as in Fig. 4.1, then the same surface can be assumed at constant potential so that $\nabla\Phi$ gives exactly the acceleration $g$. In the surroundings of $L$ the surface can be assumed flat. This actually is not just a geometrical approximation but has to do with the Coriolis parameter which is a function of latitude. The flat Earth approximation at this stage corresponds to keeping the Coriolis parameter constant with latitude. In this case, we can write Eq. (4.4) as

$$\begin{aligned} -2\Omega v \sin\varphi + 2\Omega w \cos\varphi &= -\frac{1}{\rho}\frac{\partial p}{\partial x} \\ 2\Omega u \sin\varphi &= -\frac{1}{\rho}\frac{\partial p}{\partial y} \\ 2\Omega u \cos\varphi &= -\frac{1}{\rho}\frac{\partial p}{\partial z} + g \end{aligned} \tag{4.5}$$

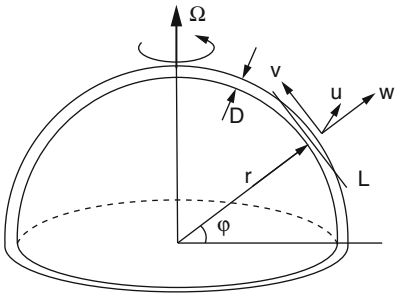



**Fig. 4.1** The approximation used for the geostrophic equilibrium. The thickness of the atmosphere has been exaggerated

At this point, we can assume that in the absence of motion, the pressure is a function only of $z$ so we can write

$$p = p_s(z) + p'(x, y, z) \qquad \rho = \rho_s(z) + \rho'(x, y, z)$$

where $p_s(z)$ and $\rho_s(z)$ are density and pressure when the atmosphere is at rest, while $p'$ and $\rho'$ are the perturbation values with respect to this state. For the atmosphere in hydrostatic equilibrium, we put $\partial p_s/\partial z = -\rho_s g$, so Eq. (4.5) becomes

$$\begin{aligned} -2\Omega v \sin\varphi + -2\Omega w \cos\varphi &= -\frac{1}{(\rho_s + \rho')}\frac{\partial p'}{\partial x} \\ -2\Omega u \sin\varphi &= \frac{1}{(\rho_s + \rho')}\frac{\partial p'}{\partial y} \\ 2\Omega u \cos\varphi &= -\frac{1}{(\rho_s + \rho')}\frac{\partial p'}{\partial z} + \frac{\rho' g}{(\rho_s + \rho')} \end{aligned} \tag{4.6}$$

We have seen that the vertical velocity can be neglected with respect to the horizontal components, so that the second left-hand term in the first part of Eq. (4.6) can be canceled. The order of magnitude of the horizontal pressure gradient is $p'/L$, and if the Coriolis force must balance, the acceleration due to pressure must be

$$\frac{p'}{L} \approx O[\rho\Omega U] \tag{4.7}$$

so we can estimate the vertical pressure gradient

$$\frac{\partial p'}{\partial z} = O\left[\frac{p'}{D}\right] = O\left[\rho\frac{\Omega UL}{D}\right] \tag{4.8}$$

In turn, we can now evaluate the ratio between the vertical components of the Coriolis acceleration and the pressure gradient acceleration:

$$\rho\frac{2\Omega u\cos\varphi}{\partial p'/\partial z} = O\left[\rho\frac{\Omega UD\cos\varphi}{\Omega UL}\right] = O\left[\frac{D}{L}\right] \ll 1 \tag{4.9}$$

We see that even in this case, in the last part of Eq. (4.6), the Coriolis acceleration can be neglected. The conclusion is that the background atmosphere is in hydrostatic equilibrium as well as the deviations from it. The magnitude of these deviations can also be evaluated from Eq. (4.8):

$$\rho' = O\left[\rho\frac{\Omega UL}{gD}\right] \Rightarrow \frac{\rho'}{\rho} = O\left[\frac{U}{\Omega L}\right]\frac{\Omega^2 L^2}{gD} \tag{4.10}$$

We notice that the ratio between densities depends only on the Rossby number and the geometrical characteristics of the fluid. In case of meteorological motions with $L = 10^3$ km, $D = 10$ km, $\Omega^2 L^2/gD = 0.1$, so that $\rho'/\rho \approx 0.1Ro \ll 1$. We have finally

$$fv_g = \frac{1}{\rho}\frac{\partial p}{\partial x}; \quad fu_g = -\frac{1}{\rho}\frac{\partial p}{\partial y}; \quad \rho g = -\frac{\partial p}{\partial z}; \tag{4.11}$$

Again these are diagnostic relations through which we can determine the wind once we know the pressure gradient or simply the pressure field. After manipulating a little, we can put Eq. (4.11) in a synthetic vector form:

$$\mathbf{V}_g = \frac{1}{f\rho}\mathbf{k} \times \nabla_z p \tag{4.12}$$

where with $\mathbf{V}_g$ we have denoted the horizontal component of the wind and with $\mathbf{k}$ the vertical unit vector. The third part of Eq. (4.11) simply says that the fluid is at rest in the vertical direction.

We can change Eq. (4.19) as a function of the geopotential we have defined as $\varphi = gz$. We have then $\nabla_z^2 \varphi = 0$ and substituting into Eq. (4.4) gives

$$\nabla_x p = -\frac{\partial p}{\partial \Phi}\nabla_p \varphi = \rho \nabla_p \varphi$$

Equation (4.12) can then be rewritten as a function of the geopotential:

$$f\mathbf{V}_g = \mathbf{k} \times \nabla_p \varphi \tag{4.13}$$

The last two relations simply tell us that the geostrophic wind is normal to the pressure or geopotential gradient. As shown in Fig. 4.2, in the northern hemisphere, the wind velocity leaves the high on the right. Both equations can be used in the calculations.

As a simple, practical example, referring to Fig. 4.2, we put $p = 454$ hPa and $p + \Delta p = 458$ hPa and assume that the isobars are separated by a distance of

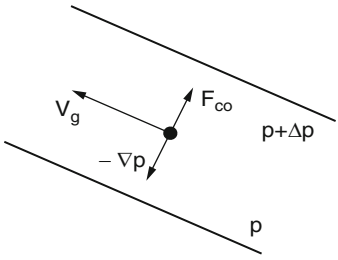



**Fig. 4.2** The equilibrium in the geostrophic motion between the pressure gradient $-\nabla p$ and the Coriolis force $F_{co}$

300 km with an average temperature of 250 K. We can calculate the density $\rho = p/RT = 4.56\ 10^4/(287\ 250) = 0.635$ kg m$^{-3}$, and for the geostrophic wind, we have $V = 400/(0.635\ 1.08\ 10^{-4}\ 300\ 10^3) = 20.5$ ms$^{-1}$ where the Coriolis parameter has been calculated for a 45° latitude and has the value $1.028\ 10^{-4}$ s$^{-1}$. The direction of the geostrophic wind is parallel to the isobars and rotates clockwise around high-pressure zones and counterclockwise around low-pressure zones.

Actually we can see some problem with this geostrophic approximation. First of all, at the equator $f = 0$ and we cannot attain any balance there. Another point is that, when we calculated the velocity divergence from Eq. (4.11), we found that it is zero in case of uniform rotation, $f = \text{cost}$. This means that geostrophic motion does not allow any space for vertical motions. Actually these considerations have much more profound implications, but to appreciate that, we need to introduce the *geostrophic streamfunction*.

### *4.1.1 The Geostrophic Streamfunction*

The most important difference between the geostrophic motion (i.e., the motion in a fluid in rotation) and the motion in a fluid at rest is that in the former it develops along isobars, while in the latter it develops along the pressure gradient. In the geostrophic approximation, once there is a pressure map, the motion can be immediately determined. This by the way may spell trouble for those who believe they can become weathermen without adequate training.

An important property of a geostrophic fluid can be found by considering Fig. 4.3, where a geostrophic flow along three isobars, separated by a small pressure difference, is sketched. We want to evaluate the mass flux through the surface of unitary height delimited by curve $C$, assuming also the fluid to be thin. We consider the elementary vector $d\mathbf{l}$ tangent to the curve $C$, while the unit vector $\mathbf{n}$ indicates the normal direction to the same curve. To evaluate the mass flux $M$ we start from the elementary flux

$$\rho_s \left( \mathbf{V}_g \cdot \mathbf{n} dl \right)$$

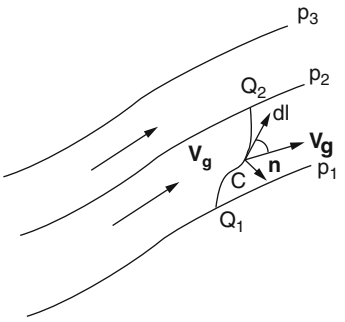



**Fig. 4.3** The evaluation of the mass flux through the line between $Q_1$ and $Q_2$ (From Pedlosky 1987)

that we integrate along the curve between points $Q_1$ and $Q_2$ which remain on the isobars $p_1$ and $p_2$, respectively. $\mathbf{V}_g$ is the geostrophic wind and $\rho_s$ is the density.

$$M = \int_{Q_1}^{Q_2} \rho_s \left(\mathbf{V}_g \cdot \mathbf{n} dl\right)$$

and substituting the value of the geostrophic wind, we obtain

$$M = \int_{Q_1}^{Q_2} \rho_s \left(\mathbf{V}_g \cdot \mathbf{n} dl\right) = \int_{Q_1}^{Q_2} \rho \left(\frac{\mathbf{k} \times \nabla p}{f}\right) \cdot \mathbf{n} dl$$

that gives the result

$$M = \int_{Q_1}^{Q_2} \frac{[\nabla p \cdot (\mathbf{n} \times \mathbf{k})]}{f} dl = -\int_{Q_1}^{Q_2} \frac{\nabla p}{f} d\mathbf{l} \tag{4.14}$$

because $\mathbf{n}$ and $\mathbf{k}$ are normal. This relation can be simplified when the performed Coriolis parameter is held constant along the integration path. In this case, we have

$$M = \frac{p\,(Q_1)}{f} - \frac{p\,(Q_2)}{f} \tag{4.15}$$

which shows how the mass flux is independent of the integration path $C$. From a simple comparison between Eqs. (4.15) and (4.12), we see that, if we define $\psi = M/\rho$, we obtain

$$u_g = -\frac{\partial \psi}{\partial y}; \quad v_g = \frac{\partial \psi}{\partial x} \tag{4.16}$$

The function $M$, which is called the *mass streamfunction*, is equivalent to the mass flux through the points $Q_1$ and $Q_2$ and is measured in kg m$^{-1}$ s$^{-1}$. The function $\Psi$ measured in m$^2$ s$^{-1}$ is called the *streamfunction*.

The result we have obtained is very elegant but has some limitations that we can understand. Because any geostrophic solution satisfies the continuity equation, this implies that the three in Eq. (4.11) are somewhat redundant. We can actually write Eq. (4.16) as

$$\rho f \frac{\partial \psi}{\partial x} = \frac{\partial p}{\partial x}; \quad \rho f \frac{\partial \psi}{\partial y} = \frac{\partial p}{\partial y}$$

If the pressure is eliminated from these two equations (by deriving the first with respect to $y$ and the second with respect to $x$ and subtracting), we have the trivial solution "zero equal zero." As a matter of fact, the definition of the streamfunction

$$f\psi = \frac{p}{\rho} \tag{4.17}$$

means that, given a pressure field, it is always possible to find the corresponding geostrophic streamfunction. In mathematical terms, this indicates a degeneracy of the solutions and so we talk about *geostrophic degeneracy*. This implies that the geostrophic solution cannot determine the fluid motion. To accomplish this, we need to take into account small deviations from geostrophy, and a very good example of this is the isallobaric wind.

### 4.1.2 The Quasi-geostrophy: The Isallobaric Wind

If we maintain the acceleration term and neglect the friction term, the equation of motion can be written as

$$\frac{D\mathbf{V}}{Dt} = -f\mathbf{k} \times \mathbf{V} - \frac{1}{\rho}\nabla_z p \tag{4.18}$$

and the wind velocity can be separated into the geostrophic and ageostrophic parts:

$$\mathbf{V} = \mathbf{V_g} + \mathbf{V_a}$$

Substituting this relation in Eq. (4.18) and expressing the geostrophic wind as in Eq. (4.12), we have

$$\frac{D\mathbf{V}}{Dt} = -f\,\mathbf{k} \times \mathbf{V}_a \tag{4.19}$$

This relation shows that the ageostrophic component is always perpendicular and "at the left" of the acceleration in the northern hemisphere. This equation can be solved for $\mathbf{V}_a$:

$$\mathbf{V}_a = \frac{1}{f}\mathbf{k} \times \frac{D\mathbf{V}}{dt}$$

Remembering the decomposition, we have

$$\mathbf{V}_a = -\frac{1}{f}\mathbf{k} \times \left[\frac{\partial}{\partial t}\left(\frac{1}{\rho f}\nabla_z p\right) + \frac{D\mathbf{V}_a}{Dt}\right] \tag{4.20}$$

As a first approximation, we can neglect the second term in parentheses so that the deviation from the geostrophic wind in terms of components is given by

$$u - u_g \approx -\frac{1}{f^2\rho}\frac{\partial^2 p}{\partial t \partial x}; \qquad v - v_g \approx -\frac{1}{f^2\rho}\frac{\partial^2 p}{\partial t \partial y} \tag{4.21}$$

The ageostrophic component of the wind depends from the rate of change of the pressure gradients, the so-called *tendency*, and the lines connecting points with the same value for the tendency are called *isallobars*. The direction is normal to

the isobars and oriented toward the isallobaric minimum. The total wind is then composed of the geostrophic component blowing parallel to the isobars and the isallobaric component. It is extremely interesting to calculate, at this point, the divergence of the total wind starting from Eq. (4.21). Considering the divergence of the geostrophic wind to be zero, we have

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=-\frac{1}{f^2\rho}\frac{\partial}{\partial t}\left(\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}\right)=-\frac{1}{f}\frac{\partial}{\partial t}\left(\frac{\partial v_g}{\partial x}-\frac{\partial u_g}{\partial y}\right) \tag{4.22}$$

We notice that in this case the divergence is not zero and the contribution is from the isallobaric terms. Also, we see that the divergence is proportional to the tendency of the geostrophic vorticity. It is easy to show that the relative vorticity is determined only by the geostrophic terms:

$$\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\cong-\frac{1}{f^2\rho}\frac{\partial}{\partial t}\left(\frac{\partial^2 p}{\partial x\partial y}-\frac{\partial^2 p}{\partial y\partial x}\right)+\left(\frac{\partial v_g}{\partial x}-\frac{\partial u_g}{\partial y}\right) \tag{4.23}$$

Comparing Eqs. (4.22) and (4.23), we see that if the time scale of the motion is $T$, then the ratio between divergence and vorticity is of the order of the Rossby number. Again for fast-rotating planets, the vorticity plays a major role in the determination of the motion.

We cannot resist at this point the temptation of a further step, obtained by eliminating the pressure between Eqs. (4.22) and (4.23):

$$\frac{\partial}{\partial t}\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)+f\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)=0 \tag{4.24}$$

This represents a particular form of the vorticity equation that will be discussed at length in the next chapters. The equation shows that the rate of change of vorticity is proportional to the horizontal divergence. If the divergence is positive, the horizontal section increases and the rotation of the fluid slows down and vice versa. Actually this is similar to the conservation of the angular momentum for a fluid. Equation (4.32) contains information on the ageostrophic components and could be used to solve for the motion in the fluid. The explicit dependence from the horizontal divergence gives information on the vertical motion.

## 4.2 The Thermal Wind

We will start by considering again Eq. (3.25) at steady state ($D\mathbf{V}/Dt=0$) so we get

$$2\Omega\times\mathbf{V}+\frac{1}{\rho}\nabla p+\nabla\varphi=0 \tag{4.25}$$

If we take the $\nabla\times$ of this equation, we obtain

$$(2\Omega\cdot\nabla)\,\mathbf{V}=\frac{1}{\rho}\nabla p\times\frac{1}{\rho}\nabla\rho \tag{4.26}$$

Again as in the case if we have a *barotropic* atmosphere, the pressure and density gradient are parallel and in this case

$$(2\Omega\cdot\nabla)\,\mathbf{V}=0 \tag{4.27}$$

because $\Omega\cdot\nabla$ is the gradient operator in the direction of $\Omega$ (i.e., $z$); we get

$$\frac{\partial\mathbf{V}}{\partial z}=0 \tag{4.28}$$

This equation expresses the Taylor–Proudman theorem that implies that in a barotropic fluid, the velocity field cannot vary in the direction of the rotation. On the other hand, for a *baroclinic* fluid, (4.26) applies and the only simplification is that the fluid may be in hydrostatic equilibrium so that $(1/\rho)\,\nabla p=-g\mathbf{k}$ so that

$$2\Omega\frac{\partial\mathbf{V}}{\partial z}=-\frac{g}{\rho}\mathbf{k}\times\nabla\rho \tag{4.29}$$

In the atmosphere, density is related to pressure and temperature so that taking the components of (4.29) with $2\Omega\rightarrow f\mathbf{k}$, we get

$$f\frac{\partial u_g}{\partial z}=-\frac{g}{T}\left(\frac{\partial T}{\partial y}\right)_p;\quad f\frac{\partial v_g}{\partial z}=\frac{g}{T}\left(\frac{\partial T}{\partial x}\right)_p \tag{4.30}$$

where the derivatives are made at constant pressure. We have used the geostrophic components because we started from the equilibrium.

The thermal wind equation relates the temperature field to the wind field and there is another simple way to obtain the same equations. Writing the components of Eq. (4.13), we have the components of the geostrophic wind as a function of the geopotential:

$$u_g=-\frac{1}{f}\left(\frac{\partial\varphi}{\partial y}\right);\quad v_g=\frac{1}{f}\left(\frac{\partial\varphi}{\partial x}\right)$$

Now we write one of these equations at two different pressure levels, and subtracting, we obtain

$$u_g\,(p_1)-u_g\,(p_0)=-\frac{1}{f}\frac{\partial\,(\delta\varphi)}{\partial y} \tag{4.31}$$

where $\delta\Phi$ is the difference between the geopotential at the two levels given by

$$\delta\Phi = R\overline{T}\ln\left(p_0/p_1\right)$$

where $\overline{T}$ is the average temperature of the layer. Substituting into Eq. (4.31), we obtain

$$u_g\left(p_1\right) - u_g\left(p_0\right) = -\frac{R}{f}\ln\left(\frac{p_0}{p_1}\right)\left(\frac{\partial\overline{T}}{\partial y}\right)_p$$

Operating with the same procedure on the other component, we have

$$v_g\left(p_1\right) - v_g\left(p_0\right) = \frac{R}{f}\ln\left(\frac{p_0}{p_1}\right)\left(\frac{\partial\overline{T}}{\partial x}\right)_p$$

and the vector form of the thermal wind equation is given by

$$\mathbf{V}\left(p_1\right) - \mathbf{V}\left(p_0\right) = \mathbf{V}_T = \left(\frac{R}{f}\right)\ln\left(\frac{p_0}{p_1}\right)\left(\mathbf{k}\times\nabla_p\overline{T}\right) \tag{4.32}$$

The thermal wind equation is a unique occasion to introduce a coordinate system we will use often in the following chapters and it is known as log pressure. In practice, the coordinate $z$ is substituted with a new coordinate $z^*$ defined as

$$z^* = -H\ln\left(\frac{p}{p_0}\right) \tag{4.33}$$

where $H$ is the scale height. The derivative with respect to pressure can be expressed as $\partial/\partial p = \left(\partial z^*/\partial p\right)\partial/\partial z^*$ and the thermal wind equation becomes

$$f\frac{\partial u_g}{\partial z^*} = -\frac{R}{H}\frac{\partial T}{\partial y};\quad f\frac{\partial v_g}{\partial z^*} = -\frac{R}{H}\frac{\partial T}{\partial x}$$

that appear in a very much simplified form.

The formal derivation of the thermal wind equation needs to be complemented with some more intuitive arguments as shown in Fig. 4.4. If we assume a temperature gradient in the positive $x$ direction, then the surfaces at constant pressure are increasingly separated with increasing $x$. The figure shows constant pressure surfaces at $p$, $p-\delta p$, $p-2\delta p$ separated by $\delta z = RT\delta p/p$. This implies that a horizontal pressure gradient is established, oriented as shown in the figure. A Coriolis acceleration that in turn requires a wind directed along the positive $y$ direction must balance the acceleration produced by such a gradient. The reason why the wind increases with altitude is because the horizontal pressure gradient increases with altitude.

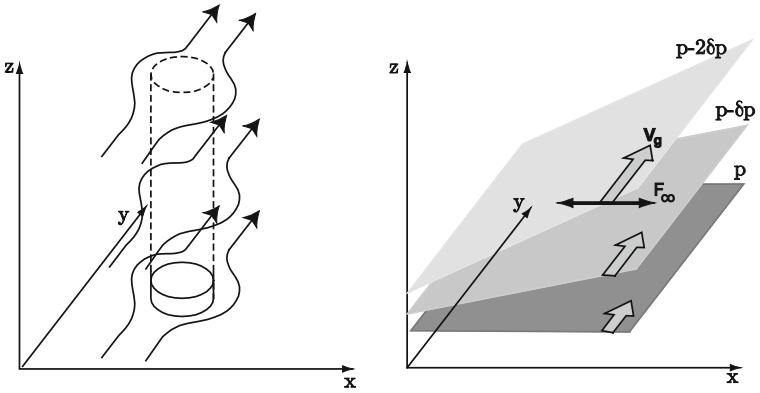



**Fig. 4.4** Change of the wind with altitude. On the *right* the case of a barotropic atmosphere where the pattern introduced by an obstacle is maintained at all altitudes (Taylor–Proudman theorem). The Taylor column is depicted in *dashed lines*. *Left* the thermal wind increases with altitude in the presence of a temperature gradient in the $x$ positive direction (Adapted from Marshall and Plumb 2008)

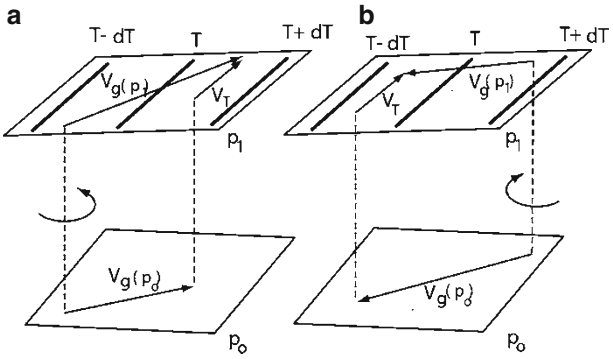



**Fig. 4.5** Temperature advection due to the thermal wind. This is parallel to isotherms (*thick solid lines*) and can be obtained as difference of the geostrophic wind at two pressure levels. In (**a**) the average geostrophic wind has a component oriented as the temperature gradient. In (**b**) the situation is opposite. Cold or warm advection is related to the direction of rotation of the wind

In the case we have just seen, the direction of the geostrophic wind does not change with altitude. This is not true in general because the temperature gradient may change with altitude. However, an interesting aspect is that, as shown in the figure, the thermal wind is parallel to the isotherms. This can be seen from Eq. (4.26) which shows how the wind is always normal to the temperature gradient and then parallel to the isotherms or the lines of equal geopotential thickness.

Equation (4.26) is also useful to predict the temperature advection (and then the arrival of cold or warm front) by looking at the behavior of the winds with altitude. Figure 4.5 shows qualitatively the behavior of the isotherms on the two isobaric

surfaces at $p_0$ and $p_1$. The thermal wind is parallel to the isotherms and can be obtained as a difference, as shown in the figure. In case (a), the average value of the geostrophic wind between the two pressure levels has a component oriented as the temperature gradient, so that the advection defined as $\mathbf{V}_g \cdot \nabla T$ is negative and the temperature in the layer tends to decrease. In case (b), the advection is positive and the temperature in the layer tends to increase.

Observing the behavior of the wind with altitude, we can outline these two situations. If the wind turns in the anticlockwise (clockwise) fashion, going from the higher to the lower pressure, then we have cold (warm) advection. This conclusion can be also obtained by considering a different expression for the thermal wind. Actually we can find the change of the geostrophic wind with height. After some algebra and using the hydrostatic equilibrium, we find

$$\frac{\partial \mathbf{V}_g}{\partial z} = \frac{g}{fT}\mathbf{k} \times \nabla_p T \tag{4.34}$$

The change of the geostrophic wind can be then approximated for a rotation $\delta\theta$ as

$$\delta \mathbf{V}_g = \left|\mathbf{V_g}\right| \delta\theta \mathbf{n}$$

where $\mathbf{n}$ is the unit vector normal to $\mathbf{V}_g$ that stays "to the left" with respect to it. As a consequence, it is easy to see after some algebra that the advection

$$-\mathbf{V}_{g\cdot} \cdot \nabla_p T = \frac{f}{g}\left|\mathbf{V_g}\right|^2 T \frac{\partial \theta}{\partial z} \tag{4.35}$$

In this way, we relate the change in the wind direction with height with the sign of the advection. The sign of the derivative of the angle will be taken as positive for a clockwise rotation and vice versa.

### *4.2.1 Thermal Wind in the Atmosphere*

Now that we have discussed at length the thermal wind, we ask if we can apply it to some real problem. Our first approach with the atmospheric data will be just to see when and where the thermal wind equation can be applied. Figure 4.6 shows in the upper part a latitude–altitude section of the annual mean of zonal temperature in the northern hemisphere. Actually usually the temperature is reported in degrees Kelvin. The figure is based on the work of A.H. Oort e E.M. Rasmussen. As we would expect, the surface temperature decreases with latitude and at each latitude decreases with altitude. The qualitative reasons are well known: the high latitudes receive less solar radiation and emit almost the same amount of infrared radiation as the low-latitude regions. The result, at least below 10 km, is a regular gradient of the temperature with latitude. If the gradient in the $y$ direction is negative, then

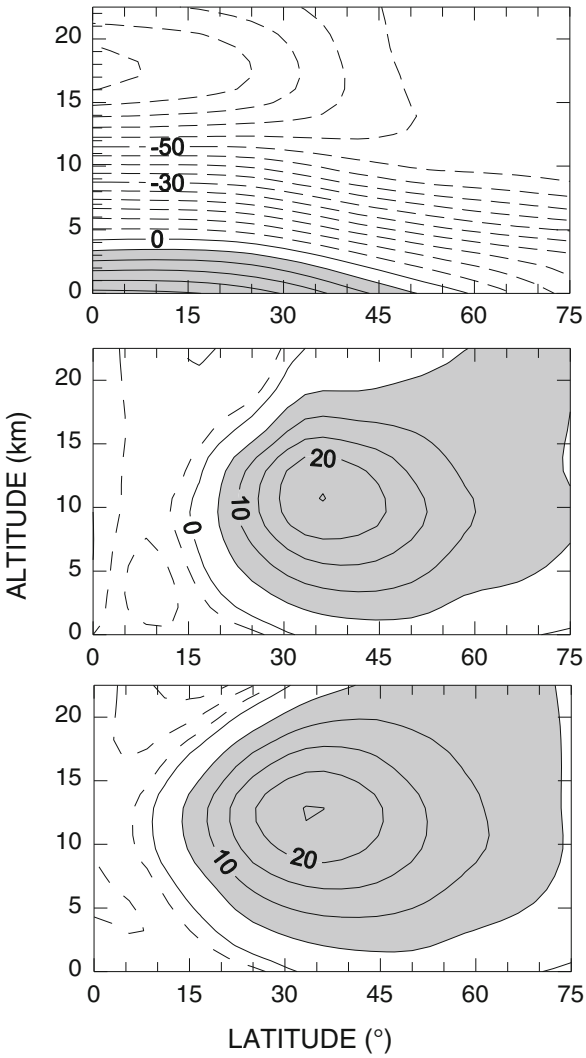



**Fig. 4.6** In the *upper* figure are shown the average yearly isotherms as a function of latitude and altitude. The interval is 5 °C. When the equation of the thermal wind is applied, the figure at the *center* is obtained which gives the isopleths of the zonal wind. This must be compared with the real data shown in the *bottom* figure. The wind is in m $s^{-1}$ and the positive values (westerly) are in *gray*. The interval is 5 m $s^{-1}$ (positive) and 2 m $s^{-1}$ (negative)

based on Eq. (4.34) the zonal wind velocity must increase with height. If we apply the thermal wind equation to the temperature field of Fig. 4.6, we obtain the figure in the middle. The real data are shown in the figure at the bottom.

What we notice is that the characteristics of the measured and the calculated zonal wind are similar but with significant differences. In particular the equatorial belt shows the largest differences. The rather spectacular closure of the isopleths of the zonal wind (isotachs) shows the so-called *jet stream*, that is, the large flow of air that blows from west to east (*westerly*). The reason for the closure is that some-where the gradient of temperature changes sign because the high tropical troposphere and low tropical stratosphere are very cold regions. These are even

colder than the high latitude regions at the same altitude. There are many reasons for the discrepancies between the observed and calculated zonal wind. First of all, not all the wind is geostrophic as we well know. Also, the Coriolis parameter tends to be zero at the equator, so that in those regions there should be other processes that contribute to the dynamical equilibrium. Finally in our equation we have so far neglected the friction term and this may be non-negligible at the surface and in the upper troposphere.

## 4.3 More About Geostrophic Wind

The applications of geostrophy will be available along the book. Here we will point out some necessary expansion of what we have studied so far.

### *4.3.1 Margules Formula*

Following our method of using what we have learned for solving some real problem, we may proceed immediately to study a very important meteorological structure, called a front. This is actually a surface that separates two regions of the atmosphere that have different characteristics. Figure 4.7 shows a very idealized front. In one case, the isotherms are represented, while in the other case, the surface of discontinuity separates just two regions at different temperatures.

In these figures, the slope of the discontinuity surface is exaggerated, and to show this, we can do a simple calculation, outlined in the figure. In practice, we can follow two paths (ABD and ACD) to calculate the temperature difference between A and D which should be the same. If we call $\Gamma_w$ the lapse rate for the warm side of the front and $\Gamma_F$ the lapse rate in the frontal zone and with $\beta$ the absolute horizontal temperature gradient, we have for the temperature difference

$$\Delta z \Gamma_w = \Delta z \Gamma_F + \beta \Delta x$$

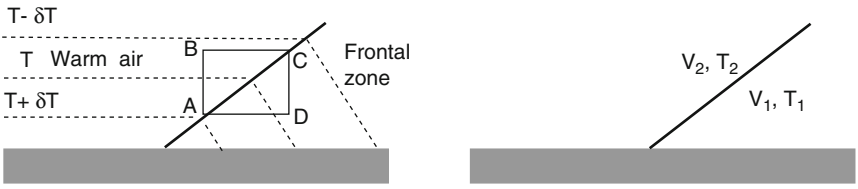



**Fig. 4.7** Schematic illustration of a front. At *left* the isothermal is shown with *dashed lines*. Notice that moving on a horizontal surface, there is a temperature discontinuity through the *thick line*. At *right* the two sides of the front are approximated as having the same temperature and wind velocity

where $\Delta z$ is the vertical distance between A and B and $\Delta x$ is the horizontal distance between B and D. The slope can be easily found to be

$$\frac{dz}{dx} = \frac{\beta}{(\Gamma_w - \Gamma_F)} \tag{4.36}$$

Assuming a vertical gradient in the warm zone of 7 °C km$^{-1}$ and zero in the frontal zone and a horizontal gradient of 10 °C per 100 km, we have a slope of about 14.3 m km$^{-1}$ that is an angle less than 1 °C.

If we stay within the frontal zone, there will be an average temperature gradient and consequently a thermal wind. If we assume the intersection of the front with the ground is at $x = 0$, then the 300 hPa level will be approximately at 8.5 km and at a horizontal distance of more than 500 km. With these data, we get a very large thermal wind at the upper level that depends strongly on the horizontal gradient.

However, although these large numbers are unrealistic, this shows that at the upper levels above a front, we may have the situation analogous of the jet stream we have found before. In a more realistic situation, we may assume that the wind velocity in the frontal zone is $v_1$ and $v_2$ in the warm front both oriented in the $y$ direction, normal to the plane $z, x$. We can assume that at the boundary the warm air and the cold air exert the same pressure:

$$p_1(x, y, z, t) = p_2(x, y, z, t)$$

and the changes in pressure on the two sides of the discontinuity surface are the same, that is, $dp_1 = dp_2$. We can then evaluate the pressure difference on the two sides through the use of the hydrostatic and geostrophic equilibrium:

$$fv_{1,2} = \frac{1}{\rho_{1,2}}\frac{\partial p}{\partial x}; \quad g = -\frac{1}{\rho_{1,2}}\frac{\partial p}{\partial z}$$

We have then

$$-\rho_1 g dz + f v_1 \rho_1 dx = -\rho_2 g dz + f v_2 \rho_2 dx$$

from which we can easily find the slope of the front (Houghton 1977):

$$g\frac{dz}{dx} = \frac{f(T_1 v_2 - T_2 v_1)}{T_1 - T_2} \tag{4.37}$$

This expression shows that the slope is a function of latitude through the Coriolis parameter and increases with increasing latitude. The slope decreases with increasing temperature difference. We see also that at the equator the air masses are stratified because the Coriolis parameter is zero there. Slopes vary between 1/50 and 1/300. Equation (4.37) is also known as Margules relation from the meteorologist

Max Margules. This relation can be changed a little considering that density goes like the inverse of temperature so that (4.37) can be written as a function of densities:

$$v_2 - v_1 = \frac{g}{f}\frac{\Delta\rho}{\rho_1}\tan\gamma \tag{4.38}$$

where we have assumed $\rho_1 \approx \rho_2$ and $\tan\gamma = dz/dx$. This equation can be expressed in terms of the Rossby radius of deformation. We recall that

$$N^2 = \frac{g}{\theta}\frac{\partial\theta}{\partial z} \approx \frac{g}{H}\frac{\Delta\rho}{\rho_1} = \frac{g'}{H} \tag{4.39}$$

where we introduced the reduced gravity $g' = g\Delta\rho/\rho$ and (4.38) can be written as

$$\Delta v = \frac{N^2 H}{f}\tan\gamma = NL_\rho \tan\gamma \tag{4.40}$$

where $L_\rho = NH/f$ is the Rossby radius of deformation. It can be shown that rotating fluid relaxes on scale which has this characteristic length. In particular, it can be seen that if $f$ tends to zero, $L_\rho$ tends to infinity, indicating that for scales smaller compared with the radius of deformation, rotation effects are negligible. On the other hand, for scales larger than $L_\rho$, the rotation effects are important. According to our definition, we could also write

$$\mathrm{L}_\rho = \frac{(g'H)^{1/2}}{f}$$

and we will see that the square root is the velocity of shallow water waves, and so the Rossby radius can be regarded also as the scale where the stratification effects balance the rotational effects.

Figure 4.8 shows in a cartoon how the Margules formula could explain the localization of the subpolar jet. In this case, the polar front is creating a discontinuity in the density.

### *4.3.2 Inertial Instability*

Another straightforward application of the geostrophic wind deals with a rather important instability in the motion of the atmosphere. We refer to Fig. 4.9 and assume a geostrophic flow that changes with latitude.

To this zonal flow corresponds a negative vorticity given by $-\partial u_g/\partial y$. We then consider an air parcel located in $A$ that moves at higher latitude of quantity $\delta y$ its zonal velocity will change according to

$$u\,(y_0 + \delta y) = u_g\,(y_0) + \left(\frac{du}{dt}\right)\delta t = u_g\,(y_0) + fv\delta t \tag{4.41}$$

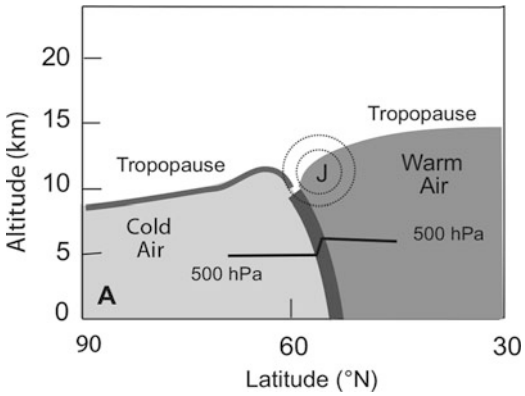



**Fig. 4.8** A cartoon indicating the generation of the subpolar jet at the discontinuity associated with the polar front

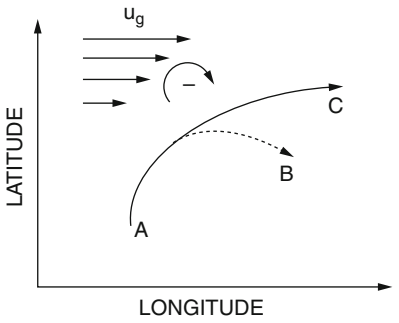



**Fig. 4.9** Illustration of the inertial instability. The geostrophic flow corresponds to a negative vorticity shown by the *rotation arrow*. *A* indicates the initial position of the air parcel

where we have assumed that the zonal acceleration is balanced by the Coriolis force. The geostrophic wind will change according to

$$u_g\left(y_0+\delta y\right)=u_g\left(y_0\right)+\left(\frac{\partial u_g}{\partial y}\right)\delta y=u_g\left(y_0\right)-\zeta_g\delta y \tag{4.42}$$

where $\zeta_g$ is the geostrophic vorticity. Subtracting (4.41) from (4.42), we have

$$u_g-u=-\left[f-\left(\frac{\partial u_g}{\partial y}\right)\right]v\delta t \tag{4.43}$$

On the other hand, the equation of motion along the y direction is given by

$$\frac{dv}{dt}+fu=\frac{1}{\rho}\frac{\partial p}{\partial y}=fu_g$$

Substituting in (4.43) and rearranging,

$$\frac{dv}{dt} = -\left(f - \frac{\partial u_g}{\partial y}\right) f\delta y$$

The acceleration in the y direction is then

$$\frac{d^2v}{dt^2} + \left(f - \frac{\partial u_g}{\partial y}\right) fv = 0 \tag{4.44}$$

The condition for instability will be

$$\left(f - \frac{\partial u_g}{\partial y}\right) \begin{cases} > 0 \text{ stable} \\ = 0 \text{ neutral} \\ < 0 \text{ unstable} \end{cases} \tag{4.45}$$

In the case of instability, the parcel will keep moving to higher latitudes (trajectory A–C). The opposite will happen when the absolute value of relative vorticity is larger than the planetary vorticity and the parcel will move toward lower latitude (trajectory A–B). This is one of the instabilities to which a fluid motion is subject and is called *inertial instability*. We will return to it when treating the general circulation of the atmosphere.

## 4.4 The Natural Coordinate System

We have already introduced the geostrophic streamfunction in Sect. 4.1.1, and now we want to illustrate the relationship between the streamfunction and the motion. A *streamline* is defined as the curve whose tangent gives the direction of the velocity at that point. This should not be confused with the trajectory, which is the curve that connects the different positions of the air parcel during the motion. Only in a few cases do the two things coincide. The streamlines can be traced on a plane (in the two-dimensional case) in such a way that their density gives an idea of the velocity: the denser the lines, the higher is the velocity. The concept of streamline is important because among other things it defines the *natural* coordinate system as shown in Fig. 4.10. The direction of the tangent defines an axis so that the other axis is normal to the tangent and remains to the left of motion direction. The absolute acceleration in this reference system is found by expressing first the velocity vector in terms of the unit vector **s** in the direction of the tangent:

$$\frac{D\mathbf{V}}{Dt} = \frac{D}{Dt}(V\mathbf{s}) = V\frac{D\mathbf{s}}{Dt} + \mathbf{s}\frac{DV}{Dt} \tag{4.46}$$

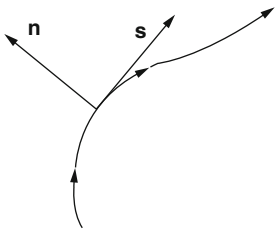



**Fig. 4.10** The natural coordinate system (Holton 1992)

It is then easy to see that

$$\frac{D\mathbf{s}}{Dt} = \frac{V}{R_t}\mathbf{n}$$

where **n** is the unit vector of the normal to s and $R_r$ is the radius of curvature of the streamline. If the acceleration is decomposed in the directions **s** and **n**, we have

$$\frac{DV}{Dt} = -\frac{1}{\rho}\frac{\partial p}{\partial s}; \quad \frac{V^2}{R_t} = -fV - \frac{1}{\rho}\frac{\partial p}{\partial n} \tag{4.47}$$

The pressure gradient along **s** has an effect only on the magnitude of the velocity while if the normal components balance, we obtain

$$\frac{V^2}{R_t} + fV - \frac{1}{\rho}\frac{\partial p}{\partial n} = 0; \quad \frac{V^2}{R_t} - fV + \frac{1}{\rho}\frac{\partial p}{\partial n} = 0 \tag{4.48}$$

the first for low pressure. The wind that results from these equilibrium equations is called gradient wind. It can be obtained by solving the second of Eq. (4.48) for $V$. We notice that $V$ is real if

$$f^2 > \frac{4}{R_t}\frac{1}{\rho}\frac{\partial p}{\partial n} \tag{4.49}$$

We see that for anticlockwise motion, that is, around a low-pressure center, condition (4.49) is always satisfied and the pressure gradient may be arbitrarily high. For an anticyclonic situation, the pressure gradient has a maximum value toward the center, and the Coriolis force will act in such a way to return to the normal situation. The opposite happens around a high-pressure center with the rising pressure at the center. If the wind velocity is very high (like around a tornado), then the Coriolis term can be neglected and the equilibrium is between centrifugal and pressure force:

$$\frac{1}{\rho}\frac{\partial p}{\partial n} = -\frac{V^2}{R_t} \tag{4.50}$$

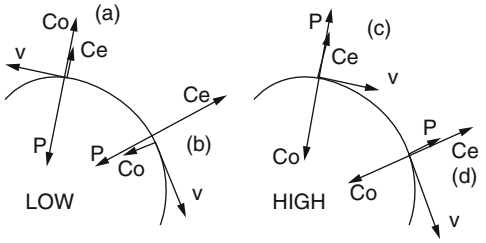



**Fig. 4.11** Possible equilibrium around low- and high-pressure centers. For the cyclonic cases, the normal situation is shown at the *upper left* (**a**): the pressure force is balanced by the Coriolis force (*Co*) and the centrifugal force (*Ce*). At the *lower left* (**b**) the anomalous situation is shown. The possible cases for high pressure are shown at the *right*

This relation corresponds to the *cyclostrophic equilibrium*. In general, this equilibrium may be found in all those cases where the curvature radius is small, but also for slow-rotating planets (e.g., Venus).

As a matter of fact, an interesting application of Eq. (4.50) gives an estimation of the latitudinal temperature gradient for Venus. If we refer to Fig. 4.11, we see that the cyclostrophic equilibrium reduces to the expression

$$\frac{u^2 \tan \varphi}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial y} \tag{4.51}$$

and expressing the pressure gradient as a function of the geopotential, we get

$$\omega_r^2 a \sin \varphi \cos \varphi = \frac{\partial \varphi}{\partial y} \tag{4.52}$$

where $\omega_r = u/a \cos \varphi$ and $a$ is the radius of the planet. Differentiating with respect to pressure and keeping in mind $\partial \varphi / \partial p = -RT/p$, we have

$$2\omega_r d\omega_r = -\frac{R}{a \sin \varphi \cos \varphi}\frac{\partial T}{\partial y} d \ln p$$

which may be integrated, giving an expression for the thermal wind relation for a slowly rotating planet:

$$\omega_r^2 (p_1) - \omega_r^2 (p_0) = -\frac{R \ln (p_0/p_1)}{a \sin \varphi \cos \varphi}\frac{\partial \overline{T}}{\partial y} \tag{4.53}$$

To simplify further, we may assume that the temperature gradient as a function of latitude is such that $\omega_r$ depends only on pressure and thus will be of the form

$$\frac{\partial T}{\partial y} = T_y \sin 2\varphi/2$$

where $T_0$ is a constant. Substituting this in Eq. (4.53), we obtain

$$\omega_r^2(p_1) - \omega_r^2(p_0) = -\frac{R\ln(p_0/p_1)}{a} T_y \tag{4.54}$$

Venus has a jet at about 60 km with velocities around 100 m $s^{-1}$ so that if we assume $p_1 = 3{\cdot}10^5$ Pa and $p_0 = 9.5{\cdot}10^6$ Pa, we have $T_y = 2.5$ K $m^{-1}$. Integrating this gradient, we obtain a difference between the pole and equator given by

$$\omega_r = \frac{u}{a\cos\varphi} = 7.7 \text{ K}$$

Notice that Venus' atmosphere is composed of carbon dioxide so that $R = 187$ J $kg^{-1}$ $K^{-1}$ while the radius is $a = 6100$ km.

The natural system also gives an interesting view of the vorticity. If $R$ is the radius of curvature of the streamline in Fig. 4.10, we can assume a local instant angular velocity $\omega$ and write the velocity as $V = \omega R$. The $x$ axis on this coordinate system corresponds to the direction of the tangent, while the $y$ axis corresponds to the normal direction (Fig. 4.12). The only component of the velocity is along $x$, so that from the definition of vorticity, we have

$$\zeta = \frac{\partial V}{\partial R} = R\frac{\partial \omega}{\partial R} + \omega \tag{4.55}$$

It can be easily shown that $R\partial\omega/\partial R = \partial V/\partial n$, so we obtain

$$\zeta = \frac{\partial V}{\partial n} + \omega = \frac{\partial V}{\partial n} + \frac{V}{R} \tag{4.56}$$

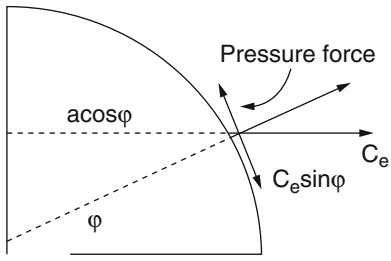



**Fig. 4.12** The cyclostrophic equilibrium in a slowly rotating planet

This equation tells us that the vorticity has two components, one related to the shear of the fluid and the other to the curvature. The sign will be positive for cyclonic vorticity and negative for anticyclonic vorticity.

## 4.5 Some Application of Circulation and Vorticity

We have just mentioned circulation and vorticity and we will find many interesting applications of this concept. In the following chapter, we will find a more rigorous formulation of these concepts.

### *4.5.1 The Sea Breeze*

We have now some knowledge of the dynamics that enable us to study apparently simple phenomena like the sea breeze. Whoever is familiar with vacations on the beach knows that baking in the sun is more easily bearable during a light, cool wind that blows from the sea toward the shore. During the night, the situation is reversed, with cool winds that now blow from the land to the sea.

A baroclinic fluid is exactly the cause of the sea breeze, as shown in Fig. 4.13. During the day, the surface of the sea warms up slowly because of the large heat capacity of the water. The surface of the land heats up faster and so a temperature difference results. To simplify things, we assume that the atmospheric column over the land has a temperature $T_2$, while over the sea the temperature is $T_1 = T_2 - \Delta T$. We also assume that the pressure at the surface is $p_0$, while at altitude $h$ is $p_1$. Then it can be easily seen that the lines of constant density are inclined, as shown in the figure. If the cold column has density $\rho$ at pressure $p$, then on the

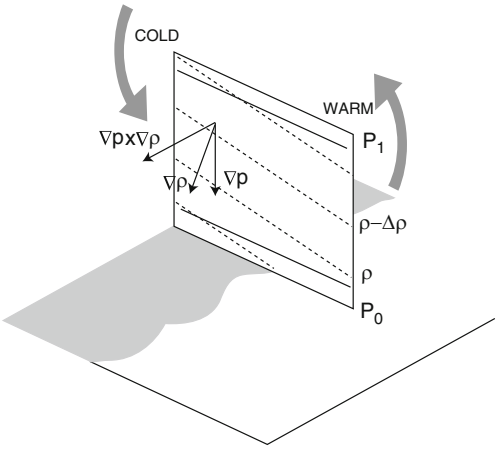



**Fig. 4.13** The sea breeze. The isobars are represented by *solid lines*, while the constant density lines are *dashed* (Adapted from Dutton 1976)

warm column, we will find the same density at pressure $p' = pT_2/T_1$. The vector representing the pressure and density gradients will form an angle, and their vector product will be oriented normal to the vertical plane. To evaluate the circulation, it is more convenient to use Eq. (3.32), so that we can write

$$\frac{DC}{Dt} = -\oint_\Gamma \frac{dp}{\rho} = \oint_\Gamma \frac{dp}{p} RT$$

where the integration path is extended to the rectangle shown in the figure on the vertical plane with height $h$ and length $L$. On this path, the pressure does not change on the horizontal, while for the vertical we have

$$\frac{DC}{Dt} = R \ln\left(\frac{p_0}{p_1}\right)(T_2 - T_1) \tag{4.57}$$

From this relation, it is possible to obtain the average acceleration on the path. This can be simply obtained by dividing Eq. (4.57) by the total length $2\,(h + L)$, so that

$$\left|\frac{DV}{Dt}\right| = R \ln\left(\frac{p_0}{p_1}\right) \frac{(T_2 - T_1)}{2\,(h + L)}$$

We can assume some typical value like $T_2 - T_1 = 10$ C, $p_0 = 1000$ hPa and the upper level at an altitude of 1 km corresponding to a pressure of 900 hPa and finally $L = 20$ km. With such values, we obtain acceleration around 25 $\mathrm{ms}^{-1}\ \mathrm{h}^{-1}$. That is much too large. The problem is that we have neglected friction which should slow down considerably the motion. It looks then that we have made a lot of effort just to get a wrong result. This is not completely true, as we will see in the next chapter, where we will show some more applications of what we have learned up to now.

### *4.5.2 Some Other Local Winds*

Sea breezes are just an example of what we call local winds. These actually are very important because they are determined mainly by the interaction of large-scale motions with local topography. This is the occasion, as an exercise, to talk about other motions in the atmosphere that can be treated with the few things about dynamics we know at this point. The first example is the motion produced by the buoyancy of air masses, which has a temperature $T_p$, while the ambient temperature is $T$. This corresponds to the *thermal updraft*. We have seen already that the acceleration in this case is simply

$$\frac{g\,(T - T_p)}{T}$$

The motion of the air parcel is opposed by a drag that can be assumed proportional to the square of the vertical velocity. The net acceleration is then

$$\frac{dw}{dt} = g\frac{(T - T_p)}{T} - C_w\frac{w^2}{z_i} \tag{4.58}$$

In this equation, $C_w$ is a drag coefficient and $z_i$ is the depth of the convective layer, which is the layer where the motion is taking place. The reason for writing the drag term in that particular form will be clear when we will talk about the boundary layer. From Eq. (4.58), we can estimate at the steady state ($dw/dt = 0$) the average vertical velocity. With an ambient temperature of 20 °C, a difference of 2 °C, and a depth of 1 km, we get $w = 3.65\ \text{ms}^{-1}$. This value has been obtained with a drag coefficient $C_w = 3$. The temperature difference and the depth of the convective layer are more representative of the common thermal you normally see on a road on a hot summer day. For a convective cloud in a thunderstorm, the temperature difference may be even 5 °C, and the depth is of the order of the entire troposphere, that is, 10 km. In this case, the velocity may reach 20 m $\text{s}^{-1}$. This is the reason why it is harmful to fly in a thunderstorm cloud.

Another interesting motion is the *katabatic wind*. As shown in Fig. 3.15, this motion is originated when a cold air mass (colder than the environment) slides downhill under the action of negative buoyancy. To study this problem, we will assume a local coordinate system, as shown in the figure. The $x$, $y$ plane is coincident with the hill slope with the $z$ axis normal. The velocity components will be oriented in the usual way. The accelerations are then

$$\frac{Du}{Dt} = g\frac{T - T_{kat}}{T}\sin\alpha + fv - C_d\frac{u^2}{h}; \quad \frac{Dv}{Dt} = -fu - C_d\frac{v^2}{h} \tag{4.59}$$

We can solve this problem neglecting any effect of the Coriolis term and assuming that the equilibrium is between the buoyancy term and the drag. This approximation can be useful when the length of the valley is such that the $v$ component does not reach too large values (Fig. 4.14). Actually, we could find a solution at the beginning, when the acceleration cannot be neglected, so that the first part of Eq. (4.59) becomes

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = g\frac{\Delta T}{T}\sin\alpha$$

that has the solution

$$u = \left(g\frac{\Delta T}{T}x\sin\alpha\right)^{1/2} \tag{4.60a}$$

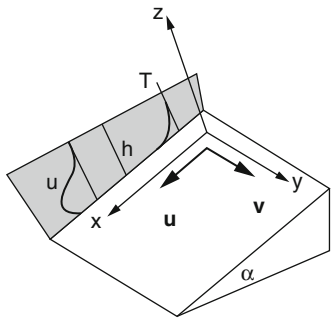



**Fig. 4.14** The geometry of the katabatic wind. On the vertical plane (in *gray*), the behavior of $T$ and $u$ with height is shown

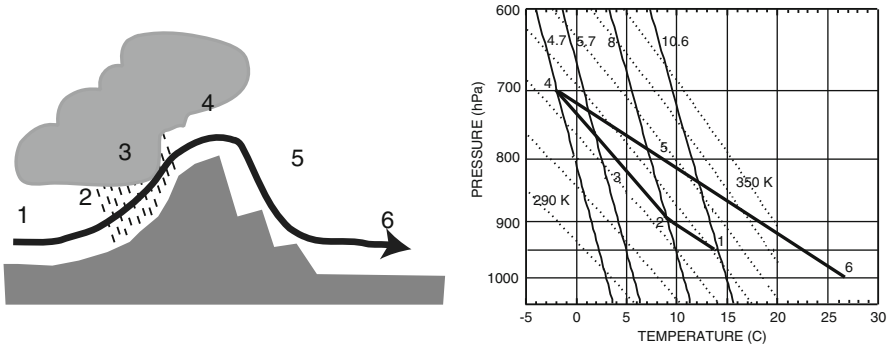



**Fig. 4.15** Thermodynamic of a foehn wind. At *left* is shown the trajectory of the air mass. The corresponding tephigram is shown at *right*

When the buoyancy equilibrates the drag term, we have

$$u = \left( g \frac{\Delta T}{T} \frac{h}{C_d} \sin \alpha \right)^{1/2} \tag{4.60b}$$

We use typical values $\triangle T = 10$ °C, $T = 10$ °C, $h = 20$ m, and $C_d = 0.005$ and we find that after 4 km the velocity has reached a value of 13.5 ms$^{-1}$ with a slope of 10°. The limit velocity given by Eq. (4.60b) is 18.9 ms$^{-1}$.

With the intensified studies of the Antarctic region, katabatic winds have become of increasing interest. Part of the continent can be the slope and velocities may reach values up to 30 ms$^{-1}$.

One last local wind we can mention with our limited knowledge is the winds generated by air parcels going over the mountains. These winds are called either foehn in Europe or chinook in North America. Their generation is schematically illustrated in Fig. 4.15, where an air mass is forced to pass over the mountain. The thermodynamic conditions of the mass are those of point 1 and are represented on the tephigram on the right. When the mass reaches the saturation point 2, it moves along the wet adiabat 3 until it reaches the highest elevation point 4. We can assume

that the air mass will lose all its water as rain along the upwind slope, so that after the top it will behave as a dry mass. In particular, descending from 4 to 5, the mass will warm following the dry adiabat 5, and finally it will reach a much higher temperature than the initial one.

The example shown starts with similar conditions as in Fig. 2.3, that is, a temperature of 14 °C at a pressure of 950 hPa and 8 g kg$^{-1}$ water vapor content.

Starting at about 900 hPa, the saturation point is reached, and the mass will follow the saturated adiabat up to 700 hPa to a mixing ratio of 4.7 g kg$^{-1}$. The difference of 3.3 g kg$^{-1}$ may be lost as rain and, if an additional 3.7 g kg$^{-1}$ are lost at the top, when the air descends, it will have a temperature of 26 °C at 1000 hPa and a relative humidity of 5 %. The net result of the entire process is rain on the windward side of the mountain and warm dry air on the lee valley.

### *4.5.3 The Rossby Waves*

A first, simple approach to the study of some of the properties of vorticity can be made starting from the equation of vorticity conservation (3.55) when the total divergence is zero:

$$\frac{D}{Dt}(\zeta + f) = 0 \tag{4.61}$$

The quantity in parentheses is called *absolute vorticity* and Eq. (4.61) tells us that for an incompressible fluid, this is a conserved quantity. Another limitation of Eq. (4.61) is that it neglects friction forces. Even so the equation can be used for an interesting application. Actually the equation of the conservation of absolute vorticity has been used in the first attempt to forecast the weather, around 1948. A very qualitative examination of this equation gives some indication about the solution, which happens to be something resembling a wave.

We consider an air parcel initially at $A$, as shown in Fig. 4.16. If the parcel moves south, the planetary vorticity will decrease (it goes like the sine of latitude), so that the relative vorticity must increase, that is, the curvature of the trajectory will be positive and the parcel will reach a minimum in latitude and then move toward north. At this point, the cycle will start again and the trajectory will have a wavelike appearance.

This conclusion can be put in mathematical terms if we solve Eq. (4.61) with some approximation. We consider a channel along a latitude circle. Its width in latitude must be such that the Coriolis parameter $f$ can be considered to change linearly with latitude, that is, $y$. In this case, we can write

$$f = f_0 + \beta y$$

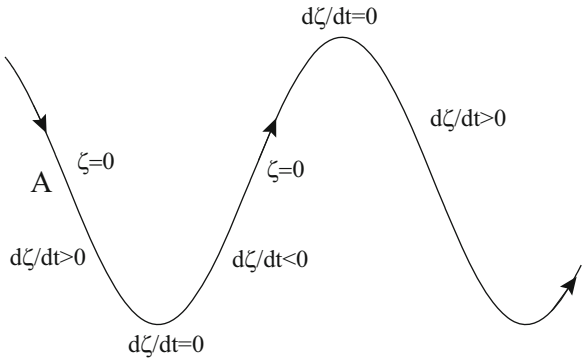



**Fig. 4.16** Trajectory of a fluid parcel conserving vorticity

where $f_0$ is some initial value and the parameter $\beta$ can be easily determined by the derivative of the Coriolis parameter

$$\frac{\partial f}{\partial y} = 2\Omega \cos\varphi \frac{\partial \varphi}{\partial y} = \frac{2\Omega \cos\varphi}{a} = \beta \tag{4.62}$$

where we have used $dy = ad\varphi$ with $a$ the Earth's radius.

This approximation is called the $\beta$-plane and if the value for $\beta$ is substituted in Eq. (4.61), we have

$$\frac{\partial \zeta}{\partial t} + u\frac{\partial \zeta}{\partial x} + v\frac{\partial \zeta}{\partial y} + v\beta = 0 \tag{4.63}$$

Such an equation has a very simple aspect, but the solution is not so obvious.

To simplify things further, we assume that in our channel there is a constant velocity current $\bar{u}$ (it could be the jet stream); then the components of the wind velocity can be written as

$$u = \overline{u} + u'; \quad v = v'$$

where the primes indicate quantities that are small deviations from $\bar{u}$. Actually we are representing the wind field as a sum of a basic flow and a perturbation. The same vorticity is a function only of perturbed quantities, and substituting in Eq. (4.63) and neglecting all those terms containing products of primed quantities, we obtain

$$\frac{\partial \zeta'}{\partial t} + \overline{u}\frac{\partial \zeta'}{\partial x} + \beta v' = 0 \tag{4.63a}$$

where

$$\zeta' = \frac{\partial v'}{\partial x} - \frac{\partial u'}{\partial y}$$

At this point, we recur to the streamfunction already defined such that

$$u' = -\frac{\partial \psi}{\partial y}; \quad v' = \frac{\partial \psi}{\partial x}$$

This, substituted into the definition of vorticity, will give

$$\zeta' = \nabla^2 \psi \tag{4.64}$$

so that Eq. (4.63) becomes

$$\frac{\partial \nabla^2 \psi}{\partial t} + \overline{u}\frac{\partial \nabla^2 \psi}{\partial x} + \beta \frac{\partial \psi}{\partial x} = 0 \tag{4.65}$$

We seek solutions for this equation that have wave characteristics. We then assume a solution of the type

$$\psi = Ae^{i(kx+ly-\omega t)} \tag{4.66}$$

This is in general a complex quantity of which we will take only the real part. The solution of Eq. (4.66) is actually a wave that propagates in the direction determined by the wave vectors $k$ and $l$ and with a frequency $\omega$. Taking only the real part, the function has the form

$$\psi = A\cos(kx + ly - \omega t) \tag{4.67}$$

This function is sketched in Fig. 4.17. The streamlines in this case are straight lines that slope toward the west according to the wave vectors. They become more dense near the flexure points of the surface and more sparse near the maxima and minima. The wind velocity can be found from Eq. (4.67) so that

$$u' = lA' \sin(kx + ly - \omega t); \quad v' = -kA' \sin(kx + ly - \omega t) \tag{4.68}$$

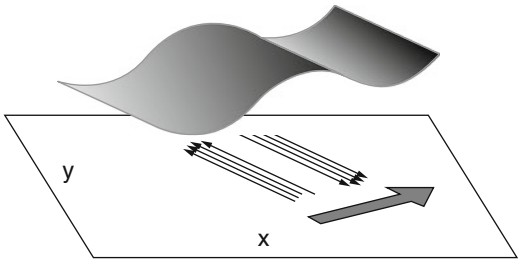



**Fig. 4.17** A Rossby wave in two dimensions. The curved surface is the streamfunction described by Eq. (7.9). On $x$, $y$, the projection of the *streamlines* (the intersection of the surface with the planes $\psi = \cos t$) with the orientation of the velocities. The *large gray arrow* gives the phase propagation

If we take a picture of the situation at some particular time $t = 0$, it is easy to see that the velocity changes sign following the maxima and minima of the streamfunction. The wave vectors are related to the wavelength by the relations

$$k = \frac{2\pi}{L_x}; \quad l = \frac{2\pi}{L_y}$$

If the propagation happens in the $x$ direction ($l = 0$), a very simple solution is obtained $v' = -kA\sin(kx - \omega t)$. The wind field can then be obtained as a sum of the current $\bar{u}$ and $v'$, as shown in Fig. 4.17.

We now substitute Eq. (4.67) into Eq. (4.68) to find which conditions the streamfunction $\psi$ must satisfy to be a solution. We have immediately for the different terms

$$\nabla^2\psi = -\left(k^2 + l^2\right)\psi; \quad \frac{\partial\psi}{\partial t} = -i\omega\psi; \quad \frac{\partial\psi}{\partial x} = -ik\psi$$

that substituted in Eq. (4.65) result in the condition

$$\omega = \overline{u}k - \frac{k\beta}{(k^2 + l^2)} \tag{4.69}$$

This represents the *dispersion relation* for the wave. This relation shows that the Rossby wave is *dispersive*, meaning that the frequency is not a linear function of the wavelength. A wave is characterized by the phase speed (we will talk at length about this later on) which is the velocity of the propagation of the perturbation (i.e., ridges and crests). This is given by

$$c = \frac{\omega}{k} = \overline{u} - \frac{\beta}{(k^2 + l^2)} \tag{4.70}$$

From this equation, we notice that the phase speed is regressive with respect to the basic wind (i.e., $\bar{u}$): an observer moving with the wind sees the wave moving in the opposite direction. If we use $L_x = 6000$ km and $L_y = 3000$ km and we adopt a 45° latitude ($\beta = 1.61\ 10^{-11}\ \mathrm{m}^{-1}\ \mathrm{s}^{-1}$), we have for the difference $c - \overline{u} \approx 18\ \mathrm{ms}^{-1}$ so that the phase velocity is usually a few meters per second. With similar arguments, it is possible to show that the frequency of the wave is of the same order of magnitude as the rotation frequency for the Earth, as would be expected for a small Rossby number. Equation (4.70) also tells us that the phase velocity increases with the wavelength, and this has some important implications for meteorology. In this case, the increase means that the phase velocity becomes more and more retrograde; in other words, in the present example, the long waves tend to move toward west.

The physical reason behind this behavior can be understood by looking at Fig. 4.18 and Eq. (4.63). According to this equation, the rate of change of vorticity depends on two terms, the advection of relative vorticity $u\partial\zeta/\partial x$ and planetary

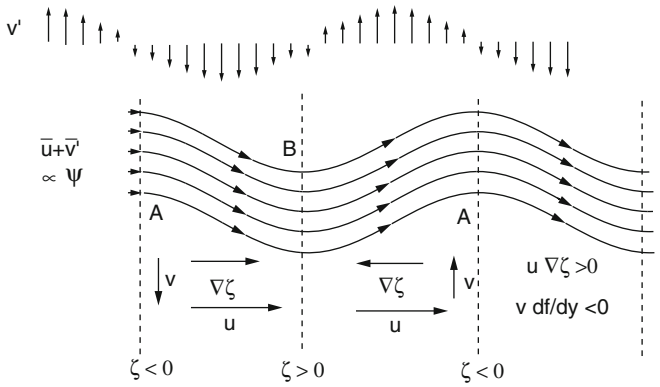



**Fig. 4.18** The composition of the velocity field given by Eq. (4.68) with the basic flow. The *sinusoidal lines* are isopleths of the streamfunction. Region A corresponds to a high-pressure area and B to a low-pressure one. The amplitude of the component *v*' is the maximum in correspondence with the flexure point of the *streamline*. Also shown are the gradients of planetary and relative vorticity (Dutton 1976)

vorticity $v\partial f/\partial y$. The same figure shows that, around the region of the maximum negative vorticity, flow is anticyclonic, as around a high-pressure zone, and vice versa for the positive maximum. In the region between high and low, the advection of relative vorticity is negative [remember that also for the vorticity the advection must be taken with the negative sign $-(u\partial\zeta/\partial x)$], while the advection of planetary vorticity is positive. The advection of relative vorticity will move the streamline pattern *downwind* (as a meteorologist would say), while the planetary vorticity will move then *upwind*. The net result will depend on which of these two tendencies prevail. We can see that in this case (dependence on $x$ only) the advection of relative vorticity will increase as the cube of the wavelength. The long waves then will slow down the phase speed or even make retrograde the streamfunction and consequently the high and low pattern.

We have mentioned a few things about this solution, and some of them may also be of practical utility. Also we hope that this has been an occasion to learn about streamfunctions and streamlines. We have also appreciated the advantages of an approach based on the conservative properties of the absolute vorticity, and we would like to see if there is something more profound about conservation principles. This may help us to be even more synthetic (reductionists, they would say today). To try this road, we need to better understand vorticity.

In this chapter, we gained territory step by step, one bite of theory and another of the applications. We finally extended the thermal wind equation to a slowly rotating planet, but to go further, we need to abandon the geostrophic stuff and try to have a more general vision through the vorticity concept.

## E.4 Examples

### *E.4.1 The Sea Breeze Circulation*

The theory given in 3.3.2 on sea breeze is correct but is limited to the case in which friction is absent. Also we have not mentioned that the direction of the sea breeze does change during the day and this may be attributed to the effect of Coriolis acceleration. We will refer to an old paper by Haurvitz published in 1947 and to Rotunno (1983).

In this paper, the reference system assumes the $x$ axis oriented perpendicular to the shoreline while the $y$ axis is parallel. The forcing function is obtained assuming that the sun shines only for half a day so that in principle the forcing function $G(t)$ should be of the form

$$G(t) = A\cos\omega t \quad \text{for} \quad -\frac{1}{2}\pi \leq \omega t \leq \frac{1}{2}\pi \tag{E.4.1}$$

and vanishing for the other half of the period. To make this function continuous, we assume the first two terms of its Fourier expansion:

$$F(t) = A/\pi + \frac{1}{2}A\cos\omega t \tag{E.4.2}$$

Then the equation of motion can be written as

$$\begin{aligned} \frac{du}{dt} - fv + ku &= -fv_g \\ \frac{dv}{dt} + fu + kv &= -fu_g \end{aligned} \tag{E.4.3}$$

where we have added a friction term in its simplest form (linear with velocity). Equation (E.3.13) can be solved very easily by introducing the complex variable:

$$W = u + iv$$

Summing the two equations, we get

$$\frac{dW}{dt} + (if + k)\,W = -fv_g + ifu_g - F(t) \tag{E.4.4}$$

$$\begin{aligned} W = &-\frac{f\left(v_g - iu_g\right)}{if + k} - \frac{A}{\pi}\frac{1}{if + k} \\ &- \frac{A}{2}\frac{\omega\sin\omega t + (if + k)\cos\omega t}{(if + k)^2 + \omega^2} + Ce^{-(if+k)t} \end{aligned} \tag{E.4.5}$$

The first term on the rhs represents the motions caused by the constant part of the pressure gradient force. In case of absence of friction, the term reduces to

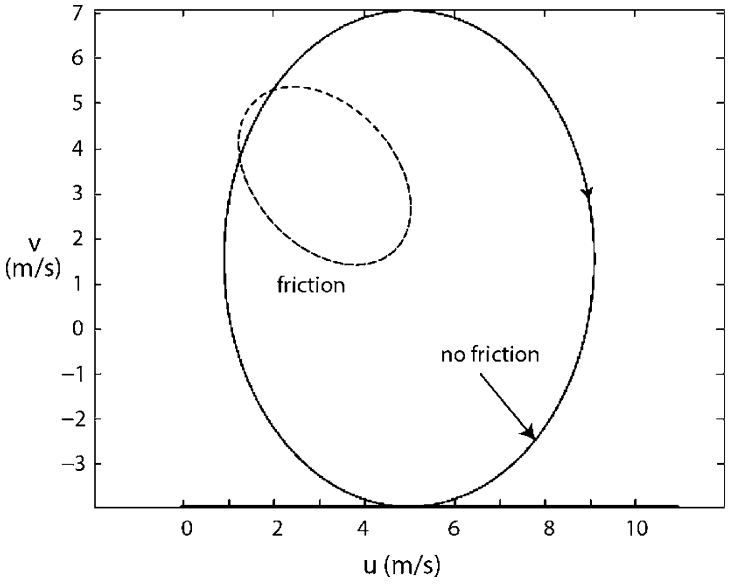



**Fig. E.4.1** The two odographs for the breezes with and without friction. See text for details

the geostrophic motion. The second and third terms are the effect of that part of pressure gradient due to the temperature difference between land and water. In the fourth term, C is an arbitrary integration constant. This constant is determined by considering that in the absence of forcing (i.e., $A = 0$ and $v_g - iu_g = 0$), $W$ should vanish and then $C = 0$. In case friction is absent, the solutions to (E.3.15) are particularly simple and are obtained by putting $k = 0$ and equating real and imaginary part:

$$\begin{aligned} u &= u_g - \frac{A}{2}\frac{\omega}{\omega^2 - f^2}\sin\omega t \\ v &= v_g + \frac{A}{f\omega} - \frac{A}{2}\frac{f}{\omega^2 - f^2}\cos\omega t \end{aligned} \tag{E.4.6}$$

We see that, superimposed to the geostrophic motion, there is a diurnal motion which has the period of the forcing. This diurnal motion goes to infinity at a latitude of 30° where $f = \omega$. In Fig. E.4.1, we compare the two cases with and without friction.

## *E.4.2 The Circulation Around Lows and Highs*

We have mentioned the possible circulations resulting around low- and high-pressure zones. We will try to be a little bit more analytic. Referring to Fig. E.4.2, we consider the equilibrium around the two pressure zones. We use the notation

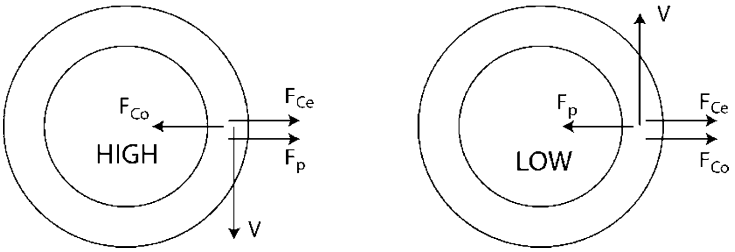



**Fig. E.4.2** Equilibrium around high- and low-pressure centers

$$F_p = -\left(\frac{1}{\rho}\right)\left(\frac{\partial p}{\partial r}\right); \quad F_{Co} = fV; \quad F_{Ce} = \frac{V^2}{r} \tag{E.4.7}$$

as the acceleration due to pressure gradient, Coriolis, and centrifugal forces. Then we get at the equilibrium

$$\frac{V^2}{r} - fV = -\frac{1}{\rho}\frac{\partial p}{\partial r} \text{ for the high pressure center} \tag{E.4.8}$$

$$\frac{V^2}{r} + fV = \frac{1}{\rho}\frac{\partial p}{\partial r} \quad \text{for the low-pressure center} \tag{E.4.9}$$

If we solve for the velocity, we get

$$V_L = \frac{fr}{2}\left[-1 + \left(1 + \frac{4V_g}{fr}\right)^{1/2}\right]; \quad V_H = \frac{fr}{2}\left[1 - \left(1 - \frac{4V_g}{fr}\right)^{1/2}\right] \tag{E.4.10}$$

for low-pressure and high-pressure centers, respectively, where we have used the notation $V_g = (1/f\rho)\,\partial p/\partial r$ as the geostrophic velocity; if we put $R_{oc} = V_g/fr$, we obtain

$$\begin{aligned} V_L &= \left(\frac{V_g}{2R_{oc}}\right)\left[-1 + (1 + 4R_{oc})^{1/2}\right] \\ V_H &= \left(\frac{V_g}{2R_{oc}}\right)\left[1 - (1 - 4R_{oc})^{1/2}\right] \end{aligned} \tag{E.4.11}$$

We see that around low-pressure centers, the gradient wind is always smaller than the geostrophic value, while the opposite happens for high-pressure centers. Besides, for the latter, it must be $R_{oc} \leq 1/4$. This means that around high pressure the gradient must be more gentle because the relation holds:

$$\left|\frac{\partial p}{\partial r}\right| \leq 0.25\rho f^2 r$$

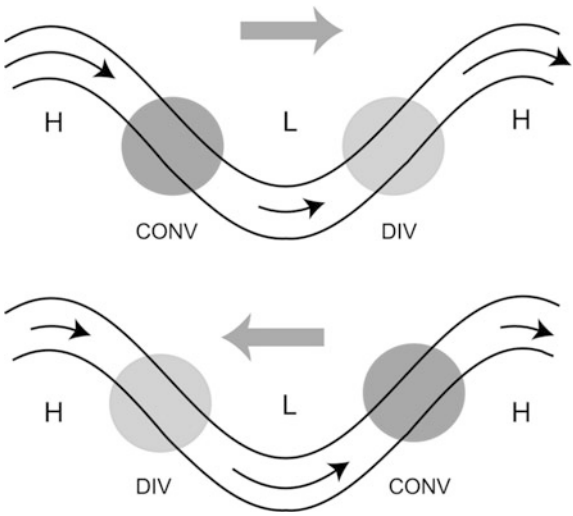



**Fig. E.4.3** The distribution of divergence due to the curvature effect of a pattern of sinusoidal isobars. East is on the *right* (Hess 1979)

This means that around a high-pressure center, the gradient cannot exceed a certain value which depends on the latitude and distance from the center.

### *E.4.3 Effects on the Propagation of Long Waves*

There is a consequence of what we just saw on the propagation of long waves in the atmosphere. Consider as shown in Fig. E.4.3 a series of isobars which have a sinusoidal pattern. The ridges of the upper wave can be associated with high-pressure zones and consistently the circulation there is anticyclonic. The trough on the other hand can be associated with a low-pressure center with cyclonic circulation. The velocity around this center is influenced by two effects, the first being that associated with (E.4.10), that is, the velocity around high pressure is larger than the geostrophic values, while the opposite is true around low-pressure centers. This difference will result in a convergence of the flow between the ridge and the trough, while there is divergence between the trough and the ridge. This is shown in Fig. E.4.3 as darker- and lighter-filled circles. The tendency for this wave will be then to move toward east because the high-pressure zone and low-pressure zone will tend to move toward east as shown by the gray arrow. Opposite to this effect, there is the one related to the change with latitude of the Coriolis parameter. We write the second of (E.4.9) as (equation is always correct with the appropriate sign for the pressure gradient)

$$V = \frac{1}{f}\left(\frac{1}{\rho}\frac{\partial p}{\partial r} - \frac{V^2}{r}\right)$$

The centrifugal term may be neglected and we see that the velocity decreases with increasing latitude (i.e., $f$ increases). The latitudinal effect is shown in the bottom portion of Fig. E.4.3 where now the situation is reversed. We have divergence between ridge and trough and convergence between trough and ridge. The results are a tendency for the wave to move westward. It is clear that such effect will dominate when the wind speeds are low and for long wavelengths because there are conditions to neglect the centrifugal term. However, we can put all this in clear mathematical terms.

We may assume a streamfunction of the form

$$\psi = \psi_0 - Uy + A \sin kx \cos ly$$

where $U$ is the background zonal wind. We need to calculate the advection of planetary and relative vorticity:

$$-u\frac{\partial \zeta}{\partial x} - v\frac{\partial \zeta}{\partial y}; \quad \beta v$$

where

$$\begin{aligned} \zeta = \nabla^2 \psi &= -\left(k^2 + l^2\right) A \sin kx \cos ly = \zeta_g \\ u = -\frac{\partial \psi}{\partial y} &= U - lA \sin kx \sin ly = U + u_g \\ v = \frac{\partial \psi}{\partial x} &= kA \cos kx \cos ly = v_g \end{aligned}$$

It can be easily verified that

$$u_g \partial \zeta / \partial x + v_g \partial \zeta / \partial y = -\left(\frac{\partial \psi}{\partial y}\right) \partial \nabla^2 \psi / \partial x + (\partial \psi / \partial x)\, \partial \nabla^2 \psi / \partial y = 0$$

so that for the advection of relative and planetary vorticity, we get

$$-u \partial \zeta / \partial x - v \partial \zeta / \partial y = -U \partial \zeta / \partial x = kU \left(k^2 + l^2\right) A \cos kx \cos ly \tag{E.4.12}$$

$$\beta v = \beta kA \cos kx \cos ly \tag{E.4.13}$$

We notice that keeping the amplitude constant for long waves (small $k$), the advection of planetary vorticity dominates, while for short waves (large $k$), the advection of relative vorticity is the dominant term. We can easily find the wavelength separating the two processes by equating (E.4.12) and (E.4.13); we get

$$\left(k^2 + l^2\right) = \frac{\beta}{U} \tag{E.4.14}$$

Using $k \approx l$, $U \approx 10 \mathrm{ms}^{-1}$, and $\beta = 1.6 \cdot 10^{-11} \mathrm{m}^{-1}\mathrm{s}^{-1}$, we get a wavelength of roughly 10.000 km.

## References[1]

### *Books*

Dutton JA (1976) The ceaseless wind: an introduction to the theory of atmospheric motion. McGraw Hill, New York
Hess SL (1979) Introduction to theoretical meteorology. Constable
Holton JR (1992) An introduction to dynamic meteorology. Academic, New York
Houghton JT (1977) The physics of the atmospheres. Cambridge University Press, Cambridge/New York
Marshall J, Plumb RA (2008) Atmosphere ocean and climate dynamics. Academic, Amsterdam
Pedlosky J (1987) Geophysical fluid dynamics. Springer, New York

### *Articles*

Haurwitz B (1947) Comments on the sea breeze circulation. J Meteor 4:1
Rotunno R (1983) On the linear theory of land and sea breeze. J Atmos Sci 40:1999

[1]This chapter is a follow-up of the previous one. Basically it has the same references with the addition of few specialized papers.

# Chapter 5
# Atmospheric Chemistry

Our plan is to introduce in the first six chapters most of the topics that will be studied in much more detail in the book. It is practically impossible to talk about atmospheric chemistry without going in some detail about the structure of the atmosphere of the Earth. For completeness but also for the sake of comparison, we will report some data also on other planets. Every once in a while, we will mention planetary atmospheres; this perspective we have indicated before may be strange: Why spend all that money to see expensive toys roving on the Martian surface when we have plenty of problems (as we will see) on Earth? The answer is that the environmentalist slogan "we have only one Earth" is a great handicap for the atmospheric physicist because it means having only one laboratory in which to acquire data, which are obtained from experiments we cannot control. In this framework, even expenses to send probes into Titan's atmosphere (which not only transmit beautiful pictures back to the Earth) are justified because some of those data may be useful to understand how the atmosphere of the Earth works at the least from a chemical point of view. Of all this, we hope, we will become aware in the course of this book.

## 5.1 Characteristics of the Atmospheres

We will start looking at those parameters which could determine the environment of the terrestrial planets (Venus, Mars, and the Earth), as listed in Table 5.1. Some of the quantities which appear in the table are obvious, and others already show some serious sign of jargon. The *solar constant* is the power received at the orbit of the planet for unit surface and its value should be related to the average temperature of the planet.

Another important parameter is the *albedo*, the fraction of the incident energy that is reflected back by the planet and its atmosphere. The *albedo* is in part due



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_5

**Table 5.1** Some atmospheric parameters for the terrestrial planets

| Quantity | Venus | Earth | Mars |
|---|---|---|---|
| Solar constant (kW $m^{-2}$) | 2.62 | 1.36 | 0.59 |
| Albedo | 0.77 | 0.3 | 0.14 |
| Cloud cover (%) | 100 | 50 | Variable |
| Cloud chemical composition | $H_2SO_4$ | $H_2O$ | $H_2O$, $CO_2$ |
| Orbit eccentricity | 0.007 | 0.017 | 0.093 |
| Inclination of the axis | 2°36' | 23°27' ± 1° | 25°12' ± 13° |
| Orbital period (days) | 225 | 365 | 687 |
| Length of the day (days) | 243 | 1 | 1.03 |
| Solar day (days) | 117 (retrograde) | 1 | 1.03 |
| Gravity acceleration (m $s^{-2}$) | 8.87 | 9.81 | 3.72 |
| Radius (km) | 6051 | 6378 | 3394 |
| Surface pressure (105 Pa) | 95 | 1 | 0.007–0.01 |
| Vertical temperature gradient (K $km^{-1}$) | 7.5 | 6 | 3 |
| Surface temperature (K) | 737 | 288 | 220 |
| $\Delta T$ (equator-pole) (K) | 5–15 | 45 | 90 |
| $\Delta h$ (topography) (km) | 13 | 9 | 25 |
| Rossby number | 20 | 0.1 | 0.1 |

to the cloud cover of the planet, that is, to what fraction of the surface is covered on the average: we can see that there is a direct connection between reflectivity and cloud cover. Venus is completely covered with clouds and has a very high albedo, while for Mars, which always has a clear sky, the albedo is essentially that of the surface. Mars however experiences worldwide dust storms and in this case its albedo increases drastically.

The chemical composition of the clouds of the different planets gives important indications as to the processes which take place in the atmosphere. Venus has clouds made of sulfuric acid and thus the atmosphere is very dry. On the other hand, the Earth is rich in water and its clouds are the proof, as may be the case for Mars. In this last case, it is interesting to note that the clouds may also be made of dry ice (i.e., solid carbon dioxide), and the process of sublimation, that is, the phase change from gas to ice, is very important on Mars during the summer. In the Martian northern hemisphere, the passage of carbon dioxide from the polar cap to the atmosphere increases the total atmospheric pressure, as shown in the table.

The orbital characteristics of a planet may be important for the long-term climatic variations. Everyone remembers having studied the Milankovitch theories of the ice age, based on the fact that some of the Earth's orbital parameters change slowly with time. In this way, the solar radiation absorbed especially at high latitude is modulated. Venus looks very stable, while both on the Earth and Mars, the inclination of the rotation axis changes within the indicated limits. On the same planets, the axis also has a precession motion, that is, the planets behave like a spinning top. On the Earth, the precession has a period of about 26,000 years, while

the inclination completes a cycle every 42,000 years. For Mars, the situation is even more complex because the amplitude of the inclination change is about 13°, while the period has different components, of the order of hundreds of thousands of years. The precession has a period of 165,000 years. Incidentally it appears that the greater orbital stability of the Earth with respect to Mars is due to the presence of a large moon.

The other parameters shown in Table 5.1 require some explanation. One of these is the Rossby number, which we will encounter again and again. For the moment, it is sufficient to say that if the wind velocity in the atmosphere has a typical value $U$, then an air mass will be transported to a distance $L$ in the time interval $L/U$. The rotation of the planet will influence the motion only if this time is of the same order (or longer) than the rotation period, $T$. The Rossby number is just the ratio between these two times and is given by

$$R_0 = \frac{UT}{L}$$

On those planets where this number is small, rotation has a great influence on atmospheric motions through the Coriolis acceleration. We should then expect that both on Mars and on Earth (but not on Venus), the motion of air masses will be strongly influenced by rotation. The significance of other data in the table, like the average level difference in the orography, the average pole to equator difference in temperature, and the vertical temperature gradient, will be explained later.

We turn now to the chemical composition of the atmospheres, shown in Table 5.2. The chemical composition is reported in terms of the volume mixing ratio, which is the ratio of the number density of a gas in the atmosphere to the atmospheric number density; in other words, it is the ratio between the number of molecules of that gas present in a given volume with respect to the total number of molecules present in the same volume. The first thing to notice is the drastic difference in the composition between the Earth's atmosphere and the atmospheres of Mars and Venus. The explanation of these differences is one of the topics we will deal with when discussing the origin of planetary atmospheres. However, it is clear that the composition of the Earth's atmosphere is influenced by the biological activity which is present only on this planet. The presence of oxygen is the most evident proof of this influence. There are a few calculations indicating that if life should disappear on Earth, then thermodynamic equilibrium would bring its atmosphere to a composition very similar to those of Mars and Venus.

Many processes that appear in the table may seem obscure at the moment, but they will become clearer as we proceed into the book. The Earth's atmosphere is composed mostly (99 %) of nitrogen and oxygen, which are completely inert gases. The remaining 1 % is made up of argon, a noble gas, and then, in even smaller quantities, carbon dioxide, water vapor, ozone, etc. However, it is just those gases with mixing ratios less than 1 % that are responsible for some of the most important properties of the Earth's atmosphere and the climate of the

**Table 5.2** The chemical composition of the atmospheres of the Earth, Venus, and Mars. Also reported are the processes which are thought to be responsible for the sources and sinks

| Gas | Mixing ratio(*) | Source | Sink |
|---|---|---|---|
| Earth | | | |
| $N_2$ (nitrogen) | 0.781 | Biologic | Biologic |
| $O_2$ (oxygen) | 0.209 | Biologic | Lithosphere-bio |
| Ar (argon) | 0.0093 | Degassing | |
| $H_2O$ (water vapor) | <0.04 | Evaporation | Condensation |
| $CO_2$ (carbon dioxide) | 0.0034 | Combustion, biologic | Biologic |
| $^{36.38}$ Ar | 0.000037 | Degassing | |
| $^{20.22}$ Ne (neon) | 0.0000182 | Degassing | |
| $CH_4$ (methane) | $1.7–3 \times 10^{-6}$ | Biologic | Photooxidation |
| $N_2O$ (nitrous oxide) | $3.1 \times 10^{-7}$ | Biologic | Photodissociation |
| CO (carbon monoxide) | $0.4–2 \times 10^{-7}$ | Photochemical | Photochemical |
| $O_3$ (ozone) | $0.1–1 \times 10^{-7}$ | Photochemical | Photochemical |
| NO, $NO_2$ (nitrogen oxides) | $0.2–5 \times 10^{-10}$ | Combustion, biologic | Photooxidation |
| $SO_2$ (sulfur dioxide) | $3 \times 10^{-10}$ | Combustion | Photooxidation |
| Venus | | | |
| $CO_2$ | 0.965 | Degassing | Carbonates for |
| $N_2$ | 0.035 | Degassing | |
| CO | 0.00002 | Photochemical | Photooxidation |
| $SO_2$ | 0.00015 | Photochemical | Formation $CaSO_4$ |
| $H_2O$ | 0.0001 | Degassing | Hydrates |
| Mars | | | |
| $CO_2$ | 0.953 | Degassing, evaporation | Condensation |
| $N_2$ | 0.027 | Degassing | Escape |
| $^{40}$ Ar | 0.016 | Degassing | |
| O2 | 0.0013 | Photochemical | Photoreduction |
| $H_2O$ | 0.0003 | Evaporation desorption | Condensation adsorption |
| $O_3$ | Photochemical | Photochemical | |
| NO | $7 \times 10^{-5}$ (120 km) | Photochemical | Photochemical |

(*) See Example E.5

planet. In turn, the climate and the environment determined by the tiny amount of these gases make life possible on our planet. For example, ozone is a gas that absorbs ultraviolet radiation of solar origin at wavelengths shorter than 300 nm, thus shielding biological tissues from potentially dangerous radiation. The stability of the environment during billions of years has contributed to the widespread diffusion of different life forms. The foregoing appears to be a circular argument, since the presence of oxygen (hence ozone) and the level of $CO_2$ in the Earth's atmosphere are due to the presence of life itself. On the other hand, the abundance of some of these trace gases is influenced by human activity (e.g., carbon dioxide and nitrogen

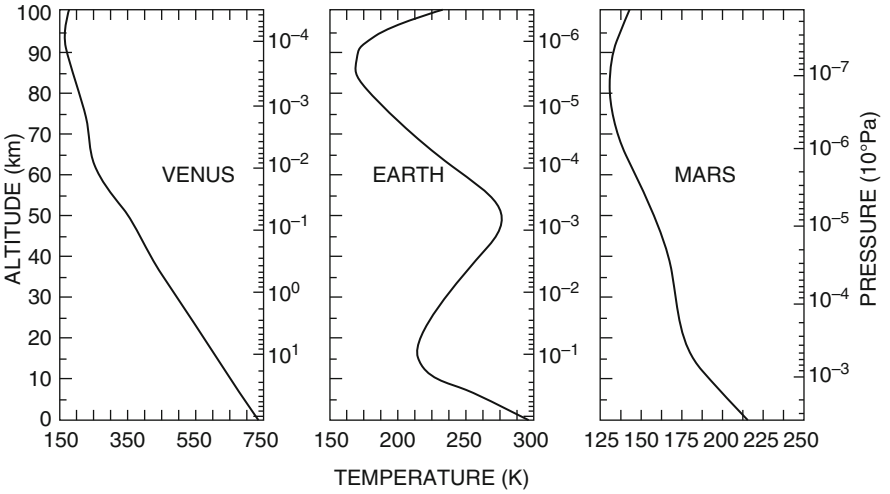



**Fig. 5.1** Temperature as a function of pressure and altitude on Venus, the Earth, and Mars. Notice the difference in surface temperature and the pressure scale

oxides) so that the environment will be influenced not only by biological activity but also by industrial activity. The atmospheric composition, that is, the mixing ratios of the different gases, determines the temperature at the surface and its rate of change with altitude. Figure 5.1 shows the temperature profile as a function of pressure for the three planets we have considered so far. The main feature of the profiles is that, neglecting the very different surface temperature, there is a constant decrease of temperature with altitude. This is not generally true for the Earth and presumably not for the thermospheres of Mars and Venus either. It is only in some lower layer of different gases that determines the temperature at the surface and its rate of change with altitude.

This behavior can be easily explained by noting that the bottom of the atmosphere is heated by the surface of the planet and very much less by the direct effect of solar radiation. In fact the surface absorbs solar radiation and infrared radiation coming from the atmosphere and heats up to the values shown in the figure. The atmosphere in contact with the surface is then heated and the warm air masses lift and expand, cooling adiabatically. We will see in a while that it is possible to evaluate rather accurately the temperature change with altitude, at least in the lower layer known as the *troposphere,* which is dominated by convective movements. Above the troposphere, the temperature change depends on the chemical composition of the atmosphere and on the distribution of clouds. The troposphere of Venus, which actually reaches up to 60 km, is filled with clouds at different altitudes whose chemical composition is very similar to a solution of sulfuric acid in water. In the case of the Earth, the troposphere also contains the highest clouds found in the atmosphere. Actually, very thin clouds like polar stratospheric clouds and noctilucent clouds are found at higher

altitudes. Equatorial clouds can reach altitudes up to 15 km (very exceptionally, up to 18–20 km). At the top of the troposphere (*tropopause*), the temperature reaches a minimum and then shows a constant increase up to about 50 km in the layer known as the *stratosphere*. These two regions, troposphere and stratosphere, are actually the most important from a practical point of view, because all human activities take place in the troposphere (including intercontinental airline flights) and the most important weather phenomena also occur in this region. On the other hand, the stratosphere is where most of the ozone gas resides. Its concentration is regulated by a very delicate equilibrium that can be perturbed by human activities.

Above the stratosphere, there is a layer, the mesosphere, where the temperature again decreases with altitude. Above this layer, the temperature increases up to values of 1000–1500 K around 300 km altitude in the regions known as the *thermosphere* and the *exosphere*. The high temperature and low density in these outermost regions give the molecules such high velocities and make the collision frequency so low that a very slow evaporation of the lightest gases occurs (the same effect we called escape in Table 5.2). For some planets and even for the Earth, the escape of light gases has a very important influence on the chemical evolution of the atmosphere.

In a very qualitative manner, we have explained the temperature variation in the troposphere. However, the atmosphere contains gases, like water vapor and ozone, that absorb solar radiation and warm the atmosphere locally; other gases, like ozone and carbon dioxide, emit infrared radiation and cool the atmosphere. When these two processes balance each other, *radiative equilibrium* is established; above the troposphere, these are the prevailing conditions. The maximum of temperature is observed around 50 km and is actually determined mainly by the UV absorption by ozone below 300 nm. At even higher altitudes, in the lower thermosphere, oxygen and nitrogen absorb the most energetic part of the UV spectrum (below 100 nm). The energy deposited amounts only to a few tens of erg $cm^{-2}$, which is nonetheless enough to heat the atmosphere to very high temperatures. The reason is that, above 100 km, the principal heat transport mechanism is the molecular conduction which is very inefficient; this implies that the very small amount of heat deposited on such low-density atmosphere is poorly dissipated and as a consequence very high temperatures result.

## 5.2 Atmospheric Composition and Chemistry

As shown in Table 5.2, the composition of the Earth's atmosphere is controlled mainly by biological processes at the least for the most abundant gas, nitrogen. On the other hand, in the case of oxygen, the biological control is negligible. The other gases like water vapor are controlled by evaporation/condensation processes. For all the other minor gases, their mixing ratio and distribution result from a number of complex chemical and physical processes.

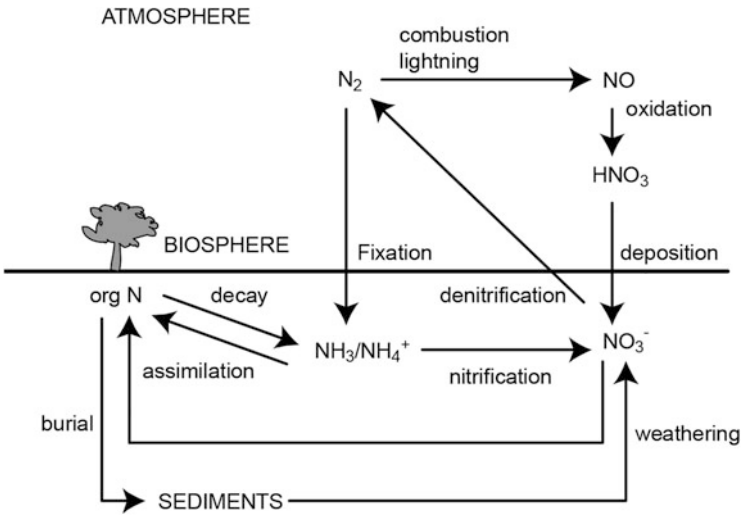



**Fig. 5.2** A simplified nitrogen cycle (Adapted from Jacob)

We start just to give an example from nitrogen. Most part of the nitrogen is contained in the atmosphere the total amount being $4 \times 10^{21}$ g, while the next larger reservoir is the land biota and soil with an amount ranging between $0.8 \times 10^{16}$ and $1 \times 10^{17}$ g. Other reservoirs include ocean and sediments ($3 \times 10^{21}$ g). Figure 5.2 shows the main processes that exchange nitrogen between reservoirs. A distinction must be made between the $N_2$ molecules which are highly stable and biological available nitrogen called fixed nitrogen. This process called fixation is produced by a class of bacteria which can reduce atmospheric $N_2$ to ammonia $NH_3$. Ammonia is assimilated as organic nitrogen by bacteria and plants and then consumed by animals. The nitrogen excreted by these animals can be mineralized to ammonium ($NH_4^+$). Ammonium can be used as energy source by bacteria and converted either to nitrite ($NO_2^-$) or nitrate ($NO_3^-$). In case of absence of oxygen in deep soil (anaerobic conditions), bacteria may use nitrates to convert organic carbon to $CO_2$ and in the process produce nitrogen. A familiar scheme (called denitrification) could be

$$6NO_3^- + 5CH_3OH \rightarrow 2N_2 + 5CO_2 + 7H_2O + 6OH$$

In this case, the organic carbon source is $CH_3OH$. Denitrification returns nitrogen to the atmosphere from the biosphere. Nitrogen can be also fixed through a quite different path. At high temperature (combustion) or in discharge (lightning), $N_2$ can form nitrogen oxide (NO) that can be further oxidized to nitric acid ($HNO_3$) which is also water soluble. In industrial regions of the world, fixation by combustion exceeds the natural fixation of $N_2$ contribution to fertilization.

On the other end, even these apparently nonbiological processes actually depend on biology. For example, combustion depends on organic material but also efficiency of lighting depends on the amount of oxygen present like $O_2$ and $CO_2$, and both of these gases are controlled in large part by biology. It is interesting to get some number on residence time for nitrogen. We ask what would happen if all the biological processes were removed. Then the only removing process would be lighting and combustion which account for about $30 \times 10^6$ ton/year. It would take then roughly 130 million years to remove all the atmospheric nitrogen. On the other hand, if we stop only denitrification and leave all the other fixation processes (industrial, biofixation, combustion), then it would take $13 \times 10^6$ years to deplete all the nitrogen. These times are so long that human activity hardly can have effect on the amount of nitrogen. Even if all the nitrogen contained in the lithosphere could be transferred to the atmosphere, the concentration would increase by 50 %.

Is this an occasion to demystify something about oxygen that most people believe is being controlled by biological activity? Actually the total amount of oxygen in the atmosphere is about $1.2 \times 10^{21}$ g while the total amount of organic carbon present in the biosphere is about $4 \times 10^{18}$ g. If this carbon would be oxidized (e.g., in case all the photosynthesis stops), it would consume only 1 % of the atmospheric oxygen. Actually oxygen in the atmosphere is controlled by very complex processes which involve the lithosphere. These processes are not completely abiotic but involve plate tectonics so they act on a time scale of a hundred million years. We will examine in more detail these processes when dealing with the cycle of carbon.

Going back to Table 5.2, we see however that on Earth most of the processes controlling atmospheric composition are biological while on other planets the processes are mostly chemical cycles independent from biology. However, on Earth, some of the processes may have changed in the course of its geological history.

## 5.3 Chemical Kinetics

Later in the book, we will present some detailed chemical mechanisms that require the knowledge of some elements of chemical kinetics. Although chemistry is one of the major courses for any physics student, it is surprising to see how few of them are familiar with this topic. For most of the gas phase reactions, the chemical rates can be grouped in three major forms:

$$\frac{d\,[i]}{dt} = \begin{cases} -J_i\,[i] \\ -k_{ij}\,[i]\,[j] \\ -l_{ij}\,[i]\,[j]\,[M] \end{cases} \tag{5.1}$$

In this expression, the square brackets indicate the number density of chemical species $i$, $j$, or $M$ ($M$ any atmospheric molecule). $k_{i,j}$ and $l_{i,j}$ are the chemical

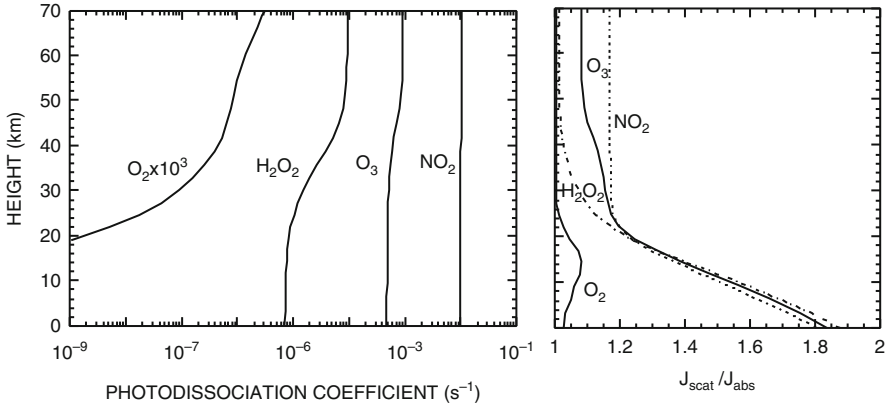



**Fig. 5.3** On the *left* the photodissociation coefficients as a function of altitude are shown for oxygen, hydrogen peroxide, ozone, and nitrogen dioxide. On the *right* is shown the ratio between the same coefficients calculated by taking into account scattering with respect to the case of pure molecular absorption

rate constants for the bimolecular and three-body reactions, respectively. The rate constant for first-order reaction $J$ is called the photodissociation coefficient and may depend both from space and time. This rate depends on the amount of absorbed solar radiation (mostly in the UV). The calculation of $J$ is very simple and we can refer to Sect. 2.2. If we express the actinic flux $I_\lambda$ as the number of photons per unit surface, unit time, and wavelength, the number of photons absorbed by a single molecule will be $\sigma_{i,\lambda} I_\lambda$ where $\sigma_{i,\lambda}$ is the absorption cross section at the wavelength $\lambda$. The quantum yield $q_{i,\lambda}$ is defined as the probability that the absorption of a single photon will produce a photodissociation event and then the photolysis rate constant is given by

$$J_{i,\lambda} = q_{i,\lambda} I_\lambda \sigma_{i,\lambda}$$

Summing over the wavelength, we obtain the photodissociation coefficient:

$$J_i = \int_\lambda q_{i,\lambda} I_\lambda \sigma_{i,\lambda} d\lambda \tag{5.2}$$

If we remember from Sect. 2.2, we always talk about energy fluxes while in this case we have the number of photons. Conversion is not so straightforward because the photon flux may contain scattering effects. Figure 5.3 shows the photodissociation coefficient for some common species in the atmosphere. It also shows that scattering effects can have a large influence on the photodissociation coefficient.

The simple expression for the chemical rates in the case of photolysis gives the occasion to introduce the concept of lifetime. We consider molecule $X$ which is photodissociated according to the mechanism

$$X + h\nu \rightarrow Y + Z \tag{5.3}$$

And the rate of reaction according to the rule (5.1) is given by

$$\frac{d[X]}{dt} = -J[X] \tag{5.4}$$

where the unit of $J$ is $s^{-1}$. If we consider a specific altitude and an hour of the day, we may use a constant value for $J$ so that from (5.4) we can get the molecular number density as a function of time:

$$[X] = [X]_0 \exp(-Jt) \tag{5.5}$$

where $[X]0$ is the initial value for the density. The concentration decay is then with a characteristic time:

$$\tau = \frac{1}{J} \tag{5.6}$$

This is a rather important parameter because it weights the photolysis against other processes.

A practical example could be the hydrogen peroxide $H_2O_2$ that may photodissociate or react with a hydroxyl radical:

$$\begin{aligned} H_2O_2 + h\nu &\rightarrow OH + OH \\ H_2O_2 + OH &\rightarrow HO_2 + H_2O \end{aligned} \tag{5.7}$$

We assume that the photolysis rate can be calculated and results as $J_{H_2O_2} = 9.6 \times 10^{-6}\ s^{-1}$. The rate of reaction for the second process in (5.7) is

$$k[OH]\{H_2O_2\} \tag{5.8}$$

where $k = 1.7 \times 10^{-12}$ molecule $cm^{-3}$ is the rate constant. The loss rates associated with reactions (5.7) are then

$$J_{H_2O_2}[H_2O_2] \quad k[OH][H_2O_2]$$

This means that a characteristic time for the reaction with hydroxyl radical is given by $k[OH]$. If we use typical concentration for $[OH] = 2 \times 10^6$ radicals $cm^{-3}$ we get $(k[OH])^{-1}. = 2.94 \times 10^5\ s = 3.4$ days that must be compared with $(9.6 \times 10^{-6}\ s^{-1})^{-1} = 1.2$ days, for the photodissociation. This means that the residence time is shorter for photodissociation which is the dominant loss process.

## 5.4 Chemistry and Transport

We rewrite the conservation equation (3.10) for the mass density of a generic component $[i]$ when in the volume we consider there are production and loss processes. It is straightforward to show that

$$\frac{D[i]}{Dt} = P_i - L_i \Rightarrow \frac{\partial [i]}{\partial t} = P_i - L_i - \nabla \cdot ([i]\,\mathbf{V}) \tag{5.9}$$

where $P_i$ and $L_i$ are the chemical production and loss rates of species $i$ and $V$ is the velocity vector. If we introduce the volume mixing ratio $X_i = [i]/[M]$, Eq. (5.9) can be rearranged for nondivergent flow:

$$\frac{\partial X_i}{\partial t} = \frac{P_i - L_i}{[M]} - \mathbf{V} \cdot \nabla X_i \tag{5.10}$$

or in Lagrangian terms

$$\frac{DX_i}{Dt} = \frac{P_i - L_i}{[M]} \tag{5.11}$$

This form is quite interesting because it shows that for an inert tracer (like dust or a noble gas), $P_i - L_i = 0$; the mixing ratio is constant in the air mass following the motion.

As an example, consider a case in which we have a constant wind in the $x$ direction, $u$ and a chemical source for $x \leq 0$, and a chemical sink with a rate $L_\mathrm{i} = [i]/\tau_i$ for $x > 0$. The continuity equation is then from (5.9):

$$\frac{\partial [i]}{\partial t} = -\frac{[i]}{\tau_i} - \frac{\partial}{\partial x}(u\,[i]) = -\frac{[i]}{\tau_i} - u\frac{\partial [i]}{\partial x} \tag{5.12}$$

For steady-state conditions ($\partial[i]/\partial \mathrm{t} = 0$), we have

$$\frac{d \ln [i]}{dx} = -\frac{1}{u\tau_i}$$

equivalent to

$$[i(x)] = [i(0)] \exp\left[\frac{-x}{u\tau_i}\right]$$

The number density decreases exponentially with a scale length given by $u\tau_i$; note that if the transport time $x/u$ is much less than $\tau_i$, then the concentration stays constant with $x$ and the species can be considered almost inert. A simple application of this concept can be found for the conversion of $SO_2$ and $NO_2$, respectively, in the acids $H_2SO_4$ and $HNO_3$ by OH. The conversion time in this case is about 3 days so that the e-folding distance $u\tau i$ is about 800 km for an average velocity of 3 $\mathrm{ms}^{-1}$. On

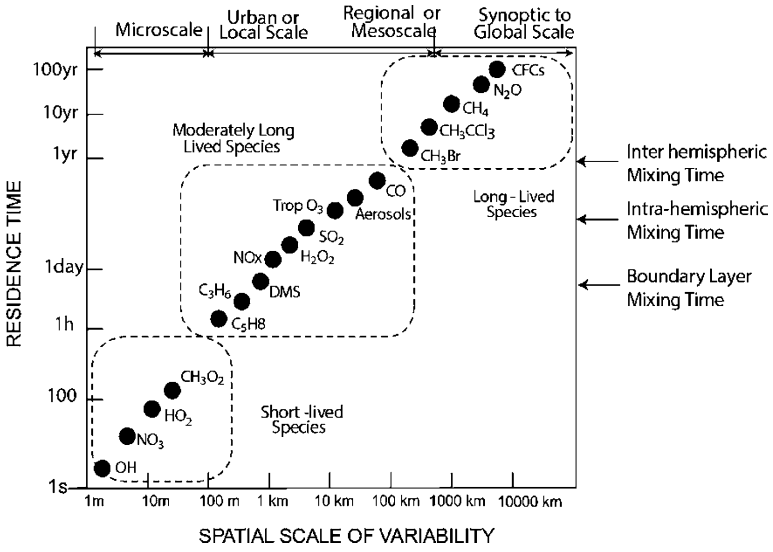



**Fig. 5.4** Spatial and temporal scales of variability for atmospheric species (Adapted from Brune 2007)

the other hand, the conversion of CO by OH into $CO_2$ takes about 3 months so that the distance is more than 20,000 km. This gives a clear significance to "regional" or "local" pollutants like $SO_2$ and $NO_2$ with respect to a "hemispheric" pollutant like CO.

It is rather interesting to examine these criteria of pollutants on different spatial and temporal scales for a number of gases present in the atmosphere. Figure 5.4 shows the temporal and associated spatial scale of variability for a number of atmospheric species. Although we do not know their name, the short-lived ones are radicals present in the lower atmosphere, while in the moderately long-lived species, we found familiar gases like ozone or carbon monoxide. The long-lived species are mostly industrial products like methyl chloroform ($CH_3CCl_3$) or chlorofluorocarbons (CFC). Others are long-lived gases like nitrous oxide ($N_2O$) or methane ($CH_4$) whose mixing ratio is strongly perturbed by the human activity. We will examine in detail the effects of these gases both on the climate of the Earth and the tropospheric and stratospheric chemistry.

To integrate the continuity equation for the various chemical species, we can generally express equations in finite difference form (see Examples). The spatial grid will have three dimensions with space $\Delta x$, $\Delta y$, and $\Delta z$, while the time step will be $\Delta t$. We rewrite Eq. (5.10) in finite form so that

$$p_i = l_i + \mathbf{V}\cdot\nabla X_i + \frac{\partial X_i}{\partial t} = \left(\frac{1}{\tau_i} + \mathbf{V}\cdot\nabla + \frac{\partial}{\partial t}\right) X_i \approx \left(\frac{1}{\tau_i} + \frac{u}{\Delta x} + \frac{v}{\Delta y} + \frac{w}{\Delta z} + \frac{1}{\Delta t}\right) X_i \tag{5.13}$$

where $p_i$ and $l_i$ are the production and chemical loss divided by the number density of the atmosphere. This expression is very useful in comparing the chemical time scale $\tau_i$ with transport time scales ($u/\Delta x$, $v/\Delta y$, $w/\Delta z$) and the time step itself. If the chemical time scale is small with respect to all the other terms in parenthesis, then expression (5.13) reduces to the chemical steady-state equilibrium for which $p_i \approx l_i$ and then

$$p_i \approx \frac{X_i}{\tau_i}, \quad \text{i.e.} \quad X_i \approx p_i \tau_i$$

The condition for chemical steady state reduces the number of equation to be integrated. We will see later that a useful concept in chemistry is the grouping of a number of components into a single family. Exchanges between members of the family can be very short while the family as such may have longer time scales comparable or longer than transport times.

A very simple example will clarify this point. The maximum simplification of the $NO_x$ family includes NO and $NO_2$, that is, $[NO_x] = [NO] + [NO_2]$. The two components are subject to the following reactions:

$$\begin{aligned} &NO + O_3 \rightarrow NO_2 + O_2 \quad \text{(rate constant k)} \\ &NO_2 + h\nu \rightarrow NO + O \quad \text{(rate constant J)} \\ &O_2 + O + M \rightarrow O_3 + M \end{aligned}$$

The conversion of NO in $NO_2$ and vice versa happens in a very short time much less than the transport time for $NO_x$. On the other hand, $NO_x$ is removed by the reaction

$$NO_2 + OH + M \rightarrow HNO_3 + M \quad \text{(rate constant l)}$$

We can assume a chemical steady state between NO and $NO_2$ and equating production and loss. From Fig. 5.3, we see that the lifetime of $NO_2$ against photolysis is only a few minutes, while for NO$x$ lifetime will be several days. Equating production and loss,

$$k\,[NO]\,[O_3] = J\,[NO_2]$$

so that we get the ratio

$$\frac{X_{NO}}{X_{NO_2}} = \frac{[NO]}{[NO_2]} = \frac{1}{J\,[O_3]}$$

On the other hand, we must use the full continuity equation for $NO_x$:

$$\frac{\partial X_{NO_x}}{\partial t} = p_{NO_x} - l\,[OH]\,[NO_2] - \mathbf{V} \cdot \nabla X_{NO_x} = p_{NO_x} - \frac{[NO_x]}{\tau_{NO_x}} - \mathbf{V} \cdot \nabla X_{NO_x}$$

where

$$\tau_{NO_x} = \left(1 + \frac{[NO]}{[NO_2]}\right) / \left(l\,[OH]\,[M]\right) \approx \left(1 + \frac{1}{J\,[O_3]}\right) / \left(l\,[OH]\,[M]\right)$$

can be calculated to be roughly 3 days.

There is another simple way to account for transport and that is the use of box models. The simplest is one dimensional and consider a column of height h where we can specify an emission rate $E$, a production $P$, a loss rate $L$, and a deposition rated $D$. The change with time of component $X$ is then

$$\frac{d\,[X]}{dt} = \frac{E}{h} + P - L - D \tag{5.14}$$

This equation can be changed if we introduce a deposition velocity vd and consider that the total derivative contains the advection term. We have

$$\frac{\partial\,[X]}{\partial t} = \frac{E}{h} + P - l_X\,[X] - \frac{v_d}{h}\,[X] - U\frac{\partial\,[X]}{\partial x} \tag{5.15}$$

where we have assumed the loss rate to be proportional to the concentration $[X]$. Passing the finite approximation, Eq. (5.15) becomes

$$\frac{\partial\,[X]}{\partial t} = \frac{E}{h} + P - l_X\,[X] - \frac{v_d}{h}\,[X] - U\frac{([X_0] - [X])}{\Delta x} \tag{5.16}$$

A very simple application of this equation is for a case where $P = D = 0$ in the region between $x$ and $x + L$, while also $E = 0$ for $x > L$. If we consider the steady-state case, for $0 \le x \le L$, Eq. (5.15) reduces to

$$\frac{E}{h} - l_X\,[X] - U\frac{\partial\,[X]}{\partial x} = 0$$

This can be easily integrated with the condition $[X(0)] = 0$ to give

$$[X] = \frac{E}{hl_X U}\left(1 - \mathrm{e}^{-\frac{l_X X}{U}}\right) \quad 0 \le x \le L$$
$$[X] = [X(L)]\,\mathrm{e}^{-\frac{l_X (x-L)}{U}} \quad 0 \le x \le L$$

This could be the case of linear source of length $L$ in a steady wind of intensity $U$.

## E.5 Examples

### *E.5.1 Units for Chemical Abundance*

The most common unit for expressing the chemical abundance of a gas is the mixing ratio by volume, that is, the fraction of volume of air occupied by that gas. This unit

is very convenient because the volume occupied by a gas at the same temperature and pressure is proportional to the number of moles. The ratio between volumes gives then the ratio between moles, that is, the mole fraction. For example, the mixing ratio of carbon dioxide ($CO_2$) is $380 \times 10^{-6}$, and this means that for every $10^6$ molecules of air, there are 380 molecules of $CO_2$. However, the same number expresses the ratio between partial pressure of the gas and total pressure of the air. For the same example, the partial pressure of $CO_2$ would be roughly 38 Pascal or 0.38 mb. Often the same mixing ratio is expressed as 380 ppm (part per million), that is, 380 molecules for each million (106) molecules of air. Other times the unit ppb (part per billion, $10^9$) or ppt (part per trillion, $10^{12}$) is used.

Another unit may be the mixing ratio by mass which is defined as the ratio between the densities of a gas and the density of air. If we indicate by $p_i$ and $p_a$ the pressure of gas and air and by $\rho_i$ and $\rho_a$ the corresponding density, we have

$$p_i = \frac{\rho_i}{M_i} RT; \quad p_a = \frac{\rho_a}{M_a} RT$$

Making the ratio, we found the relation between the mass mixing ratio and the volume mixing ratio:

$$\chi = \frac{\rho_i}{\rho_a} = \frac{p_i M_a}{p_a M_i} \tag{E.5.1}$$

If we apply this simple formula to the oxygen and nitrogen in the atmosphere, we have a mass mixing ratio for $N_2$ of 80.8 % and 19.03 % for $O_2$. Remember that the volume mixing ratios are 78 % and 21 %, respectively.

The mass mixing ratio is very convenient to find the total mass of specific constituents with constant altitude mixing ratio. Consider, for example, carbon dioxide with a volume mixing ratio of 380 ppm. This corresponds according to (E.5.1) to a mass mixing ratio of

$$\chi_{CO_2} = 380 \times 10^{-6} 28.9/44 = 249.6 \times 10^{-6}$$

The total mass of $CO_2$ in the atmosphere is then

$$m_{CO_2} = \chi_{CO_2} m_{\text{atm}} = 4\pi r^2 p_0 \chi_{CO_2}/g$$

where $r$ is the radius of the Earth (6370 km), $p_0$ the surface pressure $1.013 \times 10^5$ Pa, and g the acceleration of gravity (9.81 $\text{ms}^{-2}$). The mass is then $1.31 \times 10^{15}$ kg.

## *E.5.2 The Chapman Model for Atmospheric Ozone*

The model presented here for the atmospheric ozone was proposed in 1930 by Sydney Chapman. The great geophysicist was a pioneer also in the theory of ozone destruction. He actually noted that because ozone being an absorber of UV could

create some problem to astronomers observing stars in the spectral region, he then suggested that it would be useful to create a "hole" in the ozone layer to facilitate the observations.

The Chapman mechanism considers only four reactions:

$$\begin{array}{ll} O_2 + h\nu \rightarrow O + O & J_1 \\ O + O_2 + M \rightarrow O_3 + M & k_2 \\ O_3 + h\nu \rightarrow O_2 + O & J_3 \\ O_3 + O \rightarrow 2O_2 & k_4 \end{array} \tag{E.5.2}$$

$J_1$ and $J_3$ are the photodissociation coefficients for molecular oxygen (below 240 nm) and ozone (below 320 nm), while $k_2$ and $k_4$ are the rate coefficients for the formation of ozone and recombination with oxygen atoms. $M$ is a third-body reaction. It can be recognized that the second ($k_2$) and third reactions ($J_3$) are much faster than the other two. This means that there is a fast interchange between O and $O_3$ and a slower one between $O_2$ and $(O + O_3)$. It is convenient to introduce a chemical called odd oxygen $O_x = O + O_3$ which is formed by $J_1$ and consumed by $k_4$.

The ratio between O and $O_3$ is fixed by chemical equilibrium between the odd oxygen family:

$$k_2 [O] [O_2] [M] = J_3 [O_3] \Rightarrow \frac{[O]}{[O_3]} = \frac{J_3}{k_2 [O_2] [M]} = \frac{J_3}{k_2 f_{O_2} n_a^2} \tag{E.5.3}$$

where we have assumed that the air density $n_a$ is the same as [M] and the $[O_2]$ is simply the mixing ratio of molecular oxygen ($f_{CO_2}$) by $n_a$.

The other balance is for the entire odd oxygen family:

$$\frac{d [O_x]}{dt} = 2J_1 [O_2] - 2K_4 [O] [O_3] \tag{E.5.4}$$

And at the equilibrium $d [O_x] / dt = 0$, we have

$$J_1 [O_2] = k_4 [O] [O_3] \Rightarrow [O_3] = \left( \frac{J_1 k_2}{J_3 k_4} \right)^{1/2} f_{O_2} {n_a}^{3/2} \tag{E.5.5}$$

This expression gives a qualitative picture of the ozone layer. The last term gives the number density decreasing with altitude (${n_a}^{3/2}$). The change with altitude depends mainly by the ratio in parenthesis. The photodissociation coefficient for oxygen is really independent from the ozone concentration, while the $J_3$ coefficient depends essentially on the ozone concentration. Constant $k_2$ has a rather complicated expression and $k_4$ has a simpler expression but depends on temperature decreasing while the temperature increases. The way to proceed to evaluate $[O_3]$ is to start from high altitude with a small ozone quantity and proceeds to lower altitudes.

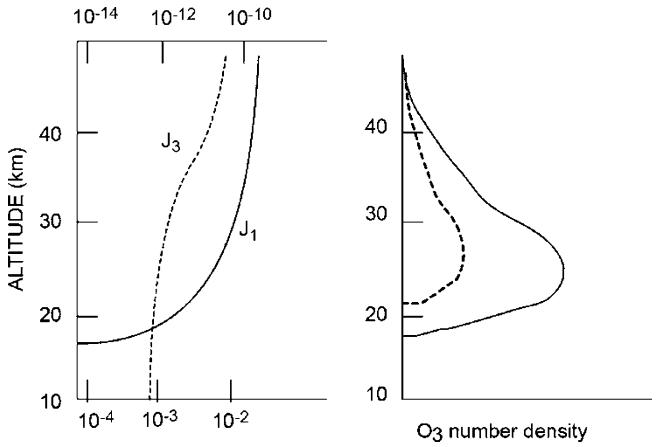



**Fig. E.5.1** Qualitative illustration of the Chapman model. On the *left* the photolysis rates for ozone ($J_3$) and oxygen ($J_1$) are reported. The $J_1$ scale is at the *top*. On the *right* the ozone number density calculated (*solid*) and observed (*dashed*) are depicted

Qualitatively the results look like Fig. E.5.1. At high altitude, ozone decreases according to the atmospheric density. At low altitude however, the photolysis rate for oxygen goes rapidly to negligible values cutting the source of odd oxygen and consequently reduces the ozone formation.

When the first ozone measurements become available (before the Second World War), they showed that the Chapman theory overestimated the ozone concentration as shown. Besides, this theory was unable to explain the ozone in the troposphere because apparently there was no source for odd oxygen there. This will be one of the problems we will confront in one of the next chapters.

### *E.5.3 Calculation of Photolysis Rate*

We make an example based on the hydrogen peroxide that photodissociates according to

$$H_2O_2 + h\nu \rightarrow 2OH \quad \lambda < 550 \text{ nm}$$

Table E.5.1 reports the data for the calculation of the photolysis rate that includes the actinic flux at intervals of 10 nm and the cross section. At this point, it is very simple to calculate the photolysis rate.

It reduces to the summation of the products between fluxes and cross section:

$$J_{H_2O_2} = \sum_{\lambda} I_\lambda \sigma_\lambda = 9.6 \times 10^{-6} s^{-1}$$

**Table E.5.1** Photolysis data for hydrogen peroxide

| Wavelength (nm) | $I(\lambda)$ (photon cm$^{-2}$ s$^{-1}$) | $\sigma$ (cm$^{-2}$ molecule$^{-1}$) |
|---|---|---|
| 295–305 | $2.66 \times 10^{13}$ | $0.71 \times 10^{-20}$ |
| 305–315 | $4.20 \times 10^{14}$ | $0.42 \times 10^{-20}$ |
| 315–325 | $1.04 \times 10^{15}$ | $0.25 \times 10^{-20}$ |
| 325–335 | $1.77 \times 10^{15}$ | $0.14 \times 10^{-20}$ |
| 335–345 | $1.89 \times 10^{15}$ | $0.08 \times 10^{-20}$ |
| 345–355 | $2.09 \times 10^{15}$ | $0.05 \times 10^{-20}$ |

The flux reported in the table is referred at the ground. Usually the calculation must be made at different levels, and in that case, the flux must be calculated taking into account the gas absorption (only gaseous absorption is considered) according to the method outlined in Sect. 2.2. In particular, we have to use the relationship (2.34) converting the flux in watt m$^{-2}$ in photons s$^{-1}$m$^{-2}$. The conversion is very simple:

$$I_\lambda = \frac{F_\lambda c}{h}\lambda$$

where $c$ is the speed of light and $h$ is the Planck constant.

### *E.5.4 Photodissociation and Vertical Transport*

In the old days when three-dimensional chemical transport models were not so popular, people used very simple one-dimensional models in which transport was confined to the vertical direction. The parameterization of this transport was very simple (we will find more about in the boundary layer chapter) so that if we had a species with mixing ratio $f$, the flux associated to the distribution was simply proportional to the gradient of the mixing ratio. This approximation assumes that molecules diffuse according to the same law used in molecular diffusion. We would get for the flux (in molecules m$^{-2}$ s$^{-1}$)

$$\Phi = -Kn_a \frac{df}{dz} \tag{E.5.6}$$

where $K$ is called the eddy diffusion coefficient (m$^2$s$^{-1}$), na is the atmospheric number density (molecules m$^{-3}$), and $f$ is the mixing ratio. It is convenient to rewrite (E.5.6):

$$\begin{aligned}\Phi &= -Kn_a \frac{d}{dz}\left(\frac{n_i}{n_a}\right) = -Kn_a\left(\frac{1}{n_a}\frac{dn_i}{dz} - \frac{n_i}{n_a^2}\frac{dn_a}{dz}\right) \\ &= -K\left(\frac{dn_i}{dz} + \frac{n_i}{H_a}\right) = -Kn_i\left(-\frac{1}{H_i} + \frac{1}{H_a}\right)\end{aligned} \tag{E.5.7}$$

where we have used the definition of scale height for the atmosphere and the species

$$\frac{1}{H_a} = -\frac{1}{n_a}\frac{dn_a}{dz}; \quad \frac{1}{H_i} = -\frac{1}{n_i}\frac{dn_i}{dz}$$

The formulation (E.5.7) is rather instructive because if $H_i < H_a$, then the flux is positive upward; that is, if the species mixing ratio decreases with height, its tendency is to mix until $H_i = H_a$. On the other hand, if the mixing ratio increases with height $H_i > H_a$, then the flux is negative (downward) and again the species has a tendency to mix with the atmosphere. These considerations were evident in (E.5.6).

Now that we have a correct expression for the flux, consider a species photodissociated with a constant $J$. From the continuity equation, we should have

$$\nabla \cdot \Phi = Jn_i \Rightarrow -K\frac{d}{dz}\left[n_i\left(-\frac{1}{H_i} + \frac{1}{H_a}\right)\right] = Jn_i \tag{E.5.8}$$

where the divergence has been reduced only to the vertical direction. (E.5.8) can be integrated assuming for $n_i$ a constant scale height so that the right-hand side becomes

$$\int_0^\infty Jn_i dz = Jn_i H_i$$

And the integration gives

$$-K\left(-\frac{1}{H_i} + \frac{1}{H_a}\right) = JH_i \Rightarrow \frac{1}{H_i^2} - \frac{1}{H_a H_i} - \frac{J}{K} = 0 \tag{E.5.9}$$

The solution for the scale height of the species results as

$$\frac{2}{H_i} = \frac{1}{H_a} + \left(\frac{1}{H_a^2} + \frac{4J}{K}\right)^{1/2} \tag{E.5.10}$$

Again we see that for a large $K$ the species is completely mixed ($H_i = H_a$), while for small $K$ the mixing ratio decreases with height. We have taken only the positive sign for obvious reasons.

### *E.5.5 A Time-Dependent Case*

If we want to calculate more accurately the concentration of species $i$, we need to recur to the complete continuity equation:

$$\frac{dn_i}{dt} + \frac{d}{dz}\left[-K\left(\frac{dn_i}{dz} + \frac{n_i}{H_a}\right)\right] = -Jn_i \tag{E.5.11}$$

When the derivative is performed, we get a diffusion equation:

$$\frac{dn_i}{dt} = K\frac{d^2n_i}{dz^2} + K\frac{dn_i}{dz} - Jn_i \tag{E.5.12}$$

This is a completely new differential equation whose solution depends on time and space $n_i(z, t)$ and it is known as the diffusion equation. We will learn how to solve it later on in the book.

## References[1]

### *Books*

Jacob D (1999) An introduction to atmospheric chemistry. Princeton University Press, Princeton, NJ, USA

Wallace JM, Hobbs PV (1977) Atmospheric science: an introductory survey. Academic, New York

### *Articles*

Brune WH (2007) In: Visconti G et al. (eds) Observing systems for atmospheric composition. Springer, New York

Prinn RG (1992) Tropospheric chemical models. In: Gille JC, Visconti G (eds) The use of EOS for studies of atmospheric physics. Elsevier, Amsterdam/New York

Prinn RG, Fegley B (1987) The atmospheres of Venus, Earth and Mars, critical comparison. Ann Rev Earth Planet Sci 15:171

[1]This chapter is an introduction to the concept that will be developed in Chaps. 16 and 18. It also contains some general information on the planetary atmospheres. The main reference may be the book by Jacob with other contributions from specialized sources.

# Chapter 6
# Introduction to Remote Sensing

In this chapter, we will introduce what is essentially new for this book and that these are the principles of remote sensing techniques. Remote sensing in general includes all the techniques that can obtain information about the nature of the surface and the atmospheres of the Earth and planets. The most ancient techniques are visual and telescopic observations on other planets well before sophisticated instrumentation would be orbited around them. In this chapter, we will concentrate on observations of the atmosphere considering that they can be made either from space or from the ground. It is very instructive to go through this introduction because in practice it applies most of the things we introduced in the earlier chapters. We will omit in this first, preliminary approach many of the technicalities and details that the interested reader could find in books on remote sensing.

## 6.1 Observations of the Atmosphere

The atmosphere can be observed either from the ground or space. These techniques may determine several characteristics like profiles of temperature and atmospheric composition and properties of clouds and aerosols, winds, water vapor pressure, and in general precipitation.

Figure 6.1 shows schematically the main techniques used in remote sensing. Besides the spaceborne measurements, there are a few ground-based techniques like radars and lidar that are shown schematically in the insert. The space-based measurements refer to three different geometries. The nadir observations look directly underneath the satellite to obtain information about the characteristics of the radiation coming from the atmosphere. The radiation can be thermal (IR) giving



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_6

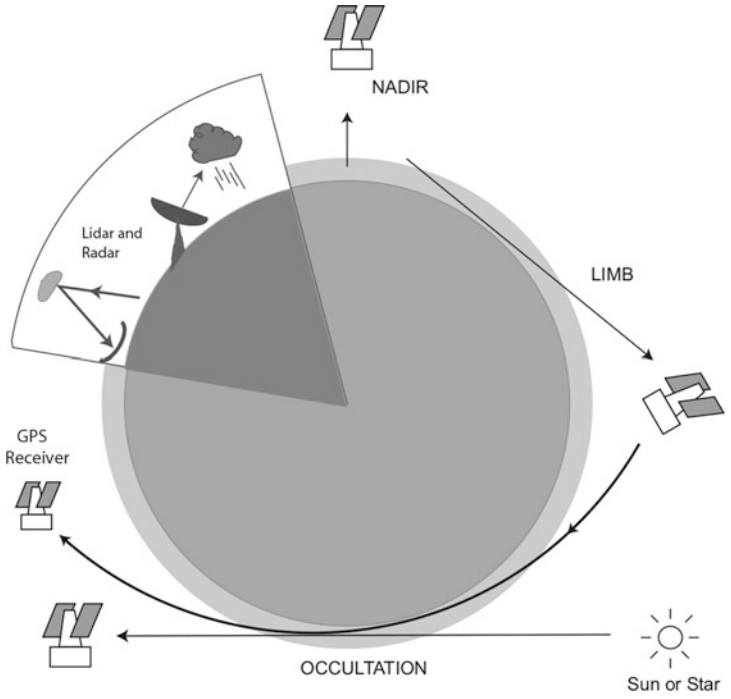



**Fig. 6.1** A synthesis of a few remote sensing techniques

information about the atmospheric composition and its thermal structure (see Fig. 2.10). However the instruments could also record backscattered solar radiation and in that case again can get information about the distribution and total gas amount. The limb observations on the other hand look again at radiation emitted along the relatively long path in such a way to enhance absorption by gases. This technique requires a correct alignment and can be used also from an aircraft or balloon platform. Finally if a radiation source is used like a star or the sun, the absorption of the radiation from that source can be used to obtain information on the composition and structure of the atmosphere. This corresponds to the occultation technique and a variation of that using microwave radiation is the radio occultation (RO). In the case depicted, a signal is emitted by a satellite that can be measured from another satellite that may be of the ground positioning system (GPS). The signal in this case may be subjected to refraction and the path can be curvilinear. The signal processed may give information about the refractive index of the atmosphere which is a function of both pressure and temperature.

The ground-based measurements shown in the figure are the radar and lidar techniques. Radar is an acronym for radio detection and ranging and consists of the emission of a pulsed beam of microwave radiation whose return can be analyzed

to have information about the characteristics of the hydrometeors. Lidar uses the same principle as radars, but in this case, the radiation emitted is a light pulse.

## 6.2 Thermal Emission Measurements

Before proceeding, we must recall the equation of radiative transfer (2.39) written in the monochromatic form:

$$\frac{dI_\nu}{d\tau_\nu} = B_\nu - I_\nu \tag{6.1}$$

A formal solution of this equation gives the monochromatic intensity of the radiation received from the sensor on the satellite (Liou 1980):

$$I_\nu(z_1) = B_\nu(T_s)\, \mathrm{e}^{-(\tau_\nu(0)-\tau_\nu(z_1))} + \int_0^{z_1} B_\nu(T(z))\, \mathrm{e}^{-(\tau_\nu(z)-\tau_\nu(z_1))} d\tau_\nu \tag{6.2}$$

In this equation, $T_s$ is the surface temperature, and $T(z)$ is the temperature at height $z$. $\tau_\nu(0)$ is the total optical thickness and $\tau_\nu(z)$ is the optical thickness at height $z$.

The exponential defines the transmission function $T(z, z') = \mathrm{e}^{\left[-\left(\tau(z)\right)-\left(\tau(z')\right)\right]}$ where $\tau(z) > \tau(z')$. Equation (6.2) can now be rewritten at the satellite altitude ($\infty$):

$$I_\nu = B_\nu(T_s)\, T_\nu(0, \infty) + \int_0^{\infty} B_\nu(T_\nu(z)) \frac{\partial T_\nu(z, \infty)}{\partial z} dz \tag{6.2a}$$

A simple solution of this equation is for the case in which the transmission function is constant and equal to unity $T_\nu(z, z_1) = 1$. In such case, because the optical thickness is zero, the satellite "sees" the surface so that we have

$$I_\nu(z_1) = B_\nu(T_s)$$

And the temperature can be obtained by simply inverting the Planck function:

$$T_s = h\nu \left[ k \ln\left(1 + \frac{2h\nu^3}{c^2 I_\nu}\right)\right]^{-1} \tag{6.3}$$

Another simple case is when an isothermal atmosphere at temperature $T_0$ overlies on a surface at the temperature $T_s$. We assume now that the atmosphere has an absorption line centered at $\nu_o$ that cannot be neglected. However we may assume that the change with frequency of the Planck function is slow (negligible) with

respect to the change around the absorption line. In this case, the integral in (6.2a) can be written as

$$\int_0^\infty B_\nu\left(T_\nu(z)\right)\frac{\partial T_\nu\left(z,\infty\right)}{\partial z}dz \approx \int_0^\infty B_\nu\left(T_\nu(z)\right)dT_\nu\left(z,\infty\right)$$
$$= B_\nu\left(T_\nu(0)\right)\left(T_\nu\left(\infty,\infty\right) - T_\nu\left(0,\infty\right)\right) + \int_0^\infty T_\nu\left(z,z_1\right)dB_\nu\left(T_\nu(z)\right)$$

The integral is zero because the Planck function is constant with altitude so that Eq. (6.2a) becomes

$$I_\nu = B_\nu\left(T_\nu(0)\right)\left[1 - T_\nu\left(0,\infty\right)\right] + B_\nu\left(T_s\right)T_\nu\left(0,\infty\right) \tag{6.4}$$

It is worth to introduce the absorptance as

$$A_\nu\left(0,\infty\right) = 1 - T_\nu\left(0,\infty\right) \tag{6.5}$$

so that (6.5) can be rewritten as

$$I_\nu = B_\nu\left(T_s\right) + \left[B_\nu\left(T(0)\right) - B_\nu\left(T_s\right)\right]A_\nu\left(0,\infty\right) \tag{6.6}$$

This simple formula says that the signal observed by the instrument will be the Planck function at the surface temperature except for the spectral range where the absorptance is significantly different from zero. In this region, if the temperature of the atmosphere is greater than the surface temperature ($B_\nu(T(0) > B_\nu(T_s)$), we will observe a "bump." While if the atmosphere is colder than the surface, we will observe an absorption dip.

A situation like that may be exemplified as in Fig. 6.2 where the spectra obtained from the Nimbus 4 are shown. In all the three spectra, the atmosphere is transparent in the 800–1250 $cm^{-1}$ (8–12 $\mu m$) so that the emission is the corresponding blackbody at 320 K, 280 K, and 180 K. In the Antarctica region, the absorption feature at 15 $\mu m$ due to $CO_2$ absorption appears in emission because there the atmosphere is warmer than the surface. These spectra show that a qualitative picture of the thermal structure could be easily obtained. We will show in one of the examples (E.6.1) that in principle a much more detailed picture of the thermal structure of the atmosphere could be obtained.

## 6.3 Ozone Measurements from Satellite

There are different methods to measure ozone from a satellite. In what follows, we will illustrate just two of them. One is based on the scattering of solar ultraviolet radiation and the other is a simple absorption method. Both these techniques

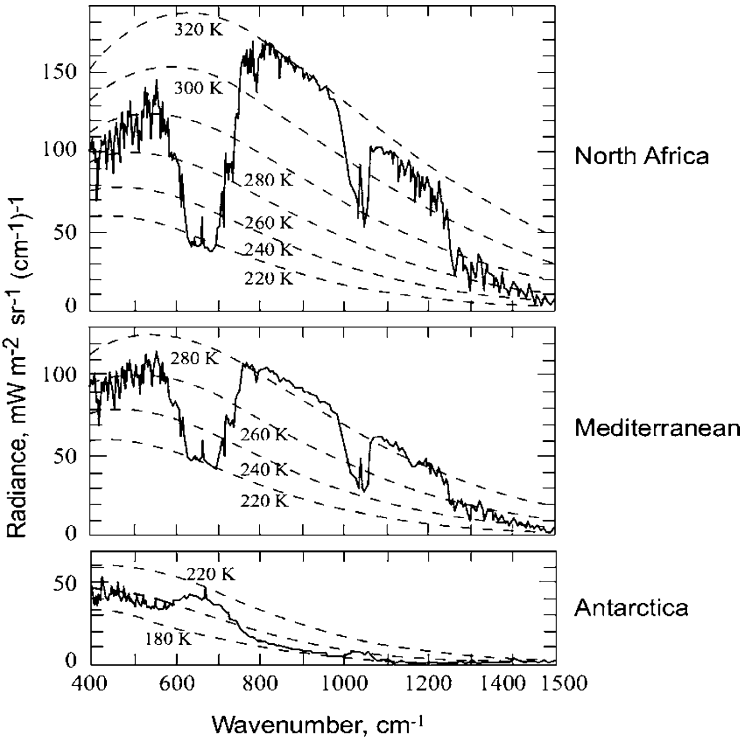



**Fig. 6.2** Spectra taken by the Nimbus 4 for different regions of the Earth. The *upper part* shows a spectrum taken over in North Africa, the Mediterranean, and Antarctica

exploit the characteristic that the ozone is a strong absorber of radiation in the ultraviolet region.

The first method is illustrated in Fig. 6.3 where the atmosphere is reduced essentially to two layers: the upper one is identified with the stratosphere and will in essence be responsible for the absorption of UV radiation by $O_3$, while the lower and denser troposphere is responsible for the reflection of the incoming solar radiation. In this simple geometry, the radiance measured by the satellite is given by

$$I(\lambda) = E_{\text{sun}}(\lambda)\, T_{O_3}(\lambda)\, R(\theta_{\text{sun}}, \theta_{\text{sat}}, R_{\text{surf}}, R_{\text{air}}) \tag{6.7}$$

where $T_{O_3}(\lambda)$ is the transmittance of the atmosphere at wavelength $\lambda$ and $R$ is the reflectance of the atmosphere and the surface. The transmittance along the slant path can be written as

$$T_{O_3}(\lambda) = \exp\left(-\int \rho k_a ds\right) = \exp\left(-\int \rho k_a \frac{dz}{\cos\theta}\right) = \left[T^{\text{vert}}{}_{O_3}(\lambda)\right]^{sec\theta} \tag{6.8}$$

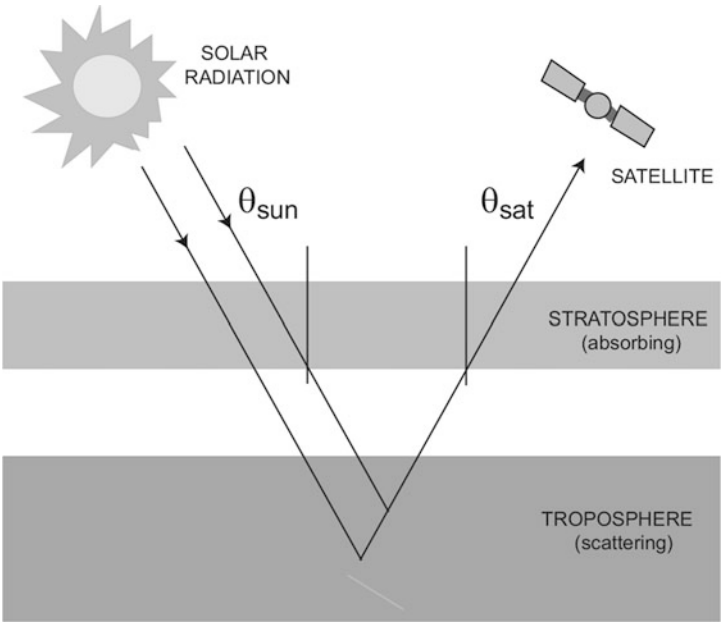



**Fig. 6.3** The geometry of the UV backscattering techniques used to measure the ozone total amount

The transmission for the vertical path can be related to the total ozone in a vertical column in the following way:

$$\left[T^{\text{vert}}{}_{\text{O}_3}(\lambda)\right] = \exp\left(-\int \rho k_a dz\right) = \exp\left(-\int k_a du\right) = \exp\left(-k_a u\right) \tag{6.9}$$

where u is total ozone quantity (molecules $\text{cm}^{-2}$) and ka is now expressed in $\text{cm}^2$. We can rewrite (6.8) in principle like

$$\begin{aligned} N_1 &= -\log \frac{I(\lambda_1)}{E_{\text{sun}}(\lambda_1)} = -\log\left[T_{\text{O}_3}(\lambda_1) R(\lambda_1)\right] \\ N_2 &= -\log \frac{I(\lambda_2)}{E_{\text{sun}}(\lambda_2)} = -\log\left[T_{\text{O}_3}(\lambda_2) R(\lambda_2)\right] \end{aligned} \tag{6.10}$$

for two nearby wavelengths $\lambda_1$ and $\lambda_2$. Also we have indicated only the dependence on wavelength of the reflectance. The difference between these two quantities gives

$$N_1 - N_2 = -\log\left[\frac{R(\lambda_1)}{R(\lambda_2)} \exp\left\{-\left[k_a(\lambda_1) - k_a(\lambda_2)\right] u\chi\right\}\right] \tag{6.11}$$

where have substituted the transmittance given in (6.9) and where $\chi$ is given by

$$\chi = \sec\theta_{\text{sun}} + \sec\theta_{\text{sat}}$$

As a first approximation, we can assume $R(\lambda_1) = R(\lambda_2)$ and we would get

$$\begin{aligned} N_1 - N_2 &= -\log\left[\exp\left\{-\left[k_a(\lambda_1) - k_a(\lambda_2)\right] u\chi\right\}\right] \\ &= \log(e)\left[k_a(\lambda_1) - k_a(\lambda_2)\right] u\chi \end{aligned} \tag{6.12}$$

Knowing the absorption coefficient $u$ could be determined. Actually the reflectance could be different at the two wavelengths so a third measurement at wavelength l3 is used to determine the reflectance at the surface ($R_{\text{surf}}(\lambda_1) \approx R_{\text{surf}}(\lambda_2)$); calculate then the Rayleigh scattering $R_{\text{air}}$ and then determine the exact values of $R(\lambda_1) = R(\lambda_2)$. This method can be adapted to obtain ozone profiles by using the fact that the absorption coefficient really depends on the pressure so that weaker absorption will sample higher pressure or lower altitude.

The methods illustrated here have been used historically on Nimbus 4 in the early 1970s and in instruments like TOMS (Total Ozone Mapping Spectrometer) and GOME (Global Ozone Monitoring Experiment) on ESA satellites.

Another way for measuring the ozone profile is the limb extinction method. As shown in Fig. 6.4, in this case, the satellite observes the radiation from the sun during sunrise and sunset with respect to the motion of the satellite. The radiation absorbed through the horizontal path is called *limb radiation*. The intensity of radiation for zero extinction (i.e., the highest altitude) is indicated with $I_\lambda{}^0$, and the intensity

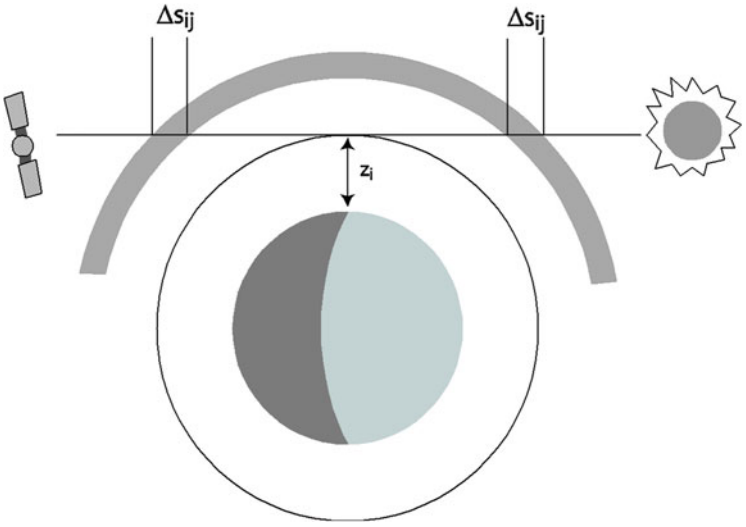



**Fig. 6.4** Geometry of the limb extinction measurement (Stephens 1994)

for the tangent altitude $z_i$ is Il(zi) and the ratio between these two intensities is the transmittance:

$$\frac{I_\lambda^0}{I_\lambda(z_i)} = T_\lambda(z_i) = \exp\left(-\int_{-\infty}^{\infty} \sigma_{\text{ext},\lambda}(s)ds\right) \tag{6.13}$$

where

$$\sigma_{\text{ext},\lambda}(s) = k_a^{ozo}(\lambda)\,\rho_{ozo}$$

is the volume extinction coefficient for ozone in the simplest case. Usually the extinction in the same wavelength range is contributed by other processes like Rayleigh scattering, aerosol, and other gas absorption. In the simplified case, we get

$$T_\lambda(z_i) = \exp\left[-\int_{-\infty}^{\infty} k_a^{ozo}(\lambda)\,\rho_{ozo}(s)ds\right] = \exp\left[-k_a^{ozo}(\lambda)\,u(z_i)\right] \tag{6.14}$$

where $u(z_i)$ is the total ozone amount along the horizontal path. The total ozone column is contributed by different spherical layers with each one responsible for the portion $\Delta s_{ij}$ as shown in Fig. 6.4. Actually then the integrand in (6.15) is given by

$$\tau_\lambda^i = \int_0^{\infty} \sigma_{\text{ext},\lambda}(s)ds = \frac{1}{2}\sum_{j=1}^{j=N} \sigma_{\text{ext},k,j}\Delta s_{i,j} \tag{6.15}$$

where $j$ is the layer index and $N$ is the highest layer to be considered. The procedure is then to obtain the optical thickness for each altitude by (6.14) and then inverting (6.16) the extinction for each layer. Knowing the absorption coefficient for each gas as a function of wavelength, it is possible to obtain the vertical profile.

## 6.4 Atmospheric Properties from Radio Occultation (RO)

The radio occultation technique has been introduced in 1965 when the signal from the Mars probe Mariner IV was analyzed to get information about the atmosphere of that planet. This technique is based on deriving the refractivity index of the atmosphere from perturbations of the radio signal transmitted by the probes. Thirty years later, the same technique was applied to obtain properties of the atmosphere of the Earth and use those data for meteorological purposes.

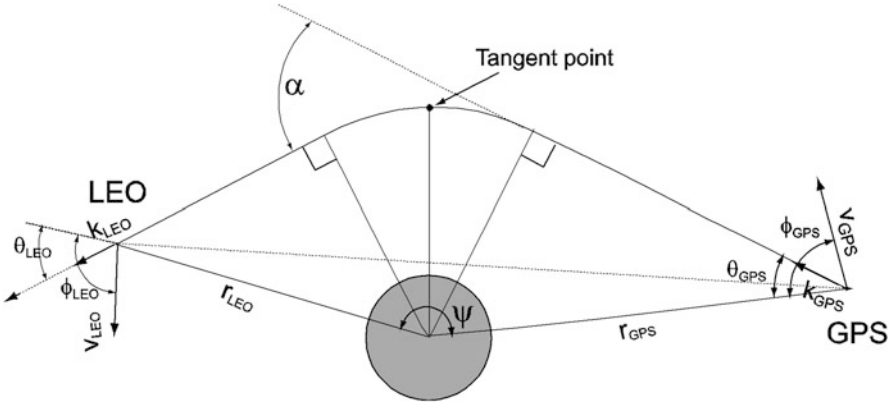



**Fig. 6.5** Radio occultation geometry

The most simple way to illustrate the method is the calculation of the bending angle $\alpha$ as illustrated in Fig. 6.5. Considering that the sum of the inner angle of a quadrilateral is $2\pi$, we have for the bending angle

$$\alpha = \theta_{LEO} + \theta_{GPS} + \psi - \pi \tag{6.16}$$

where LEO refers to low Earth orbit and GPS to global positioning system being the satellite in geostationary orbit. The GPS observable is expressed as a function of the radial velocity between the emitter (i.e., the LEO satellite) and the transmitter (i.e., the GPS satellite):

$$D = -f_e v_\rho / c \tag{6.17}$$

where $f_e$ is the nominal transmitter frequency, $c$ is the speed of light, and $v_\rho$ is the radial velocity. Always referring to the figure, the velocity can be expressed as a function of the unit vectors of the rays and the velocities so that we get

$$D = \frac{fe}{c} \left( \mathbf{v}_{\mathrm{GPS}} \cdot \mathbf{k}_{\mathrm{GPS}} - \mathbf{v}_{\mathrm{LEO}} \cdot \mathbf{k}_{\mathrm{LEO}} \right)$$

which is equivalent to

$$D = \frac{fe}{c} \left[ v_{\mathrm{GPS}} \cos\left(\phi_{\mathrm{GPS}} - \theta_{\mathrm{GPS}}\right) - v_{\mathrm{LEO}} \cos\left(\phi_{\mathrm{LEO}} - \theta_{\mathrm{LEO}}\right) \right] \tag{6.18}$$

From the figure, looking at the two segments normal to the rays, we get easily the Bouguer relation:

$$n_{\mathrm{GPS}} r_{\mathrm{GPS}} \sin\theta_{\mathrm{GPS}} = n_{\mathrm{LEO}} r_{\mathrm{LEO}} \sin\theta_{\mathrm{LEO}} = \cos t \tag{6.19}$$

The angles $\phi_{\text{GPS}}$ and $\phi_{\text{LEO}}$ are obtained from the dot product of the velocity and position vectors and after that the other angles can be obtained using (6.19) and (6.20): at this point, the bending angle can be evaluated.

We now observe that in the absence of any refracting media, we would get a relation between radius and refracting angle a little bit different from that of the Bouguer formula:

$$\begin{aligned} &\text{straight line} \Rightarrow r\sin\theta_s = a_s \\ &\text{true ray} \Rightarrow nr\sin\theta = a \end{aligned} \tag{6.20}$$

Differentiating both equations and subtracting, we get

$$dnr\sin\theta + ndr\,(\sin\theta - \sin\theta_s) + nr\,(\cos\theta d\theta - \cos\theta_s d\theta_s) = 0$$

Because the bending angle at the highest layers does not exceed 0.03°, it is possible to approximate the previous expression as

$$dnr\sin\theta + nr\cos\theta\,(d\theta - d\theta_s) = 0$$

The differential of the bending angle is just $d\theta$–$d\theta_s$ so we get

$$d\alpha = -\frac{\sin\theta}{\cos\theta}d\,(\ln(n)) = -\frac{a}{\sqrt{n^2r^2 - a^2}}d\,(\ln(n))$$

where we have expressed the trigonometric functions using (6.21). The bending angle $\alpha$ is obtained integrating the previous expression for the path toward GPS and the path toward LEO:

$$\alpha(a) = -a\left[\int_a^{x_{\text{GPS}}} + \int_a^{x_{\text{LEO}}}\right]\frac{1}{\sqrt{x^2 - a^2}}\frac{d\ln(n)}{dx}dx \tag{6.21}$$

where $x = nr$. Assuming that the contribution from LEO is negligible and considering the spherical symmetry, the bending angle is obtained as

$$\alpha(a) = -2a\int_0^{\infty}\frac{1}{\sqrt{x^2 - a^2}}\frac{d\ln(n)}{dx}dx \tag{6.22}$$

Using the Abel transform, this integral can be inverted to obtain

$$\ln(n(x)) = \frac{1}{\pi}\int_0^{\infty}\frac{\alpha(a)}{\sqrt{a^2 - x^2}}\,d\alpha \tag{6.23}$$

Once the refractive index is determined, it is possible to obtain profiles of pressure and density according to the following scheme. The refractivity is defined as

$$N = 10^6\,(n-1) = k_1\frac{P_d}{T} + k_2\frac{P_v}{T} + k_3\frac{P_v}{T^2} = \alpha\frac{P}{T} + k_3\frac{P_v}{T^2} \tag{6.24}$$

where $P_d$ and $P_v$ are the partial pressures of dry air and water vapor, $T$ is the absolute temperature, and $k_i$, $i = 1,2,3$ are the refractivity constants. Two cases can be illustrated, the dry air case and the general case.

For the first case, where the moist term can be neglected (above 10 km) so that (6.25) reduces to

$$N = k_1 \frac{P_d}{T}$$

the equation of state gives the density:

$$\rho = \frac{P_d M}{RT}$$

which has the same dependency of refractivity index. Next, from hydrostatic equilibrium, $P_d$ can be obtained knowing $\rho$:

$$\frac{\partial P_d}{\partial z} = -g\rho$$

Finally the temperature can be obtained from pressure and air density. In the general case, the main difference with the previous approach is to take into account the presence of water vapor. There are algorithms that may help to retrieve the water vapor partial pressure.

## 6.5 A Few Things About Radar

Without being too detailed, we can describe radar as a transmitter of a pulse of electromagnetic waves which is scattered by the target. A receiving antenna, as shown in Fig. 9.8, then captures the scattered power. Actually most of the time, the receiver and the transmitter antennas are the same. In this case, the radar is called *monostatic*, in contrast to the *bistatic* radar.

If we refer to Fig. 9.8, we can see that, if the radar emits isotropically a power $P_t$ at a distance $R_t$, it will produce a power flux given by

$$S_t = \frac{P_t}{4\pi R_t^2}$$

As we know, however, the power is actually emitted along a narrow beam, so that the flux will be increased by an amount that we call $G$, the gain of the antenna. Then a target having an area $A_t$ will intercept a power:

$$P_t = G\frac{P_t A_t}{4\pi R_t^2}$$

The target will radiate this power back to the receiver at distance $R_r$ where the antenna of area $A_e$ will gather a power (there is no gain in this case):

$$P_t = G\frac{P_t A_t A_e}{16\pi^2 R_t^2 R_t^2}$$

It is possible to see that the receiving area and the gain are related by

$$G = 4\pi A_e/\lambda^2$$

where $\lambda$ is the wavelength. The received power is then

$$P_r = P_t\frac{G^2\lambda^2}{(4\pi)^3 R_t R_r}A_t$$

Actually the area of the target is really a too simplistic approximation because the target does not scatter isotropically. It is convenient then to introduce a scattering cross section $\sigma$ such that

$$P_r = P_t\frac{G^2\lambda^2}{(4\pi)^3 R^4}\sigma \tag{6.25}$$

where we have also considered a more common monostatic radar. Equation (9.49) corresponds to the so-called radar equation for a single target of cross section $\sigma$ (Fig. 6.6).

When the radar is used in meteorology or in general to study the atmosphere, it is no longer possible to talk about single or hard targets. In this case, we have a

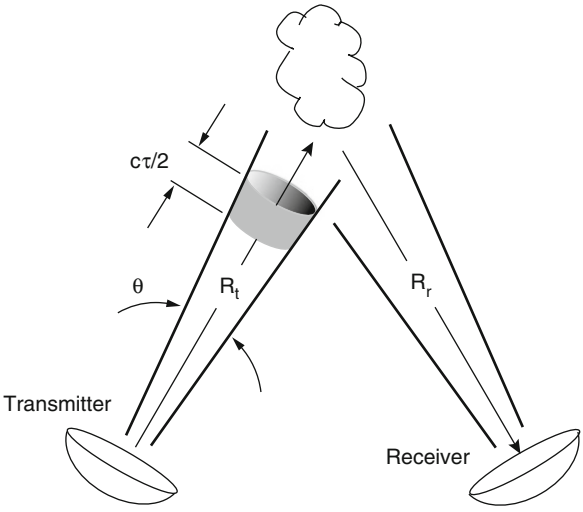



**Fig. 6.6** The geometry for the radar equation

distributed target that consists of many particles (like raindrops) that are illuminated at the same time. The simplest thing is then to substitute in Eq. (9.49) the sum of the cross sections of the single particles with the provision to average the return signal over a time interval long enough to smooth out the fluctuations in the target. It is convenient at this point to introduce a scattering cross section per unit volume $\sigma_i$ for each $i$ – particle. As shown in Fig. 9.8, the volume sampled by the radar beam is given by

$$V_m = \frac{\pi}{4}\frac{c\tau}{2}R^2\theta^2$$

where $\tau$ is the pulse length. Then the cross section will be

$$\sigma = V_m \sum \sigma_i$$

The radar equation will become

$$\overline{P}_r = P_t \frac{G^2\lambda^2}{(4\pi)^3 R^2}\frac{\pi c\tau}{8}\theta^2 \sum \sigma_i \tag{6.26}$$

Again the real world is rather different and the radar beam, rather than having a conical shape, has rather a Gaussian shape, so that Eq. (6.27) changes again to become

$$\overline{P}_r = P_t \frac{G\lambda^2 c\tau}{1024\,(\ln 2)\,R^2} \sum \sigma_i \tag{6.27}$$

When the radar wavelength is large with respect to the diameter $D$ of a scattering particle, the cross section can be assumed as that corresponding to Rayleigh scattering:

$$\sigma_i = \frac{\pi^5 {D_i}^6}{\lambda^4}|K|^2 \tag{6.28}$$

where

$$|K|^2 = \left(m^2 - 1\right) / \left(m^2 + 2\right)$$

and $m$ is the complex refractive index. This quantity is a function of wavelength and temperature. Actually the sum performed on the diameters is substituted with a measurable quantity that is called $Z$, the radar reflectivity factor defined as

$$Z = \sum D_i^6 = \int_0^\infty N(D)D^6 dD \tag{6.29}$$

The reflectivity can be put in relation with the precipitation rate $r$, usually through a power law:

$$Z = ar^b \tag{6.30}$$

The radar equation can be expressed directly in terms of the reflectivity by substituting Eqs. (9.52) and (9.53) in Eq. (9.51) so that the reflectivity becomes

$$10 \log Z = 10 \log \overline{P}_r + 20 \log R + C \tag{6.31}$$

where $C$ is a constant determined by the radar characteristics. Sometimes the reflectivity is then expressed in decibel as

$$dBz = 10 \log Z \tag{6.32}$$

This is calculated when the reflectivity is measured in $mm^6/m^3$. For example, widely accepted constants for Eq. (9.54) give

$$Z = 200r^{1.6} \tag{6.33}$$

with the rate expressed in mm/h. Considering a heavy rain of about 20 mm/h, we would have a reflectivity of 2.4 $10^4$ or $dBz$ 43.8. The sensitivity of radars is such that at a range of 15 km, they can detect a rainfall rate of about 0.1 mm/h. This means that weather radar can detect rain but not clouds; to accomplish this, radar operating at short wavelengths must be used (1 cm or less).

Different relationships can be found for other forms of precipitation like snow. However, a much more efficient way to discriminate the different precipitation types is to use a polarized radar signal. Polarized radar can send out an electromagnetic pulse that is polarized either in the vertical or the horizontal direction. Two reflectivities are obtained and their ratio can be formed, denoted as *differential reflectivity*:

$$Z_{DR} = 10 \log (Z_{HH}/Z_{VV}) \tag{6.34}$$

where $Z_{HH}$ is the reflectivity from the horizontal polarized signal and echo and $Z_{VV}$ is the same for the vertical signal. It is rather obvious to think that the geometrical form of the raindrops is influenced by their vertical motion, and especially large drops may be flattened. The depolarization factor will be then around unity for small drops (1 mm) and will grow larger for drops of several mm. For large ice particles like hail or graupel, the differential reflectivity will be roughly zero because they have the tendency to tumble and so give the same signal in the two directions. Figure 9.9 shows what is possible to obtain using modern radar. The top panel shows normal radar echo which is reflectivity that can be related to the precipitation rate. The middle panel shows the $Z_{DR}$ quantity defined before and the bottom panel takes into account the phase of the polarized signal (Fig. 6.7).

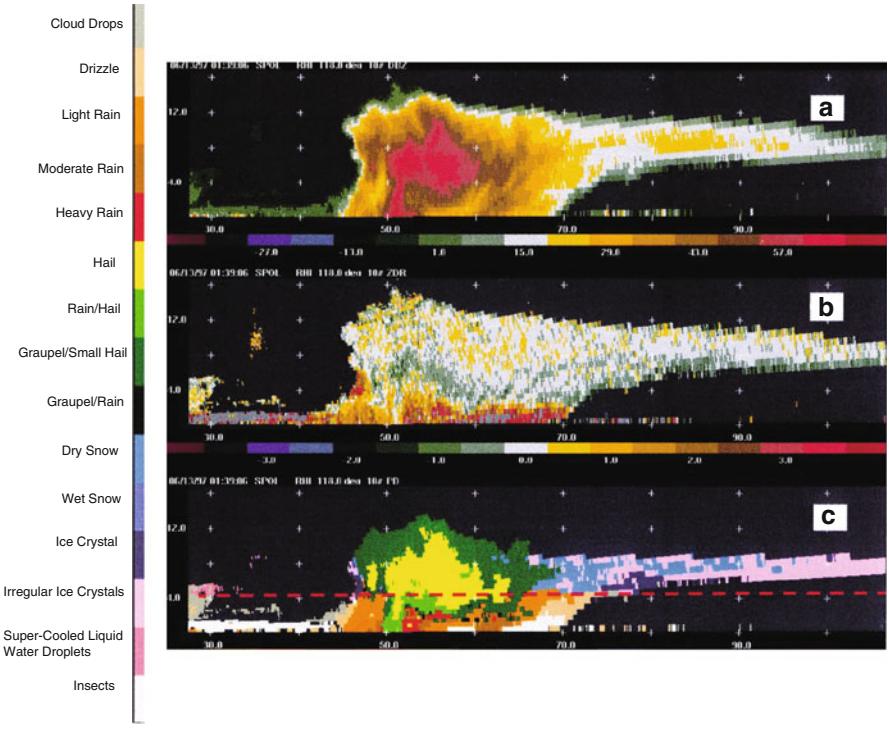



**Fig. 6.7** Three different radar images from the same thunderstorm cloud. At the *top* is shown the dBz echo, while at the *center* the $Z_{DR}$ is shown. At the *bottom*, the most sophisticated technique shows directly the different phases or physical component of the cloud (Images courtesy of J. Keeler, National Center for Atmospheric Research)

Although it is not possible to show the color code, the bright region at the top of the cloud is made by ice crystals that change gradually to hail, graupel, and different kinds of snow. In particular, the bright core at the center is essentially graupel, while the darker zones near the ground are heavy rain. The kind of data shown in the figure is obtained from a class of radar known as NEXRAD (Next-Generation Radar) that is now extensively used in a network in the USA and has dramatically improved the forecast capabilities.

## 6.6 Lidar Measurements

If the microwave beam is substituted with a light beam (possibly a laser), we have an instrument for remote sensing called lidar (light detection and ranging). In this case, a beam of light is sent in the atmosphere and the scattering medium sends back a signal that when analyzed can give information on some atmospheric parameter.

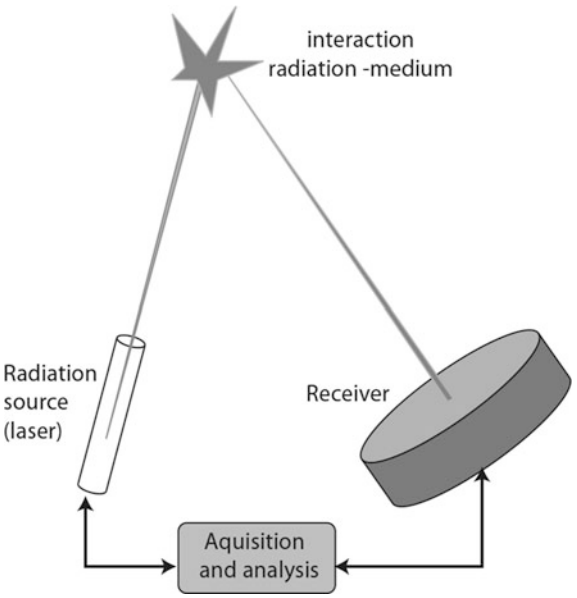



**Fig. 6.8** The layout of a lidar system (Visconti et al. 2007)

A general scheme for a lidar is given in Fig. 6.8. The transmitter usually is a laser whose characteristics (wavelength, pulse length, etc.) depend on the

specific type of measurement. The receiver is usually a telescope or a Fresnel lens that collects the light, while the acquisition system actually operates the firing of the source and maintains the timing with the received data.

The lidar equation relates the received photon counts (or signal in general) to the transmitted signal. In an interval $\Delta t$ from range interval between $s$ and $s + \Delta s$ with $\Delta s = c\Delta T/2$ ($c$ is the speed of light), we have for the number of photons received

$$N_S(\lambda_0, \lambda, s) = N_0(\lambda_0)\, T(\lambda_0, s)\, [\beta(\lambda_0, \lambda, s)\, \Delta s]\, T(\lambda, s)\, \frac{d\Omega}{4\pi}\, \eta(\lambda, \lambda_0)\, G(s) \quad (6.35)$$

where $s$ is the range; $\lambda_0$ is the emitted wavelength; $\lambda$ is the received wavelength; $T(\lambda_0, s)$ and $T(\lambda, s)$ are the transmission functions on the way up and down, respectively; $b(\lambda,\lambda_0,s)$ is the backscattering function; $\eta(\lambda, \lambda_0)$ is the efficiency in the collecting system; $G(s)$ is the geometry factor; and $d\Omega$ is the solid angle subtended by the telescope.

The transmission functions are assumed to follow the Beer–Lambert law so they may be written as

$$\begin{aligned} T_{\text{mol}}(s) &= \exp\left(-\int_0^s \sigma_{\text{mol}}^{\lambda} n_{\text{mol}}(s) ds\right) \\ T_{\text{aer}}(s) &= \exp\left(-\int_0^s \left[\int_0^{\infty} dr \pi r^2 Q_{\text{ext}}(r, m\lambda)\, n_{\text{aer}}(s, r)\right] ds\right) \\ T_{\text{abs}}(s) &= \exp\left(-\int_0^s \sigma_{\text{abs}}^{i} n_{\text{abs}}^{i}(s) ds\right) \end{aligned} \quad (6.36)$$

where the $s$ refers to the absorption cross sections for molecular scattering and absorption, while for the aerosol the transmission function depends on the total extinction cross section (the quantity in square bracket) where $r$ is the radius of the particle.

The volume backscattering coefficient ($\beta$) can be defined in general as

$$\beta\left(\lambda, \lambda_0, s\right) = \sum_i \frac{d\sigma_i\left(\lambda_0, \lambda\right)}{d\Omega} n_i(s) \tag{6.37}$$

And in case the scatterers are aerosols:

$$\beta_{\text{aer}}^{\lambda}(s) = \frac{1}{4\pi} \int_0^{\infty} dr \pi r^2 Q_{bck}\left(r, m, \lambda\right) n_{aer}\left(s, r\right) \tag{6.38}$$

where $Q_{\text{bck}}$ is the Mie backscattering efficiency of aerosol particle with radius $r$ and refractive index $m$ and $n_{\text{aer}}$ $(s, r)$ is the size distribution of the particles. As for the geometrical collecting efficiency if the telescope has surface $A$ then the solid angle in (6.36) it is simply

$$s^2 d\Omega = 4\pi A \tag{6.39}$$

Notice that the collecting angle is proportional to the collecting area while it decreases with the distance squared.

The physical processes which determine the signal are either elastic or inelastic scattering processes (Rayleigh, Mie, or Raman) that may interact with absorption processes or may relate to the properties of the absorption line. Also resonant fluorescence may be important.

## E.6 Examples

### *E.6.1 Refractive Index of Air*

We have worked out this problem at microscopic level in Chap. 2, but here we would like to explain how to arrive to Eq. (2.18). We start by considering that inside a dielectric the effective electric field is given by

$$\mathbf{E}' = \mathbf{E} + \mathbf{P}/3\varepsilon_0 \tag{E.6.1}$$

where $\mathbf{P}$ is the polarization vector and $\varepsilon_0$ is the dielectric constant. We then have for the polarization vector

$$\mathbf{P} = \alpha N \mathbf{E}' = \alpha N \left(\mathbf{E} + \mathbf{P}/3\varepsilon_0\right)$$

where $N$ is the number of molecules per unit volume and $\alpha$ is the polarizability. And then it follows that

$$\mathbf{P} = \alpha N \varepsilon_0 \mathbf{E} / (1 - \alpha N/3) \tag{E.6.2}$$

To define the refractive index, we have to find the dielectric constant $\varepsilon$ using the electric displacement vector:

$$\mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} = \varepsilon_0 \mathbf{E} + \mathbf{P} \Rightarrow \varepsilon_r = 1 + \mathbf{P} \cdot \mathbf{E} / \varepsilon_0 E^2 \tag{E.6.3}$$

The index of refraction n is then

$$n = \frac{c}{v} \approx \sqrt{\varepsilon_r} = \sqrt{1 + \frac{\mathbf{P} \cdot \mathbf{E}}{\varepsilon_0 E^2}} \tag{E.6.4}$$

Substituting for P obtained from (E.6.2), we obtain

$$N\alpha = 3\frac{n^2 - 1}{n^2 + 2} \approx 2\,(n - 1) \tag{E.6.5}$$

considering that $n$ is close to unity. This is called the Clausius–Mossotti equation. We get immediately

$$n = 1 + N\alpha/2$$

This means that the refractive index of a gas depends on the volume concentration of the gas. For a gas mixture, it follows that

$$n - 1 = \frac{1}{2} N \sum_i y_i\,(n_i - 1) \tag{E.6.6}$$

where $y_i$ is the mole fraction of species $i$ and $m_i$ is its refractive index. The quantity $n - 1$ is known as refractivity. For air, (E.6.6) can be written as

$$n - 1 = k_1 \frac{p}{T} z_a^{-1} + k_2 \frac{e}{T} z_w^{-1} + k_3 \frac{e}{T^2} z_w^{-1}$$

where $e$ is the partial pressure of water vapor, $p$ is the partial pressure of dry air, $T$ is the absolute temperature, $k_i$ are constants, and $z_{a,w}$ are deviations from the ideal gas law.

## *E.6.2 The Abel Transform*

The Abel transform can be introduced as the projection of a circularly symmetric object as shown in Fig. E.6.1. Suppose the different shades of gray represent density

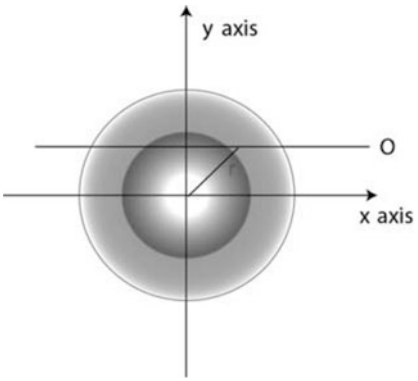



**Fig. E.6.1** The observation geometry of an object with spherical symmetry

which is a function only of the distance from the center of the circles, r. An observer in O will see

$$F(y) = \int_{-\infty}^{\infty} f(r)dx = \int_{-\infty}^{\infty} f\left(\left(x^2 + y^2\right)^{1/2}\right) dx \tag{E.6.7}$$

where

$$dx = rdr/\left(r^2 - y^2\right)^{1/2}$$

Being $f(r)$ symmetric, (E.6.7) can be written as

$$\int_{-\infty}^{\infty} f(r)dx = 2\int_{0}^{\infty} f(r)dx$$

Again substituting the expression for $dx$ as a function of $r$, we obtain the definition of the Abel transform:

$$F(y) = \int_{y}^{\infty} \frac{f(r)rdr}{(r^2 - y^2)} \tag{E.6.8}$$

The inverse Abel transform can be obtained from

$$f(r) = -\frac{1}{\pi}\int_{r}^{\infty} \frac{dF}{dy}\frac{dy}{(y^2 - r^2)^{1/2}} \tag{E.6.9}$$

A verification of this relation can be obtained by integrating by parts (E.6.8). We have

$$F(y) = -2\int_{y}^{\infty} f'(r)\sqrt{(r^2 - y^2)}dr$$

where $f'(r)$ is the derivative of $f(r)$. Differentiating we get

$$\frac{dF}{dy} = 2y\int_y^\infty \frac{f'(r)}{\sqrt{r^2 - y^2}}dr$$

that can be substituted in (E.6.9) to have

$$-\frac{1}{\pi}\int_r^\infty \frac{dF}{dy}\frac{dy}{(y^2 - r^2)^{1/2}} = \frac{1}{\pi}\int_y^\infty \int_r^\infty \frac{-2y}{[(y^2 - r^2)\,(s^2 - y^2)]^{1/2}} f'(s)dsdy$$

The last integral can be solved with Fubini's theorem:

$$\int_r^\infty \int_r^s \frac{-2y}{\pi[(y^2 - r^2)\,(s^2 - y^2)]^{1/2}} dyf'(s)ds = \int_r^\infty (-1)f'(s)ds = f(r)$$

## References[1]

### *Books*

Liou KN (1980) An introduction to atmospheric radiation. Academic, New York
Stephens GL (1994) Remote sensing of the lower atmosphere: an introduction. Oxford University Press, Oxford
Visconti G et al (eds) (2007) Observing system for atmospheric composition. Springer, New York

[1]This is a very preliminary introduction to remote sensing and the references for radiative transfer are those of Chap. 2 with some more specific publication.

# Chapter 7
# The Atmospheric Motions

In Chap. 3 we introduced the concept of circulation and vorticity, and we discovered the baroclinic atmosphere and the "production of vorticity." In Chap. 4 we saw that even the thermal wind can be obtained as a consequence of vorticity generation. It looks like the vorticity will be a long-term companion, so it may be worth it to go deeper in to its physical meaning.

Another peculiarity we need to explore in depth is its conservative properties. Just like the potential temperature, also the vorticity, under some special conditions, may become a tracer for the atmospheric motion. We naturally think vorticity as a property related to rotation of the fluid. Actually we can regard vorticity as an extension of the angular momentum concept introduced for rigid body rotation. This extension should take into account the fact that not all points within the fluid have the same angular velocity. We must then talk about differential rotation, which becomes quite familiar when we deal with the rotation of tea or coffee in a cup (a small one for espresso).

## 7.1 The Thermodynamic Equation

What is called the *thermodynamic equation* is simply a rewriting of the first principle of thermodynamics taking into account the atmospheric motions.

The starting point is then

$$\Delta Q = C_v \Delta T + p \Delta \alpha \tag{7.1}$$

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_7) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_7

We have dealt with this equation and the symbols before. The same equation can be rewritten as a function of the specific heat at constant pressure and becomes

$$\Delta Q = C_p \Delta T - \alpha \Delta p$$

Dividing through $\Delta T$ and going to the limit, we have

$$\frac{dQ}{dt} = C_p \frac{DT}{Dt} - \alpha\omega \tag{7.2}$$

where $\omega$ is the vertical velocity in pressure coordinates. Notice that the $Q$ derivative is not total because, as we know, $\Delta Q$ is not an exact differential. In this coordinate system, the total derivative of the temperature can be expressed in terms of the horizontal advection:

$$\frac{\partial T}{\partial t} = -\mathbf{V} \cdot \nabla_p T + \omega \left( \frac{\alpha}{C_p} - \frac{\partial T}{\partial p} \right) + \frac{1}{C_p} \frac{dQ}{dt} \tag{7.3}$$

This equation shows that the local temperature change is contributed to by the advection, the vertical motion, and the heat delivered to the system. If the motion is upward ($\omega < 0$), the first term in parentheses is the expansion work and so also the cooling of the air parcel; the second term is the vertical temperature advection. If $\omega$ $\partial T/\partial p < 0$ and the temperature decreases with altitude, then the contribution to the local change is a warming. The opposite happens if the motion is downward. The last term (also called *diabatic*) is the energy exchange due to processes like absorption and emission of radiation and condensation or evaporation of water vapor.

A very interesting simplification of Eq. (7.3) can be obtained rewriting the term in parentheses as a function of the potential temperature simply noting that $\ln T = \ln \theta + (R/C_p) \ln (p/p_0)$ and differentiating

$$\frac{1}{T}\frac{\partial T}{\partial p} = \frac{1}{\theta}\frac{\partial \theta}{\partial p} + \frac{R}{C_p p} \quad \Rightarrow \quad \frac{\partial T}{\partial p} = \frac{T}{\theta}\frac{\partial \theta}{\partial p} + \frac{\alpha}{C_p}$$

Substituting in Eq. (7.3) we have

$$\frac{\partial T}{\partial t} = -\mathbf{V} \cdot \nabla_p T + \omega S_p + \frac{1}{C_p} \frac{dQ}{dt} \tag{7.4}$$

where $S_p = -T\partial \ln \theta / \partial p$. For convenience we introduce a *static stability parameter* $\sigma$ defined as

$$\sigma = \frac{R}{p} S_p = -\frac{\alpha}{\theta}\frac{\partial \theta}{\partial p} \tag{7.5}$$

This is somewhat related to the consideration we made in Chap. 1, so that when the potential temperature does not change with altitude (and then with pressure) the atmosphere has a neutral stability. We have for the temperature

$$\frac{\partial T}{\partial p} = \frac{\alpha}{C_p} \Rightarrow \frac{\partial T}{\partial z} = -\frac{g}{C_p} = -\Gamma_d \tag{7.6}$$

It is interesting at this point to write the first principle in another form by deriving Eq. (7.1) with respect to time:

$$\dot{Q} = C_v \frac{DT}{Dt} + p\frac{D\alpha}{Dt} \tag{7.7}$$

where $\dot{Q} = dQ/dt$ is the net heating rate. The specific volume can be eliminated using the gas equation:

$$p\frac{D\alpha}{Dt} + \alpha\frac{Dp}{Dt} = R\frac{DT}{Dt}$$

that, substituted in Eq. (7.7), gives

$$\frac{\dot{Q}}{T} = \frac{C_p}{T}\frac{DT}{Dt} - \frac{\alpha}{T}\frac{Dp}{Dt}$$

where we have divided through by the temperature. The left-hand side represents now the entropy ($s$) change, so we write

$$\frac{Ds}{Dt} = \frac{C_p}{T}\frac{DT}{Dt} - \frac{R}{p}\frac{Dp}{Dt} = C_p\frac{D\ln T}{Dt} - R\frac{D\ln p}{Dt} \tag{7.8}$$

Notice again that we now write the total derivative for the entropy because it is an exact differential. This expression can be compared with the rate of change of the potential temperature an exact differential.

$$C_p\frac{D\ln\theta}{Dt} = C_p\frac{D\ln T}{Dt} - R\frac{D\ln p}{Dt} = \frac{Ds}{Dt} \tag{7.9}$$

From this equation we can define the *specific entropy* (per unit mass):

$$s = C_p \ln\theta \tag{7.10}$$

This is simply another way to express the thermodynamic equation.

These results have been obtained by simply working on the first principle, playing around with the total derivative; nevertheless, they give us a rather different perspective on the thermodynamics of a fluid in motion.

## 7.2 The Isentropic Coordinate System

In the isentropic coordinate system, the potential temperature $\theta$ substitutes for the vertical coordinate $z$. The advantages of such a system are many and can be seen when we represent the data of Fig. 4.6 in this new system (see Fig. 7.1).

Some algebraic work is necessary to rewrite the equations of motion in this new system. However, all this work is worth the trouble because we will use this new system a lot, especially when dealing with stratospheric dynamics. We can start by calculating in this system the elementary mass

$$\delta M = \rho \delta A \delta z = \delta A \left(-\frac{\delta p}{g}\right) = \frac{\delta A}{g}\left(-\frac{\partial p}{\partial \theta}\right)\delta\theta = \sigma \delta A \delta\theta \tag{7.11}$$

In this case $\sigma$ is the equivalent density, so that multiplied by the equivalent volume $\delta A \delta\theta$ gives the mass

$$\sigma = -\left(\frac{1}{g}\right)\frac{\partial p}{\partial \theta} \tag{7.12}$$

In this reference system, the vertical velocity corresponds to the vertical movement of the isentropic surfaces, that is,

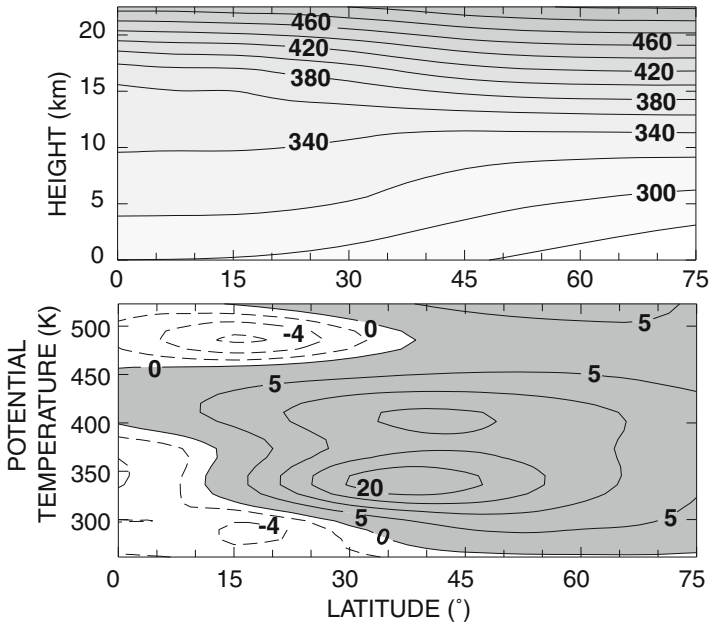



**Fig. 7.1** The potential temperature as a function of the altitude and latitude (*top*). Notice how the isentropes slope toward the equator in the troposphere. *Below* the zonal wind is represented as a function of latitude and potential temperature. Positive value is *shaded* and the data are the same used to draw Fig. 4.6

$$\dot{\theta} = \frac{D\theta}{Dt} \tag{7.13}$$

so that the total derivative is given by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + \dot{\theta}\frac{\partial}{\partial \theta} \tag{7.14}$$

In order to write the equations of motion in this coordinate system, we need to express the pressure gradient which for a generic component is given by

$$\left(\frac{\partial p}{\partial x}\right)_\theta = \left(\frac{\partial p}{\partial x}\right)_z + \left(\frac{\partial p}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_\theta \tag{7.15}$$

so that the acceleration is given by

$$-\frac{1}{\rho}\left(\frac{\partial p}{\partial x}\right)_\theta = -\frac{1}{\rho}\left(\frac{\partial p}{\partial x}\right)_z + g\left(\frac{\partial z}{\partial x}\right)_\theta \tag{7.16}$$

It is interesting to note that the pressure gradient can be greatly simplified if we obtain the derivative of the pressure from the definition of potential temperature

$$\left(\frac{\partial p}{\partial x}\right)_\theta = \frac{p}{T}\frac{C_p}{R}\left(\frac{\partial T}{\partial x}\right)_\theta = \rho C_p \left(\frac{\partial T}{\partial x}\right)_\theta \tag{7.17}$$

This expression, substituted in Eq. (7.16), gives for the pressure gradient

$$-\frac{1}{\rho}\left(\frac{\partial p}{\partial x}\right)_z = -\frac{\partial}{\partial x}\left(C_p T + gz\right)_\theta \tag{7.18}$$

The quantity

$$\Psi = \left(C_p T + gz\right)_\theta \tag{7.19}$$

is denoted as the *Montgomery streamfunction*, while the quantity $gz$ may be substituted by the geopotential. The equations of motion in this system can be written as

$$\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V}\cdot\nabla_\theta \mathbf{V} + \dot{\theta}\frac{\partial \mathbf{V}}{\partial \theta} = -f\mathbf{k}\times\mathbf{V} - \nabla_\theta \Psi \tag{7.20}$$

We can obtain a very interesting relation between the Montgomery streamfunction and the hydrostatic equilibrium. Deriving the just-defined $\Psi$ with respect to the equilibrium temperature, we have

$$\frac{\partial \psi}{\partial \theta} = C_p \frac{\partial T}{\partial \theta} + g\frac{\partial z}{\partial \theta} \tag{7.21}$$

Substituting the derivative of $T$, we obtain

$$\frac{\partial \Psi}{\partial \theta} = C_p \frac{T}{\theta} \tag{7.22}$$

In the same system, we can find the continuity equation starting from the definition Eq. (7.12). The conservation of mass requires that

$$\frac{D(\delta M)}{Dt} = \frac{D(\delta x \delta y \delta \theta \sigma)}{Dt} = 0$$

where $\delta x \delta y$ is the surface of the base. We have then

$$\delta y \delta \theta \sigma \delta \left(\frac{Dx}{Dt}\right) + \delta x \delta \theta \sigma \delta \left(\frac{Dy}{Dt}\right) + \delta x \delta y \sigma \delta \left(\frac{D\theta}{Dt}\right) + \delta x \delta y \delta \theta \left(\frac{D\sigma}{Dt}\right) = 0$$

and dividing through by $\delta x \delta y \delta \theta$

$$\sigma \frac{\delta u}{\delta x} + \sigma \frac{\delta v}{\delta y} + \sigma \frac{\delta}{\delta \theta} \dot{\theta} + \frac{\delta \sigma}{\delta t} + \mathbf{V} \cdot \nabla_\theta \sigma + \dot{\theta} \frac{\delta \sigma}{\delta \theta} = 0$$

so that

$$\sigma \nabla_\theta \cdot \mathbf{V} + \sigma \frac{\partial \dot{\theta}}{\partial \theta} + \frac{\partial \sigma}{\partial t} + \mathbf{V} \cdot \nabla_\theta \sigma + \dot{\theta} \frac{\partial \sigma}{\partial \theta} = 0$$

and the equation of continuity becomes

$$\frac{\partial \sigma}{\partial t} + \nabla_\theta \cdot (\sigma \mathbf{V}) + \frac{\partial}{\partial \theta}\left(\sigma \dot{\theta}\right) = 0 \tag{7.23}$$

One of the reasons we insist so much on these isentropic coordinates is that they are very useful in the study of the dynamics of the stratosphere. For the same reason, it is worthwhile to introduce the vorticity equation in the same coordinate system, because it is another occasion to present a new definition for the potential vorticity, which we will use a lot in the study of the stratosphere.

### *7.2.1 The Vorticity Equation in Isentropic Coordinates*

The vorticity equation can be obtained formally by taking the curl of the Eq. (7.20) so that after some laboring we get

$$\begin{aligned} \frac{\partial \zeta}{\partial t} = & -\mathbf{V} \cdot \nabla_\theta \zeta - v\beta - \frac{D\theta}{Dt} \frac{\partial \zeta}{\partial \theta} \\ & - \nabla_\theta \cdot \mathbf{V} (\zeta + f) + \mathbf{k} \cdot \frac{\partial \mathbf{V}}{\partial \theta} \times \nabla \left(\frac{D\theta}{Dt}\right) + \mathbf{k} \cdot \nabla \times \left(\frac{\mathbf{F}_r}{\rho}\right) \end{aligned} \tag{7.24}$$

We need to note that in this equation, the relative vorticity is defined a little differently because the derivative must be calculated on isentropic coordinates:

$$\zeta = \left(\frac{\partial v}{\partial x}\right)_\theta - \left(\frac{\partial u}{\partial y}\right)_\theta = \mathbf{k} \cdot \nabla_\theta \times \mathbf{V} \tag{7.25}$$

In the case of an adiabatic atmosphere without friction, the vorticity equation assumes a very simple form because the derivative involving the potential temperature goes to zero:

$$\frac{\partial (\zeta + f)}{\partial t} + \nabla_\theta \cdot (\zeta + f)\,\mathbf{V} = 0 \tag{7.26}$$

Equation (7.26) is rather interesting because it has the form of a continuity equation in which the local change of the absolute vorticity is balanced by the divergence within the volume in isentropic coordinates.

This analogy can be exploited further. In the adiabatic case, the continuity equation can be written as

$$\frac{\partial \sigma}{\partial t} + \nabla_\theta \cdot (\sigma \mathbf{V}) = 0 \tag{7.27}$$

and after some manipulation Eq. (7.58) becomes

$$\frac{\partial}{\partial t}\frac{(\zeta + f)}{\sigma} + \nabla_\theta \cdot \left(\frac{(\zeta + f)}{\sigma}\mathbf{V}\right) = 0 \tag{7.28}$$

Keeping in mind the definition for $\sigma$, we can find a new definition for the potential vorticity:

$$P = -(\zeta_\theta + f)\frac{\partial \theta}{\partial p} \tag{7.29}$$

where $\zeta_\theta$ is the relative vorticity defined on an isentropic surface. Equation (7.60) can be the interpreted as a conservation law

$$\frac{DP}{Dt} = 0 \tag{7.30}$$

This equation has a rather simple interpretation with the help of Fig. 7.2. We can figure out a vortex tube contained within two isentropic surfaces at $\theta$ and $\theta + \Delta\theta$ separated by a pressure difference $\Delta p$. The conservation of vorticity requires that if the pressure difference between the isentropic surface increases, then the absolute vorticity must also increase and vice versa.

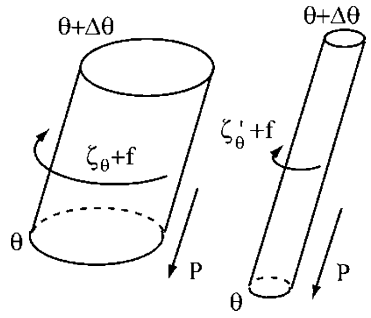



**Fig. 7.2** The conservation of potential vorticity for an elementary fluid element. The basis of the cylinder is at constant potential temperature (Pedlosky 1987)

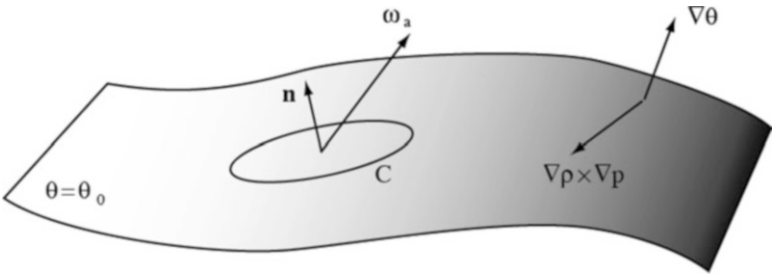



**Fig. 7.3** The computation of absolute circulation on an isentropic surface

## 7.3 The Ertel Potential Vorticity

Until now we have developed different conservation laws for different forms of potential vorticity, which are based on the Helmholtz criteria that we have mentioned at the beginning of this chapter. Also, we have discussed the potential vorticity mainly for a barotropic fluid.

The generalization of the conservation law can be made with the example illustrated in Fig. 7.3. We consider an isentropic surface, and on it we take a closed curve $C$. The vorticity equation allows us to write

$$\frac{D}{Dt}\int_A \boldsymbol{\omega}\cdot\mathbf{n}dA = \int_A \left(\nabla\rho\times\nabla p/\rho^2\right)\cdot\mathbf{n}dA \tag{7.31}$$

where $A$ is the area within the curve. For an adiabatic atmosphere, the potential temperature is conserved; thus, the surface defined by $\theta = \theta_0$ always contains the same fluid elements and as a consequence is a material surface. This means that also curve $C$ is a material curve and remains on the surface. Now because the potential temperature is only a function of pressure and density, the gradient for $\theta$ can be written as

$$\nabla\theta = \frac{\partial\theta}{\partial p}\nabla p + \frac{\partial\theta}{\partial\rho}\nabla\rho$$

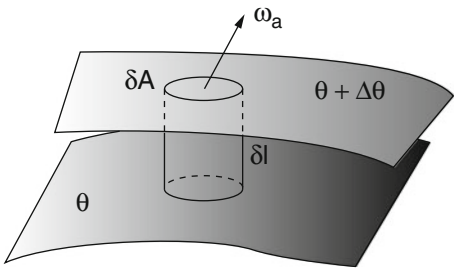



**Fig. 7.4** The elementary mass $\delta m$ between two isentropic surfaces

This means the gradient of potential temperature is always normal to the vector product $\nabla\rho \times \nabla p$. This vector remains always on the isentropic surface so that the right-hand side of Eq. (7.31) is zero. We obtain then a result very similar to the conservation of absolute circulation.

$$\frac{D}{Dt}\int_A \boldsymbol{\omega}_a \cdot \mathbf{n} dA = 0 \tag{7.32}$$

If the curve $C$ is small enough, Eq. (7.32) can be written as

$$\frac{D}{Dt}\left(\boldsymbol{\omega}_a \cdot \mathbf{n}\delta A\right) = 0 \tag{7.33}$$

where $\delta A$ is the area of the surface included by curve $C$ and $\boldsymbol{\omega}_a$ is the average value of the vorticity in the elementary volume contained within the two isentropic surfaces, as shown in Fig. 7.4. The mass of the element is given by $\delta m = \rho\delta A\delta l$, so that $\delta A = \delta m/\rho\delta l$, and expressing $\delta l$ as a function of the gradient

$$\delta l = \frac{\Delta\theta}{|\nabla\theta|}$$

and substituting into Eq. (7.33) we have

$$\frac{D}{Dt}\left(\frac{\boldsymbol{\omega}_a \cdot \mathbf{n}}{\rho}|\nabla\theta|\frac{\delta m}{\Delta\theta}\right) = 0 \tag{7.34}$$

where $\delta m$ and $\Delta\theta$ are constant along the motion. We also notice that

$$\nabla\theta = \mathbf{n}\,|\Delta\theta|$$

and Eq. (7.33) becomes

$$\frac{D}{Dt}\left(\boldsymbol{\omega}_a \cdot \frac{\nabla\theta}{\rho}\right) = 0 \tag{7.35}$$

The quantity between parentheses is known as *Ertel potential vorticity*. If we think about it a little more, we understand that Eq. (7.35) is actually a generalization of Eq. (7.30) because there we have only considered the vertical component for the vorticity. Equation (7.35) does not require, as we will see in a while, a barotropic atmosphere but has a meaning very similar to Eq. (7.30): every time the isentropic surfaces get nearer the potential, vorticity must decrease, and vice versa.

Equation (7.35) can be obtained with a formal derivation (see, e.g., Pedlosky's textbook), and it is possible to show that Eq. (7.35) can be generalized to any scalar quantity $\lambda$ that substitutes for the potential temperature. The only condition is that $\lambda$ is only a function of pressure and density and has a conservation law similar to the potential temperature for adiabatic motion. The formal derivation also assures that the atmosphere does not need to be barotropic. It is convenient to write the Ertel potential vorticity in a more familiar form:

$$\Pi = \frac{\nabla \times \mathbf{V} + f\mathbf{k}}{\rho} \cdot \nabla s \tag{7.36}$$

where we have used the entropy as an exact differential, function of pressure, and temperature. The Ertel vorticity will have important applications for stratospheric dynamics, and we will try to anticipate the reasons for that by introducing ozone gas as a tracer. We will start however with some more familiar examples.

### *7.3.1 The Application of the Potential Vorticity*

If we open a well-known textbook, like J. Holton's, we find, since the first edition, that a possible application of the conservation of potential vorticity is the study of the geophysical flux over mountain ranges. In practice we are interested in what happens when a zonal current interacts with an obstacle like a mountain. This problem is illustrated by recurring to another popular textbook (by Robert Brown), where the most frequently used figure is one like Fig. 7.5, where the case of a westerly current is shown. The hypothesis is that the airflow, even if incompressible, is deformed only on the lower surface. To understand better the situation, we may think of a river: then we see that the surface of the river is rather flat, and we do not know, looking at the surface, what is the river bottom. However, on a closer look, we can discover this by studying the circulation. It is then clear that, as the current approaches the mountain barrier, the thickness of the fluid decreases and, consequently, also the absolute vorticity must decrease. If the fluid initially has a positive relative vorticity, as shown in the figure, the required decrease will take place if the current deviates to the south. In this way there are two effects: the relative vorticity will decrease together with the planetary vorticity $f$, because the fluid is moving to lower latitudes. At this point the situation is quite similar to the one we have studied in Fig. 7.1 for the generation of the Rossby waves, so that the current will behave like a giant

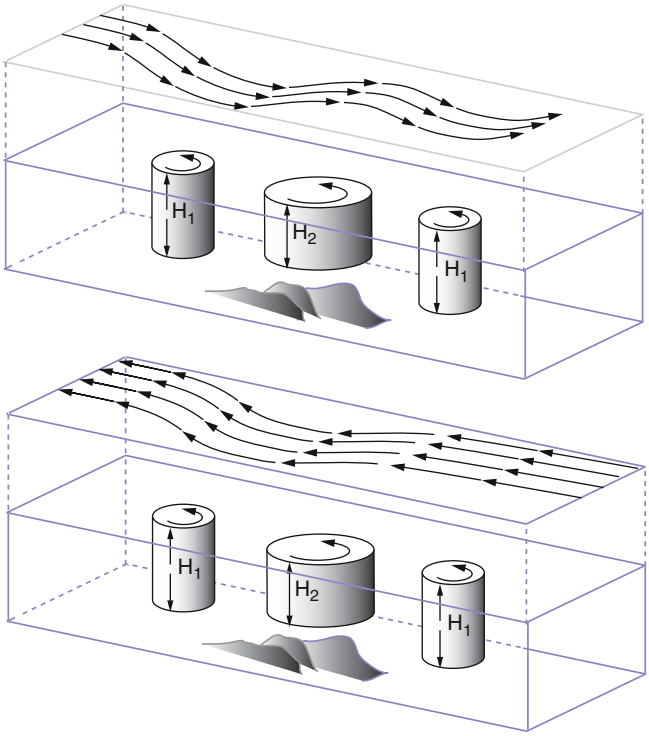



**Fig. 7.5** The flow of a zonal current over a mountain barrier. In the *upper* part the current is westerly and in the *lower part* it is easterly. The evolution of a vortex cylinder is shown. The initial depth, $H_1$, is reduced to $H_2$ over the mountain so that absolute vorticity must reduce before returning to its initial value (Brown 1991)

snake downstream. Among other things, this implies the generation of low- and high-pressure regions in the downwind region of the mountain barrier.

The situation for the current coming from the east (easterly) is completely different. Also in this case the vorticity must decrease but the problem now is how. Actually if the current moves north, the relative vorticity will decrease, while the planetary vorticity will increase. The flux will continue to be deflected north to the point that the current will be "reflected" by the mountain barrier.

If the current moves south, the relative vorticity and the planetary vorticity will increase so that again the current will be reflected. At this point, according to many textbooks, we have a miracle, and, that is, for some mysterious reason, the current will "feel" the obstacle before arriving at it. In this way, deflecting very carefully to the south, the planetary vorticity will decrease more rapidly than the relative vorticity, with the net result of an overall decrease of the absolute vorticity $\zeta + f$. In this way the vorticity will have just about the right value south of the mountain, and from that it can go back to the initial condition, that is, pure zonal flow. It is reasonable to ask how an air mass can feel at a distance the presence of an obstacle.

One possible way to explain this unusual behavior is that, based on what we learned in Sect. 7.2, the Rossby waves are generated on a westerly flow but not on an easterly flow, for the simple reason that in this case the absolute vorticity is not conserved. For an easterly flux then, the flux will simply "go over" the barrier, making a little or negligible bump on its surface. But even for the westerly flux, this simple analysis may have pitfalls. Actually it is not sure that a stationary situation could be reached, as the figure and various considerations would imply. To understand a little better, we need to reconsider the Rossby waves, but before that we want to give an example of the conservation of Ertel potential vorticity.

### 7.3.2 *Ozone and Vorticity*

We will talk about ozone later on in this book; however, it is worth to show that one simple application of Ertel vorticity was known long before the stratosphere and ozone would become topics for newspapers. Actually ozone was studied before and after the Second World War because meteorologists thought it could give indications on the "upper air level" circulation, as they said at that time. The first measurements showed that high ozone levels were correlated with low-pressure zones, while low ozone levels would occur in high-pressure zones. These correlations have been confirmed by more sophisticated measurements, and we can give here a simple explanation based on what we have learned so far.

We start by defining the ozone columnar density $N$ which is the total number of molecules contained in a vertical column of unit section above a certain height $z$ and which is measured in molecules per unit surface:

$$N(z) = \int_z^\infty n(z)dz \tag{7.37}$$

where $n(z)$ is the ozone density measured in molecules per unit volume. This expression can be rewritten as a function of the mass mixing ratio $\chi$ between ozone density and atmospheric density:

$$N(z) = \int_z^\infty \chi(z)\frac{\rho(z)}{M}dz \tag{7.38}$$

where $\rho$ is the air density and $M$ the ozone molecular mass. Using the hydrostatic equilibrium, we can write

$$N(p_s) = -\int_{p_s}^0 \frac{\chi(p)}{Mg}dp \tag{7.39}$$

The total ozone is now referred to the level above the surface pressure $p$. The pressure increment can now be expressed as a function of the potential temperature $\theta$ so that we obtain

$$N(p_s) = -\int_{\theta_s}^{\infty} \frac{\chi(\theta)}{Mg} \frac{\partial p}{\partial \theta} d\theta \tag{7.40}$$

From Eq. (7.67) the Ertel potential vorticity is given by

$$\Pi = -g(\zeta + f) \frac{\partial \theta}{\partial p} \tag{7.41}$$

that substituted into Eq. (7.40) gives

$$N(p_s) = \int_{\theta_s}^{\infty} \frac{\chi(\theta)}{\Pi(\theta)} \zeta_\theta d\theta \tag{7.42}$$

This is the equation that relates the total ozone to the vorticity. In the adiabatic motion, the Ertel potential vorticity is conserved, and in any case the residence time for ozone below 20 km is very long. The residence time is the average time spent by an ozone molecule in the atmosphere before being destroyed. If this time is long with respect to the characteristic time of the circulation, the ozone may be considered a tracer for the motion like one of those radioactive substances injected in the bloodstream to study the circulation in the human body. Data show also that the ratio $\chi/\Pi$ is roughly constant also below 20 km and that most of the ozone is found below the same altitude with no large variation with latitude. If the correlation between mixing ratio and vorticity is valid below a certain potential temperature $\theta_p$, the (7.42) can be written as

$$N \approx \overline{\left(\frac{\chi}{\Pi}\right)} \int_{\theta_t}^{\theta_p} \zeta_\theta d\theta + N_T + N_U \tag{7.43}$$

where the over bar is indicating the average value, $\theta_t$ the value at the tropopause, $N_T$ the ozone contained in the troposphere, and $N_U$ the ozone above the level of $\theta_p$ where the mixing ratio and the vorticity are no longer correlated.

For a low-pressure zone in the lower stratosphere (10–15 km), vorticity increases, and the vortex tubes must stretch, and the contribution of the integral will increase the total ozone. At the same time, the tropopause height will lower; conversely, a ridge in the flow near the tropopause will show a lower value for the total ozone. What remains to be shown is what relation exists between surface and upper level low (or ridge): that is just too early.

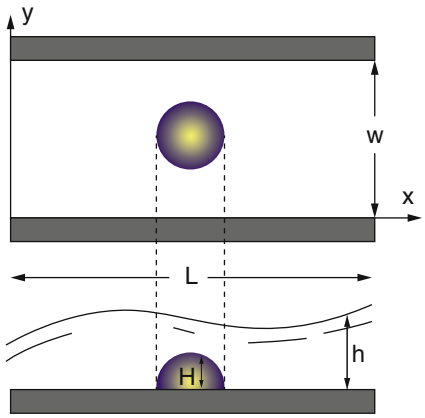



**Fig. 7.6** A schematic representation of the channel in which there is zonal flow. Indicated in the coordinate system are the width and the length of the channel (Kasahara 1966)

### *7.3.3 More on Rossby Waves*

Our starting point is the equation for the conservation of potential vorticity, but this time it is specialized to a very particular situation as shown in Fig. 7.6. This is actually a channel of width $W$ in which there is fluid with an average depth $h$. At the center of this channel, we put a barrier of a specified shape. The obstacle will be at the center in order to reduce the influence of the boundaries. In these conditions, the equation for the conservation of the vorticity becomes

$$\frac{D}{Dt}\left(\frac{\zeta + f}{h - H}\right) = 0 \tag{7.44}$$

The depth of the fluid $h$ can be then considered proportional to the geopotential height so that the components of the geostrophic wind are

$$u = -\frac{g}{f}\frac{\partial h}{\partial y} \quad v = \frac{g}{f}\frac{\partial h}{\partial x} \tag{7.45}$$

Without the central barrier, the flows has a zonal velocity $\bar{u}$ so that integrating the first of the (7.45), we obtain the depth of the fluid at the geostrophic equilibrium

$$\overline{h} = h_0 - \left(\frac{\overline{u}}{g}\right)\int_0^y f dy \tag{7.46}$$

The surface of the fluid is then a plane sloping along $y$. When the fluid is above the obstacle, we can think that $\overline{h}$ is perturbed by a small quantity so $h = \overline{h} + h'$. From now on we neglect the prime and rewrite the Eq. (7.44) as

$$\frac{D}{Dt}(\zeta + f) - \frac{\zeta + f}{\overline{h} + h - H}\frac{D}{Dt}\left(\overline{h} + h - H\right) = 0 \tag{7.47}$$

If we now calculate the relative vorticity, it will be contributed only by the perturbed term because

$$\zeta = g\nabla^2 h/f$$

Developing (7.79) we neglect all terms of higher order. The coefficient of the total derivative on the second term of Eq. (7.47) is a constant and has the value

$$\frac{(\zeta + f)}{(\overline{h} + h - H)} = \frac{f}{\overline{h}}$$

because we assume that at the entrance of the channel, the relative vorticity is zero. The ratio $f/\overline{h}$ can then be substituted by $f_0/h_0$. Equation (7.47) can be written as

$$\frac{\partial \zeta}{\partial t} + \overline{u}\frac{\partial \zeta}{\partial x} + v\beta - \frac{f_0}{h_0}\left(v\frac{\partial \overline{h}}{\partial y} + \frac{\partial h}{\partial t} + \overline{u}\frac{\partial h}{\partial x} - \overline{u}\frac{\partial H}{\partial x} - u\frac{\partial H}{\partial x} - v\frac{\partial H}{\partial y}\right) = 0$$

If we use Eq. (7.47) and neglect small terms, we have

$$\frac{\partial}{\partial t}\left(\zeta - \frac{f_0}{h_0}h\right) + \overline{u}\frac{\partial}{\partial x}\left(\zeta - \frac{f_0}{h_0}h\right) + v\left(\beta + \overline{u}\frac{f_0}{g}\right) = -\frac{f_0}{h_0}\overline{u}\frac{\partial H}{\partial x}$$

where we have neglected terms like $u\partial H/\partial x$ and $v\partial H/\partial y$. Now the vorticity can be expressed as a function of the depth to obtain

$$\frac{\partial}{\partial t}\left(\nabla^2 - \lambda^2\right)h + \overline{u}\frac{\partial}{\partial x}\left(\nabla^2 - \lambda^2\right)h + \frac{\partial h}{\partial x}\left(\beta + \overline{u}\lambda^2\right) = -\overline{u}\lambda^2\frac{\partial H}{\partial x} \tag{7.48}$$

where

$$\lambda = \frac{f_0}{\sqrt{gh_0}} \tag{7.49}$$

This equation as it stands cannot be solved immediately, but it can be instructive to put it in a form similar to Eq. (4.64), neglecting for the moment the obstacle with the right-hand side equal to zero. In this case the solution is again a Rossby wave of the form

$$h = \mathrm{Re}\left\{Ae^{i(kx+ly-\omega t)}\right\} \tag{7.50}$$

That substituted in the "homogeneous form" of Eq. (7.81) gives the dispersion relation

$$c = \overline{u} - \frac{\beta + \lambda^2\overline{u}}{k^2 + l^2 + \lambda^2} \tag{7.51}$$

This equation is very similar to Eq. (4.70) but with an important difference because it behaves normally for a westerly flux ($\overline{u} > 0$). In this case the phase velocity of the wave can be positive, negative, or zero according to the conditions

$$\beta + \lambda^2 < k^2 + l^2 + \lambda^2; \quad \beta + \lambda^2 > k^2 + l^2 + \lambda^2; \quad \beta + \lambda^2 = k^2 + l^2 + \lambda^2$$

In the case of easterly flux ($\overline{u} < 0$), the phase velocity is always negative (i.e., we have a retrograde wave). However, stationary waves are not possible because if we put in Eq. (7.51), $c = 0$, we arrive at the condition

$$k^2 + l^2 = \frac{-\beta}{\overline{u}}$$

This means that at the least one of the wave number must be imaginary and then the wave amplitude will decrease in a distance of the order of

$$\sqrt{\frac{\beta}{\overline{u}}}$$

This conclusion remains true also when the obstacle is present. Westerly flux may produce stationary waves but not easterly flux. It is worth to spend a few more words for the parameter $\lambda$. The inverse is called *Rossby deformation radius* that we have already mentioned in the first chapter. If we look better, it is the product of a fall velocity for the characteristic time related to the rotation. It represents then the maximum distance for which the Coriolis force balances the effects of the gravitational adjustment (or of pressure).

We have left the solution of Eq. (7.48) up in the air and it is time to proceed.

## 7.4 The Non-stationary Solutions

Discussing the stationary solutions, we have learned more about what happens when a flow of air goes over an obstacle. For the discussion to be complete, we need to examine the non-stationary case. For this we start from the equation of motion by referring to Fig. 7.6. The continuity equation can be written immediately just by invoking the conservation of mass

$$\frac{D\,(h-H)}{Dt} = 0$$

That gives

$$\frac{\partial h}{\partial t} + u\frac{\partial\,(h-H)}{\partial x} + v\frac{\partial\,(h-H)}{\partial y} = 0 \tag{7.52}$$

For the accelerations along $x$ and $y$, we have

$$\begin{aligned}\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}&=-g\frac{\partial h}{\partial x}+fv\\ \frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}&=-g\frac{\partial h}{\partial y}-fu\end{aligned}\tag{7.53}$$

For the numerical integration, these equations can be cast in a slightly different form that can be obtained by using the zero divergence. Then Eq. (7.52) can be written

$$\frac{\partial\left(h-H\right)}{\partial t}+\frac{\partial}{\partial x}\left[u\left(h-H\right)\right]+\frac{\partial}{\partial y}\left[v\left(h-H\right)\right]=0\tag{7.54}$$

If we take into account that the mass flux in the channel is constant, we have

$$\frac{D}{Dt}\left[u\left(h-H\right)\right]=0$$

so that the left-hand side of Eq. (7.53) can be written as

$$\begin{aligned}&\frac{Du}{Dt}=\frac{1}{h-H}\frac{D}{Dt}\left[u\left(h-H\right)\right]=\\ &\quad\frac{1}{h-H}\left\{\frac{\partial}{\partial t}\left[u\left(h-H\right)\right]+u\frac{\partial}{\partial x}\left[u\left(h-H\right)\right]+v\frac{\partial}{\partial y}\left[u\left(h-H\right)\right]\right\}\end{aligned}$$

And again using the divergence of the velocity, the first part of Eq. (7.53) becomes

$$\begin{aligned}&\frac{\partial}{\partial t}\left[u\left(h-H\right)\right]+\frac{\partial}{\partial x}\left[u^2\left(h-H\right)+\frac{g}{2}(u-H)^2\right]+\frac{\partial}{\partial y}\left[uv\left(h-H\right)\right]=\\ &\quad g\left(h-H\right)+fv\left(h-H\right)\end{aligned}\tag{7.55}$$

In synthetic form the equations of motions can then be written as

$$\frac{\partial U}{\partial t}+\frac{\partial F}{\partial x}+\frac{\partial G}{\partial y}=Q\tag{7.56}$$

where

$$U=\begin{bmatrix}m\\ n\\ \phi\end{bmatrix}\quad F=\begin{bmatrix}\frac{m^2}{\phi}+g\frac{\phi^2}{2}\\ \frac{mn}{\phi}\\ m\end{bmatrix}\quad G=\begin{bmatrix}\frac{mn}{\phi}\\ \frac{m^2}{\phi}+g\frac{\phi^2}{2}\\ n\end{bmatrix}\quad Q=\begin{bmatrix}-g\phi H_x+fn\\ -g\phi H_y+fm\\ 0\end{bmatrix}\tag{7.57}$$

and

$$m = u\,(h - H)\,, \quad n = v\,(h - H)\,, \quad \phi = h - H \tag{7.58}$$

In this case we notice that $m$ and $n$ are the components of the momentum and $\varphi$ is the thickness of the fluid above the mountain.

The system of Eq. (7.57) is solved in an Appendix and below we will discuss those solutions. These equations are also known as *shallow water*. For the above treatment, we have been inspired by the work of Akira Kasahara.

### 7.4.1 Numerical Solutions of a Flow Above an Obstacle: The Stationary Case

At this point the question may have come to mind that problems in atmospheric physics very seldom can be solved analytically in detail, and this suspicion may be true. On the other hand, it may have some advantage in the sense that, early on, an atmospheric physicist has to learn how to live with computers and numerical analysis. In this specific case, Eq. (7.81) can only be solved numerically not only for the time-dependent case $h(x, y, t)$ but even for the stationary case. This solution corresponds to making the time derivative zero and in this case

$$\nabla^2 h + \frac{\beta}{\overline{u}} h = -\lambda^2 H \tag{7.59}$$

This is what is called an elliptic partial differential equation, and it can be solved with a standard numerical program. However, the solution makes sense only in case of easterly flux, because a westerly flux produces waves and nobody assures us that those are stationary. We have taken a program from a text of numerical analysis and have solved the equation. We have used the same orography as that illustrated in Fig. 7.6, with a radius $b = 1800$ km and a channel of $60° \times 150°$ (latitude per longitude). The zonal current has a velocity of 20 m $s^{-1}$. The boundary conditions are such that the function $h(x, y)$ is zero at the northern and southern boundaries and small at the eastern and western boundaries.

The result of the integration is shown in Fig. 7.7, and at first sight it is quite surprising because the isopleths of $h(x, y)$ are circular, that is, the same as the isopleths of the obstacle $H$. The reason for this is to be found in the conservation of potential vorticity that we write for an element following the fluid

$$\mathbf{V} \cdot \nabla \left( \frac{f + \zeta}{\overline{h} - H} \right) = 0 \tag{7.60}$$

where $\mathbf{V}$ is the vector representing the perturbation to the velocity field. Equation (7.60) tells us that far from the obstacle, the relative vorticity is zero. In this case the

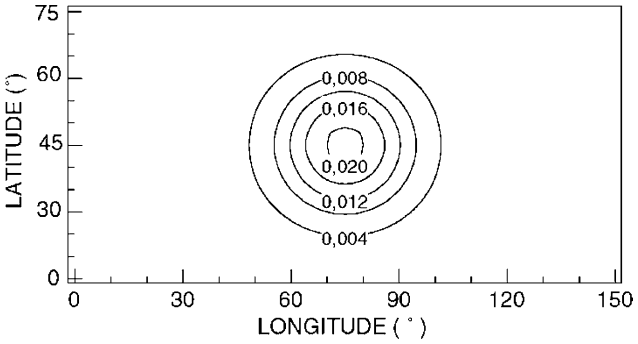



**Fig. 7.7** The height contours expressed in km for an easterly flux over the obstacle depicted in Fig. 7.6. The wind is clockwise

vector **V** is always normal to the gradient of $h - H$ and so to the obstacle. Physically we could easily understand that the isopleths of the obstacle are impenetrable and then the fluid can only go around it.

Actually such a show of mathematical culture was not necessary because we could reach the same conclusion using paper and pencil. This however is a proof that our numerical method is giving us the correct results.

The same case can be worked out when we keep the Coriolis parameter constant with latitude. In this case $\beta = 0$ and we have a Poisson equation. It can be seen that also in this case, the isopleths follow the obstacle but they change considerably near the boundaries. In the case of a constant Coriolis parameter, the conservation of potential vorticity can be obtained only at the expenses of the relative vorticity, so that the absolute change of this quantity must be larger. The atmospheric column coming from the east is compressed, and to maintain the potential vorticity, we must assume an anticyclonic relative vorticity moving toward south. Once over the top of the mountain, we have the inverse process, with a column that is stretched and a vorticity that becomes cyclonic.

The last problem left at this point is the non-stationary case.

### *7.4.2 Numerical Solutions of a Flow Above an Obstacle: The Non-stationary Case*

The numerical solutions of the equations in the non-stationary case are not very simple, and there are no programs available in books or elsewhere. For this reason we report in the examples the listing of a program with a minimum of explanation. After about 30 years, we have solved the same problem as Kasahara and among other things, we have found that at that time, draftsmen did their part in the process of solving equations. Today the same solutions are plotted without any pity by one of the several available software programs you can find in any laptop computer (Fig. 7.8).

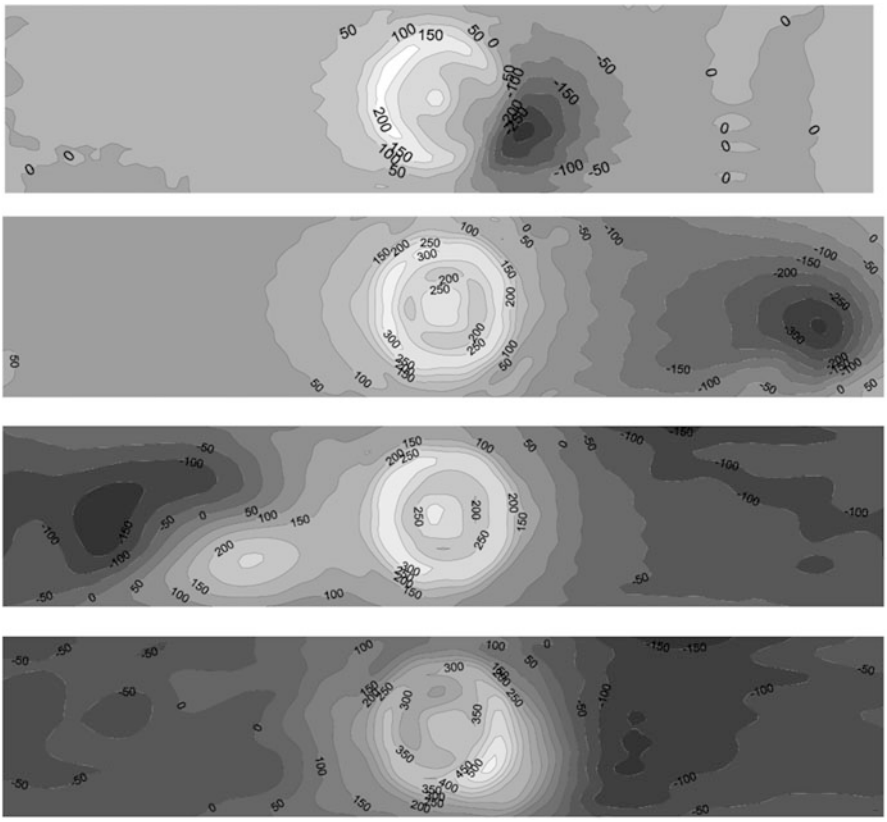


**Fig. 7.8** The non-stationary solution of a flow above an obstacle. The changes of the fluid level (in m) with respect to the initial conditions are shown. The four maps show the situation after 2, 8, 24, and 20 days from the start of the simulation. The flux is westerly at 20 m $s^{-1}$ and in this case the Coriolis parameter is constant with latitude

The problem has been solved for a channel centered at a latitude of 45° with a width of 7100 km. The dimensions of the obstacle are the same as in the stationary case. In Fig. 7.6 the changes in the fluid levels are shown at different times after the beginning of the simulation. In the case depicted, the Coriolis parameter is constant with latitude.

We notice that the level of the fluid corresponding to the obstacle starts to have a bulge and, even with some variability, maintains in that region a high-pressure zone. Downstream from the maximum altitude of the obstacle, a low-pressure zone is established (a negative bulge). This low pressure slowly tends to migrate eastward to the point that, after the fourteenth day, it reappears on the western side. The drift velocity of this low-pressure zone is then the same as the basic flow.

Figure 7.9 shows what happens when the Coriolis parameter is variable with latitude. In this case it is rather difficult to give a simplistic explanation, as given in

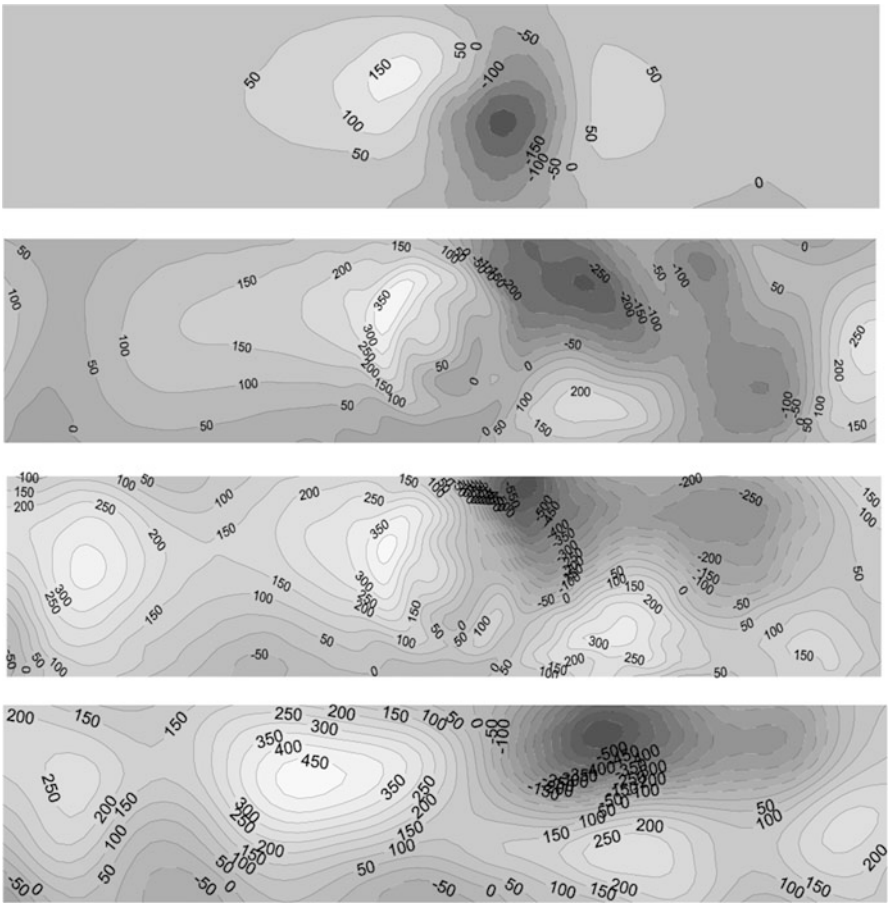


**Fig. 7.9** The same as Fig. 7.8 but in this case the Coriolis parameter is a function of latitude

the Holton textbook. The solution shows actually an initial drift of the low-pressure zone toward southern latitudes and the formation of waves. These however do not show any stationary behavior, and their wavelength is roughly one third of the length of the latitude circle (wave number 3). With these waves we may associate pressure features that move with the fluid. In this case also the high-pressure zone near the obstacle seems rather stationary.

At this point we have done enough work with our meager knowledge on a problem that has great importance for meteorology in general. Now we are almost ready to tackle a central problem of atmospheric dynamics, that is, the explanation of the general circulation of the atmosphere. We still need to learn some more about boundary layers and gravity waves.

## 7.5 Quasi-Geostrophic Vorticity

We have introduced in Chap. 3 many concepts regarding the vorticity and its conservation. In particular we can rewrite Eq. 3.50 considering that the total divergence is zero. We have then

$$\frac{D}{Dt}(\zeta + f) = (\zeta + f)\frac{\partial w}{\partial z} \tag{7.61}$$

To solve this equation, we have to eliminate the vertical velocity. Again we write a perturbation form of the (7.61) by putting, $u = \overline{u} + u'$, $v = v'$, $w = w'$, so that

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\zeta' + v'\frac{\partial f}{\partial y} - f\frac{\partial w'}{\partial z} = 0 \tag{7.62}$$

We then consider the thermodynamic Eq. (7.9)

$$\frac{D}{Dt}\ln\theta = \left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x}\right)\theta + w\frac{\partial\theta}{\partial z} = 0 \tag{7.63}$$

The perturbation form of this equation with $u = \overline{u} + u'$, $w = w'$ and $\theta = \overline{\theta} + \theta'$

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\theta'}{\overline{\theta}} + \frac{w'}{\overline{\theta}}\frac{\partial\overline{\theta}}{\partial z} = 0 \tag{7.64}$$

The ratio between the perturbation and the average value of the potential temperature can be expressed as

$$\frac{\theta'}{\overline{\theta}} = \frac{\rho'}{\overline{\rho}}$$

This is tantamount to assuming that the only changes in temperature arise from buoyancy effects, and this is known as *Boussinesq approximation*. By substituting in the Eq. (10.33), we get considering that $N^2/g = \overline{\theta}^{-1}\left(\partial\overline{\theta}/\partial z\right)$

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\rho'}{\overline{\rho}} - \frac{N^2}{g}w' = 0 \tag{7.65}$$

This equation can be used to eliminate the vertical velocity although as a further assumption we need to consider that the density fluctuations are also in hydrostatic equilibrium.

$$\frac{1}{\overline{\rho}}\frac{\partial p'}{\partial z}+\frac{\rho'}{\overline{\rho}}g=0 \tag{7.66}$$

We can now differentiate (7.65) with respect to $z$ and substitute the result in the Eq. (7.62), while to eliminate density we use (7.66)

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\left(\zeta'+\frac{f}{\overline{\rho}N^2}\frac{\partial^2 p'}{\partial z^2}\right)+v'\beta=0 \tag{7.67}$$

where $\beta$ is the latitudinal gradient of the Coriolis parameter. Introducing the streamfunction $\psi=p'/f\overline{\rho}$ and substituting in the Eq. (7.67), we get

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\left(\nabla^2\psi+\frac{f^2}{N^2}\frac{\partial^2\psi}{\partial z^2}\right)+\beta\frac{\partial\psi}{\partial x}=0 \tag{7.68}$$

This expression is known as the *quasi-geostrophic vorticity equation* because, except for the divergence term, we use the value of the geostrophic wind in the definition of $\psi$. Actually the meaning of this equation is far more reaching. We can define a potential vorticity as

$$q=\overline{q}+q'=\nabla^2\psi+f+\frac{f^2}{N^2}\frac{\partial^2\psi}{\partial z^2} \tag{7.69}$$

so that, because of (7.68), we have $\partial\overline{q}/\partial y=\beta$ and obtain the conservation equation with $\mathbf{V}_g$, the geostrophic wind

$$\left(\frac{\partial}{\partial t}+\mathbf{V}_g\cdot\nabla\right)q=0 \tag{7.70}$$

We had already defined the potential vorticity. We can see that this definition is equivalent because it can be shown to be a particular form of the Ertel potential vorticity

$$\frac{(\zeta+f)\,\partial\theta}{\partial z}$$

Actually when the difference in potential temperature is fixed, the potential vorticity will change as the distance between the corresponding isentropic surfaces changes. Consider now a perturbation of the potential temperature such that $\theta=\overline{\theta}+\theta'$; we have then for the Ertel potential vorticity

$$(\zeta+f)\,\theta_z=\overline{\theta}_z\,(\zeta+f)\left(1+\theta'_z/\overline{\theta}_z\right) \tag{7.71}$$

where subscripts indicate derivatives. At this point we notice that $\overline{\theta}_z = N^2\overline{\theta}/g$ and also $\theta'_z/\overline{\theta}_z = -\rho'_z/\overline{\rho}$. Substituting in (7.71) we have

$$\overline{\theta}_z\,(\zeta + f)\left(1 + \theta'_z/\overline{\theta}_z\right) = \overline{\theta}_z\,(\zeta + f)\left(1 - \frac{g}{N^2}\frac{\rho'_z}{\overline{\rho}}\right)$$

And using hydrostatic equilibrium we have the equivalent of Eq. (7.68)

$$\overline{\theta}_z\,(\zeta + f)\left(1 - \frac{g}{N^2}\frac{\rho'_z}{\overline{\rho}}\right) = \overline{\theta}_z\,(\zeta + f)\left(1 + \frac{1}{N^2\overline{\rho}}\frac{\partial^2 p'}{\partial z^2}\right)$$

If we consider that at middle latitudes $\zeta \ll f$ we have a form similar to (7.69)

$$\overline{\theta}_z\,(\zeta + f)\left(1 + \frac{1}{N^2\overline{\rho}}\frac{\partial^2 p'}{\partial z^2}\right) \approx \overline{\theta}_z\left(\zeta + f + \frac{f^2}{N^2}\frac{\partial^2 \psi'}{\partial z^2}\right)$$

Equation (7.69) will be used when dealing with the stratosphere.

### *7.5.1 The Equation of Quasi-Geostrophic Potential Vorticity*

In the log–pressure coordinate system in addition to the vertical coordinate

$$z^* = -H \ln\left(\frac{p}{p_s}\right)$$

we also assume that the density changes according to the relation

$$\rho_o\left(z^*\right) = \rho_s \exp\left(-\frac{z^*}{H}\right)$$

where $\rho_s$ and $H$ are the average surface density and average scale height, respectively. The pressure changes also in the same way. A few variables need to be redefined like the vertical velocity

$$w^* = \frac{Dz^*}{Dt}$$

which we write as follows:

$$w^* = -H\frac{D}{Dt}\ln\left(\frac{p}{p_s}\right) = -\frac{H}{p}\frac{Dp}{Dt} = -\frac{H\omega}{p} \tag{7.72}$$

In the same coordinate system, the continuity equation becomes

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}-\frac{\partial}{\partial p}\left(\frac{pw^*}{H}\right)=\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\rho_0 w^*\right) \tag{7.73}$$

Because we have

$$-\frac{\partial}{\partial p}\left(\frac{pw^*}{H}\right)=-\frac{w^*}{H}-\frac{p}{H}\frac{\partial w^*}{\partial p}=-\frac{w^*}{H}+\frac{\partial w^*}{\partial z^*}$$

At this point we note that the total derivative becomes

$$\frac{D}{Dt}=\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}+w^*\frac{\partial}{\partial z^*} \tag{7.74}$$

and the derivative, with respect to the geometric height $z$, need to be transformed as follows

$$\frac{\partial}{\partial z}=\frac{\partial z^*}{\partial z}\frac{\partial}{\partial z^*}=\frac{H\rho_0 g}{p}\frac{\partial}{\partial z^*} \tag{7.75}$$

From Eq. (7.75) we see how the derivative with respect to $z$ or $z^*$ is the same only for an isothermal atmosphere at a temperature that coincides with the global averaged value. In particular we can redefine the Brunt–Väisälä frequency, which we indicate as $N_*^2$ in this coordinate system:

$$N_*^2=\frac{g}{\overline{\theta}}\frac{\partial\overline{\theta}}{\partial z^*}=\frac{g}{\overline{\theta}}\frac{\partial\overline{\theta}}{\partial z}\frac{\partial z}{\partial z^*}=N^2\frac{p}{H\rho g} \tag{7.76}$$

The interested reader may go to the Appendix for the formal derivation where we show that once the quasi-geostrophic potential vorticity is defined as

$$q=\overline{q}+q' \tag{7.77}$$

where

$$q'=\nabla^2\psi+\frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\frac{f^2\rho_0}{N_*^2}\frac{\partial\psi}{\partial z^*}\right);\quad \overline{q}=f \tag{7.78}$$

we have

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)q'+\frac{\partial\overline{q}}{\partial y}\frac{\partial\psi}{\partial x}=0 \tag{7.79}$$

That is equivalent to the conservation equation:

$$\frac{Dq}{Dt} = 0 \tag{7.80}$$

This equation, even if more general, assumes that the Coriolis parameter $f$ is constant with latitude. The effects of a variable $f$ are felt for distances of the order of 1000 km. On the same horizontal scales, we will assume $\bar{u}$ to be a weak function of $x$ and $z$ so that we can write

$$\psi = -\overline{u}\,(x, z) + \psi'$$

and from the definition of quasi-geostrophic potential vorticity, we have

$$q = \nabla^2\psi + f + \frac{1}{\rho_o}\frac{\partial}{\partial z^*}\left(\frac{f^2\rho_o}{{N_*}^2}\frac{\partial\psi}{\partial z^*}\right)$$

so that

$$\begin{aligned} q' &= \nabla^2\psi' + \frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\frac{f^2\rho_0}{N_*^2}\frac{\partial\psi'}{\partial z^*}\right) \\ \frac{\partial\overline{q}}{\partial y} &= \beta - \frac{\partial^2\overline{u}}{\partial y^2} - \frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\frac{f^2\rho_0}{N_*^2}\frac{\partial\overline{u}}{\partial z^*}\right) \end{aligned} \tag{7.81}$$

These quantities satisfy the conservation equation

$$\left(\frac{\partial}{\partial t} + \frac{\overline{u}\partial}{\partial x}\right)q' + v'\partial\overline{q}/\partial y = 0 \quad \text{where} \quad v' = \frac{\partial\psi'}{\partial x} \tag{7.82}$$

## 7.6 Potential Vorticity Inversion

A problem arises in considering the conservation of potential vorticity. If $h$ is the depth of the fluid, this can be defined as $\zeta_a/h$. The conservation is expressed as

$$\frac{D}{Dt}\left[\frac{1}{h}\left(f + \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)\right] = 0 \tag{7.83}$$

If we indicate the potential vorticity with $q = \zeta_a/h$, the conservation can be expressed as

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + \mathbf{u}\cdot\nabla q = 0 \tag{7.84}$$

If the geostrophic approximation is valid (i.e., the Rossby number is small), we have

$$u = -\frac{\partial \psi}{\partial y}; \quad v = \frac{\partial \psi}{\partial x}$$

And Eq. (7.83) becomes

$$\frac{D}{Dt}\left[\frac{1}{h}\left(f + \nabla^2 \psi\right)\right] = 0; \tag{7.85}$$

where $\zeta = \nabla^2 \psi$ is the relative vorticity. Given the vorticity the streamfunction can be obtained by solving the Poisson equation, and then the velocity components are determined. This is called *vorticity inversion*. The integration of Eq. (7.83) is then a process of time stepping plus inversion: Equation (7.84) may be integrated for one time step, and the velocity component can be obtained. A very simple example due to D.J. Raymond can help to clarify the matter. Consider a fluid of thickness $h$ over a base of variable thickness $d$. The two components of the velocities are

$$u = -\frac{g}{f}\frac{\partial (h+d)}{\partial y}; \quad v = \frac{g}{f}\frac{\partial (h+d)}{\partial x} \tag{7.86}$$

so that Eq. (7.83) becomes

$$q = \frac{1}{h}\left[\frac{\partial}{\partial x}\left(\frac{g}{f}\frac{\partial (h+d)}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{g}{f}\frac{\partial (h+d)}{\partial y}\right) + f\right]$$

We assume now that thickness changes are small so that we can write $h = h_0 (1 + \eta)$, and putting $q_0 = f/h_0$ we get

$$q = q_0\left[1 - \eta + L_\rho^2 \nabla^2\left(\eta + \frac{d}{h_0}\right)\right] \tag{7.87}$$

where

$$L_\rho = \frac{\sqrt{g h_0}}{f}$$

is the Rossby deformation radius. If we define a perturbation potential vorticity $q' = q - q_0$, we have

$$\frac{q'}{q_0} = L_\rho^2 \nabla^2\left(\frac{\eta + d}{h_0}\right) - \eta \tag{7.88}$$

Within the limits of this linearized version, we can proceed using the following steps. (1) For a given initial distribution of vorticity $q$, obtain the distribution of thickness $\eta$ solving the Eq. (7.88). (2) Use Eq. (7.86) to obtain the velocities. (3)

Use these velocities to advect the vorticity using (7.84). This procedure should be repeated until a satisfactory convergence is achieved. One example of inversion will be given in the following paragraph.

### 7.6.1 A Periodic Potential Vorticity Anomaly

We assume that the potential vorticity perturbation is given by

$$q' = \varepsilon q_0 \sin(kx) \tag{7.89}$$

with $\varepsilon$ and $k$ constants. We also assume that the bottom of the fluid is flat, $d = 0$, so that Eq. (7.88) becomes

$$L_\rho^2 \nabla^2 \eta - \eta = \varepsilon \sin(kx) \tag{7.90}$$

We assume a solution of the form $\eta = \eta_0 \sin(kx)$, and upon substitution we have

$$\eta_0 = -\frac{\varepsilon}{\left(1 + k^2 L_\rho^2\right)}$$

And the thickness becomes

$$h = h_0 \left(1 - \frac{\varepsilon \sin(kx)}{1 + k^2 L_\rho^2}\right) \tag{7.91}$$

And the geostrophic velocities are

$$u = 0; \quad v = -\frac{fkL_\rho^2 \epsilon \cos(kx)}{1 + k^2 L_\rho^2} \tag{7.92}$$

And the absolute vorticity becomes

$$\zeta_a = f\left(1 + \frac{k^2 L_\rho^2 \varepsilon \sin(kx)}{1 + k^2 L_\rho^2}\right) \tag{7.93}$$

The potential vorticity can be then calculated as $\zeta_a/h$, and we get

$$q = \frac{\zeta_a}{h} = q_0 \left(1 + \left(\frac{k^2 L_\rho^2}{1 + k^2 L_\rho^2} + \frac{1}{1 + k^2 L_\rho^2}\right) \varepsilon \sin(kx)\right) \tag{7.94}$$

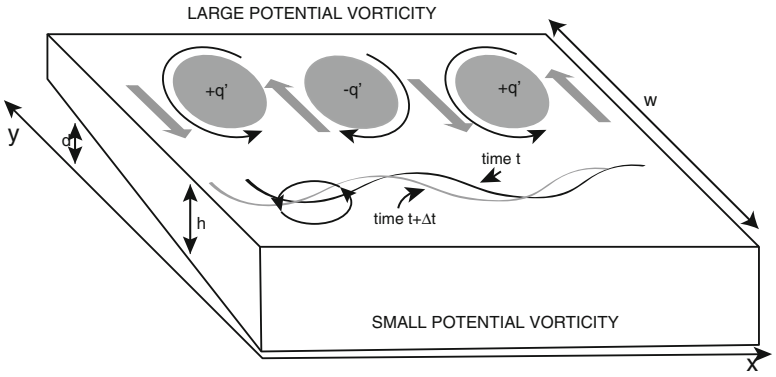



**Fig. 7.10** The generation of a Rossby wave from the displacement of a parcel of air in north south direction. The wavy lines indicates the propagation of the Rossby waves

The perturbation potential vorticity is then

$$q' = q - q_0 = q_0 \left( \frac{k^2 L_\rho^2}{1 + k^2 L_\rho^2} + \frac{1}{1 + k^2 L_\rho^2} \right) \varepsilon \sin(kx) = q_0 \varepsilon \sin(kx) \qquad (7.95)$$

This demonstrates that the initial perturbation is completely recovered. The two terms on the right-hand side of (7.95) represent a perturbation in the potential vorticity and a perturbation in the thickness of the fluid. As a consequence if $k^2 L_\rho^2 \gg 1$ (the horizontal scale of the vorticity perturbation much longer than the Rossby radius), the second term will dominate. On the other hand if $k^2 L_\rho^2 \ll 1$ (the horizontal scale of the vorticity perturbation much shorter than the Rossby radius), the first term will dominate.

### *7.6.2 Rossby Waves and Vorticity Inversion*

We have studied Rossby waves for several pages now. The inversion of potential vorticity gives us the occasion to discuss Rossby waves from a very original point of view. Consider as in Fig. 7.10 a channel of width $w$ with the bottom tilted in such a way that the fluid at rest is thinner at the north rim and thicker at the south. When a parcel moves from north to south in order to keep the potential vorticity constant, it must acquire positive relative vorticity (anticlockwise rotation), while the opposite happens when it moves northward. A northward flow between the gaps tends to reduce the potential vorticity while the opposite happens for southward flow. Examination of Fig. 7.10 shows that both negative and positive anomalies will move to the left. This is actually the generation of a Rossby wave.

We assume then that starting from the south ($y = 0$), the bottom will rise linearly with $y$ according to

$$d = h_0 \mu y \tag{7.96}$$

Because the surface is at rest, we impose that $h + d = h_0 (1 + \eta) + h_0 \mu y = h_0$ so that $\eta = -\mu y$ at rest. In general the level at rest will be perturbed so that

$$\eta = -\mu y + \eta^* \tag{7.97}$$

In order to maintain the linearization condition $|\eta| \ll 1$ so that $\mu w \ll 1$. We have

$$h + d = h_0 \left(1 + \eta^*\right)$$

So that the velocities are

$$u = -\frac{g h_0}{f} \frac{\partial \eta}{\partial y} \quad v = \frac{g h_0}{f} \frac{\partial \eta}{\partial x}$$

And the potential vorticity takes the form

$$q = \frac{f}{h_0 (1 - \mu y)} \approx \frac{f}{h_0} (1 + \mu y)$$

To which we must add a perturbation $q^*$

$$q = q_0 (1 + \mu y) + q^* \tag{7.98}$$

The Eq. (7.84) becomes

$$\frac{\partial q^*}{\partial t} + v \frac{\partial (q_0 \mu y)}{\partial y} = \frac{\partial q^*}{\partial t} + g \mu \frac{\partial \eta^*}{\partial y} = 0 \tag{7.99}$$

While Eq. (7.88) becomes

$$L_\rho^2 \nabla^2 \eta^* - \eta^* = \frac{q^*}{q_0} \tag{7.100}$$

Solution to Eqs. (7.99 and 7.100) can be assumed to be wavy in character and limited at the boundaries $y = 0, w$. A possible solution is of the form

$$\begin{aligned} \eta^* &= \eta_0^* \sin (\pi y / w) \exp [i (kx - \omega t)] \\ q^* &= q_0^* \sin (\pi y / w) \exp [i (kx - \omega t)] \end{aligned} \tag{7.101}$$

Substituting in (7.99 and 7.100) we have a homogeneous linear system

$$-\omega q^* + kg\mu\eta^* = 0$$
$$\frac{q^*}{q_0} + \left[1 + L_\rho^2\left(k^2 + \pi^2/w^2\right)\right]\eta^* = 0$$

The only nontrivial solutions exist when the determinant is zero. We have

$$\omega = -\frac{kL_\rho^2 f\mu}{1 + L_\rho^2\left(k^2 + \pi^2/w^2\right)} \tag{7.102}$$

This dispersion relation can be understood better if the quantities are normalized to the Rossby radius and the Coriolis parameter according to

$$\overline{k} = L_\rho k \quad \overline{\omega} = \frac{\omega}{f} \quad \overline{w} = \frac{w}{L_\rho} \quad \overline{\mu} = \mu L_\rho$$

And the dispersion relation becomes

$$\overline{\omega} = -\frac{\overline{k}\overline{\mu}}{1 + \overline{k}^2 + \pi^2/\overline{w}^2} \tag{7.103}$$

As expected the phase velocity $\omega/k$ is negative (directed in the negative x direction) while the group velocity $\partial\omega/\partial k$ stays negative up to a critical value $k = k_c \approx 1.5$ and then becomes positive. This means that in the short wavelength limit, the group velocity is opposite to the phase velocity.

## 7.7 Scaling of the Shallow Water Equations

In Chap. 4, we have given some argument to simplify the equations of motion based on the so-called scale analysis, that is, compare the order of magnitude of different terms. This is a rather powerful method and gives a very interesting physical insight about the equation of motion.

### *7.7.1 Scaling of the Equations of Motion*

The simplest case is the shallow water system that can be written as

$$\frac{D\mathbf{V}}{dt} + \nabla\phi + f\mathbf{k}\times\mathbf{V} = 0 \tag{7.104}$$

where $\phi = gh$ and we also assume that the Coriolis parameter is constant. We also write the continuity equation

$$\frac{Dh}{Dt} + h\left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial x}\right) = 0$$

which translate into

$$\frac{\partial \phi}{\partial t} + \mathbf{V} \cdot \nabla \phi + \phi \nabla \mathbf{V} = 0 \tag{7.105}$$

Now we let a time scale $T = L/V$ and the order of magnitude of the different terms in Eq. (7.104) are

$$\begin{array}{llll} \dfrac{\partial V}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V} + \nabla \phi + f\mathbf{k} \times \mathbf{V} = 0 \\ \dfrac{V^2}{L} & \dfrac{V^2}{L} & fV \\ RofV & RofV & fV \end{array} \tag{7.106}$$

where we have used the definition of Rossby number $\mathrm{Ro} = V/fL$. If the Rossby number is small, the first two terms can be neglected, and the equilibrium is between the Coriolis term and the pressure gradient force:

$$\nabla \phi \sim fV \tag{7.107}$$

Equation (7.105) can be analyzed if we assume the $\phi$ field to be decomposed as

$$\phi = \overline{\phi} + \phi' \tag{7.108}$$

With $\overline{\phi}$ a constant. Substituting in Eq. (7.107) gives

$$\phi' \sim fLV \tag{7.109}$$

If we carry out the scaling for Eq. (7.105), we get

$$\begin{array}{llll} \dfrac{\partial \phi'}{\partial t} + \mathbf{V} \cdot \nabla \phi' + \overline{\phi} \nabla \cdot \mathbf{V} + \phi' \nabla \cdot \mathbf{V} = 0 \\ fV^2 & fV^2 & \overline{\phi} V/L & fV^2 \\ RoF\overline{\phi}\dfrac{V}{L} & RoF\overline{\phi}\dfrac{V}{L} & \overline{\phi} V/L & RoF\overline{\phi}\dfrac{V}{L} \end{array}$$

where we have introduced the rotational Froude number $F = f^2L^2/\overline{\phi}$ that can also be written as

$$F = \frac{f^2L^2}{\overline{\phi}} = \left(\frac{L}{L_\rho}\right)^2 \tag{7.110}$$

where $L_\rho = \sqrt{gh}/f$ is the Rossby radius of deformation.

Assuming $F \leq 1$ and neglecting terms of the order of Ro in (7.106), we have the geostrophic equilibrium

$$\nabla\phi + \mathbf{k} \times \mathbf{V} = 0 \quad \Rightarrow \quad \mathbf{V} = f^{-1}\mathbf{k} \times \nabla\phi \tag{7.111}$$

And zero divergent velocity

$$\overline{\phi}\nabla \cdot \mathbf{V} = 0 \tag{7.112}$$

The last two equations are simply diagnostic and they do not contain time-dependent fields.

### 7.7.2 Scaling of the Vorticity and Divergence Equations

The divergence equation can be obtained by simply taking the divergence of the equation of motion (7.104) and putting $\delta = \nabla \cdot \mathbf{V}$. We get

$$\frac{\partial\delta}{\partial t} + \nabla \cdot (\mathbf{V} \cdot \nabla\mathbf{V}) + \nabla^2\phi - f\zeta = 0 \tag{7.113}$$

While the vorticity equation

$$\frac{\partial\zeta}{\partial t} + \mathbf{V} \cdot \nabla\zeta + (f + \zeta)\,\nabla \cdot \mathbf{V} = 0 \tag{7.114}$$

We have shown in several places that the geostrophic wind is a good approximation so that it is useful to develop a set of equations which use this result. We divide the wind velocity in the *rotational* and *divergent* part as follows:

$$\mathbf{V} = \mathbf{V}_\psi + \mathbf{V}_\chi \quad \mathbf{V}_\psi = \mathbf{k} \times \nabla\psi \quad \mathbf{V}_\chi = \nabla\chi \tag{7.115}$$

where $\psi$ is the streamfunction for the nonrotational part of the wind, and $\chi$ is the velocity potential for the divergent part of the wind. Then we get

$$\zeta = \mathbf{k} \cdot \nabla \times \mathbf{V} = \nabla^2\psi \quad \delta = \nabla \cdot \mathbf{V} = \nabla^2\chi \tag{7.116}$$

We will scale the velocity as follows:

$$\mathbf{V}_\psi \sim V \quad V_\chi \sim R_1 V \tag{7.117}$$

We start from the continuity equation:

$$\begin{gathered}\frac{\partial \phi'}{\partial t} + \mathbf{V}_\psi \cdot \nabla \phi' + \mathbf{V}_\chi \cdot \nabla \phi' + \phi' \nabla \cdot \mathbf{V}_\chi + \overline{\phi} \nabla \cdot \mathbf{V}_\chi = 0 \\ RoF\overline{\phi}\frac{V}{L} \quad RoF\overline{\phi}\frac{V}{L} \quad R_1 RoF\overline{\phi}\frac{V}{L} \quad R_1 RoF\overline{\phi}\frac{V}{L} \quad R_1\overline{\phi}\frac{V}{L}\end{gathered} \tag{7.118}$$

For the last term to be balanced, it must be

$$R_1 \leq Ro \tag{7.119}$$

The vorticity equation scaled can be written as

$$\begin{gathered}\frac{\partial \zeta}{\partial t} + \mathbf{V}_\psi \cdot \nabla \zeta + \mathbf{V}_\chi \cdot \nabla \zeta + f\delta + \zeta\delta = 0 \\ \frac{V^2}{L^2} \quad \frac{V^2}{L^2} \quad R_1\frac{V^2}{L^2} \quad \frac{R_1}{Ro}\frac{V^2}{L^2} \quad R_1\frac{V^2}{L^2}\end{gathered} \tag{7.120}$$

where we have used $\zeta \sim V/L$ and $\delta \sim R_1 V/L$. The divergence equation it is scaled as

$$\begin{gathered}\frac{\partial \delta}{\partial t} + \nabla \cdot \left(\mathbf{V}_\psi \cdot \nabla \mathbf{V}_\psi\right) + \nabla \cdot \left(\mathbf{V}_\psi \cdot \nabla \mathbf{V}_\chi\right) + \nabla \cdot \left(\mathbf{V}_\chi \cdot \nabla \mathbf{V}_\psi\right) \\ R_1\frac{V^2}{L^2} \quad \frac{V^2}{L^2} \quad R_1\frac{V^2}{L^2} \quad R_1\frac{V^2}{L^2} \\ + \nabla \cdot \left(\mathbf{V}_\chi \cdot \nabla \mathbf{V}_\psi\right) + \nabla^2 \phi' - f\zeta = 0 \\ R_1^2\frac{V^2}{L^2} \quad \frac{1}{Ro}\frac{V^2}{L^2} \quad \frac{1}{Ro}\frac{V^2}{L^2}\end{gathered} \tag{7.121}$$

From Eq. (7.120) we see that

$$R_1 \leq Ro \tag{7.122}$$

The ratio between divergence and vorticity can now be written as

$$\frac{|D|}{|\zeta|} = R_1 \tag{7.123}$$

In the examples these results will be shown to be compatible with the Rossby wave solution. Finally we can arrive at the simplified solution. We assume $F \sim 1$ and

$Ro \ll 1$ so that based on (7.122) and (7.11) we have $R_1 = Ro$. Now in the scaled equations, we drop all terms of the order Ro or smaller and we get the simplified equations:

$$\frac{\partial \phi'}{\partial t} + \mathbf{V}_\psi \cdot \nabla \phi' + \overline{\phi}\delta = 0 \tag{7.124}$$

$$\frac{\partial \zeta}{\partial t} + \mathbf{V}_\psi \cdot \nabla \zeta + f\delta = 0 \tag{7.125}$$

$$\nabla^2 \phi' - f\zeta = 0 \tag{7.126}$$

The last of these equations gives the solution for the streamfunction $\psi = \phi'/f$, $\zeta = \nabla^2 \psi$, and consequently the rotational part of the wind:

$$\mathbf{V}_\psi = \mathbf{k} \times \nabla \psi = (1/f)\, \mathbf{k} \times \nabla \phi' \tag{7.127}$$

which is the geostrophic wind. Equations (7.124, 7.125, 7.126, and 7.127) are called the quasi-geostrophic equations because they use the quasi-geostrophy in all terms except in the divergence which is zero. If the divergence is eliminated between (7.124) and (7.125), we get

$$\left(\frac{\partial}{\partial t} + \mathbf{V}_\psi \cdot \nabla\right)\left[\zeta - \left(\frac{f}{\overline{\phi}}\right)\phi'\right] = 0 \tag{7.128}$$

This is the quasi-geostrophic potential vorticity equation and it will show that this coincides with Eq. (7.44).

A diagnostic equation for the divergence can be obtained from Eqs. (7.124 and 7.125); this equation becomes

$$\nabla^2 \delta - \left(\frac{f^2}{\overline{\phi}}\right)\delta = \left(\frac{1}{f\overline{\phi}}\right)\mathbf{k} \times \nabla \phi' \cdot \nabla\left(\nabla^2 \phi'\right) \tag{7.129}$$

A reasonable value for F for cyclone scale motions is $L = 10^6 \mathrm{m}, f = 10^{-4} \mathrm{s}^{-1}$ and $\overline{\phi}^{1/2} \sim 300\ \mathrm{ms}^{-1}$ is $F = 0.1$. However if $F \sim Ro \ll 1$, then (7.119) gives $R_1 = Ro^2$, and the vorticity Eq. (7.120) becomes

$$\frac{\partial \zeta}{\partial t} + \mathbf{V}_\psi \cdot \nabla \zeta = 0 \tag{7.130}$$

which is the nondivergent barotropic vorticity equation. The result is again the geostrophy of the wind.

## E.7 Examples

### *E.7.1 Ertel Potential Vorticity in a Barotropic Fluid*

Suppose we have a two-dimensional motion in $x$ and $y$, while the vertical motions are absent

$$w = \frac{Dz}{Dt} = 0 \tag{E.7.1}$$

When the fluid is barotropic we are free to choose $\lambda$ (Eq. (7.35)) to be the coordinate $z$ in which case the potential vorticity is simply

$$\frac{\omega_a \cdot \nabla\lambda}{\rho} = \frac{\omega_a \cdot \nabla z}{\rho} = \frac{\omega_a \cdot \mathbf{k}}{\rho} = \frac{\zeta_a}{\rho} = \frac{\zeta + f}{\rho} \tag{E.7.2}$$

And the quantity will be conserved as absolute vorticity.

### *E.7.2 Conservation of Potential Vorticity*

We now consider a shallow layer of a homogeneous fluid. Because the density is constant, the total divergence is zero:

$$\nabla \cdot \mathbf{V} = 0 \tag{E.7.3}$$

At the upper surface ($z = h$), we have

$$w = \frac{Dh}{Dt} \tag{E.7.4}$$

While at the bottom ($h = h_b$), the velocity must be zero:

$$w = \frac{Dh_b}{Dt} = u\frac{\partial h_b}{\partial x} + v\frac{\partial h_b}{\partial y} \tag{E.7.5}$$

The ratio between the vertical and horizontal velocity will be of the same order of magnitude of $H/L$, with $H$ and $L$ vertical and horizontal scale, respectively. In this approximation we can assume the horizontal velocity to be independent of $z$ so that Eq. (E.7.3) can be integrated to give

$$w = -z\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + A\,(x, y, t) \tag{E.7.6}$$

where $A(x, y, t)$ is a function to be determined. This can be done by applying the boundary condition at the bottom (E.7.5)

$$w = -(z - h_b)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + u\frac{\partial h_b}{\partial x} + v\frac{\partial h_b}{\partial y} \tag{E.7.7}$$

And at the top (E.7.4)

$$w = \frac{Dh}{Dt} - (h - h_b)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + u\frac{\partial h_b}{\partial x} + v\frac{\partial h_b}{\partial y} \tag{E.7.8}$$

And with $H = h - h_b$ we have

$$\begin{aligned} &\frac{Dh}{Dt} = -H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + u\frac{\partial h_b}{\partial x} + v\frac{\partial h_b}{\partial y} \\ &\frac{Dh}{Dt} - \frac{Dh_b}{Dt} = -H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \quad \Rightarrow \quad \frac{DH}{Dt} + H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0 \end{aligned} \tag{E.7.9}$$

The last equation is equivalent to

$$\frac{\partial H}{\partial t} + \frac{\partial}{\partial x}(uH) + \frac{\partial}{\partial y}(vH) = 0 \tag{E.7.10}$$

Equation (E.7.9) combined with (E.7.7) gives

$$w = \frac{(z - h_b)}{H}\frac{DH}{Dt} + \mathbf{v} \cdot \nabla h_b \tag{E.7.11}$$

We now introduce the scalar quantity:

$$\lambda = \frac{z - h_b}{H} \tag{E.7.12}$$

The rate of change of this quantity is zero:

$$\frac{D\lambda}{Dt} = \frac{w}{H} - \frac{1}{H}\frac{Dh_b}{Dt} - \frac{1}{H^2}(z - h_b)\frac{DH}{Dt} = 0$$

So that is actually a "tracer." The relative vorticity can be written explicitly:

$$\omega = \mathbf{i}\left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right) + \mathbf{j}\left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right) + \mathbf{k}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)$$

The hypothesis is that the horizontal velocities are independent of $z$ so that the first two terms can be neglected because they involve only horizontal derivative of $w$. The conservation of Ertel potential voracity becomes

$$Q = \zeta_a \mathbf{k} \cdot \nabla \frac{z - h_b}{H} = \frac{\zeta_a}{H} = \frac{\zeta + f}{H}$$

This is actually the potential vorticity we have introduced before.

### *E.7.3 Scaling and Vorticity Inversion*

Consider the conservation of potential vorticity in the shallow water equation written in the form

$$Q = \frac{\zeta + f}{H + h} \tag{E.7.13}$$

where $H$ and $h$ are the mean and perturbation fluid depth, respectively. Considering that $h/H \ll 1$ we may write

$$Q = \frac{\zeta + f}{H\,(1 + h/H)} \approx \frac{(\zeta + f)}{H}\left(1 - \frac{h}{H}\right)$$

And if we expand the Coriolis parameter

$$f = f_o + \beta y$$

we get

$$QH = \zeta + f_0 + \beta y - f_0 h/H \tag{E.7.14}$$

The inversion can be obtained introducing a streamfunction $\psi$ so that

$$u = -\frac{\partial \psi}{\partial y} = -\frac{g}{f}\frac{\partial h}{\partial y} \quad v = \frac{\partial \psi}{\partial x} = \frac{g}{f}\frac{\partial h}{\partial x}$$

so that $\psi = gh/f$ and $\zeta = \nabla^2 \psi$ and Eq. (E.7.14) becomes

$$QH = \nabla^2 \psi + f_0 + \beta y - f_0 \psi / H$$

The quantity

$$q = QH - f_0 = \nabla^2 \psi + \beta y - f_0 \psi / H \tag{E.7.15}$$

is materially conserved, that is,

$$\frac{D}{Dt}\left[\nabla^2\psi + \beta y - f_0\psi/H\right] = 0$$

Different assumptions go in to this equation. First of all we have neglected the relative vorticity with respect to f (small Rossby number). We have also assumed that $h \ll H$, that is, the deformation scale is the same as the scale of the motion.

### *E.7.4 Rossby Waves in Shallow Water*

Consider the equation of motions and the continuity equation (Gill 1982)

$$\begin{aligned}
&\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + g\frac{\partial h}{\partial x} = 0\\
&\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + g\frac{\partial h}{\partial y} = 0\\
&\frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y} - h\left(\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right)
\end{aligned} \tag{E.7.16}$$

We assume that the basic state provides a velocity $\bar{u}$ from a deformation of the height $H$:

$$\overline{u} = -\frac{g}{f}\frac{\partial H}{\partial y} \tag{E.7.17}$$

We also assume that the velocity is perturbed as well as the height, $h$:

$$u = \overline{u} + u', \quad v = v' \quad h = H + h'$$

And this perturbation depends only on y. We get easily

$$\begin{aligned}
&\frac{\partial u}{\partial t} + \overline{u}\frac{\partial u}{\partial x} - fv + g\frac{\partial h}{\partial x} = 0\\
&\frac{\partial v}{\partial t} + \overline{u}\frac{\partial v}{\partial x} + fu = 0\\
&\frac{\partial h}{\partial t} + \overline{u}\frac{\partial h}{\partial x} + H\frac{\partial u}{\partial x} + v\frac{\partial H}{\partial y} = 0
\end{aligned} \tag{E.7.18}$$

where we have dropped terms of second order together with the ($'$). We may assume a solution of the form

$$u = u_0 \exp ik\,(x - ct)\,, \quad v = v_0 \exp ik\,(x - ct)\,, \quad h = h_0 \exp ik\,(x - ct)$$

Substituting in (E.7.17) we get a linear homogeneous system:

$$\begin{aligned} &(\overline{u} - c)\, iku_0 - fv_0 + gikh_0 = 0 \\ &fu_0 + ik\,(\overline{u} - c)\, v_0 = 0 \\ &ikHu_0 + \left(\frac{\partial H}{\partial y}\right) v_0 + ik\,(\overline{u} - c)\, h_0 = 0 \end{aligned} \tag{E.7.19}$$

whose determinant put to zero gives the equation

$$(\overline{u} - c)^3 - \left(gH + \frac{f^2}{k^2}\right)(\overline{u} - c) - \frac{fg}{k^2}\frac{\partial H}{\partial y} = 0$$

In case the current velocity is zero then according to (E.7.18) also $\partial H/\partial y = 0$ and the solutions are

$$c = 0, \quad c = \pm\sqrt{gh + f^2/k^2} \tag{E.7.20}$$

When the Earth is not rotating, the phase velocity of the wave is the same of the shallow water gravity wave. An interesting case is when in the first Eq. (E.7.17) we assume geostrophic equilibrium. In this case applying the same method, we obtain a solution for the phase velocity:

$$c = \overline{u} + \frac{(f/H)\,\partial H/\partial y}{k^2 + (f^2/gH)} \tag{E.7.21}$$

This velocity can be related to the potential vorticity gradient because $\overline{q} = f/H$ so

$$\frac{\partial \overline{q}}{\partial y} = -\frac{1}{H^2}\frac{\partial H}{\partial y} = \frac{f^2 U}{gH^2}$$

And the phase velocity is given by

$$c = \overline{u} - \frac{H\partial \overline{q}/\partial y}{k^2 + (f^2/gH)} \tag{E.7.22}$$

An interesting conclusion is to find in this case the ratio between the divergence and the vorticity. In this case the ratio of these two quantities is very easy to compute:

$$\frac{|D|}{|\zeta|} = \left| \frac{\partial u/\partial y}{\partial v/\partial x} \right|$$

For the inertial gravity waves with $\overline{u} = 0$, we have from the second of (E.7.18) that

$$v = -\frac{ifu}{kc}$$

And using the appropriate expression for $c$, we get

$$\left| \frac{D}{\zeta} \right| = \sqrt{1 + k^2 g H f^{-2}} \tag{E.7.23}$$

This quantity is always greater so that in inertial gravity waves divergence dominates the vorticity.

For the Rossby solution we assume geostrophic equilibrium so that

$$v = \frac{g}{f}\frac{\partial h}{\partial x} \quad \frac{\partial v}{\partial x} = \frac{g}{f}\frac{\partial^2 h}{\partial x^2} = -\frac{gk^2}{f}h \tag{E.7.24}$$

To get the expression for the derivative of the $u$ component, we get over the last two Eqs. (E.7.18)

$$\frac{\overline{u}}{H}\frac{\partial h}{\partial x} + \frac{\partial u}{\partial x} + \frac{v}{H}\frac{\partial H}{\partial y} = 0$$

Deriving with respect to x and substituting $\partial v/\partial x$ from the second of (E.7.17) and $\partial H/\partial y$ from (E.7.17), we get

$$\frac{\partial^2 u}{\partial x^2} - \left(\frac{f^2}{gH}\right) u = \frac{\overline{u}}{H}\frac{\partial^2 h}{\partial x^2} \tag{E.7.25}$$

And again writing a wave solution, we have

$$\frac{\partial u}{\partial x} = \frac{ik\overline{u}gh}{gH + (f^2/k^2)}$$

The ratio of divergence to vorticity is then

$$\left| \frac{D}{\zeta} \right| = \frac{(\overline{u}k/f)}{(k^2 gH/f^2) + 1} \tag{E.7.26}$$

The numerator of this expression is the Rossby number so that in a Rossby wave the ratio between divergence to vorticity is less than the Rossby number as shown by (7.123).

### *E.7.5 Flow Over an Obstacle: The Numerical Solution*

The program that follows solves the potential vorticity equation in the non-stationary case. The program refers to a domain $3200 \times 1600$ km centered at 45° latitude north with an obstacle at the center that has a form of a spherical dome with a radius of 400 km and a height of 2 km. The equation to be solved is of the form

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = 0 \tag{E.7.27}$$

where the quantities **U** and **F**(**U**) are matrices as we have seen in the shallow water equation. The first step in solving (E.7.27) is to transform it in a difference equation that means to introduce a spatial grid and a difference in time. To simplify the notation, we will forget about the vector form and indicate with $U_i^n$ the variable calculated at the *n-th* time in the *i-th* point. A similar notation will be used also for $F$. Equation (E.7.27) becomes

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} = -\frac{F_{i+1}^n - F_{i-1}^n}{2\Delta x}$$

And from this equation the value of the variables at the time step $n+1$ can be calculated explicitly

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{2\Delta x}\left(F_{i+1}^n - F_{i-1}^n\right) \tag{E.7.28}$$

It can be shown that such an equation needs a very short time step to be solved and in any case may be unstable. For example, if $U$ is a concentration, it may become negative. A quite efficient method, known as Lax–Wendroff, is based on approximating the solution to Eq. (E.7.28) with a Taylor series stopped to the second order:

$$U_j^{n+1} = U_j^n + \Delta t\left(\frac{\partial U}{\partial t}\right)_j^n + \frac{\Delta t^2}{2}\left(\frac{\partial^2 U}{\partial t^2}\right)_{,}^n + \cdots \tag{E.7.29}$$

The time derivative can be replaced with space derivative using Eq. (E.7.27):

$$\frac{\partial^2 U}{\partial t^2} = -\frac{\partial}{\partial t}\frac{\partial F}{\partial x} = -\frac{\partial}{\partial x}\frac{\partial F}{\partial t}$$

In particular we may assume the flux to be of the form $F = AU$ so that we obtain

$$\frac{\partial^2 U}{\partial t^2} = -\frac{\partial}{\partial x}\frac{\partial F}{\partial t} = -\frac{\partial}{\partial x}A\frac{\partial U}{\partial t} = \frac{\partial}{\partial x}A\frac{\partial F}{\partial x} \tag{E.7.30}$$

So that in the case of $A$ constant

$$U_j^{n+1} = U_j^n - \tfrac{1}{2}A\tfrac{\Delta t}{\Delta x}\left(U_{j+1}^n - U_{j-1}^n\right) + \tfrac{1}{2}\left(A\tfrac{\Delta t}{\Delta x}\right)^2\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right) \tag{E.7.31}$$

This equation actually is equivalent to a two-step program that can be generalized also in the case that A is not constant. The starting point is the calculation of the intermediate value in the $x$, $t$ space, that is,

$$U_{j+1/2}^{n+1/2} = \tfrac{1}{2}\left(U_{j+1}^n + U_j^n\right) - \tfrac{\Delta t}{2\Delta x}\left(F_{j+1}^n - F_j^n\right) \tag{E.7.31a}$$

The values obtained for $U$ are used to calculate the intermediate flux values

$$F_{j+1/2}^{n+1/2},\ F_{j+1/2}^{n+1/2}$$

that are used to obtained the final values for the variable $U$

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x}\left(F_{j+1/2}^{n+1/2} - F_{j-1/2}^{n+1/2}\right) \tag{E.7.31b}$$

We can easily see that in case $F = AU$, substitution of Eq. (E.7.31a) in Eq. (E.7.31b) gives back Equation (E.7.31).

This method suggested in the original paper by Kasahara is used in the program that is in the attached CD. For the notation the interested reader may go back to the original paper which does not have the program.

## References[1]

### *Books*

Brown RA (1991) Fluid mechanics of the atmosphere. Academic, San Diego
Gerald CF, Wheatley PO (1989) Applied numerical analysis. Addison-Wesley, Reading
Gill AE (1982) Atmosphere–ocean dynamics. Academic, London
Holton JR (1992) An introduction to dynamic meteorology. Academic, New York
Houghton JT (1977) The physics of the atmospheres. Cambridge University Press, Cambridge/New York
Lindzen RS (1990) Dynamics in atmospheric physics. Cambridge University Press, Cambridge/New York

[1] Again classical textbooks are the basis for this chapter, like Bluestein, Holton, Gill Houghton, and Pedlosky. Some controversy about flow over an obstacle can be settled by going through Kasahara's paper.

Oort AH (1983) Global atmospheric circulation statistics, 1958–1973. NOAA Professional paper 14, U.S. Printing Office
Pedlosky J (1987) Geophysical fluid dynamics. Springer, New York
Raymond DJ (2012) Lecture Notes. Physics Department and Geophysical Research Center, New Mexico Tech, Socorro
Vallis GK (2006) Atmospheric and oceanic fluid dynamics. Cambridge University Press, Cambridge

## *Articles*

Kasahara A (1966) The dynamical influence of orography on the large scale motion of the atmosphere. J Atmos Sci 23:259
Vallis GK (1996) Potential vorticity inversion and balanced equations of motion for rotating and stratified flows. Q J R Meteorol Soc 122:291

# Chapter 8
# The Planetary Boundary Layer

It is natural to think that the atmosphere rotates with the same velocity as the solid Earth although it is not at all obvious why this should happen. In general, we could say that the Earth drags the atmosphere because of the friction between the surface and the atmospheric layers near it. A very important role is accomplished by the orography with the mountain chains which constitute one of the major points of resistance. Again, it is not possible to make any generalizations because, as we have seen already, the mountain range may be where high- and low-pressure zones are located on the opposite sides, so that the drag must take into account the meteorological situation. The layer that directly interacts with the surface is called the *planetary boundary layer* and, except for the "planetary" part, is a very familiar concept for people interested in fluid dynamics.

## 8.1 Turbulence and Diffusion

For those students interested in starting from scratch, we recommend reading the second volume of *The Feynman Lectures on Physics*, where the last chapter is *The Flow of Wet Water*. Actually, when the molecules of a fluid interact with a solid surface, those directly in contact are "at rest," and mathematically this translates into the "no-slip condition." At some distance from the surface, the fluid will move freely, and the region where the transition takes place from zero velocity to free movement is called the *boundary layer*. To go to zero on the surface, the velocity must decrease from the upper limit of the boundary layer downward; the surface acts as a sink for the momentum. As we have seen in Sect. 5.2.4, the viscous stress has the dimension of pressure and is proportional to the velocity gradient according to the relation

$$\tau = \mu \frac{\partial v}{\partial z}$$



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_8

The viscosity that appears in this relation is the molecular viscosity and thus we talk about the *laminar boundary layer*. In the scale analysis we have performed in Sect. 6.3, we have neglected the viscous forces, but now we can examine things with a little more detail. The force per unit mass is given by

$$\nu \frac{\partial^2 v}{\partial z^2}$$

so that the ratio between the acceleration and the viscous resistance is given by the so-called Reynolds number. In terms of scale analysis, we have

$$R_e = \frac{U^2}{L} \frac{L^2}{\nu U} = \frac{UL}{\nu} \tag{8.1}$$

According to the Feynman book (even if in a different context), this number establishes a transition between laminar and turbulent flow. It is rather difficult to give a definition for turbulence and it is simpler to describe what looks like. For example, the fundamental characteristic is that the fluid motion is rather irregular and the deterministic methods are no longer applicable. From this derives the use of statistical methods, and the most recent approach is based on the chaos theories. Another characteristic of the turbulence is its capability to diffuse and to transport in this way momentum, heat, and mass. Based on these transport characteristics, it is possible to give a more physically based interpretation of the passage from the laminar to the turbulent regime. To accomplish this, we need to make again a temporary detour and talk about *heat diffusion.*

In a gas, heat can be transported only by the molecular diffusion. In that case a diffusion law, obtained from the energy conservation, regulates the temperature distribution within the gas

$$\frac{\partial}{\partial t} \left( \rho C_p T \right) = -\nabla \cdot F_h \tag{8.2}$$

where $F_h$ indicates the heat flux that in case of molecular diffusion is simply proportional to the temperature gradient through a diffusion coefficient $D'$

$$F_h = -D' \nabla T \tag{8.3}$$

This substituted into Eq. (8.2) gives the *diffusion equation*

$$\frac{\partial T}{\partial t} = \nu \nabla^2 T \tag{8.4}$$

where the cinematic viscosity $\nu$ has been substituted by the molecular diffusivity. The substitution is justified by the fact that the coefficients have the same order of magnitude.

We follow now an example reported by Tennekes and Lumley. Suppose we are in a room with dimension of the order of $L$. In this room we have a heat source (like a radiator), and, when the air in the room is at rest, heat is transferred only by the molecular diffusion. A simple dimensional analysis of Eq. (8.4) gives

$$\frac{\nabla T}{t_m} = \nu \frac{\nabla T}{L^2} \tag{8.5}$$

So we have a characteristic time for diffusion

$$t_m = \frac{L^2}{\nu} \tag{8.6}$$

which is the time it would take for the heat to be transported across the room. If we assume $L$ be of the order of 5 m and $\nu = 2{\cdot}10^{-5}$ m$^2$ s$^{-1}$, we obtain a characteristic time of 10 days, which is rather disappointing (ah! Physics applied to home appliances!) so that some faster mechanism must exist to transport heat.

A possible alternative mechanism is that the radiator may heat the air in contact with it by conduction. This air for buoyancy, in turn, will produce a circulation in the room that for a characteristic velocity $U$ and dimension $L$ will have a characteristic time

$$t_t = \frac{L}{U} \tag{8.7}$$

The evaluation of $U$ can be made in a very simple way. For a temperature change $\Delta T$, the acceleration of the air parcel is of the order of $g\Delta T/T$, and if the heated layer has a thickness, then the energy gained per unit mass in the layer will be $gh\Delta T/T$. We assume $\Delta T = 10$ °C e $h = 10$ m and we have an energy of the order of 0.03 m$^2$ s$^{-2}$ from which we get a velocity of about 0.24 m s$^{-1}$. The average velocity actually may be only a few centimeters per second, and nevertheless the characteristic time will be only a few minutes. It is easy to see then how the Reynolds number is just the ratio between the diffusion time due to the turbulence and the time due to the molecular diffusion:

$$R_e = \frac{t_m}{t_t} = \frac{UL}{\nu} \tag{8.8}$$

In the example just made, this number is of the order of 60,000. The transition between laminar and turbulent flux happens when the Reynolds number becomes larger than unity.

The reason for this can be seen in another way. The gas in contact with the surface due to the no-slip condition is at rest. The molecules reach gradually the velocity of the fluid through a shear in a condition of free flux. The velocity shear is a source of vorticity and is diffused in the fluid only if the molecular diffusion time

is shorter than the turbulent time ($t_m < t_t$). When this condition is no longer satisfied, the vorticity remains confined in the shear layer that then detaches and generates turbulence

In analogy we can introduce at this time a turbulent diffusion coefficient $K$, such that the Reynolds number is the ratio between the two coefficients. In this approximation

$$K = UL \tag{8.9}$$

It is rather interesting at this point to translate all these dimensional considerations to the atmosphere, where we need to consider the effects of the Coriolis acceleration. The characteristic time in this case is of the order of $f^{-1}$. If the molecular diffusion were responsible for the boundary layer, the thickness of it would be

$$H_m^2 \approx \frac{\nu}{f} \tag{8.10}$$

which corresponds at middle latitudes to $H_m \approx 0.4$ m. On the other hand, if $\nu$ is substituted by $K$ we get

$$H_t^2 \approx \frac{K}{f} \approx \frac{LU}{f} \tag{8.11}$$

where $L \approx U/f$, so that

$$H_t \approx \left(\frac{U}{f}\right) \tag{8.12}$$

At this point we need to establish the value for $U$, that is, the average velocity generated by the turbulence. Usually it is assumed that such velocity is about 1/30 of the wind velocity, so that for a wind of 10 m s$^{-1}$ we obtain a thickness of about 3 km. The observed value is about 1 km and so we are roughly in the right range. We need to consider however the uncertainties in the diffusion coefficient. As we will see in a while, the estimation for the coefficient 1–5 m$^2$ s$^{-1}$ so that the lower limit for the thickness of the layer is about 200 m.

The thickness of the boundary layer is somewhat related to the altitude where the wind coincides with the geostrophic value. We can imagine that if there is a resistance to the motion, the Coriolis force is no longer able to equilibrate the pressure gradient. The wind direction will no longer be parallel to the isobars but will tend to rotate toward the low-pressure center. This tendency will be more definite near the surface and will attenuate near the top of the boundary layer. It is time then to apply the things we learned on the friction at Sect. 5.2.

## 8.2 Turbulent Friction

The acceleration due to friction is given by

$$\mathbf{F}_r = \nu \frac{\partial^2 \mathbf{V}}{\partial z^2} \tag{8.13}$$

The first thing we notice from this relation is that, if the shear is constant (i.e., the velocity changes linearly with height), the friction acceleration is zero. The coefficient has been taken to coincide with the cinematic viscosity, but actually it must be substituted with something more complex like $K$. This is what we plan to do now: to elaborate a simple theory on how to determine the coefficient based on an elementary vision of the turbulence.

We start from the notion that the velocity vector can be separated in a time-averaged value $\langle \mathbf{V} \rangle$ and a fluctuation from this value $\mathbf{V}'$ that should represent the small-scale turbulence. Without going too deep we can assume that the characteristic times on which $\langle \mathbf{V} \rangle$ changes are very long with respect to those on which $\mathbf{V}'$ is averaged. So we have

$$\mathbf{V} = \langle \mathbf{V} \rangle + \mathbf{V}' \quad \text{with} \quad \langle \mathbf{V}' \rangle = 0 \tag{8.14}$$

The quantities representing deviation from the temporal average in meteorology are indicated with a prime and are called "eddy." We will discuss them a great deal in Chap. 10. The simplest way to see how eddies have some relation with the viscous stresses is to write the equations of motion taking into account the fact that each component of the velocity can be written as an averaged value to which we add a fluctuation. The net effect on the acceleration term will depend on the correlation that results between eddy quantities of the different components:

$$\left[\frac{\partial}{\partial t} + \left(\langle u \rangle + u'\right)\frac{\partial}{\partial x} + \left(\langle v \rangle + v'\right)\frac{\partial}{\partial y} + \left(\langle w \rangle + w'\right)\frac{\partial}{\partial z}\right]\left(\langle u \rangle + u'\right)$$
$$-f\left(\langle v \rangle + v'\right) = -\frac{1}{\rho}\frac{\partial \langle p \rangle}{\partial x} - \frac{1}{\rho}\frac{\partial \langle p' \rangle}{\partial x} + F_x \tag{8.15}$$

where with $F_x$ we have indicated the component of the viscous forces along $x$. We can now average (8.15) to obtain

$$\frac{\partial \langle u \rangle}{\partial t} + \langle u \rangle \frac{\partial \langle u \rangle}{\partial x} + \langle v \rangle \frac{\partial \langle u \rangle}{\partial y} + \langle w \rangle \frac{\partial \langle u \rangle}{\partial z} - f \langle v \rangle$$
$$= -\frac{1}{\rho}\frac{\partial \langle p \rangle}{\partial x} + \nu \nabla^2 \langle u \rangle - u'\frac{\partial u'}{\partial x} - v'\frac{\partial u'}{\partial y} - w'\frac{\partial u'}{\partial z} \tag{8.16}$$

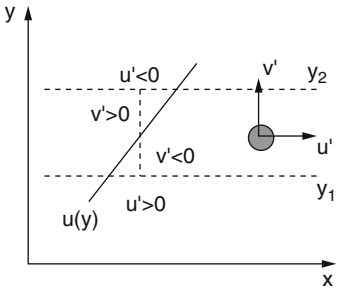



**Fig. 8.1** Illustration of a correlation term. With the air parcel shown, we can associate a momentum flux of the $u'$ component in the direction $y$ (Pedlosky 1987)

In this operation, we have eliminated all terms in which an eddy appears isolated, and we have suspended any judgment on the last three terms that we can write as

$$-\frac{\partial u'u'}{\partial x}+u'\frac{\partial u'}{\partial x}-\frac{\partial u'v'}{\partial y}+u'\frac{\partial v'}{\partial y}-\frac{\partial u'w'}{\partial z}+u'\frac{\partial w'}{\partial z} \tag{8.17}$$

So that considering that the divergence must be zero also for the eddy components, Eq. (8.16) becomes, carrying out the average on the last three terms,

$$\begin{aligned}&\frac{\partial\langle u\rangle}{\partial t}+\langle u\rangle\frac{\partial\langle u\rangle}{\partial x}+\langle v\rangle\frac{\partial\langle u\rangle}{\partial y}+\langle w\rangle\frac{\partial\langle u\rangle}{\partial z}-f\langle v\rangle\\&\quad=-\frac{1}{\rho}\frac{\partial\langle p\rangle}{\partial x}+v\nabla^2\langle u\rangle-\frac{\partial\langle u'u'\rangle}{\partial x}-\frac{\partial\langle u'v'\rangle}{\partial y}-\frac{\partial\langle u'w'\rangle}{\partial z}\end{aligned} \tag{8.18}$$

To understand how the correlation terms must be interpreted, we may refer to Fig. 8.1. The quantity $\langle u'v'\rangle$ represents the momentum flux in the $y$ direction (i.e., the $x$ component of momentum is transported in the normal direction). If this flux increases with $y$, then, in the region between $y_1$ and $y_2$, the average velocity will decrease in the same region, that is, if the result is

$$\frac{\partial\langle u'v'\rangle}{\partial y}>0 \quad \text{then} \quad \frac{\partial\langle u\rangle}{\partial t}<0$$

We see then that the function of the correlation terms is the same as the friction terms because they can actually slow down the motion. The real difference is that the same terms can also accelerate the motion if they provide momentum, so that we are in presence of a negative viscosity.

The eddy terms present a problem that is known as closure. In practice, we have written equations for the changes of average quantities expressed as a function of eddy terms. To solve these equations, we have to find a relation between the correlation terms and the average quantities. Figure 8.1 gives a hint as this is where we have also shown the shear of the zonal wind. If the air parcel moves in the positive $y$ direction ($v'>0$) and the wind shear is positive

$$\frac{\partial \langle u \rangle}{\partial y} > 0$$

the air parcel will end in a region with higher zonal velocity so that $u' < 0$. The opposite happens for $v' < 0$, and this means that the term $\langle u'v' \rangle$ is always negative, so that in the presence of a shear, we always have

$$\frac{\langle u'v' \rangle \, \partial \langle u \rangle}{\partial y} < 0 \tag{8.19}$$

If we change the sign of the wind shear, it is easy to see $\langle u'v' \rangle > 0$, and the shear being negative, the sign of Eq. (8.19) does not change. We can then assume that the momentum transport is essentially proportional to the wind shear and, remembering the notation for the stress, we have

$$\tau_{zx} = -\rho \langle u'w' \rangle = \mu_e \frac{\partial \langle u \rangle}{\partial z} \quad \tau_{zy} = -\rho \langle v'w' \rangle = \mu_e \frac{\partial \langle v \rangle}{\partial z} \tag{8.20}$$

So that the accelerations will be

$$\begin{aligned} F_x &= \frac{1}{\rho} \frac{\partial \tau_{zx}}{\partial z} = \frac{1}{\rho} \frac{\partial}{\partial z} \rho \langle u'w' \rangle = \frac{1}{\rho} \frac{\partial}{\partial z} \mu_{ex} \frac{\partial \langle u \rangle}{\partial z} \\ F_y &= \frac{1}{\rho} \frac{\partial \tau_{zy}}{\partial z} = \frac{1}{\rho} \frac{\partial}{\partial z} \rho \langle v'w' \rangle = \frac{1}{\rho} \frac{\partial}{\partial z} \mu_{ey} \frac{\partial \langle v \rangle}{\partial z} \end{aligned} \tag{8.21}$$

where with $\mu_{ex}$ and $\mu_{ey}$ we have indicated the eddy viscosities which are analogous to the molecular viscosity, only with much higher values. At this point, before going to the applications of the above relations, we have to understand a little better this viscosity.

### *8.2.1 The Mixing Length*

An eddy can be imagined as a region of the fluid with roughly the same properties as the momentum, for example. After drifting for some distance, this region of the fluid will mix with another region and will exchange some properties with it. At this point the analogy with the molecular diffusion is complete, considering that the molecules exchange momentum through collisions and the diffusion coefficient is somewhat related to the average velocity and the mean free path.

Suppose then, for analogy, that an eddy moves upward by a certain amount $\xi'$ before mixing. We can expect that the deviations on the velocity are given by

$$u' = -\xi' \frac{\partial \langle u \rangle}{\partial z} \quad v' = -\xi' \frac{\partial \langle v \rangle}{\partial z} \tag{8.22}$$

where $\xi' > 0$ is for upward movements and vice versa for downward movements. Based on these relations the flux of zonal momentum in the vertical direction is given by

$$\langle u'w' \rangle = -\langle w'\xi' \rangle \frac{\partial \langle u \rangle}{\partial z} \tag{8.23}$$

The problem remains for the estimation of $w'$ in terms of averaged quantities. One hypothesis in this case is that the turbulence is isotropic, meaning that the deviations from the mean of the three components are of the same order of magnitude:

$$u' \approx v' \approx w' \tag{8.24}$$

Also we assume that they are of the same order of the average value. These assumptions allow us to write the vertical component $w'$ as in Eq. (8.22) so that

$$w' \approx -\xi' \left| \frac{\partial \langle \mathbf{V} \rangle}{\partial z} \right|$$

where with $|\mathbf{V}|$ we have indicated the average of the horizontal total wind. Substituting in Eq. (8.23), we have

$$\langle u'w' \rangle = -\left\langle \xi'^2 \right\rangle \left| \frac{\partial \langle \mathbf{V} \rangle}{\partial z} \right| \frac{\partial \langle u \rangle}{\partial z} = -K_e \frac{\partial \langle u \rangle}{\partial z} \tag{8.25}$$

which is of the same form as Eq. (8.20). The coefficient $K_e$ has the same function as a cinematic viscosity and can be expressed as

$$K_e = \left\langle \xi'^2 \right\rangle \left| \frac{\partial \langle \mathbf{V} \rangle}{\partial z} \right| = \langle l^2 \rangle \left| \frac{\partial \langle \mathbf{V} \rangle}{\partial z} \right|$$

where $l$ is called the mixing length and is given by

$$l = \left( \left\langle \xi'^2 \right\rangle \right)^{1/2} \tag{8.26}$$

This derivation, which may look too formal, does not give any clue as to how to calculate $K_e$, but simply says that the larger the dimension of the eddy and the wind shear, the greater the diffusion coefficient.

Before closing this part it is worthwhile to mention how we expect the turbulence to be generated in the boundary layer. Mainly responsible is the vertical shear, which produces vorticity and then mixing, at least on the vertical plane. This mixing implies energy dissipation at the expense of the horizontal flux. A criterion to establish under what conditions the turbulence may arise is to compare the potential and kinetic energy gained by the parcel due to the vertical acceleration. For a deviation $w'$ we associate a kinetic energy per unit mass given by

$$\frac{1}{2}w'^2 \approx \frac{1}{2}l^2\left(\frac{\mathrm{d}\langle u\rangle}{\mathrm{d}z}\right)^2$$

If the parcel moves an altitude interval $\Delta z$, the energy gained due to the buoyancy acceleration is

$$g\frac{\Delta\theta}{\theta}\Delta z$$

These two energies will be equal when

$$\frac{1}{2}\left(\frac{l}{\Delta z}\right)^2 \approx \frac{g}{\theta}\frac{d\theta}{dz}\Big/\left(\frac{du}{dz}\right)^2 \tag{8.27}$$

The left-hand side of this equation represents the *Richardson number*:

$$Ri = \frac{g}{\theta}\frac{d\theta}{dz}\Big/\left(\frac{du}{dz}\right)^2 \tag{8.28}$$

We notice that the Richardson number is always positive for a stable atmosphere. Comparing Eqs. (8.27) and (8.28), we see that turbulence will be likely when the change in altitude is larger than the mixing length. The condition is then $Ri < 1$ and from empirical data the onset of turbulence is assumed when the Richardson number is less than 0.25.

## 8.3 The Surface Layer

We generically talked about the planetary boundary layer and gave the impression that it is a relatively simple structure. Actually it is a rather complicated object and to understand this we may refer to the thermal structure of the boundary layer. Figure 8.2 shows qualitatively the time evolution of the boundary layer in terms of changing temperature profile. We start at noon when the heating of the surface is such that the convective mechanism has invaded all the boundary layer, so that the *mixed layer* is a well-mixed region. It may grow by entraining air from the region above the boundary layer, that is, the free atmosphere. This layer reaches the maximum extension at noon. In the afternoon the strength of the convection decreases, and after sunset the turbulence in the mixed layer decays, giving rise to the *residual layer* which is a somewhat weakened version of the mixed layer. As the night progresses, the bottom part of the residual layer is transformed into a stable layer. This is characterized by a stable temperature inversion that suppresses the turbulence. The wind may grow from negligible to geostrophic values at the top of the boundary layer and the resulting strong shear may produce an occasional turbulence burst in the residual layer.

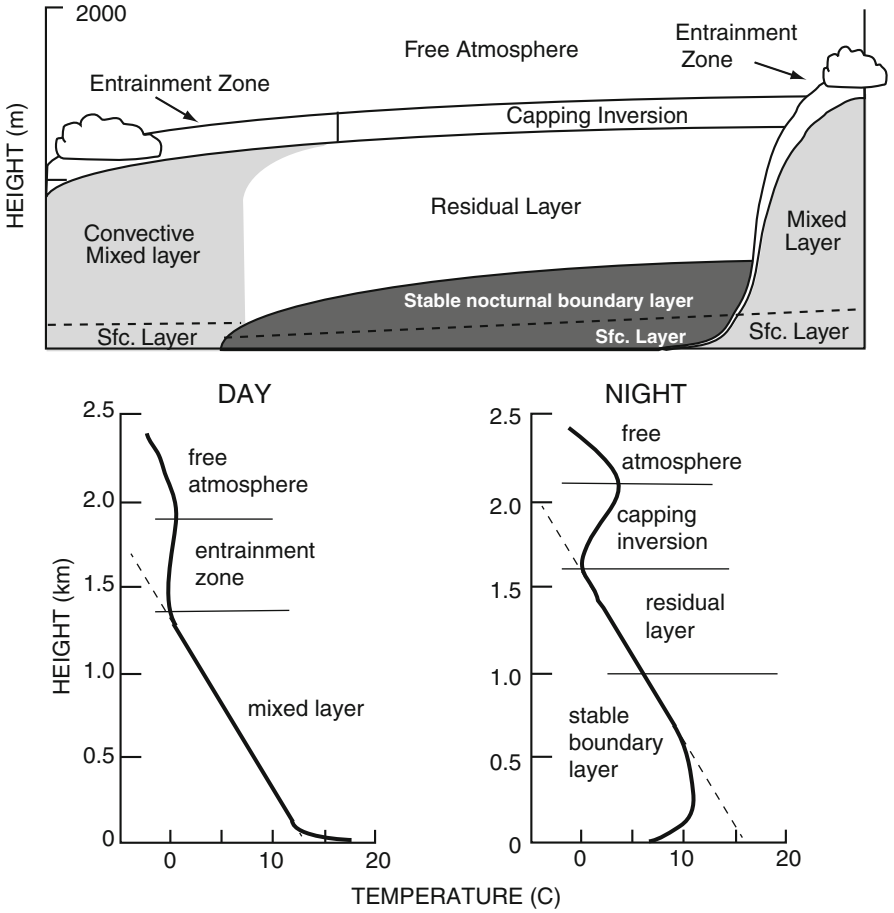



**Fig. 8.2** The *top* of the figure shows the qualitative evolution of the boundary layer. At the *bottom* two temperature profiles are shown, representative of the conditions at noon and at night (From Stull 1988)

The bottom part of Fig. 8.2 shows two schematic temperature profiles referring to daytime (early afternoon) and to the nighttime. One thing we can notice is that the atmospheric temperature at the bottom of the layer is strongly influenced by the surface temperature. During the day, the surface is warmer than the atmosphere, so that the first layer is heated by conduction and the temperature reaches gradually the mixed layer values. During the night, the surface cools and again by conduction the atmospheric temperature decreases toward the surface. In the first case the bottom layer is rather unstable, while at night we may have a temperature inversion and then a very stable situation. It is assumed that the bottom 10 % of the mixed layer may be involved in to these phenomena and to this region the name*surface layer* is given.

For this layer at the surface, we can assume a scale velocity $u_*$, simply defined on a dimensional basis and given by

$$u_* = \left(\frac{\tau}{\rho}\right)^{1/2} \tag{8.29}$$

This velocity is also called friction velocity and $\tau$ is the appropriate stress. From the definition in Eq. (8.20) we have

$$u_*^2 = \left|\langle u'w' \rangle\right|$$

so that

$$u_*^2 = K_e \frac{\partial \langle u \rangle}{\partial z} \tag{8.30}$$

At this point we need the mixing length theory to estimate the diffusion coefficient. The simplest assumption is to take the mixing length proportional to the height according to $l \approx kz$. In this case we have

$$K_e = (kz)^2 \frac{\partial \langle u \rangle}{\partial z}$$

Substituting in Eq. (8.30), we have

$$\frac{\partial \langle u \rangle}{\partial z} = \frac{u_*}{kz} \tag{8.31}$$

which can be integrated to give a logarithmic law for the change of the velocity with height

$$\langle u \rangle = \left(\frac{u_*}{k}\right) \ln\left(\frac{z}{z_0}\right) \tag{8.32}$$

The quantity $z_0$ is called the *roughness length* and depends on the height where $\langle u \rangle$ is zero, while $k$ is called the von Karman constant, from the name of one of the founding fathers of aerodynamics. The value of the roughness length depends on the roughness of the surface and may change from 0.0002 for the sea surface to values greater than 2 at the center of large cities.

The relation of Eq. (8.31) can be very useful for calculating the wind at higher altitude when it is known at the height of the anemometer. For example, if we imagine to measure a wind 5 m $s^{-1}$ at 10 m over a field with roughness length 0.5 m, we can find from Eq. (8.32) a wind of 6.5 m $s^{-1}$ at 250 m which is a typical height of a smokestack. We will see in a while that such simple calculations are rather important in a first rough calculation, of the distribution of the effluents from a smokestack: again very earthly matters for the atmospheric physicist.

A very interesting consequence of Eq. (8.31) is that if we multiply both sides by $u_*$ and take into account Eq. (8.30), we get for the stress

$$\tau = \rho \left(k u_* z\right) \frac{\partial \langle u \rangle}{\partial z}$$

This is actually a flux of the horizontal momentum per unit volume in the vertical direction. This relation can be extended to the sensible heat flux simply substituting for the mass per unit volume the heat content per unit volume, so we get

$$H = -\rho C_p \left(k u_* z\right) \frac{\partial \theta}{\partial z} \tag{8.33}$$

This equation can be understood on the basis that the quantity in parenthesis is actually the diffusion coefficient and we are simply writing that the heat flux is proportional to the temperature gradient. Similar to the momentum flux, also the heat flux is constant across the surface layer, so that from Eq. (8.33) we can get the temperature difference between a layer extending between $z_1$ and $z_2$:

$$\Delta\theta = \frac{H}{\rho C_p k u_*} \mathrm{In}\left(\frac{z_2}{z_1}\right) \tag{8.34}$$

We can have an idea of the temperature difference across a layer 1.5 m thick. Assuming $H = 100$ W m$^{-2}$ and $u_* = 0.3$ m s$^{-1}$ we get a difference of about 5 °C.

In a similar way we can obtain the latent heat flux by writing first the water vapor flux:

$$F = -\left(k u_* z\right) \frac{\partial \rho_v}{\partial z}$$

with $\rho_v$ the density of water vapor. If we multiply this by the latent heat $L$, we get the latent heat flux $L_H$:

$$L_H = -L \left(k u_* z\right) \frac{\partial \rho_v}{\partial z} \tag{8.35}$$

We define then the Bowen ratio between the sensible and latent heat flux that we denote with $\beta$:

$$\beta = \frac{H}{L_H} = \frac{\rho C_p}{L} \left(\frac{\partial \theta}{\partial z} \Big/ \frac{\partial \rho_v}{\partial z}\right) \tag{8.36}$$

The latent and sensible heat fluxes are very important in determining the surface energy budget, as we will see later, because they help to dissipate the excess heat of the surface. To simplify matters, the fluxes are parameterized in terms of easily measurable quantities and drag coefficient that can be defined as

$$C_D = k^2 \left[ \ln \left( \frac{z}{z_0} \right) \right]^{-2} \tag{8.37}$$

This corresponds to the assumption of a direct proportion between the velocity and the friction velocity like

$$\langle u^2 \rangle = C_D u_*^2$$

This relation is assumed to be true only at 10 m so that Eq. (8.37) can be used to calculate the drag coefficient once the von Karman constant and the roughness length are known.

We will treat in detail the surface heat balance later. We simply note that without the contribution of the sensible and heat fluxes the surface of the Earth would be much hotter.

## 8.4 The Ekman Layer

Above the surface layer we find a mixed layer where the wind should go from the values we have just found to the geostrophic values. This region is also called the *Ekman layer*. To study this layer, we can start from Eq. (8.21) and write again the condition for zero acceleration, when the acting forces are due to the pressure gradients, the Coriolis term, and the friction. We have immediately

$$\begin{aligned} fv - \frac{1}{\rho}\frac{\partial p}{\partial x} + K_e \frac{d^2 u}{dz^2} &= 0 \\ fu + \frac{1}{\rho}\frac{\partial p}{\partial y} - K_e \frac{d^2 v}{dz^2} &= 0 \end{aligned} \tag{8.38}$$

where for simplicity we have omitted the average parentheses. The pressure gradients can be substituted by the values of the geostrophic wind so we obtain

$$\begin{aligned} K_e \frac{d^2 u}{dz^2} + f\left(v - v_g\right) &= 0 \\ K_e \frac{d^2 v}{dz^2} - f\left(u - u_g\right) &= 0 \end{aligned} \tag{8.39}$$

We may assume the wind does not change with altitude (barotropic atmosphere), and multiplying the second by $i = \sqrt{-1}$ and then adding, these equations can be solved. We get

$$K_e \frac{d^2 (u + iv)}{dz^2} - if\,(u + iv) = -if\left(u_g + iv_g\right)$$

This has the solution

$$(u + iv) = A \exp\left[(if/K_e)^{1/2} z\right] + B \exp\left[-(if/K_e)^{1/2} z\right] + u_g + iv_g \tag{8.40}$$

The integration constants can be determined by imposing the conditions

$$u = v = 0 \quad \text{for} \quad z = 0$$

$$u \to u_g; \quad v \to v_g \quad \text{for} \quad z \to \infty$$

For simplicity we may assume the geostrophic wind only in the $x$ direction so that $v_g = 0$ and we have

$$(u + iv) = -u_g \exp\left[-\left(\frac{f}{2K_e}\right)^{1/2} (1 + i)\, z\right] + u_g \tag{8.41}$$

We can then equate real and imaginary parts on both sides:

$$u = u_g \left(1 - \mathrm{e}^{-\gamma z} \cos \gamma z\right); \quad v = u_g \mathrm{e}^{-\gamma z} \mathrm{sen} \gamma z \tag{8.42}$$

with $\gamma = (f/2K_e)^{1/2}$. This solution is known as the *Ekman spiral* and can be understood by referringto Fig. 8.3. This figure shows the curve that connects all the tips of the velocity vector whose components are given by Eq. (8.42). The vector rotates clockwise until it coincides with the geostrophic wind at the top of the boundary layer. What is happening is that, for a constant pressure gradient, the friction slows down the air parcel so that the Coriolis acceleration no longer balances

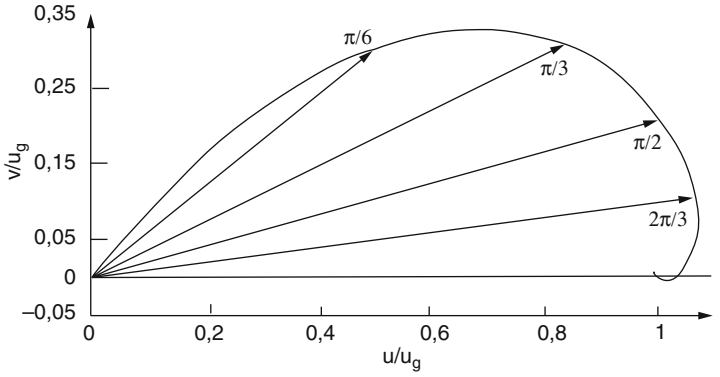



**Fig. 8.3** The Ekman spiral. The velocity vector is shown at different altitudes that can be obtained by the argument of the trigonometric functions in Eq. (8.42). Altitude is decreasing from the larger to the smaller arguments

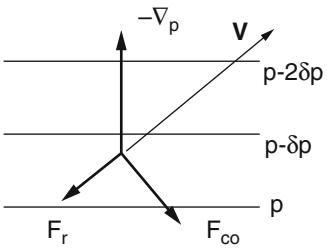



**Fig. 8.4** The equilibrium in the boundary layer between pressure $\nabla p$, Coriolis $F_{co}$, and friction forces $F_r$

the acceleration due to pressure. This balance is shown in detail in Fig. 8.4, where we see that now the resultant between the friction and the Coriolis forces balances the pressure gradient force. The result, quite evident from Figs. 8.3 and 8.4, is that within the Ekman layer there is a wind component in the direction of the pressure gradient and this means that the wind is no longer geostrophic. In these conditions, we can estimate the vertical extension of the Ekman layer. From Eq. (8.42), we see that the conditions $u = u_g$ and $v = 0$ are satisfied when the altitude has as a value $D_e$ such that

$$D_e\gamma = D_e\left(\frac{f}{2K_e}\right)^{1/2} = \pi$$

that gives the thickness of the Ekman layer $D_e$. With typical values for $f$ at middle latitude and with $K_e \approx \mathrm{m}^2\ \mathrm{s}^{-1}$ we obtain $D_e = 1$ km. The value of the diffusion coefficient is consistent with Eq. (8.26) because for a vertical shear of 5 m $\mathrm{s}^{-1}$ $\mathrm{km}^{-1}$ we obtain a mixing length of $l \approx 30$ m that is small with respect to the depth of the boundary layer.

The simple theory we have illustrated however does not take into account the surface layer or the layer that extends from the surface to the anemometer level, where the wind velocity is given by Eq. (8.32). We may interpret (8.42) in a slightly different way, that is, that at the anemometer level, the wind has two components, the geostrophic one and the ageostrophic component:

$$u_a = -u_g e^{-\gamma z}\cos\gamma z$$

This must be the total wind velocity at the top of the surface layer. The general solution for the entire boundary layer can be obtained by introducing a complex notation for the velocity vector:

$$W = (u + iv) - \left(u_g + iv_g\right) \tag{8.43}$$

With this notation, Eq. (8.39) reduces to

$$K_e = \frac{\partial^2 W}{\partial z^2} - ifW = 0 \tag{8.44}$$

with the general solution

$$W = C_1 e^{-(1+i)\gamma z} + C_2 e^{(1+i)\gamma z} \tag{8.45}$$

If we denote with $h$ the thickness of the surface layer and with $W(h)$ the velocity at the top of it, then

$$W = W(h) \exp\left[-(1+i)\,\gamma\,(z-h)\right] \tag{8.46}$$

For continuity, the stress at the base of the Ekman layer must be equal to the surface stress. At the base of the Ekman layer, the stress is parallel to the wind shear:

$$\frac{\partial\,(u+iv)}{\partial z} = \frac{\partial W}{\partial z}$$

so that it must be parallel to $(u + iv)$ at the top of the surface layer. Differentiating Eq. (8.46) and putting $W(h) = W_h$, we have

$$\left(\frac{\partial W}{\partial z}\right)_h = -W_h\gamma\,(1+i) = W_h\gamma 2^{1/2} e^{-3i\pi/4} \tag{8.47}$$

This relation shows that at the base of the Ekman layer the angle between the shear and the wind is 135°. The meaning of the relation between the geostrophic wind and the wind in the surface layer can be clarified by looking at Fig. 8.5. In this figure the geostrophic wind is directed along the $x$ axis. According to Eq. (8.43) the wind $W$ is the difference between the wind in the surface layer and the geostrophic wind. The direction of the wind in the surface layer is then constant and its intensity changes according to Eq. (8.32). We need to determine now the value for $\alpha_s$. We refer to Fig. 8.4 and apply a simple trigonometric rule we find

$$\frac{|W_h|}{\sin\alpha_s} \Rightarrow |u+iv|_h = 2^{1/2} V_g \sin\left(\frac{\pi}{4} - \alpha_s\right) = V_g\,(\cos\alpha_s - \sin\alpha_s)$$

From the same figure and the previous relations we have

$$W_h = 2^{1/2} V_g \sin\alpha_s \exp i\left(\frac{3\pi}{4} + \alpha_s\right)$$

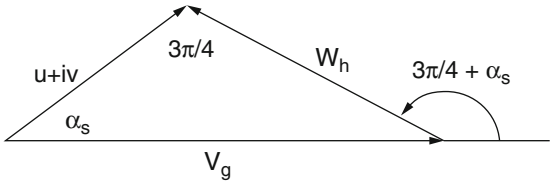



**Fig. 8.5** The vector relationship between the geostrophic wind and the wind in the surface layer. *Symbols* are the same as used in the text (Bluestein 1992)

that, once substituted in Eq. (8.43), gives the solution

$$u + iv = V_g \left(1 + 2^{1/2} \sin \alpha_s e^{-\gamma(z-h)} e^{i[(3\pi/4)+\alpha-\gamma(z-h)]}\right) \quad z \geq h \tag{8.48}$$

It is evident that this equation coincides with the one found for the Ekman spiral when $\alpha_s = -3\pi/4$. In the surface layer we must have, according to Eq. (8.33)

$$u + iv = \frac{u_*}{k} \ln \frac{z}{z_0} e^{i\alpha_s} \quad z \leq h \tag{8.49}$$

The complex exponential has been introduced to take into account the phase. Because the two vectors must coincide for $z = h$, we must have

$$\frac{u_*}{k} \ln \frac{h}{z_0} = V_g (\cos \alpha_s - \sin \alpha_s) \tag{8.50}$$

Also the diffusion coefficients must be the same at $z = h$

$$K_e = k u_* h \tag{8.51}$$

And the last condition to be satisfied is that the stress must be the same at the boundary:

$$|\tau| = \left|\rho K_e \frac{\partial (u + iv)}{\partial z}\right| = \left|\rho K_e \frac{\partial W}{\partial z}\right| = 2^{1/2} \rho K_e |W_h| \gamma \tag{8.52}$$

Substituting for

$$|W_h| = 2^{1/2} V_g \sin \alpha_s$$

and for $\gamma$ and putting $|\tau| = \rho u_*^2$, we have the relation

$$(2fK_2)^{1/2} V_g \sin \alpha_s = u_*^2 \tag{8.53}$$

The quantities that characterize the boundary layer are now

$$V_g,\ u_*, h, \alpha_s, K_e, z_0$$

However, the equations are not enough to determine all of these. It is clear that $V_g$, $h$, and $z_0$ are obtained from observations or from specific characteristics of the surface. Once these data are known, Eqs. (8.50), (8.51), and (8.53) can be used to determine the remaining quantities.

## 8.5 The Secondary Circulation

One of the most important results of the previous section is that the wind within the boundary layer has a component that remains to the left of the geostrophic wind. That means that, if the geostrophic wind is caused by a pressure difference, the circulation in the boundary layer is roughly that shown in Fig. 8.6. We notice that there is a divergence around the high-pressure zone and the convergence around the low-pressure zone. These motions on the horizontal plane correspond to a vertical motion, like the one shown in Fig. 8.7. It is clear that the divergent part corresponds to upward motion in the boundary layer and vice versa. For continuity reasons we will find divergence above the low-pressure zone and convergence above the high-pressure zone.

As it is schematically shown in the figure normal to the geostrophic wind, there will be a mass flux that we can evaluate through the integral extended to the boundary layer:

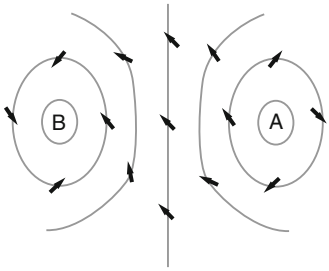



**Fig. 8.6** A sketch of the circulation in the boundary layer around high- and low-pressure zones; notice the difference with respect to the geostrophic wind

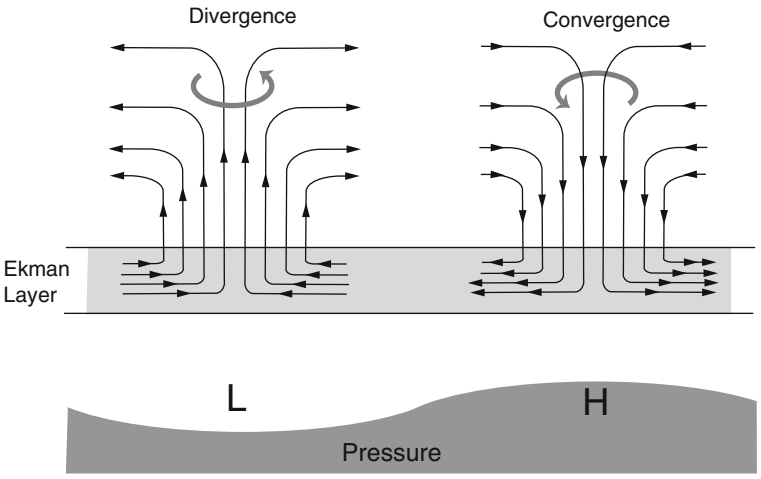



**Fig. 8.7** A sketch of the circulation in the boundary layer around high- and low-pressure zones, in the vertical plane. The pressure is shown schematically below. A convergence (divergence) in the Ekman layer corresponds a divergence (convergence) in the free atmosphere (From Holton 1992)

$$M = \int_0^{D_e} \rho v \, dz = \int_0^{D_e} \rho u_g \mathrm{e}^{-\pi z/D_e} \sin\left(\frac{\pi z}{D_e}\right) dz \tag{8.54}$$

We have substituted for the $y$ component of the velocity its value in the Ekman layer. This equation can be put in relation with the vertical velocity if we recur to the continuity equation:

$$\frac{\partial}{\partial z}(\rho w) = -\frac{\partial}{\partial x}(\rho u) - \frac{\partial}{\partial y}(\rho v)$$

In this relation, $u$ is independent from $x$ so that substituting in Eq. (8.54) and noting that $w = 0$, for $z = 0$ we obtain, assuming a constant density

$$(\rho w)_{D_e} = -\frac{\partial}{\partial y}\int_0^{D_e} \rho u_g \mathrm{e}^{-\pi z/D_e} \sin(\pi z/D_e)\, dz = -\frac{\partial M}{\partial y} \tag{8.55}$$

The vorticity in the Ekman layer

$$\frac{\partial u_g}{\partial y} = -\zeta_g$$

is assumed constant so that the integral of Eq. (8.54) has the value $(1 + e^{-\pi})\, D_e/2\pi$ and the vertical velocity at the top of the boundary layer becomes

$$w_{D_e} = \zeta_g (K_e/2f)^{1/2} \tag{8.56}$$

where we have assumed $(1 + e^{-\pi}) \approx 1$.

From Fig. 8.7, we can see that in the Ekman layer a secondary circulation is generated that damps the vorticity around the lows and highs. This can be easily seen if we conserve the divergence term in the vorticity equation:

$$\frac{D(\zeta + f)}{Dt} = -f\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = -f\frac{\partial w}{\partial z}$$

This equation can be easily integrated from the top of the boundary layer to the top of the tropopause keeping $f$ constant:

$$\int_{D_e}^{H} \left(\frac{D\zeta}{Dt}\right) dz = f\,(w(H) - w\,(D_e)) - fw\,(D_e)$$

where $H$ is the altitude of the tropopause. Substituting the vertical velocity of Eq. (8.56) and assuming the rate of vorticity change constant, we have

$$\frac{D\zeta_g}{Dt} = -\frac{f}{(H - D_e)} w\,(D_e) = -\left(\frac{fK_e}{2H^2}\right)^{1/2} \zeta_g \tag{8.57}$$

This relation gives a decaying time for the vorticity:

$$\tau_e = H\left(\frac{2}{fK_e}\right)^{1/2} \tag{8.58}$$

With $H = 10$ km, $f = 10^{-4}$ s$^{-1}$, and $K_e = 10$ m$^2$ s$^{-1}$, we obtain a characteristic time of about 4 days. That is roughly the lifetime of a cyclone.

The secondary circulation is also a well-known classical problem, like the problem of the leaves in a teacup (or little foam in an espresso), and it useful to clarify further the meaning of the secondary circulation.

### *8.5.1 Spin-Down in a Teacup*

To put in the proper framework this problem, we consider the equation of motion written in vector form:

$$\frac{D\mathbf{V}}{Dt} + 2\Omega \times \mathbf{V} = -\left(\frac{1}{\rho}\right)\nabla p + \nu\nabla^2\mathbf{V} \tag{8.59}$$

In this case, the pressure takes into account all the forces and these are

$$p = P + \phi + \frac{\rho\Omega^2 r^2}{2}$$

where $P$ is the pressure within the fluid. The reference system that we choose is such that the $z$ axis coincides with the rotation axis, while $x$ and $y$ are normal to it. We may assume that at time $t = 0$, the angular velocity on the lateral wall of the cup changes by a small quantity $\varepsilon\Omega$, either in acceleration (spin-up) or deceleration (spin-down). At this point, the pressure forces are no longer balanced by the centrifugal forces so that the fluid starts to move in the radial direction with a velocity of the order of $\varepsilon\Omega r$. This motion happens within a boundary layer whose thickness can be found by equating the Coriolis acceleration to the one produced by viscous forces:

$$2\Omega u = \nu\left(\frac{\partial^2 v}{\partial z^2}\right); \quad -2\Omega v = \nu\left(\frac{\partial^2 u}{\partial z^2}\right) \tag{8.60}$$

where $\nu$ is the cinematic viscosity. To satisfy these equations and considering that in the boundary layer we must have $u \approx v$, the thickness will be of the order

$$d \approx \left(\frac{\nu}{2\Omega}\right)^{1/2}$$

An estimation of the time it takes to establish the boundary layer can be made through a scale analysis of the equation of motion:

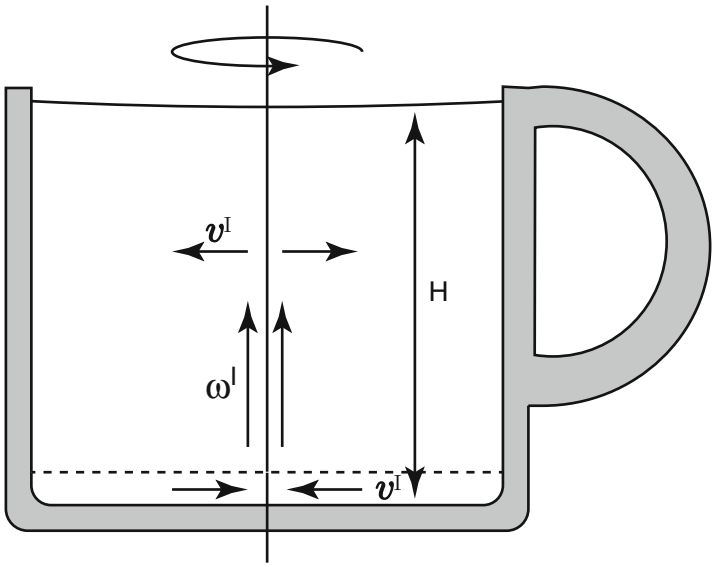



**Fig. 8.8** The secondary circulation in a teacup (Baker 1966, 1968)

$$\left(\frac{dv}{dt}\right) + 2\Omega u = \nu\left(\frac{\partial^2 v}{\partial z^2}\right) v_r \approx \epsilon\Omega r$$

from which we easily find the time to be of the order of $\Omega^{-1}$.

If we consider a teacup, as in Fig. 8.8, using the continuity equation on a radial plane we have

$$\frac{\partial w}{\partial z} = -\frac{\partial v_r}{\partial r}$$

where $v_r$ is the radial velocity. Integrating the above equation across the thickness of the boundary layer, we have

$$w \cong -\int_0^d \left(\frac{\partial v_r}{\partial r}\right) dz \approx \epsilon\Omega\left(\frac{\nu}{2\Omega}\right)^{1/2}$$

Within the fluid the velocities can be found using the continuity equation with the boundary conditions existing at the top and the bottom of the fluid:

$$w^I = -\left(\frac{\partial v_r}{\partial r}\right) z + \cos t$$

We assume that the fluid is contained between the free surface where $z = H$ and the bottom ($z = 0$); thus we have

$$w^I = 0 \quad \text{at} \quad z = H, \quad w^I = \varepsilon\left(\frac{\nu\Omega}{2}\right)^{1/2} \quad \text{at} \quad z = 0$$

from which we can get the constant of integration and the radial velocity gradient:

$$w^I = \epsilon\left(\frac{\nu\Omega}{2}\right)^{1/2}\left(1 - \frac{z}{H}\right) \quad v_r^I = \epsilon\left(\frac{\nu\Omega}{2}\right)^{1/2}\left(\frac{r}{H}\right) \tag{8.61}$$

This result shows that the vertical velocity seems to be independent from the radius

The spin-down time can be obtained from another consideration. If we refer to Fig. 8.8, we can see that the secondary circulation transports fluid with low angular momentum to regions of high angular momentum. If we now consider a ring of fluid of radius $R$ and mass $\delta m$, the angular momentum is $J = \delta m \Omega r^2$. Neglecting the friction within the fluid, the conservation of the angular momentum requires that

$$\frac{\Delta\Omega}{\Omega} = -\frac{2\Delta r}{r} \approx \frac{\varepsilon\Omega}{\Omega}$$

In the spin-down time, the radial distance traveled will be

$$\Delta \mathrm{r} \approx \frac{\varepsilon r}{2} \approx v_r^I \quad \tau_{\mathrm{spin-down}} \approx \varepsilon\left(\frac{\nu\Omega}{2}\right)^{1/2}\left(\frac{r}{H}\right)\tau_{\mathrm{spin-down}}$$

so that the spin-down time will be

$$\tau_{\mathrm{spin-down}} = \frac{H}{(2\nu\Omega)^{1/2}} \tag{8.62}$$

This relation is very similar to that of Eq. (8.58) but has been obtained in a way that can be understood also by students in a general physics course. Naturally, everything we said is valid for an ideal cup where the radial dimension must be much larger than the vertical dimension so that the effect of the friction on the lateral wall can be neglected.

## 8.6 Turbulent Diffusion from Discrete Sources

It may look a bit strange to go from a classical problem like the tea leaves in a cup to a very ugly problem like the plume from a smoke stack. However, as atmospheric physicists, we should not be ashamed by such a variety of interests because this simply confirms as important is our role in society.

A plume of pollutants is formed any time a source, like a chimney, emits continuously. The plume may have different forms and behaviors depending on the environmental conditions of the atmosphere. In the absence or in light winds, the plume is dispersed vertically, while with the wind blowing horizontally, the pollutants are mainly dispersed downwind. Although the problem is not simple, engineers (mainly) have found simple ways to evaluate the concentration of pollutants in the presence of a discrete source. We are going to justify in the next few paragraphs this approach.

### *8.6.1 The Characteristics of Smoke Plumes*

The shape of a smoke plume is very much related to the temperature gradient of the atmosphere and the stability. We can have a very rough idea of the problem simply by looking at Fig. 8.9. In this figure we have on the right the profile of the ambient temperature (that we denote $T_{\text{air}}$) as compared to the lapse rate (presumably adiabatic) of a parcel that is freely moving in the atmosphere. The situations depicted in the figure assume that a constant wind is blowing from left to right. In case (a), the two temperatures coincide and we are in a neutral situation. In this case, the plume is dispersed almost symmetrically downwind and assumes the characteristic form of a cone. In (b), the situation is that of a stable atmosphere. We must remember that we are in the boundary layer and the temperature may increase all the way up to the top of the layer. The smoke from the stack will not diffuse in the vertical direction and will move essentially in the horizontal direction and this corresponds to *fanning*. In case (c) the atmosphere may be unstable below a certain

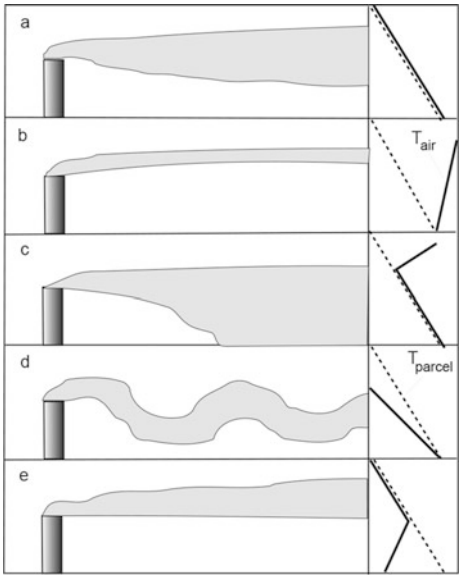



**Fig. 8.9** The various shapes of a smokestack plume. The *right side* of the figure shows the behavior of the temperature in the background atmosphere (*solid line*) and that of the air parcel according to an adiabatic lapse rate (*dashed line*). Case (**a**) corresponds to coning, case (**b**) to fanning, case (**c**) to fumigation, case (**d**) to looping, and case (**e**) to lofting

altitude and stable above. The level of inversion indicates the thickness of the mixed layer. In this case the diffusion will be inhibited above the mixed layer, while it will be very energetic below it and the effluent will reach the ground. This is the situation that corresponds to *fumigation* which is the most dangerous of all because large concentrations of pollutants may result at ground level. Case (d) is in some sense opposite to case (b) because the atmosphere is now unstable in the entire boundary layer. In this case the typical size of an eddy is larger than the width of the plume, with the result that the plume will be moved up and down, assuming the shape of a snake: this is what is called *looping*. Finally, in case (e), which corresponds to *lofting*, the inversion is at the ground up to some level, so that the plume cannot diffuse below and will grow mainly above the inversion. The shape of the plume is important because it may give a rough idea of the concentration of pollutant we should expect. It would be very complicated to evaluate the shape illustrated in Fig. 8.9 because it may depend on stability parameters and on the winds. Also, the boundary layer structure and the wind change during the day, and this adds some more complications. We can, however, calculate some average situation that may not give the exact distribution of the concentration from the sources but would be enough to estimate whether we are within prescribed limits. This average situation may refer also to different environmental conditions.

We may start from a simple one-dimensional problem and assume that the $x$ direction coincides with the downwind direction. Thus, we consider a particle leaving the source and moving in the $x$ direction under the action of turbulence and wind. We may define the probability density function $F(x)$ such that the probability of finding the particle between point $x$ and $x + dx$ is given by $F(x)dx$. If the particle does not disappear, it must be

$$\int_{-\infty}^{+\infty} F(x)dx = 1 \tag{8.63}$$

In the same way, we may assume a probability density in the lateral direction $y$ as $G(y)$ and in the vertical direction $z$, $H(z)$ and because the probabilities are independent, we have

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(x)G(y)H(z)dxdydz = 1 \tag{8.64}$$

We define at this point the concentration of the pollutant as $\chi(x, y, z)$ measured, for example, as mass per unit volume, so that if the strength of the source is $Q$, we have

$$\chi\,(x, y, z) = QF(x)G(y)H(z) \tag{8.65}$$

because the amount of mass in the elementary volume is $\chi(x, y, z)dxdydz$. Consequently, the integral of Eq. (8.74) becomes

$$Q = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \chi\,(x, y, z)\,dxdydz \tag{8.66}$$

This is the case of an instantaneous point source that may not look very realistic, but actually it may correspond to the case in which a pollutant is released accidentally or, for example, from a ruptured vehicle. In this case, the source is not point-like but at the least is of finite size.

Another class of problems has to do with a continuous point source, this being the most common case, like a chimney that emits fumes. This problem can still be included in the previous framework by substituting for the continuous point source a number of discrete instantaneous sources, each one of strength $Q\mathrm{d}t$, where now $Q$ must have different dimensions. We may further assume that in the wind direction there is no spread due to diffusion and the probability function at a distance from the source $ut$ (with $u$ the wind velocity) is simply given by

$$F(x) = \frac{1}{udt} \tag{8.67}$$

And using Eq. (8.75), we have

$$\chi\,(x, y, z) = QdtF(x)G(y)H(z) = \frac{QG(y)H(z)}{udt}dt = \frac{Q}{u}G(y)H(z) \tag{8.68}$$

At this point, in order to solve the problem, we need to find an explicit dependence of the probability functions. This may be very simple if we assume that the spread of the plume is due to a process of Fickian diffusion which we have found in previous chapters (Fig. 8.10).

### *8.6.2 The Gaussian Plume*

When the diffusion flux can be assumed to be proportional to the gradient of the concentration for a constant wind in the $x$ direction, we have from the continuity equation

$$\frac{\partial \chi}{\partial t} + u\frac{\partial \chi}{\partial x} = \frac{\partial}{\partial x}\left(K_x\frac{\partial \chi}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y\frac{\partial \chi}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z\frac{\partial \chi}{\partial z}\right) \tag{8.69}$$

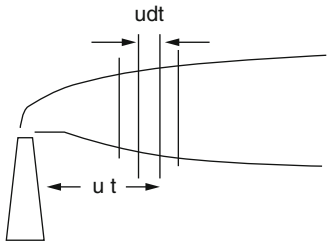



**Fig. 8.10** The approximation of a continuous source with a succession of point sources (From Blackadar 1997)

For a reference system moving with the wind, we have

$$\frac{\partial \chi}{\partial t} = \frac{\partial}{\partial x}\left(K_x \frac{\partial \chi}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial \chi}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial \chi}{\partial z}\right) \tag{8.70}$$

that can be simplified when the eddy diffusion coefficients $K_x, K_y, K_z$ are constant:

$$\frac{\partial \chi}{\partial t} = K_x \frac{\partial^2 \chi}{\partial x^2} + K_y \frac{\partial^2 \chi}{\partial y^2} + K_z \frac{\partial^2 \chi}{\partial z^2}$$

This equation has an analytical solution for the case of an instantaneous point source that releases a total mass $Q$ at time $t = 0$ and at the origin $x = y = z = 0$. In this case, the solution has a form of a Gaussian:

$$\chi = \frac{Q}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \exp\left(-\frac{z^2}{2\sigma_z^2}\right) \tag{8.71}$$

where

$$\sigma_x^2 = 2K_x t; \quad \sigma_y^2 = 2K_y t; \quad \sigma_z^2 = 2K_z t; \tag{8.72}$$

It is easy to see that Eq. (8.81) is in the form of Eq. (8.65) if we put

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right); \quad G(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \tag{8.73}$$

$$H(z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left(-\frac{z^2}{2\sigma_z^2}\right)$$

Once we now have the single probability functions, we can use Eq. (8.68) to find the solution for a continuous source in a mean wind $u$. We have

$$\chi = \frac{Q}{u} G(y)H(z) = \frac{Q}{2\pi u\sigma_y\sigma_z} \exp\left[-\left(\frac{y^2}{2\sigma_y^2} + \frac{z^2}{2\sigma_z^2}\right)\right] \tag{8.74}$$

This equation does not contain explicitly the $x$ dependence. However, it is clear that the width of the plume increases with increasing time and then the corresponding $x$ coordinate is simply given by $x = ut$.

Equation (8.74) is very simple and it can be extended without any particular precaution to the real world: it is actually used by engineers to calculate how much pollutant one may expect from a smokestack. The equation may look the same as Eq. (8.71) but the calculation for $\sigma$,s is a little bit different. The analytical solution gives time dependence for $\sigma$ as $t^n$ where $n = 0.5$. Observations give a range for $n$ between 0.75 and 1. An empirical approach is adopted such that, although the

form of the equation is the same, $\sigma_x$ and $\sigma_y$ are related to the stability of the atmosphere. This is because the $\sigma$ coefficients are proportional to the standard deviation of the horizontal wind direction and this in turn is related to stability. Again, there are empirical approaches that assume a number of stability classes, for the atmosphere according to the wind values and insolation parameters. Once the class is established, tables are used to find the standard deviation as a function of the downwind distance $x$.

We really do not want to spend too much time on this topic, but we need to mention that the source in reality is not at the ground but at the height of the smokestack, $h_c$. To obtain the right formula, we simply change the origin of the $z$ coordinate so that Eq. (8.74) becomes

$$\chi = \frac{Q}{u} G(y)H(z) = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp\left[ -\left( \frac{y^2}{2\sigma_y^2} + \frac{(h_c - z)^2}{2\sigma_z^2} \right) \right] \tag{8.75}$$

This opens another problem because this formula cannot be used for $z < 0$, but it is not clear if it can be used for $z > 0$. Actually, it depends on what we assume is the fate of the particles that reach the ground. If the particles are reflected by the surface, then the vertical flux at $z = 0$ is zero. Then the solution in this case is equivalent to adding a virtual source symmetric with respect to the real one at the point of coordinates 0, 0, $-h_c$. Then Eq. (8.75) becomes

$$\chi = \frac{Q}{u} G(y)H(z) = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp\left[ -\left( \frac{y^2}{2\sigma_y^2} + \frac{(h_c - z)^2}{2\sigma_z^2} + \frac{(h_c + z)^2}{2\sigma_z^2} \right) \right] \tag{8.76}$$

It is then clear that the concentration above the ground is larger in case the particles are reflected.

A last detail is that fumes from smokestacks may be considerably warmer than the surrounding air so that they have positive buoyancy. This may have the effect of lifting the height of the source, and actually, in Eq. (8.76) to the quantity $h_c$, we should add a term like

$$\Delta h = 1.6 F^{-1/3} u^{-1} x_f^{2/3}$$

Again, this is an empirical formula where $F$ is called the buoyancy parameter and is related to the heat released from the stack, the exhaust velocity of the fumes, and the radius of the stack tip. Simple calculations based on this formula show that the effect of buoyancy may be important and even double the height of the stack. A very comprehensive treatment of these problems may be found in the book by G. Kiely.

Once more we have concluded a chapter with everyday-life examples. This is exceptionally good because it means we are dealing with reality. However, we should not get too euphoric because the problem of transport and transformation of pollutants is much more complex than can be contained in an engineering manual. Gradually, but determined, we will try to open the way to a complete treatment of this very important problem.

## E.8 Examples

### *E.8.1 Boundary Layer in the Ocean*

In the ocean, there exists a boundary layer where currents are forced by the winds. In this case, we solve the same equations (8.3) with somewhat different boundary conditions. We impose in the atmosphere ocean interface some stress, while at the bottom of the layer where friction is negligible, we impose the velocity of the current to coincide with the "geostrophic" value:

$$\begin{aligned} z = 0; \quad & K_e \frac{\partial u}{\partial z} = \tau_x, \quad K_e \frac{\partial v}{\partial z} = \tau_y \\ z \to -\infty; \quad & u = u_g, \quad v = v_g \end{aligned} \tag{E.8.1}$$

The solutions with these boundary conditions can be found in a similar way as illustrated before, and they are (Vallis 2006)

$$\begin{aligned} u &= u_g + \frac{\sqrt{2}}{f} \gamma e^{z\gamma} \left[ \tau_x \cos\left(z\gamma - \frac{\pi}{4}\right) - \tau_y \sin\left(z\gamma - \frac{\pi}{4}\right) \right] \\ v &= v_g + \frac{\sqrt{2}}{f} \gamma e^{z\gamma} \left[ \tau_x \sin\left(z\gamma - \frac{\pi}{4}\right) + \tau_y \cos\left(z\gamma - \frac{\pi}{4}\right) \right] \end{aligned} \tag{E.8.2}$$

These relations have a very interesting interpretation. When $z = 0$ the value of $v$ is negative (with $\tau_y = 0$) and this means the initial current is deviated to the right according to the Coriolis force. This is quite different from the atmospheric case when the direction of the wind is determined by a balance between the Coriolis force and the pressure gradient force. In that case the deviation is toward a low-pressure zone. The Ekman spirals in the atmosphere and the ocean are quite symmetric although the values of the winds and currents are obviously not the same (Fig. E.8.1).

When $z$ goes to infinity (toward the bottom of the ocean) the terms in parenthesis go to zero so that the values of the current coincide with the deep current.

### *E.8.2 The Transfer of Sensible and Latent Heat*

After a nostalgic plunge into classical physics we will make an attempt to combine the practice with the theory to see which mechanisms should be able to transfer energy and water from the surface to the atmosphere and vice versa. Robert Dickinson has been a pioneer in this field.

The first step may be to find more practical formulas than Eq. (8.35) and then to evaluate the fluxes. Eq. (8.33) can be read as

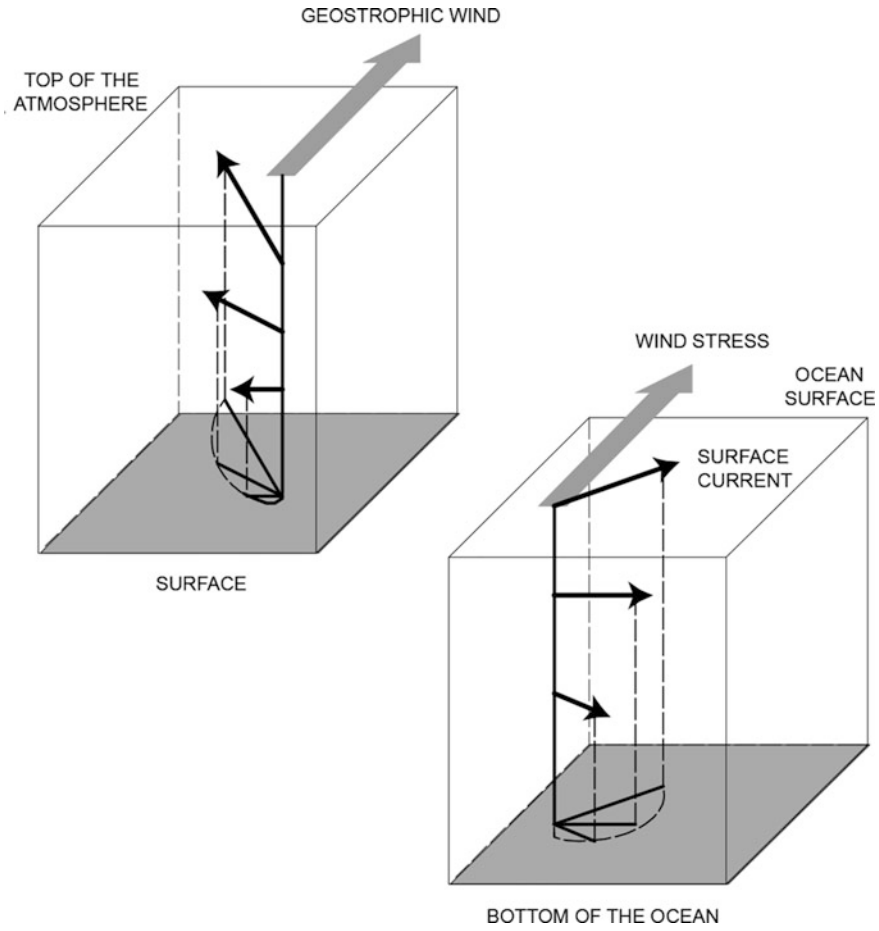



**Fig. E.8.1** The Ekman layer in the atmosphere (*top*) and the ocean (*bottom*). Notice that at the top of the atmosphere the wind coincides with the geostrophic value and goes to zero at the surface. For the ocean the difference between the actual current and its geostrophic value is plotted

$$H = \rho C_p C_{DH} V (T_s - T_a)$$
$$E = \rho C_{DE} V (q_s - q_a) \tag{E.8.3}$$

where, beside the known symbols $C_{DH}$, is the drag coefficient for the heat transfer and $C_{DE}$ is the correspondent for water vapor. $V$ is the wind velocity while $T_a$ and $T_s$ are the air temperature near the surface and the surface temperature; $q_a$ and $q_s$ are the correspondent values for the humidity. As we have shown in (8.37), $C_{DH}\ V$ is the same as $ku_* \ln(z_1/z_2)$. The second part in Eq. (8.63) is actually the flux of water vapor, so that in order to get the latent heat flux, we need to multiply it by the

latent heat. The difference between the soil and air humidity can be assumed to be heat saturation:

$$q_s - q_a \approx \left(\frac{\partial q^{SAT}}{\partial T}\right)(T_s - T_a)$$

To have a rough idea on the numbers we assume $C_{DH} \approx C_{DE} = 2.75 \times 10^{-3}$ so that with a velocity of 6 $\mathrm{ms}^{-1}$ we get

$$\rho C_p C_{DH} V \approx 20 \ \ \mathrm{Wm}^{-2}\mathrm{K}^{-1}$$

We put also

$$B_e^{-1} = L\left(\frac{\partial q^{SAT}}{\partial T}\right)/C_p$$

and show that this is just the inverse of the Bowen ratio that is equal to 1 when $T = 7$ °C. The value for ${B_e}^{-1}$ changes from 5 for tropical hot days to 0.1 for the polar regions.

The total contribution of the latent and sensible fluxes is given by

$$L_H + H = \rho C_p C_{DH} V \left(1 + B_e^{-1}\right)(T_s - T_a) = \lambda_s (T_s - T_a) \tag{E.8.4}$$

where

$$\lambda_s = \rho C_p C_{DH} V \left(1 + B_e^{-1}\right).$$

Through Eq. (E.8.4) we can evaluate in a more realistic way the difference between the temperature of the surface and that of the air. In Chap. 3 we mentioned that the temperature of the surface is determined by the radiative and convective fluxes. At this point, we can specify the convective fluxes to be the sum of the latent and sensible fluxes so that the equilibrium reads

$$L_H + H = F_{Rn} \tag{E.8.5}$$

where $F_{Rn}$ is the net radiative flux given by

$$F_{Rn} = F_{\mathrm{sol}} + F_{IR}^{\downarrow} - \sigma T_s^4$$

with symbols that are self-explanatory.

For a temperature of 25 °C we have ${B_e}^{-1} = 3$ and then $\lambda_s = 80$ W $\mathrm{m}^{-2}$ $\mathrm{K}^{-1}$. The maximum values of the net flux (at mid-latitudes and at the summer solstice) are of the order of 600 W $\mathrm{m}^{-2}$ so that from Eqs. (E.8.5) and (E.8.4) we have $T_\mathrm{s} - T_0 \approx 7.5$ °C. At night the net flux depends on many other parameters like the

surface temperature, cloudiness, and humidity. For a clear summer night, a typical value may be

$$F_{Rn} \approx -0.2\sigma T_s^4 \approx -80 \ \text{W m}^{-2}$$

Thus, the surface is losing energy and the surface temperature then is lower than the atmospheric temperature $T_s - T_0 \approx -1$ °C.

It is interesting at this point to consider another term in the energy budget of the surface and that is the heat lost due to the heat conduction in the soil. The heat flux entering the soil (in W m$^{-2}$) is given by

$$F_{\text{SOIL}} = -\rho_s C_s k_s \left.\frac{\partial T}{\partial z}\right|_{z=0} \tag{E.8.6}$$

where $\rho_s$ and $C_s$ are the density and thermal capacity of the soil while $k_s$ is the thermal diffusivity measured in m$^2$ s$^{-1}$. Solving the diffusion equation it is possible to obtain the temperature gradient

$$\frac{\partial T}{\partial t} = k_s \frac{\partial^2 T}{\partial z^2}$$

We assume a forcing (i.e., the time dependence for the temperature) proportional to $\exp(i\nu t)$ and we obtain the solution

$$T(z) = \overline{T}_s + \left(T_s - \overline{T}_s\right) \exp\left[(-i\nu/k_s)^{1/2} z\right] \tag{E.8.7}$$

where $\overline{T}_s$ is the diurnal averaged value. The flux is then

$$F_{\text{SOIL}} = -\rho_s C_s (k_s \nu)^{1/2} \left(T_s - \overline{T}_s\right) \exp\left(i\pi/4\right) \tag{E.8.8}$$

The meaning of the complex exponential in Eq. (E.8.8) is that the heat flux in the soil leads the temperature maximum by about 3 h. The soil flux has values that depend very much on the soil humidity, which may be between 4 and 20 W m$^{-2}$ for temperature difference of the order of 1 °C.

### *E.8.3 The Fluxes in the Presence of Vegetation*

We cannot resist the temptation to show how wide must be the knowledge of the atmospheric physicist: he must know even about plant physiology.

We can generalize Eq. (E.8.3) by referring to Fig. E.8.2, where the transfer of sensible and latent heat between the soil and the atmosphere happens through "resistances." Actually, the first part of Eq. (E.8.3) can be written as

$$H = \frac{T_s - T_r}{r_a} \rho C_p \tag{E.8.9}$$

where $T_s$ and $T_r$ are the soil and the air temperature, respectively. We also set

$$r_a^{-1} = C_{DH} V$$

which is a real aerodynamic resistance and is actually the same term as in Eq. (E.8.4). As for the latent heat transfer, the potential difference is again proportional to the difference between the saturation pressure at the soil temperature and the vapor pressure of the air $e_r$. However in this case, the resistance is taken to be variable so that the latent heat flux becomes

$$L_H = \beta \left[ \frac{e_s^* (T_s) - e_r}{r_a} \right] \frac{\rho C_p}{\gamma} \tag{E.8.10}$$

where $\gamma$ is an appropriate dimensional conversion . The function of $\beta$ is that of establishing what portion of the soil humidity is actually available. If we refer to Fig. E.8.2 we notice that in this case we adopt the so-called bucket model. In practice, we assume that the soil contains a certain amount of water (measured as mass for unit surface) corresponding to $W_{\max}$. If, for example, the quantity of water coming from precipitation $W$ is larger than this maximum value, the difference results as runoff. The variation introduced with $\beta$ is such that it is zero for $W = 0$ and increases linearly with increasing $W$ up to the value 1 for $W = W_{\max}$. Actually this model does not take into account the vegetation that may cover the soil. In Fig. E.8.3 the vegetation canopy is sketched as an enlarged stomata. The idea is that now the transfer is a two-step process, from soil to canopy and from canopy to the atmosphere. In this way it is possible to take into account a number of factors that include the specific type of vegetation and the fraction of the soil covered by it. The conductance $g_s$ of a single leaf to the water or vapor passage is a rather complicated function that can be written as

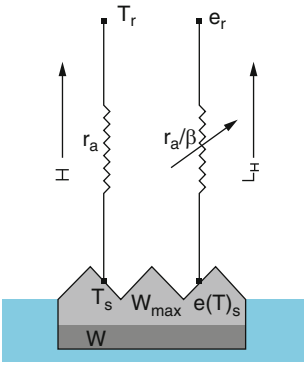



**Fig. E.8.2** Transfer of latent and sensible heat between soil and atmosphere (Dickinson 1983; Sellers 1992)

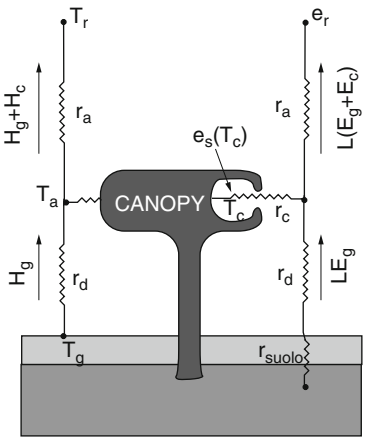



**Fig. E.8.3** The transfer of water and heat between the soil and the atmosphere in the presence of vegetation (Dickinson 1983; Sellers 1992)

$$g_s = g_s\,(\mathrm{PAR})\,[f\,(\delta_e)\,f(T)f\,(\Psi_l)] \tag{E.8.11}$$

In this equation $g_s$(PAR) is the only part of the conductance that is a function of the quantity of light received for photosynthesis (PAR = photosynthetically active radiation), while the other term depends on the vapor deficiency, the temperature of the leaf and the hydraulic potential. The dependence of the conductance on the amount of light is easily understood because it must be somewhat obvious that the transfer of water vapor is facilitated in the presence of radiation.

There are few experiments to establish the validity of an empirical formula like Eq. (E.8.11). However actually the procedure is to find an average value for the canopy conductance $g_c = {r_c}^{-1}$ in such a way that in Eq. (E.8.10) the aerodynamic effects can be completely separated from those of the water vapor transfer through the vegetation. The simplest way is to write

$$LE = \beta \left[ \frac{e_s^*\,(T_s) - e_r}{r_a + r_c} \right] \frac{\rho C_p}{\gamma} \tag{E.8.12}$$

Typical values are $r_a \approx 10$ s m$^{-1}$ and $r_c \approx 100$ s m$^{-1}$ so that in normal conditions and in the presence of vegetation, the latent heat fluxes are much smaller than those given by Eq. (8.70). This simply states something we knew and that is the vegetation holds the humidity of the soil. The total resistance increases as we have seen before because now the transfer of water is a two-step process.

We have simply made a very short introduction to the complex problem of the fluxes between the surface and the atmosphere. We have completely neglected the role of carbon dioxide which is important for the radiative transfer but also for the photosynthetic processes. We need also to think about all the problems related to the capture of rain by the vegetation and the transfer of solar radiation within the canopy.

Naturally there are many scientists now that are interested in the interaction of the vegetation with the atmosphere. Thus we hope than during springtime, when you look at those beautiful green hills, you will have the impression that the vegetation it is just the skin through which the Earth breathes. This could be the most rewarding result after reading this chapter.

### *E.8.4 The Kolmogorov Spectrum*

We have already mentioned the parallelism between eddies and colliding molecules in a gas with the difference that eddies play the same role much more efficiently. The example is the mixing of sugar in a cup of coffee. If we let the sugar melt and mix only by molecular motions, it may take forever, while if we use a spoon to agitate the coffee, the mixing will be much faster. The Kolmogorov idea is that the energy of the stirring is "cascaded" to smaller and smaller eddies until the molecular scale is reached. The energy is indicated with $\varepsilon$ so that it has the dimension of

$$\frac{\text{Energy}}{\text{Time}} = \text{m}^2\text{s}^{-3} \tag{E.8.13}$$

This apparently is independent of time because $\varepsilon$ could be defined as the energy per unit mass and unit time.

Then we want to study how the energy is distributed at different scales so we assume that the energy spectrum take the form

$$E(k) = g\,(\varepsilon, k) \tag{E.8.14}$$

With k being the wave number with the condition that

$$\int E(k)\mathrm{d}k = \overline{E} \tag{E.8.15}$$

We see $E(k)$ must have the dimension of $\text{m}^3\text{s}^{-2}$ being energy per unit wave number. From (E.8.14), it follows that because $k$ has no time dependence, the only way to represent the time dependence of $E(k)$ is a term like $(\varepsilon)^{2/3}$. Then it is very simple to get (Fig. E.8.4):

$$E(k) = K\varepsilon^{2/3}k^{-5/3} \tag{E.8.16}$$

An equivalent way to derive the same result is to introduce the turnover $\tau_k$ defined as the time it takes for a parcel with velocity $v_k$ to traverse a distance $1/k$. From dimensional considerations

$$v_k = [E\,(k)\,k]^{1/2}$$

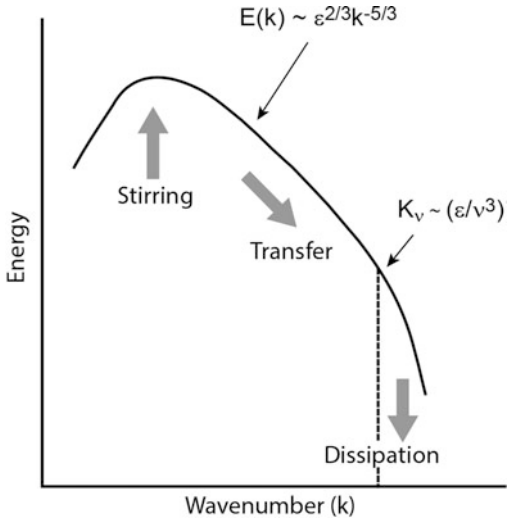



**Fig. E.8.4** The qualitative Kolmogorov spectrum. Energy is supplied at rate $e$ and is cascaded at the smaller scales

so we get

$$\tau_k = [v_k k]^{-1} = \left[k^3 E(k)\right]^{-1/2} = k^{-2/3}\epsilon^{-1/3} \tag{E.8.17}$$

The dissipation rate is then

$$\varepsilon \sim v_k^2/\tau_k = kE(k)/\tau_k \tag{E.8.18}$$

And using (E.8.17) we get (E.8.16).

With the same qualitative argument, it is possible to establish at which spatial scale the viscosity dissipation becomes important. The viscosity dissipation term is of the order of (with $\nu$ viscosity)

$$\nu\nabla^2\mathbf{u} \approx \nu U/L^2 \approx \nu U k^2 \approx U/\tau_k^{\nu}$$

And we get for the characteristic turnover time

$$\tau_k^{\nu} \sim \frac{1}{k^2\nu} \tag{E.8.19}$$

Equating this value to Eq. (E.8.17), we get the dissipation wave number $k_\nu$

$$k_\nu \sim \left(\varepsilon/\nu^3\right)^{1/4}, \quad L_\nu \sim \left(\nu^3/\varepsilon\right)^{1/4} \tag{E.8.20}$$

It is interesting to find the energy dissipation rate given by

$$\dot{E} = \frac{1}{V}\int \nu\mathbf{v}\cdot\nabla^2\mathbf{v}dV \approx \nu k_\nu^2 v_{k_\nu}^2 \tag{E.8.21}$$

where the velocity $v_k^2 \sim \varepsilon^{2/3} k^{-2/3}$ is from $v_k = [E(k)\, k]^{1/2}$. We get

$$\dot{E} \sim \nu k_\nu^2 \varepsilon^{2/3} k_\nu^{-2/3} \sim \varepsilon \tag{E.8.22}$$

One we have substituted $k_\nu$ from (E.8.20). The surprising result is that the dissipation rate is independent of the viscosity. This means that even if viscosity tends to zero, the dissipation rate remains constant. However, if we look at the dissipation length scale given by (E.8.20), we see that this decreases with decreasing viscosity so that the scale at which dissipation becomes important becomes smaller and smaller. In the atmospheric boundary layer, the horizontal scale is of the order of 100 m and the velocity fluctuations are of the order of $10^{-2}$ m so that the dissipation rate is roughly $U^3$/L and that is $10^{-6}/10^2 \times 10^{-8}$ $m^2 s^{-3}$. If we use (E.8.20), we have a dissipation scale of roughly 3 mm with a viscosity of $10^{-6}$ $m^2 s^{-1}$. We have followed Vallis (2006).

## References[1]

### *Books*

Bluestein HB (1992) Synoptic – dynamic meteorology in midlatitudes, vol 1: Principles of kinematics and dynamics. Oxford University Press, New York

Blackadar AK (1997) Turbulence and diffusion in the atmosphere. Springer, Berlin/New York

Dickinson RE (1983) Land surfaces processes and climate. In: Saltzmann B (ed) Theory of climate. Academic, New York

Holton JR (1992) An introduction to dynamic meteorology. Academic, New York

Pedlosky J (1987) Geophysical fluid dynamics. Springer, New York

Sellers PJ (1992) Biophysical models of land surface processes. In: Trenberth KE (ed) Climate system modeling. Cambridge University Press, Cambridge

Stull RB (1988) An introduction to boundary layer meteorology. Kluwer, Dordrecht/Boston

Tennekes H, Lumley JL (1972) A first course in turbulence. MIT Press, Cambridge, MA

Vallis GK (2006) Atmospheric and oceanic fluid dynamics. Cambridge University Press, Cambridge

### *Articles*

Baker JD (1966) Demonstrations of fluid flow in a rotating system. Am J Phys 34:67

Baker JD (1968) Demonstrations of fluid flow in a rotating system II: the spin up problem. Am J Phys 36:980

Tandon A, Marshall J (2010) Einstein's tea leaves and pressure systems in the atmosphere. Phys Teach 48:292

[1]This chapter is essentially the same of the previous edition. The very specialized books by Blackadar and Stull are the main references.

# Chapter 9
# Aerosols and Clouds

Through this entire book, we will discuss aerosols and clouds. We started in Chap. 2 when we dealt with condensation or supersaturation. The water vapor, in order to condense, has to find a support that may be wettable and on which an initial raindrop can grow. In the absence of such a support, condensation is very unlikely to occur. Most of the time, the support is provided by a tiny particle that may have a different chemical composition and to which we refer to as condensation nuclei. These particles are so small as to be invisible and represent the lower end in terms of dimension of the atmospheric aerosols. Strictly speaking, these include also raindrops and other liquid (or ice) particles. A good point to start then is to consider the entire population of aerosols and drops, as shown in Fig. 9.1.

## 9.1 Sources of Atmospheric Aerosols

Aerosols are a very important component of the atmosphere. Besides the strong interaction they have with clouds, they may influence atmospheric chemistry and the amount of radiation absorbed by the atmosphere and reaching the ground.

It is interesting to know which are the main sources of aerosols in the atmosphere, as summarized in Table 9.1 taken from Hobbs. This table refers only to tropospheric aerosols and groups the sources in two large classes. The spatial source refers mainly to particles produced in the atmosphere.

The principal mechanism for producing aerosols from the oceans is through sea sprays, and consequently, they have a composition that is very similar to sea salt. Aerosols of this origin may contain NaCl (sodium chloride), KCl (potassium chloride), $CaSO_4$ (calcium sulfate), and $(NH_4)_2\ SO_4$ (ammonium sulfate) and they are very hygroscopic. From the table, it can be seen that also freshwater may produce aerosol, but with negligible strength with respect to oceans.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_9

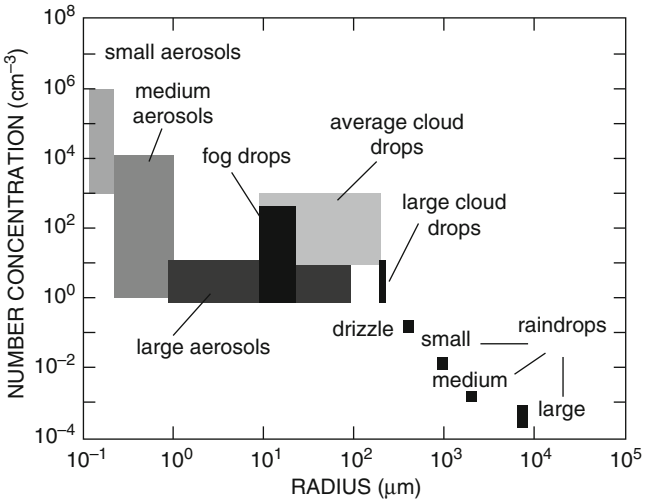



**Fig. 9.1** The distribution in terms of number concentration and radius for atmospheric aerosols and cloud particles (Figure Adapted from Jacobson 1999)

**Table 9.1** Global strengths of the aerosol particles in the troposphere

| Source | Strength ($Tg$ year$^{-1}$) |
|---|---|
| Surface sources | |
| Oceans and freshwater bodies | ∼1000–2000 |
| Crust and cryosphere | ∼2000 |
| Biosphere and biomass burning | ∼450 |
| Volcanoes | ∼15–90 |
| Spatial sources | |
| Gas to particle conversion | ∼1300 |
| Clouds | ∼3000 |
| Extraterrestrial | ∼10 |

Crust and cryosphere production (i.e., ice in the polar region) includes mainly dust from the deserts. Other sources include soils but also snow. These aerosols have a typical mineral composition.

Aerosols originating from the biosphere include all the organic particles, even pollen or microorganisms and spores. The origin of such particles may be natural and industrial or anthropogenic. One of the main sources is again the ocean. Aerosols originating from burning biomass are soot and fly-ash particles, their composition being very near to that of elemental carbon. In this case, it is very difficult to separate the natural (i.e., spontaneous or wildfires) from the anthropogenic source.

Volcanic eruptions may contribute in different ways to the aerosol population. Solid particles include silicates and metallic oxides and gases like sulfur dioxide ($SO_2$) that can be converted to aerosol particles. The solid particles may have a very short lifetime in the atmosphere, while in order to have a substantial production of

sulfate particles (gas to particles conversion GPC), the eruption must be quite large and sent gases directly above the tropopause. Sulfur dioxide has a very short lifetime in the troposphere because of chemical processes in gaseous and aqueous reactions.

Among the spatial sources, one of the most important is GPC.

We will study this process in some detail, especially in connection with stratospheric chemistry. In this case, as we have mentioned before, sulfur dioxide is converted through a series of chemical reactions into sulfuric acid which can be dissolved into water drops, forming a very important class of aerosols.

The aerosol source related to clouds has been only recently recognized. In the precipitation process, the clouds can be both a source and a sink for aerosols. During the evaporation process, the clouds can leave behind a considerable amount of aerosols.

The extraterrestrial source due to the influx of meteorites or dust can be easily neglected.

## 9.2 The Size Distribution of Atmospheric Aerosols

Starting from Fig. 9.1, we have implicitly assumed that their number concentration and their dimension characterize aerosols. Actually these properties can be defined by introducing the size distribution. If we denote with $n_i$ the number of particles per unit volume with radius $r_i$, then the total number of particles will be

$$N = \sum_i n_i (r_i) \tag{9.1}$$

Assuming a continuous distribution, we can introduce a density $n(r)$ that represents the number of particles per unit volume with radius between $r$ and $r + dr$ so that

$$n(r) = \frac{dN}{dr} \tag{9.2}$$

The function $n(r)$ is called the *size distribution* and in this case the total density is obtained by integrating Eq. (9.2) with respect to the radius. The concept of size distribution is extremely useful because it can be used to find the average quantities related to the distribution, like the average radius:

$$\bar{r} = \frac{\int_0^\infty rn(r)dr}{\int_0^\infty n(r)dr} = \frac{\int_0^\infty rn(r)dr}{N} \tag{9.3}$$

Each quantity related to the distribution, like the total surface or the total volume occupied by the particles, can be treated in the same way.

The size distributions are very different. Usually, however, because of the range of variation of both density and radius, logarithmic scales are preferred. One of the simpler distributions relates linearly to the logarithms of number concentration and the radius:

$$\log\left[\frac{dN}{d\log r}\right] = c - \beta \log r \quad \Rightarrow \frac{dN}{d\log r} = Cr^{-\beta} \tag{9.4}$$

After some manipulation, it is very easy to show that this is equivalent to the distribution:

$$n(r) = Cr^{-(1+\beta)} \tag{9.5}$$

Another distribution that is often used both for tropospheric and stratospheric aerosols is called *lognormal* and is given by

$$\frac{dN}{d\ln r} = \frac{N_0}{\sqrt{2\pi}\ln\ \sigma} \exp\left\{\left[\frac{(\ln r - \ln R)^2}{2(\ln\ \sigma)^2}\right]\right\} \tag{9.6}$$

This expression can be put in another form using $\mu = \ln R$ and $\sigma' = \ln\sigma$ and remembering that $dN/d\ln r = rn(r)$ we get

$$n(r) = \frac{N_0}{r\sigma'\sqrt{2\pi}} \exp\left[-\frac{(\ln r - \mu)^2}{2\sigma'^2}\right] \tag{9.6a}$$

where $N_0$, $\sigma$, and $R$ are the total number of particles, the semidispersion, and the mean radius characterizing the distribution. This expression shows that the distribution is Gaussian except for the $r$ at the denominator which introduces an asymmetry in the distribution. This distribution, as shown in Fig. 9.2, is used to describe several kinds of tropospheric aerosols. For each type of aerosol shown, the distribution is a superposition of three different lognormal distributions. For each size distribution, only the distribution for particles with radiuses larger than 0.001 μm is shown.

The most abundant are urban aerosols and their source is pollution or, more specifically, industry, traffic, and so on. Their total density is about $1.5{\cdot}10^5$ cm$^{-3}$, with the smaller mode being the most abundant. Rural aerosols are the next most abundant type. The shape of their size distribution reveals that the main contribution to them may come from urban aerosols. Their total density is of the order of $10^4$ cm$^{-3}$. Again the shape shows that continental aerosols may also be a source of rural aerosols. Their total density is roughly half that of the rural type, and they characterize the lowest layers of the atmosphere in remote continental regions.

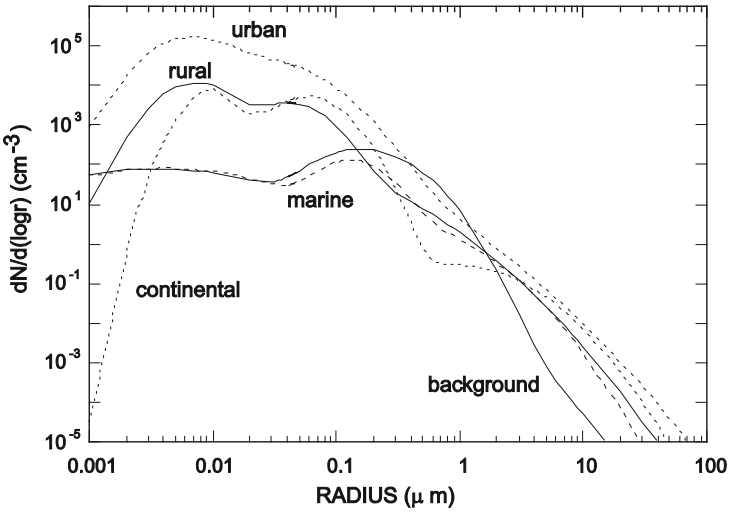



**Fig. 9.2** The size distribution of different types of tropospheric aerosols

Background aerosols refer usually to the population found above clouds. Their total density is rather small and does not exceed 200 $cm^{-3}$. Finally, maritime (or marine) aerosols are given by the sum of background and sea salt. The total number is again of the order of 200 $cm^{-3}$.

Once the size distribution is known, it is possible to obtain the surface or volume distribution with radius. For example, if we consider particles with radius $r$, the surface of a number $dN$ is given by

$$dS = 4\pi r^2 dN$$

so that the surface distribution is given by

$$\frac{dS}{d(\log r)} = 4\pi r^2 \frac{dN}{d(\log r)} = 4\pi C r^{-(\beta-2)} \tag{9.7}$$

And in a similar way for the volume

$$\frac{dV}{d(\log r)} = \frac{4\pi r^3}{3} \frac{dN}{d(\log r)} = \frac{4\pi}{3} C r^{-(\beta-3)} \tag{9.8}$$

for the size distribution (9.5). The effect of the multiplication by the square or the cube of the radius is to enhance the influence of the large particles on the distribution. For example, the secondary peaks in the distributions of Fig. 9.2 will now be the highest. A lognormal distribution can be approximated with Eq. (9.5) for very large radiuses.

## 9.3 Nucleation and Growth

Before we can proceed to study aerosols in some detail, we must introduce the concept of nucleation and growth. This can be done in a simple way by starting out with pure water (or water vapor for that matter) and then extending these ideas to more complicated processes. The nucleation or the initial aggregation can result in either drops or ice crystals.

All the processes related to aerosol formation are studied in the realm of *microphysics*, as if physics should change names according to the dimension of the objects under study. The germination of an aerosol particle may start with nucleation, that is, the condensation of two or more chemical species (*heteromolecular* nucleation) or of a single species (*homomolecular* nucleation). The germ can form on an existing particle (heterogeneous) or in a saturated environment.

As shown in Fig. 9.3, the processes that form aerosols however are much more complex. If we limit to the process of gas to particle conversion, we will start from condensable gases that may nucleate and give rise to ultrafine aerosols. This may coagulate and grow to a somewhat larger size until much complex processes enter the game. Dust and sea salt may provide condensation nuclei which may facilitate nucleation and further grow until the aqueous phase chemistry determines the interaction between aerosol and clouds.

We need to keep the discussion at a very simple level, and this can be done by starting with water vapor and its possible transformation.

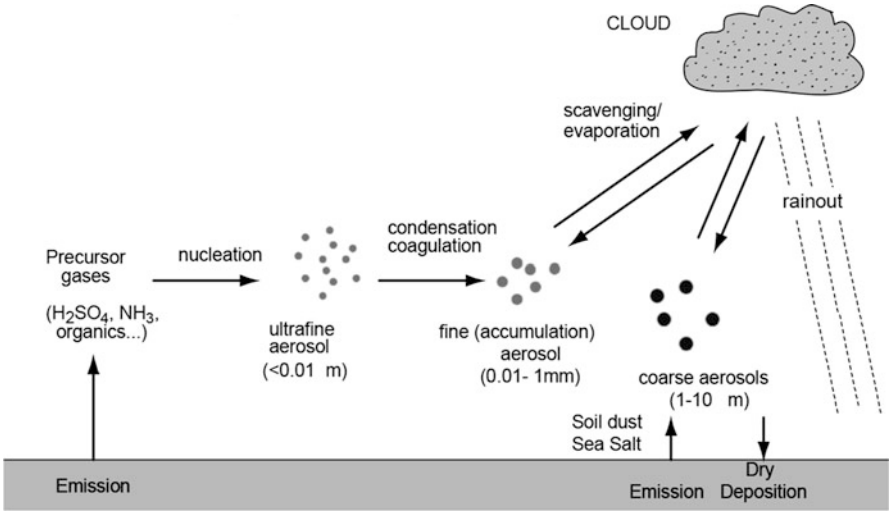



**Fig. 9.3** A cartoon summarizing the processes which form aerosols (Jacob 1999)

### 9.3.1 Nucleation from Water Vapor Condensation

A very simple theory can be formulated for water vapor. If we consider a droplet of radius $r$, then in order to grow we need to provide energy in two different ways. If the droplet grows in size, then we must spend some work to overcome the surface tension which is proportional to the surface of the drop. Also we need to compress (or expand) the water vapor from the ambient pressure to the saturation. So we have

$$\Delta E = A\sigma - nVkT \ln (e/e_s) \tag{9.9}$$

Here $A$ is the surface of the droplet, $\sigma$ is the work required to increase the surface of a unit, $n$ is the molecular density of the drop, and $k$ is the Boltzmann constant. Notice that the work of compression can be either positive or negative depending on the saturation ratio. This relation (some call it the change in *Gibbs free energy*) may be written as a function of the droplet radius:

$$\Delta E = 4\pi r^2 \sigma - (4\pi/3)\, r^3 nkT \ln (e/e_s) \tag{9.10}$$

The sign convention can be understood by considering that, if the saturation ratio is less than one, then the energy change is positive: in this case, we must provide energy to the system. On the other hand, if the saturation ratio is greater than one, the droplet will grow spontaneously if the radius is greater than a critical value $r_c$ which we find by putting $\partial \Delta E/\partial R = 0$, giving

$$r_c = 2\sigma/nkT \ln (e/e_s) \tag{9.11}$$

A very simple calculation with this relation gives, for a droplet of 0.01 μm radius, a saturation ratio of 1.125 which is a relative humidity of 112.5 %. Unfortunately, these are not routine values of relative humidity found in the atmosphere, and consequently even small embryos cannot grow. Droplets that have reached the critical radius may capture an additional molecule and become supercritical. If we invert relation (9.11), we get

$$e_s (r) = e_s (\infty) \exp (2\sigma/rnkT) = e_s (\infty) \exp (2\sigma/r\rho_L R_v T)$$

This result shows that if the particle grows in size, the saturation pressure decreases and the rate of growth (which depends on the difference between $e - e_s$) increases so that supercritical droplets grow spontaneously. From statistical thermodynamics, the nucleation rate per unit volume is given approximately by

$$J = 4\pi r_c^2 \frac{e}{\sqrt{2\pi mkT}} Zn \exp\left(-\frac{4\pi r_c^2 \sigma}{3kT}\right) \tag{9.12}$$

The nucleation rate is defined as the rate at which supercritical droplets are formed. In Eq. (9.12), $k$ is the Boltzmann constant, $n$ is the number density of water molecules, and $Z$ is a numerical factor of the order of $10^{-2}$. As reference, a significant rate of homogeneous nucleation is 1 $cm^{-3}$ $s^{-1}$. Saturation ratio measured experimentally for this nucleation rate ranges around 3–6 increasing with decreasing temperature. This confirms the simple calculation made before.

The situation changes considerably if we consider heterogeneous nucleation. In this case, water vapor condenses on a preexisting particle. The simplest case is when the constituents of this particle are soluble in water. In this case, the saturation pressure on the surface of the particle is lowered because the number of water molecules on the droplet surface is reduced. The fractional reduction of the vapor pressure is given by

$$f = e'/e \tag{9.13}$$

We now consider a solution droplet of radius $r$ that contains a mass $m_s$ of dissolved salt with molecular weight $M_s$. If each molecule of the salt dissociates into $i$ ions, the effective moles of the salt in the droplet will be $im_s/M_s$. If the density of the solution is $\rho'$ and the molecular mass of water $M_w$, then the number moles of pure water in the droplet is

$$\left(4\pi r^3\rho' - m_s\right)/M_w$$

And the mole fraction of pure water is

$$f = \frac{\left(\frac{4}{3}\pi r^3\rho' - m_s\right)/M_w}{\left(\frac{4}{3}\pi r^3\rho' - m_s\right)/M_w + im_s/M_s} \tag{9.14}$$

If we substitute from Eq. (9.12) the vapor pressure in Eq. (9.13) and then equate the quantity obtained in Eq. (9.14), we have

$$\frac{e'}{e_s} = \left[\exp\frac{2\sigma'}{n'kTr}\right]\left[1 + \frac{im_sM_w}{M_s\left(\frac{4}{3}\pi r^3\rho' - m_s\right)}\right]^{-1} \tag{9.15}$$

Now we have put primes on everything because we refer to the quantity in solution. This equation can be used to calculate the relative humidity needed to grow small drops that may have a condensation nucleus inside and should be compared directly with Eq. (9.12). The comparison can be simplified if we consider that both the exponential term and the fraction within parentheses are very small. If we neglect the solute mass with respect to the mass of the droplets, we get

$$\left[1 + \frac{im_sM_w}{M_s\left(\frac{4}{3}\pi r^3\rho' - m_s\right)}\right]^{-1} \cong \left[1 + \frac{3im_sM_w}{4\pi r^3M_s\rho'}\right]^{-1} = 1 - \frac{b}{r^3} \tag{9.16}$$

where $b = 3im_sM_w/4\pi M_s\rho'$. On the other hand, the exponential can be written as

$$\exp\frac{2\sigma'}{n'kTr} \cong 1 + \frac{2\sigma'}{\rho' R_v Tr} = 1 + \frac{a}{r} \tag{9.17}$$

where $a = 2\sigma'/\rho' R_v T$. These terms can be expressed numerically as

$$a = 3.3 \cdot 10^{-5} T^{-1} \text{ (cm)} \qquad b = 4.3im_s/M_s \left(\text{cm}^3\right)$$

and it is rather obvious that they are quite small. As a consequence, Eq. (9.15) can be expressed as

$$\frac{e'}{e_s} = 1 + \frac{a}{r} - \frac{b}{r^3} \tag{9.18}$$

The term on the right, proportional to the inverse of the radius, is referred to as the "curvature effect" and expresses the decrease of the saturation ratio with increasing radius of the droplet. The last term is referred to as the "solution effect" and shows the decrease of the saturation ratio with the increasing mass of the solution.

A graphic representation of Eq. (9.18) or (9.15) is given in Fig. 9.4, and the curves shown are known as *Köhler curves*. From Eq. (9.18), it can be easily worked out that the maximum for the curves are given by

$$r = \left(\frac{3b}{a}\right)^{1/2} \qquad \frac{e'}{e_s} = 1 + \left(\frac{4a^3}{27b}\right)^{1/2} \tag{9.19}$$

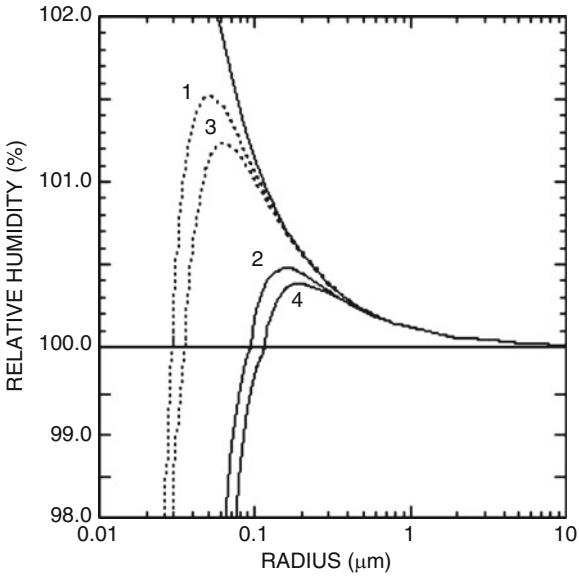



**Fig. 9.4** Variation of the relative humidity and supersaturation for droplets of ammonium sulfate and salt. The *upper curve* shows the relative humidity for a droplet of pure water. The other curves (*1*) and (*3*) are ammonium sulfate with $10^{-16}$ and $10^{-17}$ g of solute, respectively; (*2*) and (*4*) are the same for salt. Notice the expansion of the scale of the relative humidity above 100 %

As an example, we consider a NaCl salt with $i = 2$, $M_s = 58.44$, and $m_s = 10^{-19}$ kg at 0 °C and then we get $a = 1.2\ 10^{-7}$ cm and $b = 1.47e^{-17}$ cm$^3$. Then $r = 0.19$ μm and the saturation ratio 1.00417 which is a relative humidity of 100.4 %. Figure 9.4 shows that the lowering of the saturation ratio is much more evident for the small droplets that can nucleate now at reasonable values of the relative humidity. In the atmosphere, the relative humidity rarely exceeds by a few percent the 100 % limit.

The droplets forming in conditions represented by the left-hand side of the Kohler curve are in equilibrium with the ambient air. As a matter of fact, if a droplet with a 0.05 mm radius and 1 % supersaturation should grow a little, the vapor pressure on their surface would grow above the vapor pressure of the ambient air and they would evaporate back to the initial size. Conversely, if the droplet should decrease in size, the vapor pressure would be below the saturation of the ambient air and the particle would grow back to the original size. Droplets on the left side of the Köhler curves are stable and correspond to *haze* particles. Droplets at the maximum of the Kohler curve, or on the right side, are in an unstable equilibrium. If they increase a little in size, the vapor pressure on them will decrease below that of the ambient air and they will continue to grow.

It appears that a droplet that goes beyond the top of the curves may grow indefinitely by condensation, but we can easily see that there are a few problems that prevent this from happening.

### 9.3.2 The Growth by Condensation

The simplest way for the droplets to grow is apparently through condensation. This happens when the vapor pressure in the environment of the droplet is greater than the saturation pressure. However, it is a very interesting exercise to show that it is very unlikely that rain can be produced through condensation.

We assume that the water vapor flow to the particle is due to diffusion. Then the mass $M$ of the droplet will change at the rate

$$dM/dt = 4\pi x^2 D d\rho_w/dx \tag{9.20}$$

where $\rho_v$ is the water vapor density at distance $x$ from the center of the spherical droplet and $D$ is the diffusion coefficient. The mass flux responsible for the growth is independent from the distance so that Eq. (9.20) can be integrated from the surface of the droplet up to a very large distance. We obtain

$$dM/dt \int_r^\infty dx/x^2 = 4\pi D \int_{\rho_\upsilon(r)}^{\rho_\upsilon(\infty)} d\rho_w \Rightarrow dM/dt = 4\pi r D \left[\rho_\upsilon(\infty) - \rho_\upsilon(r)\right] \tag{9.21}$$

Equation (9.21) can be easily written as the rate of change of the droplet radius as a function of the saturation pressures and the density of liquid water $\rho_l$

$$\frac{dr}{dt} = \frac{1}{r} D \frac{\rho_\upsilon(\infty)}{\rho_l e(\infty)} [e(\infty) - e(r)] \tag{9.22}$$

This relation shows that the rate of growth is proportional to the inverse of the radius so that the small particles will grow faster than the large ones. Rearranging Eq. (9.22) and defining the supersaturation $S$ of the ambient air as

$$S = e(\infty) - e(r)/e(\infty) \cong e(\infty) - e_s/e_s,$$

we have

$$r\frac{dr}{dt} = D\frac{\rho_\upsilon(\infty)}{\rho_l} S = G_l S \qquad \text{with} \qquad G_l = D\frac{\rho_\upsilon(\infty)}{\rho_l} \tag{9.23}$$

The factor $G_l$ can be considered constant for a given environment so that Eq. (9.23) can be easily integrated to give

$$r = (2G_l S t)^{1/2}$$

The diffusion coefficient is of the order of $2\ 10^{-5}$ m$^2$ s$^{-1}$ at the pressure of 1000 kPa, and the density of water vapor may be expressed as a function of the mass mixing ratio $w$. With a mixing ratio of 0.002, we have $G_l \cong 5{\cdot}10^{-11}$ m$^2$s$^{-1}$. With a supersaturation of 1 %, the droplet will reach 1 μm in 1 s and will take about 1 h to reach 60 μm. Figure 9.1 shows that raindrops could never be formed in a reasonable time with the condensation mechanism. There are other considerations that hint that condensation is not responsible for the formation of raindrops. Given a mixing ratio $w$ and the number of condensation nuclei for unit volume ($n_{\text{CCN}}$), it is easy to show that the maximum radius for droplets growing by condensation is simply

$$r = (3w\rho/4\pi\rho_l n_{\text{CCN}})^{1/3}$$

With the observed range for the variables, this expression gives droplets in the size range of tens of microns which are still too small. Also, the distribution of the droplets would be essentially monodisperse with all the droplets having the same radius.

### *9.3.3 Droplet Growth by Collision and Coalescence*

We have seen that condensation will hardly produce droplets large enough to initiate rain. A much more efficient mechanism can be related to the fact that, if a droplet gets large enough, the forces that keep it from falling (electrical or aerodynamic) are overcome by the gravitational force. During the fall, the droplet may collide with smaller droplets and capture them and grow in size (the old story about the big fish eating the smaller one).

First of all, we need to refresh our memory about the terminal fall speed which is reached when the viscous force on the droplet is equilibrated by the weight of the particle. The drag force on a droplet of radius $r$ is given by

$$F_R = \pi C_D r^2 w^2 \rho / 2 \tag{9.24}$$

where $C_d$ is the drag coefficient, $\rho$ is the density of air, and $w$ is the vertical velocity. The drag coefficient can be expressed in terms of the Reynolds number $R_e$:

$$R_e = 2\rho w r / \mu \tag{9.25}$$

with $\mu$ as the dynamic viscosity. Then Eq. (9.24) becomes

$$F_R = 6\pi \mu r w \, (C_D R_e / 24) \tag{9.26}$$

Equating this force to the gravitational force acting on the particle, we have for the terminal velocity

$$w = \frac{2}{9} \frac{r^2 g \rho_l}{\mu \, (C_D R_e / 24)} \tag{9.27}$$

When the Reynolds number is small, we can use the formula we learned in school (Stokes' solution) which corresponds to having the terms in parentheses in Eq. (9.26) or (9.27) put to one. Then we have

$$w = \frac{2}{9} \frac{r^2 g \rho_l}{\mu} = k_1 r^2 \tag{9.28}$$

where $k_1 \cong 1.2\ 10^6\ \mathrm{cm}^{-1}\ \mathrm{s}^{-1}$. This relation (also known as Stokes' law) shows that the terminal velocity is proportional to the square of the radius and can be applied up to particles with about a 30 mm radius. For large spheres and high Reynolds number, the drag coefficient becomes about 0.45 and the dependence on the radius is now on the square root according to

$$w = 2.2 \cdot 10^3 (\rho_0 / \rho)^{1/2} r^{1/2} \tag{9.29}$$

with the velocity expressed in cm $\mathrm{s}^{-1}$. To have an idea of the numbers, a spherical raindrop of 0.1 mm will fall to about 0.3 m $\mathrm{s}^{-1}$ with a Reynolds number of 3.6, while a 5 mm drop will fall at 9 m $\mathrm{s}^{-1}$ with a Reynolds number of 6000.

At this point, we should be convinced that particles with different radiuses fall at different velocities, the larger going faster. The capture geometry looks like that in Fig. 9.5, where a drop of radius $R$ will overtake, and possibly capture, a smaller drop of radius $r$. Actually, even if geometry does not give the smaller drop a chance, if it is within the range of the larger radius, physics gives a different result.

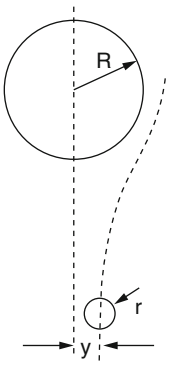



**Fig. 9.5** The geometry for collision between two drops with the impact parameter $y$ shown

The air will slipstream around the larger drop and the capture will depend on the critical distance, called the impact parameter. All the drops falling within the drop parameters will collide with the larger drop but will not necessarily be captured.

The collision efficiency will be the ratio between the area swept by the impact parameter and the possible total area swept by the droplets:

$$E = y^2/(R + r)^2 \tag{9.30}$$

The collision efficiency is not easy to calculate and has very small values for small ratios *r/R* and values greater than unity for droplets of the same size. However, even if one would be able to calculate $E$ exactly, then it would still be uncertain if after the collision there is coalescence. We need to introduce another quantity, called *coalescence factor*, that is actually the fraction of collision that results in coalescence. The product of the coalescence factor and the collision efficiency is called the collection efficiency $E_c$. The coalescence factor is unity for a very small *r/R* ratio and then falls off rapidly.

We can now proceed to a very simple calculation, assuming that a drop of radius $R$ falls at terminal speed through a population of smaller droplets. The volume swept per unit time will be proportional to the relative speed of the droplets:

$$\pi (R + r)^2 \left[u(R) - u(r)\right]$$

where $u$ is the terminal velocity for the droplets. If $n(r)$ is the number of droplets, with radius between $r$ and $r + dr$, they will be captured at a rate

$$\pi (R + r)^2 \left[u(R) - u(r)\right] n(r) E_c\,(R, r)\, dr$$

The total change of volume of the larger drop is obtained by integrating all over the drop sizes

$$\frac{dV}{dt} = \int_0^R \pi (R + r)^2 \frac{4}{3} \pi r^3 \left[u(R) - u(r)\right] n(r) E_c \Big(R, r\Big) dr$$

This expression is very much simplified if we assume that $r << R$ so that for the velocities $u(R) >> u(r)$. We can easily obtain in this approximation the increase in the radius of the collecting drop as a function of liquid water content $w_l$ (in kg m$^{-3}$):

$$\frac{dR}{dt} = \frac{u(R) w_l E}{4\rho_l} \tag{9.31}$$

where we have substituted $E$ for $E_c$. This equation shows that the size increases linearly with the radius of the collecting drop (the terminal velocity is linear with the radius). We may then expect that, up to a certain size, growth by condensation will dominate and after that the coalescence will pick up.

We have assumed that the relative velocity coincides with the terminal velocity of the larger drop. However, the relative velocity can be decreased if an updraft (i.e., a vertical current) is present. If we assume $w$ to be the value of the vertical velocity, the change in altitude of the collecting drop will be

$$\frac{dh}{dt} = w - u$$

so that, writing $dR/dt = (dR/dh)\,(dh/dt)$, we have from Eq. (9.31)

$$\frac{dR}{dh} = \frac{u w_l E}{4\rho_l\,(w - u)} \tag{9.32}$$

This expression can be integrated between the base and the top of the cloud:

$$\int_0^H w_l dh = 4\rho_l \int_{R_0}^{R_H} \frac{w - u}{uE} dR$$

From this, we can obtain the altitude $H$ for which the drop has an assigned value for the radius $R_H$ when the liquid water content is constant:

$$H = \frac{4\rho_l}{w_l} \int_0^H \frac{w}{uE} dR - \int_{R0}^{rH} \frac{dR}{E} \tag{9.33}$$

This equation gives a rather clear idea of the growth process by collision within a cloud. When the drops are small ($w >> u$), the first integral dominates the second and $H$ increases with $R_H$ which means the drop grows by being carried upward. When the drop is large enough, then $w < u$ and the altitude $H$ decreases with increasing $R_H$, meaning the size of the drops will increase by falling. If it gets large enough, it will pass through the base of the cloud and become rain; otherwise it can break up in smaller drops. Each one of the fragments can start the process anew either by going up or down.

We assumed implicitly that falling drops encounter a constant and uniform population of smaller drops. However, dealing with collisions, we should know that this problem must be treated statistically.

### 9.3.4 The Statistical Growth

We start by noting that a proper way to see the growth through collisions is to think that a drop spectrum (i.e., size distribution) initially determined by condensation and diffusion changes its shape by random collision between droplets. Consider again that the volume swept by a drop of radius $R$ will be $\pi R^2 u(R)$ and similarly for a drop of radius $r$. In order for the drops to collide, they must be contained in the volume $\pi (R + r)^2 [u(R) - u(r)]$. The aerodynamic effect, the collision efficiency,

will reduce this volume to an amount:

$$K(R, r) = \pi (R + r)^2 [u(R) - u(r)] E(R, r) \tag{9.34}$$

that we will denote the *coagulation coefficient*. All the previous treatments for continuous growth could be reformulated in terms of probability, interpreting Eq. (9.34) as the probability that a drop of radius $R$ will coalesce with a drop of radius $r$. Actually the formulation based on the volumes is simpler than the one based on the radius. The coagulation coefficient is then reformulated as

$$H(V, v) = K\left[(3V/4\pi)^{1/3}, (3v/4\pi)^{1/3}\right] \tag{9.35}$$

Consider now a drop spectrum described by $n(v)$ that is the average number of drops with volumes between $v$ and $v + dv$. According to the definition of coagulation coefficient, the rate of coalescence is given for drops included in the size interval by

$$n(v)\, dv \int_0^\infty H(V, v)\, n(V)\, dV$$

Actually this rate will decrease the number of drops in the size interval because it accounts for the capture by both large particles and small particles. However, an increase may be attributed to all the pairs of smaller particles with volumes amounting to $v$. The rate will be

$$\frac{\partial}{\partial t} n(v)\, dv = \frac{1}{2} dv \int_0^v H\left(\delta, v'\right) n\left(v'\right) n(\delta)\, dv' - n(v)\, dv \int_0^\infty H(V, v)\, n(V)\, dV \tag{9.36}$$

where $\delta = v - v'$ and the factor ½ is to prevent any capture to be counted twice. Equation (9.36) can be solved numerically (i.e., with a computer program), but also analytical solutions for simple cases are known. One simple solution refers to a monodisperse initial population (particles with the same radius) and is known as the Smoluchowski solution. It shows that the final state corresponds to a spreading of the initial Dirac-$\delta$ distribution. An analytical solution can be found also for the case of preserving size distribution. Here, the final state shows a drift of the initial population toward larger sizes.

The coagulation mechanism is quite general and of great importance in studying stratospheric aerosols. Sometimes the coagulation coefficient is also called the kernel because it is the quantity appearing in the integral of Eq. (9.36) and establishes the interactions between different size distributions.

## 9.4 Formation and Growth of Ice Crystals

When the temperature within a cloud is colder than 0 °C, then in theory freezing can take place. Actually the phase transition into ice may happen either by direct freezing of a liquid drop or by a direct sublimation from the vapor phase to the solid. Both these processes, which may be regarded as homogeneous nucleation, are very unlikely and can take place only at very low temperatures (less than −40 °C) or extreme supersaturation or both.

As in the case of liquid droplets, the nucleation is much easier if a substrate exists. Usually the presence of condensation nuclei may raise the temperature threshold for ice formation. Among the pure substances, the highest temperature refers to ice (0 °C), while all the other substances start ice nucleation between −4 and −12 °C. This is the reason why some of these substances (like silver iodide) are used to seed clouds in the hope that they will facilitate ice formation and then rain.

Once ice crystals are formed, they may grow initially by condensation and much faster than the liquid drops, because at the same temperature the saturation pressure for ice is lower. The saturation ratio with respect to the ice can be written as

$$S_i = e/e_i = (e/e_s)\,(e_s/e_i) = S\,(e_s/e_i) \tag{9.37}$$

where $S$ indicates the saturation with respect to water. As we have shown in Chap. 1, the saturation pressure over ice can be approximated as

$$e_i = Ae^{-B/T}$$

where $A = 3.41{\cdot}10^9$ kPa and $B = 6.13{\cdot}10^3$. The difference with respect to water is that we have to use the heat of sublimation $L = 2.83{\cdot}10^6$ J kg$^{-1}$. Then a direct comparison can be made with the saturation pressure over water and we obtain

$$\frac{e_s}{e_i} = \exp\left[\frac{L_s - L}{R_v T_0}\left(\frac{T_0}{T} - 1\right)\right] \tag{9.38}$$

where the symbols are the same used in Chap. 2. This expression shows that the saturation pressure over water exceeds that over ice for temperatures below 273 K. The quantity $(e_s/e_i) - 1$ is shown in Fig. 9.6. Although the values on which the figure is based do not take into account the temperature dependence of the latent heat, it is clear that the supersaturation with respect to ice is quite large. We can then imagine that inside the clouds the ice deposition will continue as long as there are droplets that, by evaporating, maintain a vapor pressure in equilibrium with respect to water.

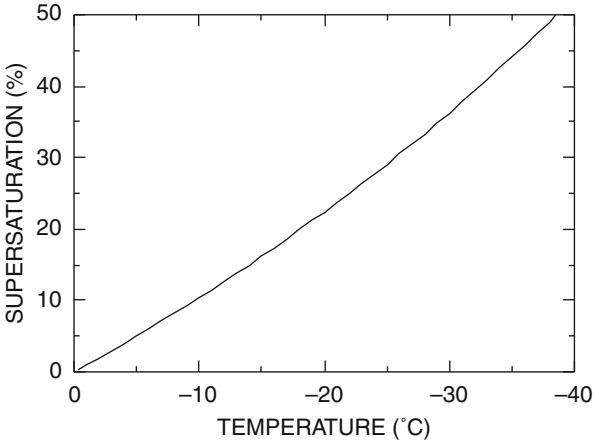



**Fig. 9.6** The supersaturation with respect to ice in an environment at equilibrium saturation with respect to water

Growth by diffusion is not so simple as in the case of liquid droplets because we cannot use the spherical symmetry. There is however a simple argument based on the fact that the diffusion equation at steady state is similar to the Poisson equation for the electrostatic potential:

$$\nabla^2 V = -\rho/\varepsilon_0 \tag{9.39}$$

In this case, the electric field is given by the Gauss theorem

$$4\pi R^2 \varepsilon_0 \left(-\nabla V\right) = Q \tag{9.40}$$

where $R$ is the radius of the region that includes the charge $Q$. If we refer to a conducting sphere of the same radius, we can relate the capacitance $C$ to the electric field and the potential:

$$4\pi R^2 \varepsilon_0 \left(-\nabla V\right) = CV$$

In complete analogy, we assume that, in case of diffusion flux, the left-hand side coincides with the total current of water molecules with number density $n$:

$$4\pi R^2 \left(-D\nabla n\right) = 4\pi CD \left(n_s - n_\infty\right)$$

where $D$ is the diffusion coefficient $n_s$ and $n_\infty$ is the number density on the surface of the crystal and far away from it. The capacitance must then be expressed in units of length. In terms of the mass densities, the growth equation can be written as

$$\frac{dm}{dt} = 4\pi CD \left(\rho_v - \rho_{vr}\right) \tag{9.41}$$

Here $\rho_v$ and $\rho_{vr}$ are the density of water vapor far away and at the surface of the crystal. The release of the heat of sublimation during crystal growth implies that the surface is warmer. The heat released during sublimation is given by $L_s dm/dt$ and this must be equal to the heat taken away by conduction:

$$4\pi CDL_s\left(\rho_v - \rho_{vr}\right) = 4\pi CK\left(T - T_r\right)$$

so that we have for the difference in density

$$\left(\rho_v - \rho_{vr}\right) = K\left(T - T_r\right) / \left(L_s D\right) \tag{9.42}$$

In this equation, the environmental conditions are represented by $T$ and $\rho_v$, while the conditions on the surface of the crystal are unknown. However, density and pressure are related by the equation of gases so that the coupled system could be integrated in principle starting from some initial conditions. A possible alternative is to linearize the dependence of the saturation pressure on the temperature. Using

$$\frac{d\rho_{vs}}{\rho_{vs}} = \frac{de_s}{e_s} - \frac{dT}{T} = \frac{L_s}{R_v}\frac{dT}{T^2} - \frac{dT}{T}$$

it is possible to obtain, after a few manipulations, that

$$\frac{\rho_{vs} - \rho_{vrs}}{\rho_{vrs}} = \left(\frac{T - T_r}{T}\right)\left(\frac{L_s}{R_v T} - 1\right) \tag{9.43}$$

The temperature difference can be expressed in terms of heat lost by the crystal and then the mass increase:

$$Ldm/dt = 4\pi rK\left(T_r - T\right) \tag{9.44}$$

This substituted in Eq. (9.43) gives an expression for the saturation densities:

$$\frac{\rho_{vs} - \rho_{vrs}}{\rho_{vrs}} = \left(\frac{L_s}{4\pi krT}\right)\left(\frac{L_s}{R_v T} - 1\right)\frac{dm}{dt} \tag{9.45}$$

We can get a similar expression for the vapor density from Eq. (9.41)

$$\frac{\rho_v - \rho_{vr}}{\rho_{vr}} = \left(4\pi CD\rho_{vr}\right)^{-1}\frac{dm}{dt} \tag{9.46}$$

Subtracting and assuming that

$$\rho_{vr} \approx \rho_{vrs} \qquad S_i = \rho/\rho_s$$

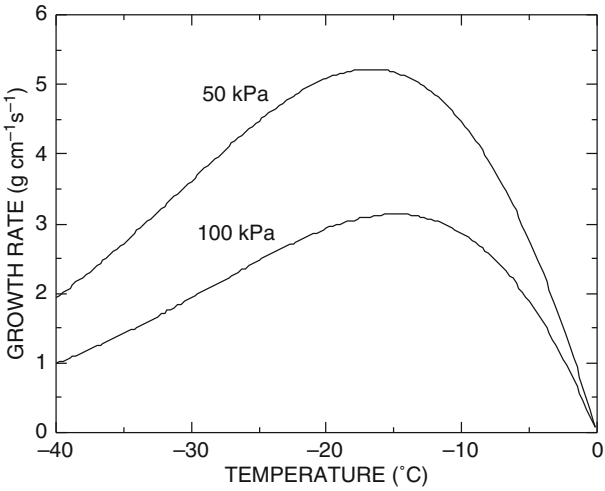



**Fig. 9.7** The growth of an ice crystal at different values of the ambient pressure (Rogers and Yau 1989)

**Table 9.2** Ice crystal shapes that form at various temperatures

| Temperature (°C) | Habit | Crystal form |
|---|---|---|
| 0 to −4 | Platelike | Thin hexagonal plates |
| −4 to −10 | Prismlike | Needles (−4 to −6 °C). Hollow columns (−4 to −10 °C) |
| −10 to −22 | Platelike | Sector plates (−10 to −12 °C). Dendrites (−12 to −16 °C). Sector plates (−16 to −22) |
| −22 to −50 | Prismlike | Hollow columns |

we have

$$\frac{dm}{dt} = 4\pi C\,(S_i - 1)\left[\left(\frac{L_s}{R_v T} - 1\right)\frac{L_s}{KT} + \frac{R_v T}{e_i\,(T)\,D}\right]^{-1} \tag{9.47}$$

In Fig. 9.7, the growth rate of an ice particle is normalized to the factor $4\pi C$. It is clear that the growth rate is a function of the pressure, while the maximum growth is obtained for a temperature around −15 °C.

After the growth has been initiated by condensation, the ice crystal may assume a number of shapes that are called habits. The most simple forms are platelike hexagonal or needle, that is, hexagonal prisms with a long aspect ratio. Starting from these basic forms, there is a variety of other shapes that seem to depend on the water vapor pressure and temperature. Table 9.2, taken from Wallace and Hobbs, lists the different habits and shapes we may expect as a function of temperature.

Once the crystal is formed, it may grow further with mechanisms that are similar to those seen for liquid drops. A very important factor is the different saturation pressure with respect to ice and water. An environment saturated with respect to water is supersaturated with respect to ice. This implies that, as the crystal grows

by deposition of vapor, it depletes the environment of vapor so that at some point the vapor pressure may fall below the saturation with respect to water. The water droplets will evaporate and replenish the environment of vapor which will be used for further growth of the ice crystal.

While the mechanisms for growth are similar to those seen for water droplets, a different vocabulary is used in the case of ice crystals. The term *accretion* indicates the capture of supercooled droplets by ice particles. *Aggregation* describes the clumping of ice crystals to form snowflakes.

Accretion is a process very similar to coalescence in that a crystal large enough may fall through a population of supercooled droplets, colliding and capturing them. As they hit the surface of the crystal, the droplets freeze, causing an ice deposit to form that is called *rime*. Because the process is very similar to coalescence, we may use a similar relationship for the growth of a mass $m_i$:

$$\frac{dm_i}{dt} = \int_0^\infty A_i \left(v_i - v_c\right) E\left(r_r, r\right) n\left(r\right) dr \tag{9.48}$$

where $A_i$ is the geometrical cross section of the crystal, $v_{i,c}$ are the terminal velocities of the ice particle and the droplet, $E$ is the collection efficiency, and $n(r)$ is the size distribution of cloud droplets. Equation (9.48) assumes that the droplets are spherical; otherwise we must use functions that describe the spectral density of the droplets. Each one of the variables in Eq. (9.48) is dependent on the growth habit of the crystal and, because during the riming process the shape changes, these parameters also change (Fig. 9.8).

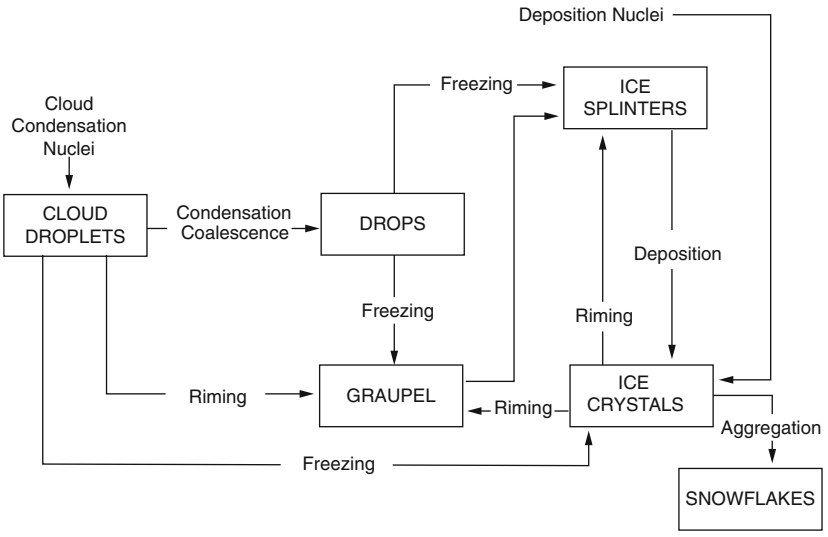



**Fig. 9.8** A very simple scheme of the interaction between different phase and water substances (Adapted from Houze 1993)

The growth of the ice crystal can be so efficient that soon the particle loses its original shape. In this case, we talk about a *graupel* particle. The density of a graupel particle is usually smaller than that of water but can be enough to make the particle fall and tumble and the shape may become quite irregular. Graupel may also function as embryos for hailstone if there is enough liquid water in the cloud and a strong updraft. The heat deposited on the hailstones surface due to sublimation may be large enough to heat the surface and limit the rate of ice formation. In this case, it is said that the hailstone grows in a wet regime.

Another way for an ice crystal to grow is aggregation, during which snowflakes are formed. Actually, snowflakes are a cluster or aggregate of ice crystals, and these may have formed through a previous process of collision or coalescence. The study and parameterization of the aggregation process are quite complex because snowflakes may assume a variety of forms and then influence terminal velocities and collection efficiency.

This section on microphysics cannot be exhaustive and the interested reader should see more specialized texts. However, the topic of precipitation is strictly connected to the weather forecast which nowadays is very much related to radar observation. Actually, recent advances in radar techniques give such nice results that they have a didactic value, that is, they illustrate very clearly the way precipitation can form and how water phase changes in the different cloud regions.

## 9.5 Stratospheric Aerosols

Until a few years ago, interest in stratospheric aerosols was confined to what was called the Junge layer. This was a thin aerosol layer in the lower stratosphere that was replenished each time a volcanic eruption would be large enough to inject appreciable quantities of sulfur dioxide directly above the tropopause.

This background aerosol layer was studied starting from the 1960s using optical radar (lidar). Beginning in 1978, first with SAM (stratospheric aerosols monitoring) and then with SAGE (stratospheric aerosols and gas experiment), the layer was studied mainly through observations with orbiting satellites. The advantage was that data from space could provide a global picture of the distribution of the aerosol load. "Luckily" enough, there were three large eruptions in a few years, Mt. St. Helens in 1980, then El Chichon in 1982, and Pinatubo in 1991. The last one happened just before the launch of the UARS (Upper Atmosphere Research Satellite).

Large eruptions, by perturbing the stratospheric aerosol layer, were thought to perturb Earth's climate. This question as of today is still unresolved, at the least for the eruptions we have known directly. New interest in stratospheric aerosols was stimulated by the hypothesis that their presence could influence stratospheric chemistry and in particular the ozone. This idea was a spin-off from the explanation of the ozone depletion in the polar region. The implications of the stratospheric aerosols are many and cannot be exhausted in a paragraph of a general textbook. We will review only a few aspects and invite to read specific papers and report on the subject, especially the ASAP, SPARC report.

### 9.5.1 The Sulfate Aerosol Layer

The stratospheric sulfate aerosol (SSA) is the main component of the background aerosol layer in the stratosphere. These particles are liquid and composed of sulfuric acid ($H_2SO_4$) and water ($H_2O$) which is the same composition as the Venus clouds. The way to get the sulfur above the troposphere is through oxidation of long-lived sulfur-bearing species like carbonyl sulfide ($COS$). Sulfur dioxide can also reach the stratosphere but only when there is a strong convection within cumulonimbus clouds. In this way, the transport is fast enough with respect to chemical processes that it destroys the troposphere. The remaining way to get sulfur in the stratosphere is through explosive eruptions. There is strong evidence that most of the stratospheric aerosols actually originate from these eruptions, and most of the time we are simply looking at the remnants of what is produced during these sporadic events. Figure 9.9 shows a possible budget for the sulfur species which contributes to the aerosol population. A major volcanic eruption (like Pinatubo) would inject a sulfur amount around 10 Tg S in the form of $SO_2$ that is about 100 times all the sources in quiescent condition. Once injected in the stratosphere, $SO_2$ is converted into sulfuric acid ($H_2SO_4$) vapor by two mechanisms:

$$\begin{aligned} SO_2 + OH + M &\rightarrow HOSO_2 + M \\ HOSO_2 + O_2 &\rightarrow HO_2 + SO_3 \end{aligned} \tag{9.49}$$

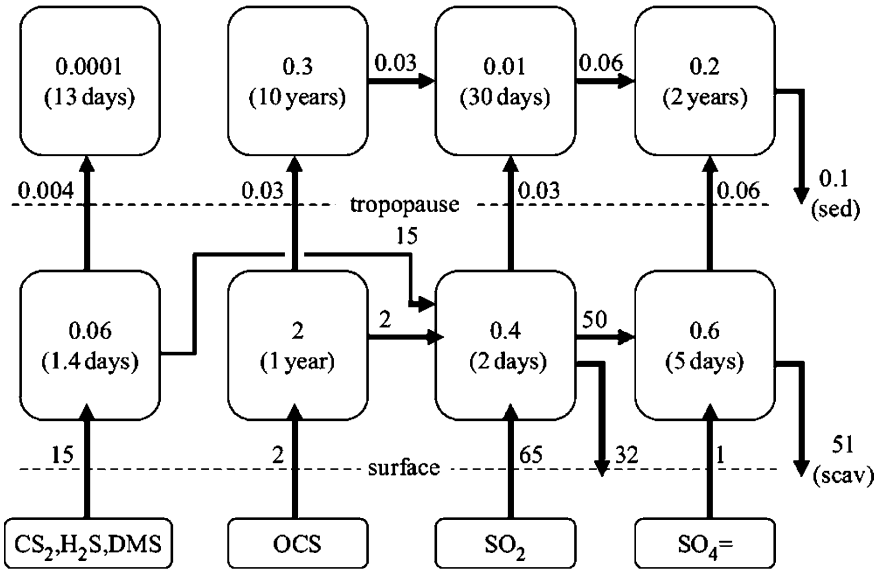



**Fig. 9.9** A rough budget (Rasch et al. 2008) of the sulfur compounds contributing to form the stratospheric sulfate layer. *Numbers inside the boxes* indicate the sulfur amount (in $Tg = 10^{12}$ g) of S. The boxes also report the lifetime for the major sink. The *numbers beside the arrows* are the net sources or sink (Adapted from Rasch et al. 2008)

Sulfur trioxide ($SO_3$) is converted to sulfuric acid by

$$SO_3 + H_2O \rightarrow H_2SO_4 \tag{9.50}$$

The conversion of sulfuric acid gas into liquid can occur by two mechanisms. The first involves the combination of sulfuric acid and water molecule (homogeneous, bimolecular nucleation) or the combination of $H_2SO_4$, $H_2O$, and $HNO_3$ molecules (homogeneous, heteromolecular nucleation) to form mainly sulfuric acid droplets. Also vapor condensation of $H_2SO_4$, $H_2O$, and $HNO_3$ can occur on preexisting particles with radius greater than 0.15 μm (heterogeneous, heteromolecular nucleation). The second mechanism is the most likely with the conversion taking about 1 month.

SSA particles have a density of the order of 10 $cm^{-3}$ and their size distribution changes considerably after an eruption. Figure 9.10 shows a possible size distribution before and after the eruption with a density of 1 $cm^{-3}$. The total area density normalized to this density is around 0.086 $\mu m^2$ $cm^{-3}$ and 1.9 $\mu m^2$ $cm^{-3}$. The area density is then proportional to N.

As can be seen from Fig. 9.10, a typical dimension of the SSA particles is less than 0.1 μm, but a more important parameter for stratospheric particles is rather the area density, that is, the surface area per unit volume. This is because most

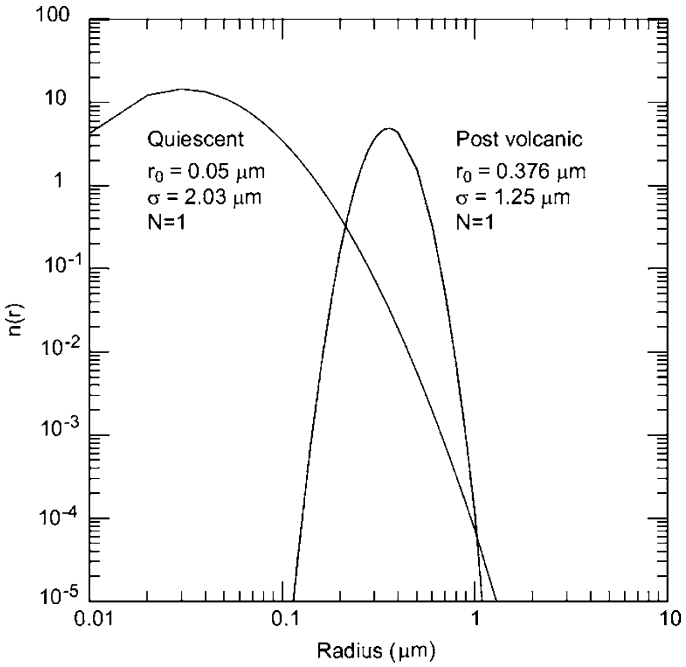



**Fig. 9.10** The size distribution of sulfate aerosol during quiescent condition and after a major volcanic eruption. Notice the change in both the standard deviation and the node radius. The distribution refers to a density $N = 1$ $cm^{-3}$

of the aerosol particles are the site for chemical reactions to occur. Typical values measured range between 1 and 10 $\mu m^2$ $cm^{-3}$.

We can refer to Fig. 1.6 to see that, for typical stratospheric conditions, that is, pressure between 100 hPa and 20 hPa and temperature between 200 and 240 K, the partial pressure of water vapor may range between $5.10^{-2}$ and $1.10^{-3}$ Pa. That figure would give sulfate particles with 70–80 % of sulfuric acid mass fraction. SSA particles are supercooled solutions, but if the temperature goes below 220 K, they may freeze and produce hydrates, the most likely being sulfuric acid tetrahydrate (SAT, $H_2SO_4 \bullet 4H_2O$). Recent data and laboratory experiments show however that freezing is very unlikely, and at low temperature, the aerosol particles may switch completely in composition to NAT (nitric acid trihydrate).

### 9.5.2 Polar Stratospheric Clouds

*Polar stratospheric clouds of type 1 (PSC-1)* have the chemical composition of NAT ($HNO_3 \bullet 3H_2O$). Actually, the chemical composition of such particles is not very well known and is mainly based on the measurement of their size distribution and on the fact that, in the lower polar stratosphere (i.e., between 20 and 24 km), there exist environmental conditions that may be favorable to their formation, with temperatures lower than 200 K. Typical dimensions of these particles are around 1 $\mu m$, while their chemical compositions reflect roughly, but not exactly, the stoichiometric ratio (54 % mass fraction of nitric acid).

*Polar stratospheric clouds of the second kind (PSC-2)* are formed when the temperature goes below 188 K so that water can freeze. These particles are very large, with a radius of the order of 10 $\mu m$, and their residence time in the stratosphere is only a few hours. The size distribution can be assumed to be lognormal with the following parameters. For PSC-1, we have $N_0 = 0.1$ $cm^{-3}$, $r_m = 0.9$ $\mu m$, and $e$ $\sigma = 1.8$ $\mu m$ while for PSC-2 we have $N_0 = 0.008$ $cm^{-3}$, $r_m = 4.5$ $\mu m$, and $\sigma = 1.7$ $\mu m$. It is to note that the distributions are well differentiated and that the aerosol density decreases with increasing average radius.

It is rather easy to show that for a lognormal distribution the surface density is proportional to the average radius squared:

$$S = 4\pi N_0 r_m^2 \exp\left[2\ln(\sigma)^2\right] \tag{9.51}$$

Based on that formula, PSC-1 are those with the larger surface density followed by PSC-2 and then SAA. It is to note however that these particles act in different regions with the SAA that are found mainly in midlatitude regions, while PSC are more frequent at high latitudes.

In recent times, the study of such particles has made important progress so that we know that the surface is not the only parameter important for heterogeneous reactions but also composition and temperature must be taken into account. Also

progress has been made on the formation of stratospheric aerosols. Initially, it was thought that PSC could form on SSA. As we have seen, SAA particles may freeze when the temperature falls below 215 K and form SAT particles. However, when the temperature decreases, also the gas solubility must be taken into account. It is possible to show that the solubility of sulfuric acid decreases drastically below 195 K so that below this temperature, more precisely in the range 192–197 K, NAT may form as a coating on preexisting sulfate particles. If the temperature decreases further, below 188–190 K, water ice may form.

Actually, the processes involved are more complex, as shown schematically in Fig. 9.11, where the stability regions for SAT, NAT, and water ice are shown. This figure is drawn for stratospheric conditions that refer to 50 hPa and 5 ppm of water vapor. In this figure, slightly different abbreviations are used, so that SAT is $H_2SO_4 \bullet 4H_2O$, SAX is the generic name for sulfuric acid hydrates, SAM is $H_2SO_4 \bullet H_2O$,

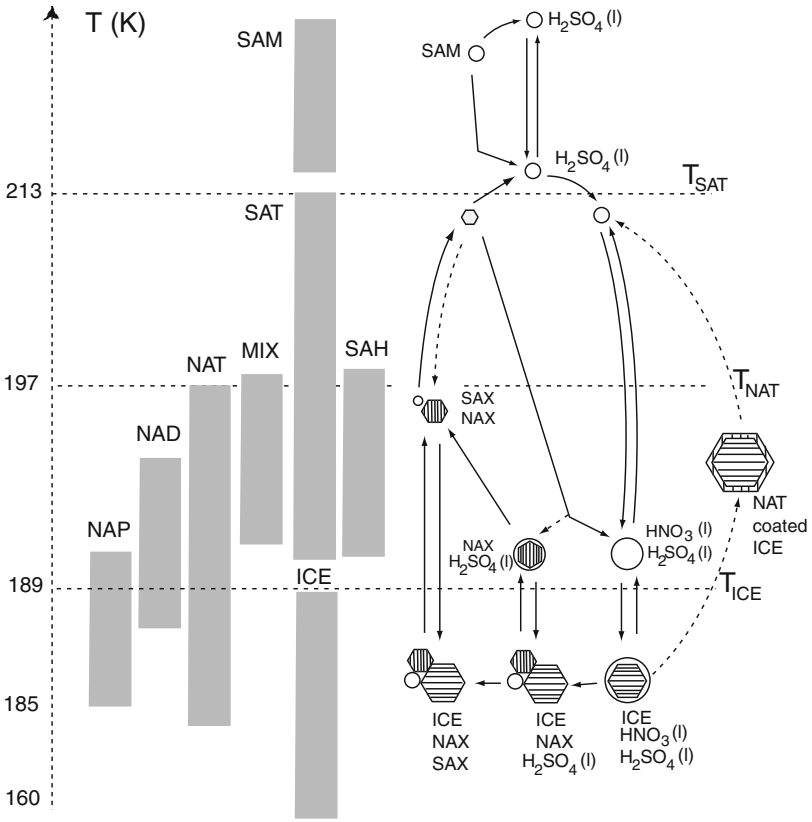



**Fig. 9.11** Stability regions of sulfuric acid and nitric acid. See text for explanation. Notice that the temperature scale is not linear. Scientific assessment of ozone depletion: (From WMO report n. 44, Scientific Assessment of Ozone Depletion)

and SAH is $H_2SO_4 \bullet 6.5\ H_2O$. NAT is for $HNO_3 \bullet 3H_2O$, NAX is again the generic nitric acid hydrate, NAD is $HNO_3 \bullet 2H_2O$, and NAP is $HNO_3 \bullet 5H_2O$.

MIX indicates a mixture of nitric and sulfuric acid according to $H_2SO_4$ $HNO_3$ $H_2O$. The (l) indicates a liquid phase. All the solid arrows indicate equilibrium phases and the dashed nonequilibrium phases.

The temperature scale is delimited by values where the different hydrates should form, starting from the highest (213 K) for SAT down to the colder 189 K for water ice. The scheme given in this figure is taken from a recent NASA report and is still very speculative. We will meet stratospheric aerosols again when dealing with stratospheric chemistry.

## 9.6 Clouds in Planetary Atmospheres

We will dedicate a later chapter to the planetary atmosphere, but it is worth to anticipate a few things about clouds in the atmospheres of the planet using some of the thermodynamics we have learned so far.

A good starting point is superimposing the phase diagram for water to the environmental conditions of the planets in the solar system as it appears in Fig. 9.12. From this, we get the indication that water clouds are in vapor phase on Earth, Venus, and Mars and mostly in the liquid phase on Earth, Jupiter, and Saturn.

To enter the detail, we will assume the general condition that clouds will form when the partial pressure of the condensing gas will be greater than the saturation pressure. We could use a more accurate form of the Clausius–Clapeyron equation studied in Chap. 1 (Eq. (1.34)). Equation (1.33) has to be integrated using a latent heat of condensation which depends on temperature. This can be obtained by a simple series expansion like

$$L = L_0 + \Delta\alpha T + \frac{\Delta\beta}{2}T^2 + O\left(T^3\right) \tag{9.52}$$

Equation (1.33) can be easily integrated:

$$\frac{de_s}{e_s} = \frac{1}{R_v}\left(L_0\frac{dT}{T^2} + \Delta\alpha\frac{dT}{T} + \frac{\Delta\beta}{2}dT\right)$$

Giving the saturation pressure dependence from temperature,

$$\ln e_s - \ln e_{s0} = \frac{1}{R_v}\left(-\frac{L_0}{T} + \Delta\alpha \ln T + \frac{\Delta\beta}{2}T\right) \tag{9.53}$$

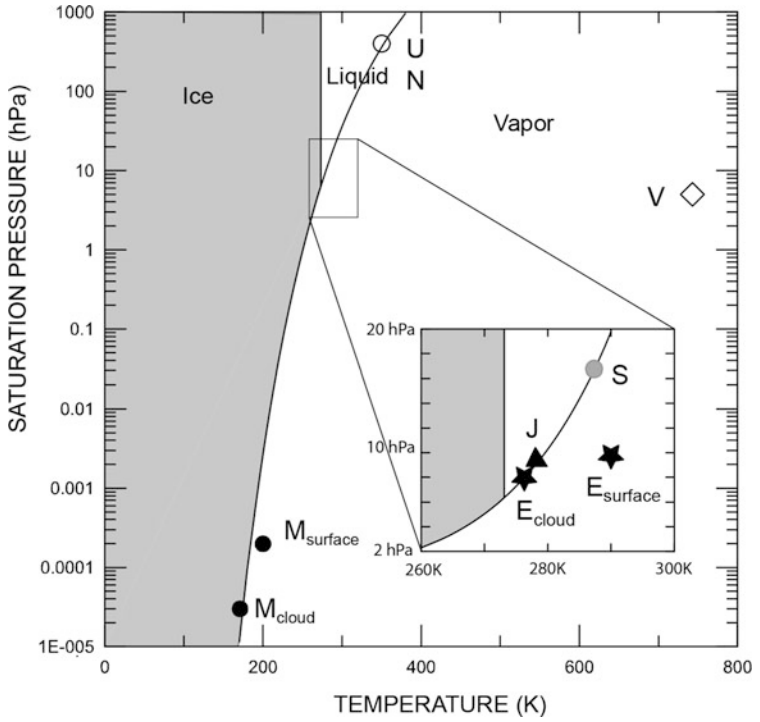



**Fig. 9.12** The phase diagram of water with the temperature–pressure range adequate for some planet of the solar system. *V* (Venus, surface), *E* (Earth surface and cloud level), *M* (Mars, surface and cloud level), *J* Jupiter, *S* Saturn, *U* Uranus, *N* Neptune (From Sanchez-Lavega et al. 2004)

**Table 9.3** Saturation vapor pressure and latent heat coefficient for some of the condensable gases in planetary atmospheres (From Sanchez-Lavega)

| Component | $\ln(e_{s0})$ | $L_0$ (J kg$^{-1}$) | $\Delta\alpha$ (J kg$^{-1}$ K$^{-1}$) | $\Delta\beta/2$ (J kg$^{-1}$ K$^{-2}$) |
|---|---|---|---|---|
| $H_2SO_4$ | 16.256 | 8.65 $10^6$ | | |
| $H_2O$ | 25.096 | 3.148 $10^6$ | | −8.7 |
| $CO_2$ | 26.1 | 639.6 | | −1.7 |
| $NH_3$ | 27.863 | 2016 | −8.88 $10^{-2}$ | |

(*)$e_{s0}$ in bar

The coefficients appearing in this equation are listed in Table 9.3 only for a few gases, that is, $H_2SO_4$ (Venus), $CO_2$ (Mars), $NH_3$ (Jupiter), and $H_2O$ for all the planets. Based on this equation, we have considered the atmosphere of Venus, Earth, Mars, and Jupiter. For each planet, we have assigned a mass mixing ratio for the gas and used the simple formula which relates the mass mixing ratio $\chi$ with the volume mixing ratio $e/p$:

$$\chi = \frac{M_v}{M_a}\frac{e}{p} \tag{9.54}$$

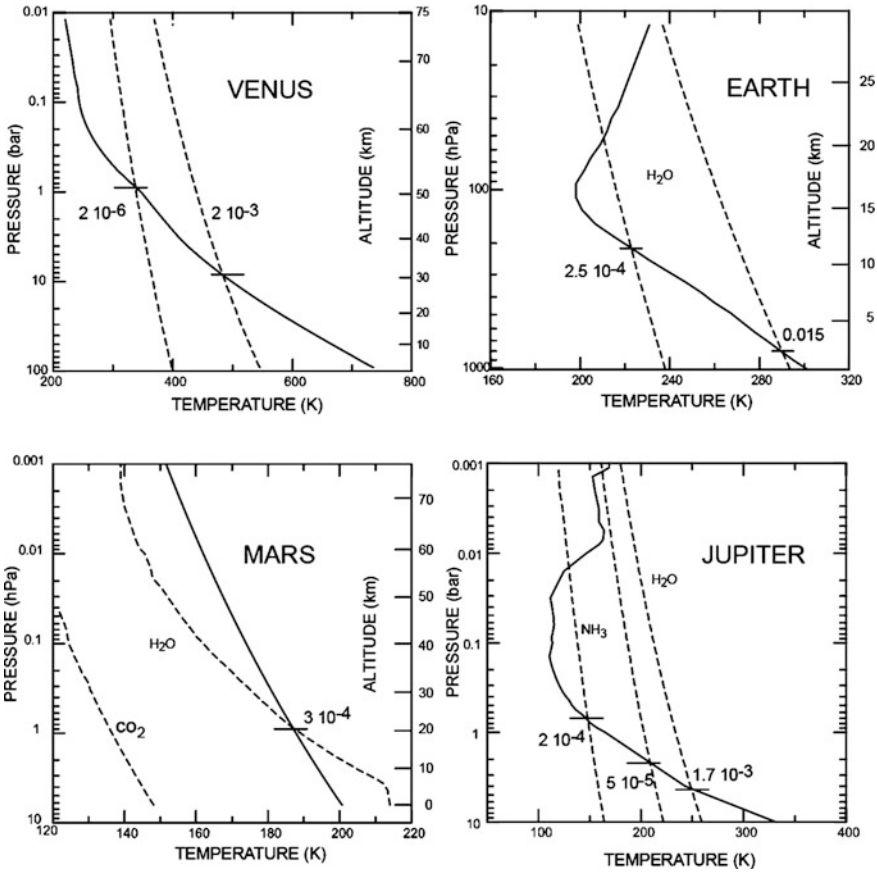



**Fig. 9.13** Condensation levels on four planets oft he solar system. See text for explanation (Adapted from Sanchez-Lavega)

where $M_v$ and $M_a$ are the molecular masses of the vapor and the atmosphere, respectively. For an assigned mixing ratio, the condensation level could be obtained for each planet as shown in Fig. 9.13. For each planet, the figure shows the temperature profile (solid line). For Venus, the condensation temperatures at two different mixing ratios for sulfuric acid ($2{\cdot}10^{-3}$, $2{\cdot}10^{-6}$) are given. The resulting condensation levels are at about 30 and 50 km, respectively, which are consistent with observed altitude range for clouds on Venus. For Earth, we have plotted the results for two different mixing ratios; the higher (15 g/kg) gives a very low condensation altitude. For Mars, besides water, we have plotted the possible condensation for carbon dioxide ice with a mixing ratio of 0.95. Apparently clouds of this kind will not form in the common environmental conditions of the atmosphere but may form in the polar atmosphere. For Jupiter, calculations have been performed for ammonia (NH3) and again this will form at pressure lower than 1 hPa.

This calculation can be further extended to have some idea about the thickness of the clouds. This can be done easily if we assume the simplified form of the Clausius–Clapeyron equation written for the cloud base:

$$e_s = e_{s,\mathrm{cl}} \exp\left(-\frac{L_i\,(T_{\mathrm{cl}} - T)}{R_v T T_{\mathrm{cl}}}\right)$$

where $e_{\,\mathrm{s,cl}}$ is the saturation pressure at the cloud base, $T_{\mathrm{cl}}$ is the temperature at the cloud base, and $L_i$ is the latent heat of the generic compound. If the temperature changes adiabatically, then we have

$$T_{\mathrm{cl}} - T = \left(g/c_p\right)\left(z_{\mathrm{cl}} - z\right)$$

Combining the last two expressions, we have

$$e_s = e_{s,\mathrm{cl}} \exp\left(-\frac{L_i\left(g/c_p\right)\left(z - z_{cl}\right)}{R_v T_{\mathrm{cl}}^2}\right) \tag{9.55}$$

so that we can define a vertical scale height for the cloud and its ratio with the atmospheric scale height:

$$H_{\mathrm{cl}} = \frac{R_v T_{\mathrm{cl}}^2 c_p}{g L_i} \Rightarrow \frac{H_{\mathrm{cl}}}{H} = c_p \frac{R_v}{R} \frac{T_{\mathrm{cl}}}{L_i} \tag{9.56}$$

From the data of Table (9.3), for the different planets, this ratio changes between 0.12 and 0.2 so that the vertical extent of the clouds is thin with respect to scale height. These are very simple considerations and refer to simple stratified clouds without taking into account any dynamical process. The mass per unit surface in the clouds can now be evaluated as

$$\rho_{\mathrm{cl}} H_{\mathrm{cl}} = \frac{\chi_{\mathrm{cl}} p_{\mathrm{cl}}}{g} \tag{9.57}$$

where $p_{\mathrm{cl}}$ is the atmospheric pressure at the condensation level and $p_{\mathrm{cl}}/g$ is the atmospheric mass which, multiplied by the mass mixing ratio $\chi_{\,\mathrm{cl}}$, gives the total condensed mass. From (9.57), the density of the cloud can be obtained.

Finally, we can return to the saturated adiabat as calculated in (E.1.4) and readapt the formula (E.1.6) for a generic component:

$$\Gamma_s = \frac{g}{c_p} \frac{\left(1 + \left(L_i w_s / RT\right)\right)}{\left(1 + \left(L_i^2 w_s / c_p R_v T^2\right)\right)} \tag{9.58}$$

Table 9.4 gives some data based on these simple calculations. These results will be reexamined when dealing with radiative transfer in the atmosphere.

**Table 9.4** Some cloud characteristics on the planets. $\chi_{vv}$, mixing ratio ($s =$ ice, $l =$ liquid); $P_{cl}$, pressure at the condensation level; $H_{cl}/H$, scale height ratio; $\rho_{cl}$, density of the cloud; $\Gamma_s/\Gamma_d$, ratio of the saturated to the dry adiabat

| | $\chi_{vv}$ | $P_{cl}$ ($10^5$ Pa) | $H_{cl}/H$ | $\rho_{cl}$(gcm$^{-3}$) | $\Gamma_s/\Gamma_d$ |
|---|---|---|---|---|---|
| *Venus* | | | | | |
| $H_2SO_4$(l) | $2\ 10^{-6}$ | 1.0 | 0.15 | $4.5\ 10^{-8}$ | 1 |
| $H_2SO_4$(l) | $2\ 10^{-3}$ | 11.3 | 0.22 | $2.4\ 10^{-4}$ | 0.88 |
| *Earth* | | | | | |
| $H_2O$ (s) | $2.5\ 10^{-4}$ | 0.30 | 0.13 | $5.3\ 10^{-7}$ | 0.957 |
| $H_2O$ (l) | 0.015 | 0.96 | 0.18 | $5.9\ 10^{-5}$ | 0.508 |
| *Mars* | | | | | |
| $CO_2$(s) | 0.95 | $2\ 10^{-4}$ | 0.17 | $4.6\ 10^{-6}$ | 0.17 |
| $H_2O$(s) | $3\ 10^{-4}$ | $1\ 10^{-3}$ | 0.14 | $2.5\ 10^{-9}$ | 0.87 |
| *Jupiter* | | | | | |
| $NH_3$ (s) | $2\ 10^{-4}$ | 0.75 | 0.13 | $1.6\ 10^{-6}$ | 0.967 |
| $H_2O$ | $5\ 10^{-5}$ | 3.2 | 0.12 | $1.2\ 10^{-6}$ | 0.990 |
| $H_2O$ | $1.7\ 10^{-3}$ | 5.7 | 0.17 | $4.4\ 10^{-5}$ | 0.867 |

## E.9 Examples

### *E.9.1 The Lognormal Size Distribution*

We start out with the expression for the lognormal distribution

$$\frac{dN}{d\ln r} = \frac{N_0}{\sqrt{2\pi}\ln\ \sigma}\exp\left\{\left[\frac{(\ln r - \ln R)^2}{2(\ln\ \sigma)^2}\right]\right\} \tag{E.9.1}$$

that can be easily written as

$$\frac{dN}{dr} = n(r) = \frac{N_0}{\sqrt{2\pi}r\ln\ \sigma}\exp\left\{\left[\frac{(\ln r - \ln R)^2}{2(\ln\ \sigma)^2}\right]\right\} \tag{E.9.1a}$$

and wish to understand the meaning of the two parameters $R$ and $\sigma$. We start out by integrating with respect to r to find the cumulative size distribution $N(r)$:

$$N(r) = \frac{N_0}{\sqrt{2\pi}\ln\ \sigma}\int_0^r \exp\left\{\frac{1}{r'}\left[\frac{(\ln r - \ln R)^2}{2(\ln\ \sigma)^2}\right]\right\} dr' \tag{E.9.2}$$

To evaluate the integral, we make the substitution:

$$y = (\ln r - \ln R)\,/\sqrt{2}\ln\sigma$$

And the integral becomes

$$N(r) = \frac{N_0}{\sqrt{\pi}} \int_{-\infty}^{y=(\ln r - \ln R)/\sqrt{2}\ln\sigma} \mathrm{e}^{-y^2} dy \tag{E.9.3}$$

If we define the error function,

$$erfz = \frac{2}{\sqrt{\pi}} \int_0^z \mathrm{e}^{-y^2} dy \tag{E.9.4}$$

The integral (E.9.3) can be written as

$$N(r) = \frac{N_0}{\sqrt{\pi}} \int_{-\infty}^{0} \mathrm{e}^{-y^2} dy + \frac{N_0}{\sqrt{\pi}} \int_0^{y=(\ln r - \ln R)/\sqrt{2}\ln\sigma} \mathrm{e}^{-y^2} dy$$

We get

$$N(r) = \frac{N_0}{2} + \frac{N_0}{2} erf\left(\frac{\ln(r/R)}{\sqrt{2}\ln\sigma}\right) \tag{E.9.5}$$

If we put $r = R$, we get

$$N(R) = N_0/2 \tag{E.9.6}$$

so that the average radius is such that half of the particles have a radius less than the average and the other half have a radius larger than the average. To understand the significance of $\sigma_{,}$ we consider a radius $r_\sigma$ such that $\sigma_g = r_\sigma/R$. For this radius, (E.9.5) becomes

$$N(r_\sigma) = \frac{N_0}{2}\left[1 + \mathrm{erf}\left(\frac{1}{\sqrt{2}}\right)\right] = 0.841 N_0$$

Thus $\sigma_g$ is the ratio of the radius below which 84 % of the particles lie to the median radius and is called the *geometric standard deviation*. For a monodisperse aerosol distribution, the median radius coincides with $r_\sigma$ and so $\sigma_g = 1$. Equation (E.9.5) also shows that $N_0 = N(\infty)$, that is, the total number of particles.

It is interesting to calculate the mean radius $\bar{r}$ for a lognormal distribution. According to the definition, we get

$$\bar{r} = \frac{1}{N_0} \int_0^\infty n(r) r dr$$

where $n(r)$ is given by (E.9.1a). We have

$$\bar{r} = \frac{1}{\sqrt{2\pi}\ln\sigma}\int_0^\infty \exp\left[-\frac{(\ln r - \ln R)^2}{2(\ln\sigma)^2}\right]dr$$

The integral can be evaluated with the result

$$\bar{r} = R\exp\left(\frac{\ln^2\sigma}{2}\right) \tag{E.9.7}$$

We see that mean radius of a lognormal distribution depends on both the average radius and the standard deviation.

## *E.9.2 A Few Things More About the Köhler Curve*

We have shown that Eq. (9.15) can be approximated by

$$\frac{e'}{e_s} \approx \exp(a/r)\left(1-\frac{b}{r^3}\right) \approx \left(1+\frac{a}{r}\right)\left(1-\frac{b}{r^3}\right) \approx 1+\frac{a}{r}-\frac{b}{r^3} \tag{E.9.8}$$

If we call this ratio supersaturation S, then the value of r which maximizes it is given by

$$dS/dr = -a/r^2 + 3b/r^4 = 0 \tag{E.9.9}$$

from which we get the value for $r^* = \sqrt{3b/a}$. Substituting this value in (E.9.8), we get

$$S^* \approx 1+\frac{a}{r^*}-\frac{b}{r^{*3}} = 1+\frac{1}{r^*}\left(a-\frac{b}{r^{*2}}\right) = 1+\left(\frac{4a^3}{27b}\right) \tag{E.9.10}$$

Figure E.9.1 shows an analysis of the Kőhler curve which reports also the values for r* and S*. The curve shown refers to a solution of 10–16 mass of ammonium sulfate.

In this case, the values are $S^* = 1.0051$ (0.51 %) and $r^* = 0.156$ μm. The figure shows that the solution effect dominates up to the supersaturation values. For radius smaller than r*, the solution droplet is in equilibrium with the environment so that if the relative humidity increases by a small amount, the droplet will grow to a larger radius. This process continues up to a relative humidity of 100 % or slightly over. If the particle grows beyond the critical radius, the saturation ratio starts to decrease and the particle will increase in size thanks to the diffusion of water vapor toward the drop.

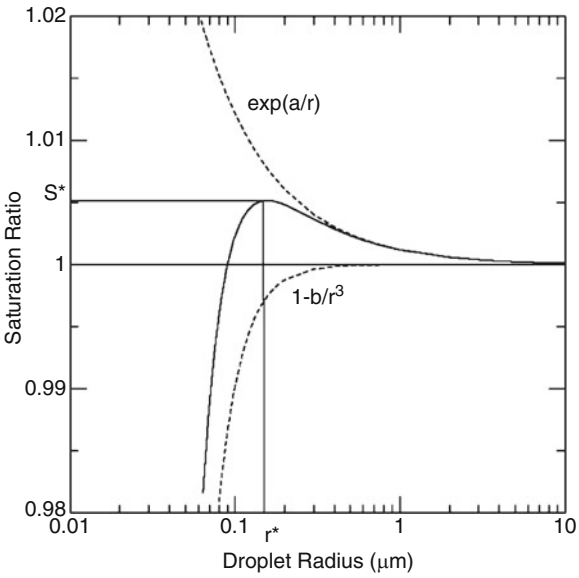



**Fig. E.9.1** Equilibrium saturation ratio of a solution droplet formed on a nucleus of ammonium sulfate with a mass 10–16 g. The values with asterisk are maximum saturation ratio ($S^*$) and the corresponding radius. The *dotted line curves* are the two terms in Eq. (E.9.8)

## *E.9.3 Sedimentation of Particles*

Particle sedimentation is treated very rarely because this process takes care of the removal of about 10 % of the atmospheric particles. The sedimentation velocity depends on two numbers; one is the Reynolds number:

$$\mathrm{Re} = 2\rho r v/\eta \tag{E.9.11}$$

where $r$ is the particle radius, $\rho$ is the atmospheric density, $v$ is the velocity, and $\eta$ is the dynamic viscosity of the atmosphere. The other is the Knudsen number:

$$Kn = \lambda/r \tag{E.9.12}$$

where $\lambda$ is the gas mean free path. Kn establishes whether the fluid can be treated as a continuum (Kn << 1) or by the statistical mechanics (Kn >> 1). We already know that the Reynolds number gives the difference between the laminar (Re << 1) and a turbulent regime (Re >> 1).

It may be convenient to introduce the time it takes for a particle to fall a scale height. We may start with the terminal velocity for a rigid sphere:

$$V = 2\rho_p g r^2/9\eta \tag{E.9.13}$$

this being valid for Re << 1 and Kn << 1. The sedimentation time for the fall is written as

$$\tau_{\text{fall}} = \frac{H}{V} = \frac{RT}{g}\frac{9\eta}{2\rho_p g^2 r^2} \tag{E.9.14}$$

For the regime with Re >> 70 and Kn << 1, we have

$$\tau_{\text{fall}} = H\left(\frac{3\rho}{40\rho_p g r}\right)^{1/2} \tag{E.9.15}$$

And finally for Kn >> 1, we have

$$\tau_{\text{fall}} = 27\pi\rho\frac{(2kT/\pi m)^{3/2}}{16\rho_p g^2 r} \tag{E.9.16}$$

In this case, k is the Boltzmann constant and m the mass of a single molecule. Notice that low Reynolds numbers correspond to small particle diameters (<20 μm), while moderate Reynolds numbers correspond to size between 20 μm and 1 mm, and large Reynolds numbers refer to particles between 1 and 5 mm. Drops up to 1 mm behave like rigid spheres, while for a larger dimension, the atmospheric resistance may modify the shape.

Most of the atmospheric aerosols have sizes comprised between 0.1 and 1 mm, and these particles have a time between 36 years and 3.28 days to fall 1 km. Typical volcanic particles have a diameter smaller than 5 mm and they will take about 2 weeks to fall 1 km.

## References[1]

### *Books*

Houze RA (1993) Cloud dynamics. Academic, San Diego
Jacob D (1999) An introduction to atmospheric chemistry. Princeton University Press, Princeton
Jacobson MZ (1999) Fundamentals of atmospheric modeling. Cambridge University Press, Cambridge
Rogers RR, Yau MK (1989) A short course in cloud physics. Butterworth/Heinemann, Oxford
Scientific assessment of ozone depletion: 1998, WMO, Report n. 44

[1]This chapter is the same as the previous edition except for the paragraph on planetary clouds where results are reproduced from the paper by Sanchez-Lavega and corrected. The main references here are the Jacobson, Rogers–Yau, and Hobbs books. The section on stratospheric particles is based on one of the NASA/WMO reports.

## *Articles*

Rasch PJ et al (2008) An overview of geoengineering of climate using stratospheric sulphate aerosols. Phil Trans R Soc A 366:4007–4037

Sanchez-Lavega A, Perez-Hoyos S, Hueso R (2004) Clouds in planetary atmospheres: a useful application of the Clausius–Clapeyron equation. Am J Phys 72:767

# Chapter 10
# Waves in the Atmosphere

In the previous chapters, we have already found wavelike dynamical manifestations in the atmosphere such as Rossby waves. These waves are a consequence of the conservation of absolute vorticity in the sense that we have considered the effect of Earth's rotation but not the presence of gravity.

Here, we will start from the other extreme and will consider a fluid in the gravity field but subject it to no rotation, and we will find a different class of waves. We are naturally interested in those waves whose frequency is comparable to the Brunt–Väisälä frequency that we found since Chap. 1 and which we may call buoyancy frequency. We neglect in this treatment the sound waves, a subject that should be familiar from elementary physics courses. In any case, sound waves usually have much higher frequencies.

Waves for which the "spring" or restoring force is buoyancy are called gravity waves. In contrast, Rossby waves are due to the existence of a vorticity gradient along the latitude. Another approximation we will employ is the linearization that we have already used to find a solution to the Rossby wave problem.

Everything will become clear as we go along, but for the moment we need to slow down a little and refresh our ideas on some of the properties of the waves.

## 10.1 Some Properties of the Waves

Phase and group velocities are the joy and nuisance of every physics student, although phase velocity may seem a simple thing. We know actually that for a wave propagating in $x$ direction, the phase can be expressed as $(kx - \omega t)$ so that the

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_10) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_10

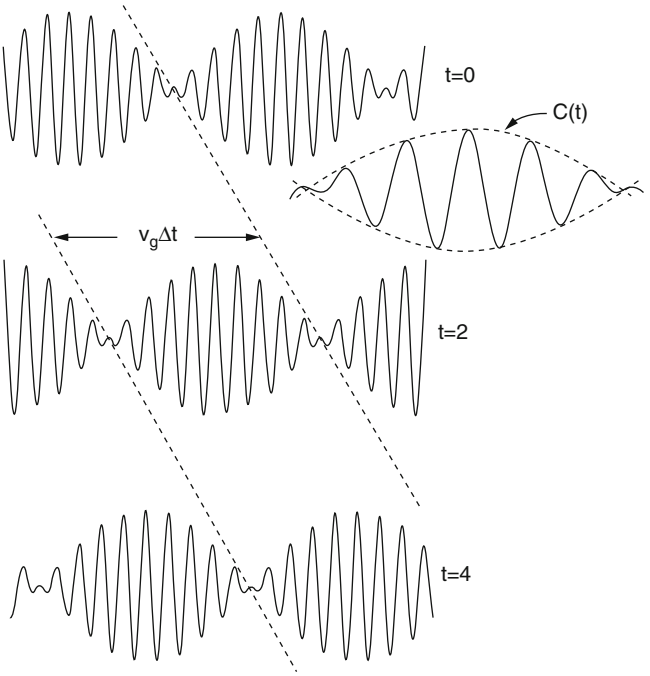



**Fig. 10.1** The graphic representation of a function given by (10.1) when the radio between high and low frequency is 10. The function is represented at regular time interval so that the movement can be related to the group velocity. The inset shows the envelope *C(t)* (Adapted from Lindzen 1990)

surface at constant phase will propagate with a velocity $\omega/k$. The group velocity can be introduced in a very simple way, for example, when studying the transmission of a low-frequency signal (a song on the radio) using a high-frequency carrier wave. We have for the signal

$$S = \cos \omega_c t + \frac{1}{2} \cos (\omega_c + \omega_m) t + \frac{1}{2} \cos (\omega_c - \omega_m) t \tag{10.1}$$

A visual representation of the signal (10.1) is given in Fig. 10.1 where the ratio between the two frequencies is 10. It can be clearly seen that the sinusoidal signal contains the highest frequency, while its amplitude is modulated at the lowest frequency. Actually, if a Fourier analysis of the signal (10.1) is carried out, we can see that it is composed of three frequencies, the carrier and two lateral bands that are at frequencies $((\omega_m - \omega_c))$ and $((\omega_m + \omega_c))$ with intensity $(1/2)^2$. We now consider the sum of two waves with different frequencies and wave numbers

$$\cos (k_1 x - \omega_1 t) + \cos (k_2 x - \omega_2 t) = 2 \cos [0.5 (\omega_1 - \omega_2) t - 0.5 (k_1 - k_2) x]$$
$$\times \cos [0.5 (\omega_1 + \omega_2) t - 0.5 (k_1 + k_2) x]$$

We also assume that $\omega_1 - \omega_2 = \Delta\omega << \omega_1$ and $(\omega_1 + \omega_2)/2 \approx \omega_1$ the same for $k$

$$\cos(k_1 x - \omega_1 t) + \cos(k_2 x - \omega_2 t) = 2\cos\left[\frac{(\Delta\omega t - \Delta k x)}{2}\right]\cos\left[\frac{(\omega t - k x)}{2}\right] \tag{10.2}$$

This may be interpreted as a wave propagation with a phase velocity $c = \omega/k$ and a modulated amplitude with maxima and minima that moves at the *group velocity* $v_g$

$$\upsilon_g = \frac{\Delta\omega}{\Delta k} \Rightarrow \frac{d\omega}{dk} \tag{10.3}$$

This very simple argument can be applied to a more general function. For example, the envelope of the wave in Fig. 10.1 is given by $\cos(\omega_c - \omega_m)t$ which can be represented by a Fourier transform

$$\cos(\omega_c - \omega_m)t = \int_{-\infty}^{+\infty} \delta(\omega(\omega_c - \omega_m)) e^{i\omega t} d\omega$$

where $\delta(\omega(\omega_c - \omega_m))$ is the Dirac function. The envelope function can then be written as

$$C(t) = \int_{-\infty}^{+\infty} B(\omega) e^{i\omega t} d\omega \tag{10.4}$$

so that the modulated function (or signal) can be written as in (10.1)

$$S(t, x) = \int_{-\infty}^{+\infty} B(\omega) e^{i[(\omega+\omega_0)t - k(\omega+\omega_0)x]} d\omega \tag{10.5}$$

If the function $B(\omega)$ is narrow enough, it can be shown (see, e.g., Lindzen)

$$\begin{aligned} S(t, x) &= \exp\{i[\omega_0 t - k(\omega_0)x]\} \int_{-\infty}^{+\infty} B(\omega) \exp\left\{i\left[\omega t - \frac{dk}{d\omega}\omega x + \right]\right\} d\omega \\ &= \exp\{i[\omega_0 t - k(\omega_0)x]\} C\left(t - \frac{dk}{d\omega}x\right) \end{aligned} \tag{10.6}$$

so that the envelope (i.e., *C(t)*) travels at the group velocity.

## 10.2 Gravity Waves in Shallow Water

As a first example of gravity waves, consider the waves produced in a shallow pond as sketched in Fig. 10.2. In this case, there are two incompressible fluids of different density, $\rho_1$ and $\rho_2$, one on top of the other and confined on a semi-infinite plane. The requirement of incompressibility eliminates the presence of sound waves.

Because the densities $\rho_1$ and $\rho_2$ are constant, the horizontal pressure gradient is independent from the depth so that we can write (Holton 1992)

$$\frac{\partial}{\partial z}\left(\frac{\partial p}{\partial x}\right) = -\frac{\partial \rho}{\partial x} g = 0 \tag{10.7}$$

If we limit ourselves to the study in the $x$, $z$, we can write the equation of motion in the lower layer of the form when we also assume no horizontal pressure gradient in the upper layer:

$$-\frac{1}{\rho_1}\frac{\partial p}{\partial x} = \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + w\frac{\partial u}{\partial z} \tag{10.8}$$

The horizontal pressure gradient arises as can be seen from Fig. 10.2 because of the different weight of fluid columns. For the lowest two points at the same height

$$p + \delta p_1 = p + \rho_1 g \Delta z;\; p + \delta p_2 = p + \rho_2 g \Delta_z$$

so that the horizontal gradient is given by

$$\frac{(p + \delta p_1) - (p + \delta \rho_2)}{\Delta x} = (\rho_1 - \rho_2)\, g \frac{\Delta z}{\Delta x} \Rightarrow g \Delta \rho \frac{\partial h}{\partial x} \tag{10.9}$$

that gives immediately the equation of motion

$$\frac{Du}{Dt} = -g \frac{\Delta \rho}{\rho_1} \frac{\partial h}{\partial x} \tag{10.10}$$

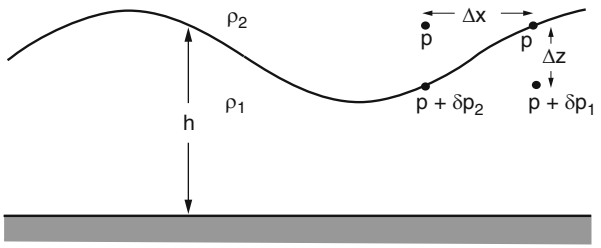



**Fig. 10.2** Waves in a two-layer fluid system

The equation of continuity on a vertical plane

$$\partial u/\partial x + \partial w/\partial z = 0$$

when integrated between 0 and $h$ with the assumption that $u$ is only a function of $x$

$$w(h) - w(0) = -h\partial u/\partial x$$

and assuming that the vertical velocity to be zero at the bottom

$$w(h) = \frac{Dh}{Dt} = \frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} = -h\frac{\partial u}{\partial x} \tag{10.11}$$

that can be differentiated with respect to $x$ to obtain

$$\frac{D}{Dt}\frac{\partial h}{\partial x} = -\frac{\partial}{\partial x}\left(h\frac{\partial u}{\partial x}\right) = -\frac{\rho_1}{g\Delta\rho}\frac{D^2u}{Dt^2}$$

At this point, we resort again to the perturbation method. Setting $h = H + h'$, $u = \overline{u} + u'$ and then neglecting second order terms, we have

$$\frac{D^2u'}{Dt^2} = -gH\frac{\Delta\rho}{\rho_1}\frac{D^2h'}{Dx^2} \tag{10.12}$$

This equation describes a wave that propagates with a velocity $c = (gH\Delta\rho/\rho_1)^{1/2}$. This should be added to the zonal background velocity $\bar{u}$ so that

$$c = \overline{u} \mp (gH\Delta\rho/\rho_1)^{1/2} \tag{10.13}$$

is the phase speed with respect an observer at rest in the coordinate system. In the case where the upper layer is air and the lower layer water, we can put $\Delta\rho = \rho_1$ so that

$$c = \overline{u} \pm (gH)^{1/2} \tag{10.14}$$

This should be the velocity of propagation of an ocean wave. If we assume an average depth of 4 km for this ocean, we have propagation velocities of the order of 200 $ms^{-1}$ which are really too high. The explanation of this rather strange result is that we have worked out a perturbation solution, that is, the vertical wavelength should be much larger than $H$, and this happens only for waves triggered by earthquakes (the famous *tsunami)*. However, waves can also be produced when the density differences are produced by differences in temperature as in the case of the thermocline. In this case, $\Delta\rho/\rho_1 \cong 0.001$ and the velocity is much more reasonable.

So far, we have neglected rotation, and the previous serves merely to illustrate the nature of gravity waves. Another point to stress is that gravity waves have nothing to do with gravitational waves. These are part of another sector of physics; it costs a lot more to detect them (actually, nobody has done that yet), and it does not cause as much damage as gravity waves, either marine or atmospheric.

## 10.3 Orographic Waves

One of the most striking examples of atmospheric waves is given by the alignment of clouds that give the impression of being the foam produced by waves breaking on the shore. This kind of wave is generated by a current flowing over a mountain range, as illustrated in Fig. 10.3.

The upper part of this figure shows how a westerly current flowing over a mountain range could trigger waves. The oscillations move the air upward and downward so that in the ridges there could be condensation and then formation of clouds on the wave ridges. Clearly, we are dealing with stationary waves, and in the reference system that moves with the wind, the surfaces of constant phase move away from the observer.

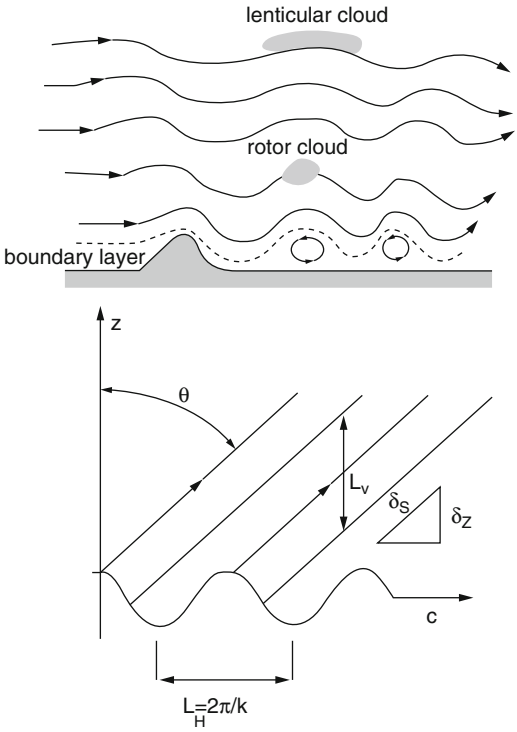



**Fig. 10.3** A general scheme of the generation of waves over orography. It is to note that the phase of the wave moves in the opposite direction of the wind. The *dashed line* limits the boundary layer. In the *lower panel*, lines at constant phase are shown when the corrugated profile moves with a velocity equal to the phase velocity $c$ (Lindzen 1970)

Another way to approach the problem is to consider the corrugated surface moving below the atmosphere. This is equivalent to studying the process in the reference system at rest with the wind. If we consider a parcel along one of the lines at constant phase, this will oscillate with a frequency that is proportional to the Brunt–Väisälä frequency $N$ because the restoring force along the vertical will be $F = -N^2\delta z$ so that the force in a direction forming the angle $\Theta$ with the vertical is

$$F = -N^2\delta z\cos\Theta$$

where $\delta z = \delta s\ \cos\Theta$ Along the direction $s$, we have the equation of motion

$$\frac{D^2\delta s}{Dt^2} = -N^2\cos^2\Theta\delta s$$

which represents an oscillation with frequency $(kc)^2 = (N\cos\Theta)^2$ that determines the angle $\Theta$. On the other hand, the same angle can be related to the horizontal and vertical wavelength considering that $\tan\Theta = L_H/L_v = l/k$. From these simple relations, it is possible to obtain the dispersion relation

$$\tan^2\Theta = \frac{1-\cos^2\Theta}{\cos^2\Theta} = \left(\frac{l}{k}\right)^2 = \frac{1-(kc/N)^2}{(kc/N)^2}$$
$$l^2 = \left(N^2/\sigma^2 - 1\right) \tag{10.15}$$

and $\sigma = kc$. Vertical propagation is possible if the vertical wave number is real which implies $N > \sigma$; otherwise, the buoyancy force is not enough to maintain the oscillation, and the amplitude of the wave decreases with altitude. At this point, the components of the phase and group velocities are easily obtained. First of all, we calculate $\sigma$

$$\sigma = \pm Nk\left(k^2 + l^2\right)^{-1/2} \tag{10.16}$$

then applying the rules we learned previously

$$\begin{aligned} c_{px} &= \sigma/k = \pm N\left(k^2+l^2\right)^{-1/2} \qquad c_{py} = \sigma/l = \pm Nk\left[l\left(k^2+l^2\right)^{1/2}\right]^{-1} \\ c_{gx} &= d\sigma/dk = \pm Nl^2\left(k^2+l^2\right)^{-3/2}\ c_{gz} = d\sigma/dl = \pm Nkl\left(k^2+l^2\right)^{-3/2} \end{aligned} \tag{10.17}$$

Before we examine in detail these relations, it is useful to remember that the phase velocity is not a vector while the phase propagation is determined by the wave vector whose components in this case are $k$ and $l$. Also, the phase velocity we have obtained is relative to the air, that is, in the presence of wind $\bar{u}$, we just add the background wind.

The positive sign is intended for eastward propagating waves with respect to the average wind. A reason for this choice of sign is that if the waves are stationary,

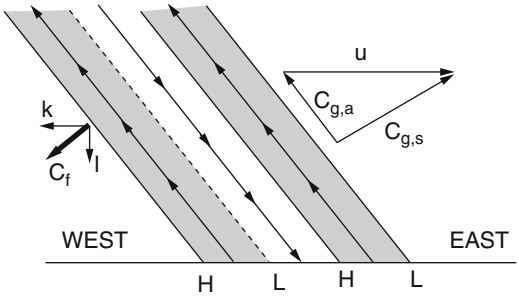



**Fig. 10.4** The relation between phase and group velocity and the air movement. The phase propagation is shown with the *thick arrow*. The shade zones are those in which parcels move away from the ground creating the alternating high ($H$) and low ($L$) pressures. The composition of the group velocity is also shown

then the phase velocity must be zero with respect to the orography so that the phase propagates toward the west with respect to the wind. At this point, things start to be interesting because we see that constant phase lines are tilted westward (see Fig. 10.3), and this requires that $k < 0$ and $l < 0$. We notice that the phase is simply $\phi = kz + ly$ so that increasing $x$ means to decrease $z$ in order for $\phi$ to remain constant. This means the phase propagates downward as shown in Fig. 10.4. The energy on the other hand must propagate from below because the energy source is at the surface. It can be seen that the same choice of wave vectors implies that the group velocity vector lies along the lines of constant phase. Therefore, the group velocity is normal to the direction of propagation of the phase. However, it can be seen from Fig. 10.4 that with respect to an observer at rest with respect to the surface the group, velocity forms a small angle with respect to the horizontal.

## 10.4 Internal Gravity Waves

Everything we have introduced so far, based on intuitive considerations, can be obtained through analytical method by recurring to the equation of motions that give a generalization of the solution at what is known as *internal gravity waves*.

When we have introduced the shallow water waves, we should have noticed that those are a particular type of waves known as *surface waves*, and these waves cannot propagate vertically. A more general class of waves is the so-called internal waves whose phase may propagate vertically. However, we will also study these waves on the vertical plane, and the starting point will be the component of the equations of motion in the horizontal and vertical directions and the continuity equation. Again, we write the equation of motion

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + w\frac{\partial u}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial x}$$

This can be linearized by putting $u = \overline{u} + u'$; $p = \overline{p} + p'$; $\rho = \overline{\rho} + \rho'$; $w = w'$ so that by neglecting higher order terms, we get (in this case, $\bar{u}$ does not depend on height)

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right) u' = -\frac{1}{\overline{\rho}}\frac{\partial p'}{\partial x} \tag{10.18}$$

The vertical equation of motion is obtained considering that while there is hydrostatic equilibrium on the average values, fluctuations in the potential temperature $\theta'$ may produce buoyancy accelerations of the order of $g\theta'/\theta$ so that the net vertical acceleration is given by

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right) w' + \frac{1}{\overline{\rho}}\frac{\partial p'}{\partial z} - g\frac{\theta'}{\overline{\theta}} = 0 \tag{10.19}$$

We are then bound to a rather inevitable complication, that is, the appearance of the potential temperature. The only way to get rid of it is to recur to the thermodynamic equation. This is particularly simple for adiabatic motions so that we have $D\ln\theta/Dt = 0$ and then

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right) \theta' + w'\frac{\partial \overline{\theta}}{\partial z} = 0 \tag{10.20}$$

where we have taken into account that $D\theta/Dt = 0$ for adiabatic motions. At this point, we can eliminate two variables, that is, pressure and $u'$. We differentiate (10.18) with respect to $z$ and (10.19) with respect to $x$ and subtract. We then use the continuity equation and obtain

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)^2 \left(\frac{\partial^2 w'}{\partial x^2} + \frac{\partial^2 w'}{\partial z^2}\right) + N^2\frac{\partial^2 w'}{\partial x^2} = 0 \tag{10.21}$$

where we have used $N^2 = \left(g/\overline{\theta}\right)\partial\overline{\theta}/\partial z$. We seek a solution of the form

$$w' = \mathrm{Re}\left[\widehat{w}e^{i\varphi}\right] \tag{10.22}$$

where $\hat{w}$ is a complex quantity and where the phase is given by $\varphi = kx + lz - \omega t$.

In this case, the horizontal wave number is real, while we should require that $l$ be complex in order to study the vertical propagation. In any case, by substituting Eq. (10.22) into (10.21), we obtain the dispersion relation

$$\widehat{\omega} = \omega - \overline{u}k = \pm Nk\left(k^2 + l^2\right)^{-1/2} \tag{10.23}$$

The quantity $\widehat{\omega}$ is the wave frequency with respect to the average wind so that the phase velocities relative to the wind in the two directions are given by $c_{px} = \widehat{\omega}/k$

and $c_{px} = \widehat{\omega}/l$ that coincide with those given before. In the same way, we could find the expressions for the group velocities.

It is interesting to note that the solution just found includes also the orographic waves. Considering that they are stationary waves, Eq. (10.21) becomes

$$\left(\frac{\partial^2 w'}{\partial x^2} + \frac{\partial^2 w'}{\partial z^2}\right) + \frac{N^2}{\overline{u}^2} w' = 0 \tag{10.24}$$

Assuming a phase like $\phi = kx + lz$, the dispersion relation becomes

$$k^2 + l^2 = N^2/\overline{u}^2 \tag{10.25}$$

which is the same relation that can be obtained from (10.17) when $c_{px} = 0$ or from (10.23) by putting $\omega = 0$. Equation (10.25) can be useful for another reason because in some cases, we may have $l^2 < 0$, that is, there is no vertical propagation of the wave because $l$ is imaginary. The solution then becomes

$$w' = \widehat{w} e^{ikx} e^{-lz} \tag{10.26}$$

This equation confirms that the energy propagates upward, and the amplitude can be found by imposing a forcing on the ground. For a sinusoidal forcing, we have $h = h_m \cos kx$ so that

$$w'(x, 0) = (Dh/Dt)_{z=0} \approx \overline{u}\partial h/\partial x = \overline{u} k h_m \sin kx \tag{10.27}$$

The real part in (10.26) evaluated at $z = 0$ (considering that $\hat{w}$ is a complex quantity) is given by

$$w'(x, 0) = \widehat{w}_r \cos kx - \widehat{w}_i \sin kx \tag{10.28}$$

compared with Eq. (10.27) gives

$$\widehat{w}_i = -\overline{u} k h_m; \quad \widehat{w}_r = 0$$

so that the solution becomes

$$w' = -\overline{u} k h_m e^{-lz} \sin kx \tag{10.29}$$

Similarly for the vertical propagation, we may write

$$w' = -\overline{u} k h_m \sin(kx + lz) \tag{10.30}$$

A vertical trapped wave is found when $\overline{u}k > N$ as can be easily seen from Eq. (10.25) because in that case the vertical wave number is imaginary. Vertical propagation is obtained for the opposite condition. The propagation of the wave

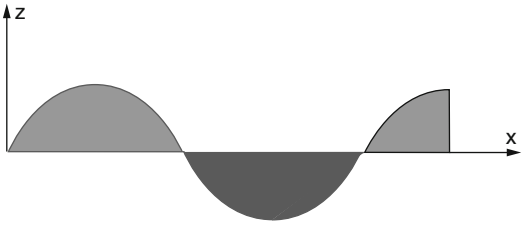



**Fig. 10.5** Example of net transport due to wave breaking. A tracer is transported *upward* (*clear shadow*) or *downward* (*dark shadow*) so that in an even number of cycles, there is no net transport. This happens only when the wave breaks (Adapted from Lindzen 1970)

is then determined by the characteristics of the topography once the wind and the frequency are fixed.

Up to now, we have learned a number of nice things concerning mainly the phase and group velocity, but we are still stuck with the old question: what do we do with gravity waves? We will see later on that these are so ubiquitous that we will need means to rule them out, and mainly gravity waves or waves in general are very important when dealing with transport processes.

To fix the ideas, let us consider waves in the sea. Far away from the shore, these are quite regular and of the same amplitude. When they approach the shore because they move in less and less water, their amplitude grows to the point that the wave becomes unstable and breaks, forming a billow. Until the wave is regular, a passive tracer (like junk) is simply moved around, for example, in the vertical direction as shown in Fig. 10.5. However, when the wave breaks, we have net transport.

In the case of the atmosphere, a gravity wave may produce oscillation in the temperature so that the lapse rate is also changed. If the amplification of the wave occurs, such a gradient may become unstable and cause the wave to break.

Another important property of gravity waves is that they can transport momentum (i.e., the correlation $u'w' \neq 0$) and so explain a few things about the motion in the upper atmosphere or in the boundary layer. However, to take into account all these effects, we need to introduce a vertical shear for the zonal velocity and the rotation. Some of this work can be done through the introduction of the three-dimensional Rossby waves.

## 10.5 Three-Dimensional Rossby Waves

The starting point for the introduction of the three-dimensional Rossby waves is the quasi-geostrophic vorticity (Eq. 7.61). In that equation, the horizontal divergence has been replaced by the change of the vertical velocity with height so that

$$\frac{D}{Dt}(\zeta + f) = (\zeta + f)\frac{\partial w}{\partial z} \tag{10.31}$$

To solve this equation, we have to eliminate the vertical velocity. Again, we write a perturbation form of the (10.31) by putting $u = \overline{u} + u'$, $\upsilon = \upsilon'$, $w = w'$, so that

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\zeta' + \upsilon'\frac{\partial f}{\partial y} - f\frac{\partial w'}{\partial z} = 0 \tag{10.32}$$

We then consider the thermodynamic Eq. (10.20) and divide it through $\overline{\theta}$

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\theta'}{\overline{\theta}} + \frac{w'}{\overline{\theta}}\frac{\partial\overline{\theta}}{\partial z} = 0 \tag{10.33}$$

The ratio between the perturbation and the average value of the potential temperature can be expressed as

$$\theta'/\overline{\theta} = \rho'/\overline{\rho}$$

This is tantamount to assuming that the only changes in temperature arise from buoyancy effects, and this is known as *Boussinesq approximation.* By substituting in Eq. (10.33), we get

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\rho'}{\overline{\rho}} - \frac{N^2}{g}w' = 0 \tag{10.34}$$

This equation can be used to eliminate the vertical velocity, although as a further assumption, we need to consider that the density fluctuations are also in hydrostatic equilibrium:

$$\frac{1}{\overline{\rho}}\frac{\partial p'}{\partial z} + \frac{\rho'}{\overline{\rho}}g = 0 \tag{10.35}$$

We can now differentiate (10.34) with respect to $z$ and substitute the result in Eq. (10.32), while to eliminate density, we use (10.34)

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\left(\zeta' + \frac{f}{\overline{\rho}N^2}\frac{\partial^2 p'}{\partial z^2}\right) + \upsilon'\beta = 0 \tag{10.36}$$

where $\beta$ is the latitudinal gradient of the Coriolis parameter. Introducing the streamfunction $\psi = p'/f\overline{\rho}$ and substituting in Eq. (10.36), we get

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\left(\nabla^2\psi + \frac{f^2}{N^2}\frac{\partial^2\psi}{\partial z^2}\right) + \beta\frac{\partial\psi}{\partial x} = 0 \tag{10.37}$$

This expression is known as the *quasi-geostrophic vorticity equation* because, except for the divergence term, we use the value of the geostrophic wind. Actually,

the meaning of this equation is far more reaching. We can define a potential vorticity as

$$q = \overline{q} + q' = \nabla^2 \psi + f + \frac{f^2}{N^2} \frac{\partial^2 \psi}{\partial z^2} \tag{10.38}$$

so that, because of (10.37), we have $\partial \overline{q}/\partial y = \beta$ and obtain the conservation equation

$$\left( \frac{\partial}{\partial t} + \mathbf{V}_g \cdot \nabla \right) q = 0 \tag{10.39}$$

We had already defined the potential vorticity in Eq. (7.36). We can see that this definition is equivalent because it can be shown to be a particular form of the Ertel potential vorticity:

$$(\zeta + f)\, \partial \theta / \partial z$$

Actually, when the difference in potential temperature is fixed, the potential vorticity will change as the distance between the corresponding isentropic surfaces changes. Consider now a perturbation of the potential temperature such that $\theta = \overline{\theta} + \theta'$. We have then for the Ertel potential vorticity

$$(\zeta + f)\, \theta_z = \overline{\theta}_z\, (\zeta + f) \left( 1 + \theta'_z / \overline{\theta}_z \right) \tag{10.40}$$

where subscripts indicate derivatives. At this point, we notice that $\overline{\theta}_z = N^2 \overline{\theta}/g$ and also $\theta'_z/\overline{\theta}_z = -\rho'_z/\overline{\rho}$. Substituting in (10.40), we have

$$\overline{\theta}_z\, (\zeta + f) \left( 1 + \theta'_z / \overline{\theta}_z \right) = \overline{\theta}_z\, (\zeta + f) \left( 1 - \frac{g}{N^2} \frac{\rho'_z}{\overline{\rho}} \right)$$

And using hydrostatic equilibrium, we have the equivalent of Eq. (10.37)

$$\overline{\theta}_z\, (\zeta + f) \left( 1 - \frac{g}{N^2} \frac{\rho'_z}{\overline{\rho}} \right) = \overline{\theta}_z\, (\zeta + f) \left( 1 + \frac{1}{N^2 \overline{\rho}} \frac{\partial^2 p'}{\partial z^2} \right)$$

If we consider that at middle latitudes $\zeta << f$ we have a form similar to (10.38)

$$\overline{\theta}_z\, (\zeta + f) \left( 1 + \frac{1}{N^2 \overline{\rho}} \frac{\partial^2 p'}{\partial z^2} \right) \approx \overline{\theta}_z \left( \zeta + f + \frac{f^2}{N^2} \frac{\partial^2 \psi'}{\partial z^2} \right)$$

Equation (10.39) will be used when dealing with the stratosphere. For the moment, we are interested to find a solution for (10.37) that we may assume to be given by

$$\psi = \mathrm{Re}\left[ \psi_0 \exp i\, (\omega t + kx + ly + mz) \right]$$

which upon substitution in Eq. (10.37) gives the dispersion relation

$$\omega = -\overline{u}k + \beta k / \left[\left(k^2 + l^2\right) + f^2 m^2 / N^2\right] \tag{10.41}$$

For stationary waves ($\omega/k = 0$), vertical propagation occurs only if $m^2 > 0$

$$m^2 = \frac{N^2}{f^2}\left[\beta/\overline{u} - \left(k^2 + l^2\right)\right] > 0$$

that is equivalent to

$$0 < \overline{u} < \beta / \left(k^2 + l^2\right) \tag{10.42}$$

This means that propagation of stationary waves is possible only for westerly winds less than a certain value. During summer, winds are easterly in the stratosphere, and Eq. (10.42) is not satisfied. During winter, however, winds are westerly and they can be weak enough to allow propagation of the longest waves (wave number 1 and 2).

This result was found for the first time by Jule Charney and Peter Drazin at the beginning of the 1960s and suggested an answer to a very important question. The very low probability of such events saves the Earth's atmosphere from a very fast evaporation. As stated in the introduction of the paper by Charney and Drazin:

> If the large – scale tropospheric motions were to propagate in this manner, then, because of their vastly greater energy, an atmospheric corona would be in all likelihood be produced. The kinetic energy density of the lower troposphere is of the order of $10^3$ erg cm$^{-3}$. If this energy were to travel upward with little attenuation and be converted into heat by friction or some other means at, say 100 km were the density is diminished by a factor of $10^{-6}$, it would raise the air temperature at 100,000 K. At such temperatures most of the atmosphere would escape the earth's gravitational field . . . .

Since that paper, more realistic calculations have shown that the effect is much less dramatic. These calculations refer to propagation in two-dimensional space (latitude–height) where the zonal wind is now also a function of the two variables. Also, starwarms (as they are called) must be intended as a rearrangement of potential vorticity and temperature, while the effect of energy dissipation is manifested at high latitude.

Even in such rare events, conditions are such that some energy penetrates the stratosphere, warming the region even by 40 °C in a few days. When this happens, we say we have a *sudden stratospheric warming.* Stratospheric warmings have a strong influence on the dynamics of the lower stratosphere and even some relationship with the occurrence of the ozone hole in the southern hemisphere. We will discuss these phenomena in some length in later chapters.

The energy absorption from propagating waves requires a few more considerations. In our calculations, in order to keep things simple, we have considered the density independent of altitude. However, a very simple correction can be introduced though a term like $\exp(z/2H)$.

This takes into account the fact that kinetic energy of the waves is proportional to the product of the density for the square of the velocity. Energy conservation then requires that the amplitude of the wave grow exponentially with height. We plan to explore matters further concerning wave absorption. In the process, we will discover that this will help us understand more about turbulence in the upper atmosphere, and at the end we will be very near to explaining why the major gases are well mixed through all the atmosphere up to about 100 km. This requires we go deeper in the physics of gravity waves.

## 10.6 The Physics of Gravity Waves

For what we have to say in the following paragraphs, it will be more convenient to write the equation of quasi-geostrophic potential vorticity in the log-pressure coordinate system that was introduced in Chap. 6 when dealing with the thermal wind equation. We will start by giving some more details on this coordinate system that is very convenient when studying the dynamics of the stratosphere and the physics of the waves.

### *10.6.1 The Equation of Quasi-geostrophic Potential Vorticity*

We will not repeat here Sect. 7.5.1 up to the point where we define

$$q = \overline{q} + q' \tag{10.43}$$

where

$$q' = \nabla^2 \psi + \frac{1}{\rho_0} \frac{\partial}{\partial z^*} \left( \frac{f^2 \rho_0}{N_*^2} \frac{\partial \psi}{\partial z^*} \right) ; \overline{q} = f \tag{10.44}$$

where the latter obeys to the equations

$$\frac{\partial \overline{q}}{\partial y} = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} - \frac{1}{\rho_0} \frac{\partial}{\partial z^*} \left( \frac{f^2 \rho_0}{N_*^2} \frac{\partial \overline{u}}{\partial z^*} \right) \tag{10.45}$$

and

$$\left( \frac{\partial}{\partial t} + \overline{u} \frac{\partial}{\partial x} \right) q' + \frac{\partial \overline{q}}{\partial y} \frac{\partial \psi}{\partial x} = 0 \tag{10.46}$$

Equivalent to the conservation equation

$$Dq/Dt = 0 \tag{10.47}$$

we have shown (Eq. 7.82) that this is equivalent to the conservation equation

$$\left(\partial/\partial t + \overline{u}\partial/\partial x\right) q' + \upsilon' \partial\overline{q}/\partial y = 0 \tag{10.48}$$

for which we assume a solution of the form

$$\psi = \Psi\left(z^*\right) \exp\left(z^*/2H\right) \exp\left[i\left(kx + ly - \omega t\right)\right] \tag{10.49}$$

which is a wave propagating in the direction $x$ with an amplitude that depends on height. Substituting this solution in Eq. (10.48), we obtain the "vertical structure" equation

$$\partial^2\Psi/\partial z^{*2} + m^2\Psi = 0 \tag{10.50}$$

Where

$$m^2 = \left(N_*/f\right)^2 \left[\left(\overline{u} - c\right)^{-1} \partial\overline{q}/\partial y - \left(k^2 + l^2\right)\right] - 1/4H^2 \tag{10.51}$$

The vertical wave number gives conditions for propagation that are a little bit different from those of the stationary three-dimensional Rossby waves. In this case, we have propagation if $m^2 > 0$ but at the level for which $\overline{u} = c$, the solution has a singularity and the wave tends to be absorbed: the altitude where this occurs is called *the critical level*. The reason for the absorption is that through this level, $m^2$ goes from positive to negative values and then the amplitude of the wave decays exponentially. This is a preliminary illustration of the interaction between waves and zonal flow. We will return to this point later in the chapter because we need to introduce now one of the most important concepts of atmospheric dynamics.

### *10.6.2 The Eliassen–Palm Flux*

When a wave is absorbed, as we will see, there may be momentum transfer and the wave may be dissipated. In this case, heat can also be produced through friction. To fix the ideas, we can multiply Eq. (10.48) through $q'$ to obtain

$$\left(\partial/\partial t + \overline{u}\partial/\partial x\right)\left(q'^2/2\right) + \left(q'\upsilon'\partial\overline{q}/\partial y\right) = 0$$

Averaging this equation zonally and assuming a stationary solution and $\partial\overline{q}/\partial y \neq 0$ we must have

$$\overline{q'\upsilon'} = 0 \tag{10.52}$$

that is, without friction and for stationary waves, the meridional flux of the potential vorticity is zero. This is a different way to express the famous Eliassen and Palm first theorem: that is, in case of stationary waves in the absence of dissipation, there is no interaction between waves and zonal flow. It can be shown through a number of manipulations (which we show in the Appendix) that Eq. (10.52) can also be expressed as

$$\rho_0 \overline{q'\upsilon'} = \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z^*} \tag{10.53}$$

where we have let

$$E_y = \rho_o \overline{\frac{\partial \psi}{\partial y}\frac{\partial \psi}{\partial x}} = -\rho_o \overline{u'\upsilon'};\ E_z = \rho_o \frac{f^2}{N_*^2}\overline{\frac{\partial \psi}{\partial x}\frac{\partial \psi}{\partial z^*}} = \rho_o \frac{fg}{N_*^2 \overline{\theta}}\overline{\upsilon'\theta'} \tag{10.54}$$

and where we have utilized $u' = -\partial\psi/\partial y,\ \upsilon' = \partial\psi/\partial x$. It can also be shown that $\partial\psi/\partial z^* = \theta' g/\left(f\overline{\theta}\right)$, $E_y$, and $E_z$ are the component of the *Eliassen–Palm (EP) flux*. Equation (10.59) states that the meridional flux of the potential vorticity is zero when the divergence of the EP flux is zero, and this happens when the waves are stationary and the background flux is purely zonal:

$$\rho_o \overline{q'\upsilon'} = \nabla \cdot \mathbf{E} = 0 \tag{10.55}$$

This equation can be actually generalized including the case when the divergence is not zero. The most obvious way is to write a general form of the conservation equation

$$\partial A/\partial t + \nabla \cdot \mathbf{E} = D \tag{10.56}$$

where $A$ is a quantity called wave action and $D$ is a term of net production that includes dissipation and adiabatic effects. Equation (10.56) can now be compared with (10.52) which we can write in a more general form

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right) q' + \upsilon' \overline{q}_y = Z' \tag{10.57}$$

where $q_y^{'}$ indicates the derivative of the average value and $Z'$ is a source term. We multiply (10.64) by $\rho_0 q'/\overline{q}_y$ to obtain

$$\rho_0 \left(q'/\overline{q}_y\right) \frac{\partial q'}{\partial t} + \rho_0 \overline{u} \left(q'/\overline{q}_y\right) \frac{\partial q'}{\partial x} + \upsilon' \rho_0 \left(q'/\overline{q}_y\right) \overline{q}_y = Z' \rho_0 q'/\overline{q}_y$$

In the zonal averaging, the second term disappears so that

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\rho_0 \frac{\overline{q'^2}}{\overline{q}_y}\right) + \rho_0 \,\overline{v'q'} = \rho_0 \,\overline{Z'q'}/\overline{q}_y$$

with the second term representing the divergence of the EP flux. A comparison with (10.56) gives

$$A = \frac{1}{2}\rho_0 \frac{\overline{q'^2}}{\overline{q}_y}; \; D = \rho_0 \overline{Z'q'}/\overline{q}_y \tag{10.58}$$

We will have other occasions to talk about wave action or wave activity, and for the moment we want to concentrate a little more on the physical meaning on the EP flux. Going back to Eq. (10.53) and assuming $q = \overline{q} + q'$, we can write

$$\frac{\partial\,(\overline{q} + q')}{\partial t} + \overline{u}\frac{\partial\,(\overline{q} + q')}{\partial x} + u'\frac{\partial\,(\overline{q} + q')}{\partial x} + v'\frac{\partial\,(\overline{q} + q')}{\partial y} = 0$$

and zonally averaging, we get

$$\frac{\partial \overline{q}}{\partial t} + \frac{\partial\,(q'v')}{\partial y} = 0 \tag{10.59}$$

In Eq. (10.56), we can substitute for the action the definition (10.65) and for the divergence of the EP flux (10.55). We get with $D = 0$

$$\overline{q'v'} = -\frac{1}{2\overline{q}_y}\frac{\partial\left(q'^2\right)}{\partial t} \tag{10.60}$$

If we now consider a wave whose amplitude is changing (i.e., $\overline{q'^2}$ changes), this implies a meridional flux of potential vorticity according to (10.60). For example, at high latitude, the zonally averaged gradient of the potential vorticity is positive, while the gradient of the flux is negative because the flux of potential vorticity must vanish at the pole. For a growing wave, this means, based on (10.60), a negative potential vorticity flux (directed toward the equator) so that according to Eq. (10.60), the zonally averaged potential vorticity decreases in time which means the zonally averaged westerly flow must decrease in time.

These considerations show that the EP flux acts directly on the zonal flux. We will return to this topic when dealing with the general circulation of the atmosphere, but we will see that the interaction between waves and mean flow is of primary importance in the stratosphere. Another very important point is that the zonal mean potential vorticity is influenced by the divergence of the EP flux and not by its absolute value. This means the EP flux could be very large and yet have no influence on the background state. For this reason, it would be very convenient to have a reference system where the effects of the eddies could be minimized. By the way, the system we have used so far is called the *Eulerian mean reference system*.

### 10.6.3 Energetics of Gravity Waves

So far, most of the concepts we have introduced like the Eliassen–Palm flux or the wave action remain rather remote from the physical point of view. A good starting point to grasp the physics underlying these concepts is through the energy transported by the waves. Before considering the energy of the waves, we rewrite again the equation of motion in our log-pressure coordinate system in the perturbation form. We drop all the asterisks and assume that $\overline{u} = \overline{u}(z)$ and restrict everything in the $x, z$ plane (assuming $v' = 0$). We have for the equation of motion

$$\frac{\partial u'}{\partial t} + \overline{u}\frac{\partial u'}{\partial x} + w'\frac{\partial \overline{u}}{\partial z} + \frac{\partial \Phi'}{\partial x} = 0 \tag{10.61}$$

In this equation, the pressure gradient has been replaced by the geopotential gradient. The geopotential $\Phi'$ is related to the streamfunction $\psi'$ by the relation $\Phi' = f\psi'$. The zonal wind is forced by the divergence of the EP flux.

$$\frac{\partial \overline{u}}{\partial t} = -\frac{1}{\rho_0}\frac{\partial}{\partial z}\left(\rho_0 \overline{u'w'}\right) \tag{10.62}$$

We have not treated this particular term, although it represents vertical transport of zonal eddy momentum (as in the boundary layer).

The thermodynamic equation (obtained in the Examples) is given by

$$\frac{\partial \Phi'_z}{\partial t} + \overline{u}\frac{\partial \Phi'_z}{\partial x} + N^2 w' = 0 \tag{10.63}$$

where the asterisk has been dropped on both $N^2$ and $w'$. Finally, we have the continuity equation

$$\frac{\partial u'}{\partial x} + \frac{1}{\rho_0}\frac{\partial\left(\rho_0 w'\right)}{\partial z} = 0 \tag{10.64}$$

Energy is introduced in the form of energy density (per unit volume) $E$. We show in the Appendix we can write a kind of conservation equation

$$\frac{\partial E}{\partial t} + \nabla \cdot \left(\mathbf{u}_g E\right) - \frac{E}{\overline{u} - c}\frac{D_g}{Dt}\left(\overline{u} - c\right) = 0 \tag{10.65}$$

where $D_{\mathrm{g}}$ indicates total derivative using the group velocity $\mathbf{u}_{\mathrm{g}}$ while $c$ is the phase velocity. This equation is equivalent to

$$\frac{1}{E}\frac{\partial E}{\partial t} + \frac{1}{E}\nabla \cdot \left(\mathbf{u}_g E\right) = \frac{1}{\overline{u} - c}\frac{D_g}{Dt}\left(\overline{u} - c\right) \Rightarrow \frac{1}{E}\frac{D_g E}{Dt} = \frac{1}{\overline{u} - c}\frac{D_g}{Dt}\left(\overline{u} - c\right)$$

and this implies the conservation law

$$\frac{\partial}{\partial t}\left(\frac{E}{\overline{u}-c}\right)+\nabla\cdot\left(\mathbf{u}_g\frac{E}{\overline{u}-c}\right)=0 \tag{10.66}$$

The wave action can then vibe defined as (see Appendix)

$$A=\rho_0\frac{\overline{u'^2}+\left(\overline{\Phi'_z}/N^2\right)}{2\left(\overline{u}-c\right)}=\frac{E}{\left(\overline{u}-c\right)} \tag{10.67}$$

and this is a conservative quantity.

This detailed derivation is given here (more precisely the Appendix) because the result is always mentioned but never shown or, if it is shown, it is through very complex arguments. Our derivation is simply adapted from an old paper of Francis Bretherton. In the presence of dissipation $D$, Eq. (10.66) can be generalized as

$$\partial A/\partial t+\nabla\cdot\left(\mathbf{u}_gA\right)=D \tag{10.68}$$

From Eq. (10.56), considering only the vertical component of the EP flux, it can be easily shown that

$$\rho_0\,\overline{u'w'}=Aw_g$$

so that the vertical eddy momentum flux is actually proportional to the wave action flux through the group velocity $w_g$. Comparing Eq. (10.56) and Eq. (10.69), we obtain

$$\rho_0\overline{u}=A \tag{10.69}$$

In this particular case, the action is equivalent to the zonal mass flux induced by the wave action.

A final consideration is that in the absence of dissipation and in the stationary case, we have from Eq. (10.68)

$$\nabla\cdot\left(\mathbf{u}_gA\right)=0 \tag{10.70}$$

which is a different form of the EP theorem in that, in the absence of dissipation, there is no interaction between zonal flow and the wave.

This section may have seemed quite boring but has been necessary to increase our insight into the physics and to give a correct formulation of the wave processes. This knowledge enables us to make an application to the problem of turbulent diffusion that we have mentioned even in Chap. 4 when dealing with the diffusion of gases in the upper atmosphere. Most of this application will be based on the concept that a vertically propagating wave is amplified to the point that it breaks down generating turbulence like breaking waves on a beach.

## 10.7 Breaking, Saturation, and Turbulence in the Upper Atmosphere

After thinking about the problem for a few years, in 1981, Richard Lindzen formulated a theory on the turbulence induced by the gravity wave breaking. We base this section on that paper except for a few minor modifications.

We may begin by looking at Fig. 10.6 which shows what may happen to a material surface (like an isentropic surface) under the influence of a vertically propagating wave. The amplitude of the wave grows to the point that in some region, the temperature gradient becomes adiabatic or superadiabatic so that the potential temperature becomes constant with height or even decreases. Under such conditions, there is a convective instability and consequently, the breaking of the wave. Another way to look at this problem is to consider the turbulence as a way of dissipating the energy of the wave and then limiting its growth. The breaking may also derive from nonlinear interactions between the different spectral components of the wave.

The phenomena we have described is very similar to a saturation process (that is why we talk about wave saturation) which is intrinsically nonlinear. This creates the first difficulty because to maintain a simple treatment, there will be the need to reduce it to a linear problem.

The assumptions we make to formulate our theory is that the frequency of the gravity wave $\omega$ is such that $f << \omega << N$ and also we will treat only two-dimensional waves so that the phase can be written as $(kx + mz - \omega t)$. The potential temperature gradient is the sum of the zonally averaged gradient and the deviation from it

$$\overline{\theta}_z + \theta'_z$$

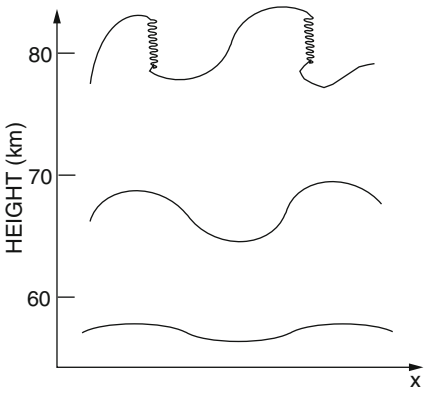



**Fig. 10.6** A scheme for the breaking of a vertically propagating gravity wave. The instabilities are shown. The curves are successive stages of a material surface (From Andrews et al. 1987)

The second term can be easily related to the geopotential gradient through the hydrostatic equilibrium $\Phi\prime_z = -\theta' g/\overline{\theta}$ so that the gradient of the potential temperature will be $\theta\prime_z = -\Phi\prime_{zz}\overline{\theta}/g$. The total gradient becomes

$$\theta_z = -\Phi\prime_{zz}\overline{\theta}/g + \overline{\theta}_z = -\Phi\prime_{zz}\overline{\theta}/g + N^2\overline{\theta}/g \tag{10.71}$$

And this will be zero when

$$\Phi\prime_{zz} = N^2 \tag{10.72}$$

A saturation condition can also be obtained through thermodynamic Eq. (10.34) which can be written as

$$ik\left(\overline{u} - c\right)\Phi\prime_z = N^2 w'$$

This equation is differentiated with respect to $z$ after multiplying it through $\rho_0$

$$k\left(\overline{k} - c\right)\Phi\prime_{zz} = N^2 m w'$$

The vertical velocity can be eliminated by the continuity equation $ku' = mw'$, and saturation will be reached when

$$u' = \left(\overline{u} - c\right) \tag{10.73}$$

At this point, we need to make a minor detour because we are calculating deviations from the zonal mean. We are actually interested in the zonal mean value of the square of these quantities, and for that we will use the relation

$$\overline{\phi\prime}^2 = |\phi|^2/2$$

This simple relation comes from the fact that the deviation from the zonal mean can be expressed as

$$\phi' = \mathrm{Re}\left\{\phi e^{ikx}\right\}$$

Our aim at this point is to find the action for the saturated wave, and so we need both the kinetic energy and the potential energy related to the eddy value of the geopotential. To this end, we eliminate $\Phi'$ and w′ between Eqs. (10.59), (10.72), and (10.73). We assume

$$w'\left(x,\ z,\ t\right) = W(z)\exp\left[k\left(x - ct\right) + z/2H\right] \tag{10.74}$$

so the derivative with respect to height is

$$w'_z = w'\left(x,\ z,\ t\right)/2H + W_z \exp\left(kx - ct + z/2H\right)$$

We are actually interested in the dependence of the amplitude on height so we will take into account only the second term on the right. Substituting $u'$ from Eq. (10.68), we obtain from the continuity and thermodynamic equation

$$-(\overline{u}-c)\,w\prime_z + w'\overline{u}_z + \Phi\prime_x = 0\;;\; ik\,(\overline{u}-c)\,\Phi\prime_z + N^2 w' = 0$$

Differentiating the first with respect to $z$ and the second with respect to $x$ and taking into account only the deviations from the zonal means (our assumption about linearity), we obtain

$$(\overline{u}-c)\,W_{zz} - W_z\overline{u}_z + WN^2/\,(\overline{u}-c) = 0$$

The term in the middle is negligible because the vertical shear of the zonal wind is small with respect to the changes in $W$. We obtain the vertical structure equation

$$W_{zz} + [N/\,(u-c)]^2 W = 0 \tag{10.75}$$

In principle, the quantity squared should depend on height, and the solution to this equation can only be obtained using the WKB approximation as shown in the Appendix. Using these results, we can find the vertical component of the group velocity. The frequency of the wave is given by

$$\upsilon = k(\overline{u}-c)^2 = kN/m$$

so that the group velocity will be

$$\partial\upsilon/\partial m = -kN/m^2 = k(\overline{u}-c)^2/2 \tag{10.76}$$

At this point, we can find the other term that makes up the wave action because in the relation between $u'$ and $w'$, we substitute for $m$

$$w' = k(\overline{u}-c)^2/N$$

This can be substituted in the thermodynamic equation to give

$$\Phi\prime_z = i\,(\overline{u}-c)\,N$$

so that at saturation, we have

$$\overline{\Phi\prime_z} = (\overline{u}-c)^2 N^2/2$$

Finally, the wave action according to Eq. 10.69 is given by

$$A = \rho\,(\overline{u}-c)\,/2 \tag{10.77}$$

Before we proceed, it is important to understand what is happening to a breaking wave. In particular, we consider the acceleration on the zonal wind. We can now calculate this term explicitly according to Eq. 10.62 because we have the eddy contribution which is

$$\rho_0 u'w' = -k\rho_0(\overline{u}-c)^3/N \Rightarrow \overline{u'w'} = -k(\overline{u}-c)^3/(2N)$$

The mean acceleration of the zonal wind is then

$$\overline{u}_t = -\frac{k(\overline{u}-c)^2}{2N}\left[\frac{(\overline{u}-c)}{H} - 3\overline{u}_z\right] \tag{10.78}$$

And if the zonal mean velocity changes slowly with altitude, we have

$$\overline{u}_t = \frac{\overline{u'w'}}{H} = -\frac{k(\overline{u}-c)^3}{2NH} \tag{10.79}$$

These considerations are valid only above the saturation altitude because below that the EP theorem holds. Above the saturation level, the acceleration on the zonal wind is such as to bring $\bar{u}$ closer to the phase velocity. From Eq. 10.79, we see that if $\overline{u} > c$, the acceleration is negative and so $\bar{u}$ decreases; the opposite happens when $\overline{u} < c$. The situation can be represented as in Fig. 10.7 where we have shown a vertically propagating monochromatic wave. The wave with a growing amplitude is saturated at level $z_s$ and is completely absorbed at $z_c$.

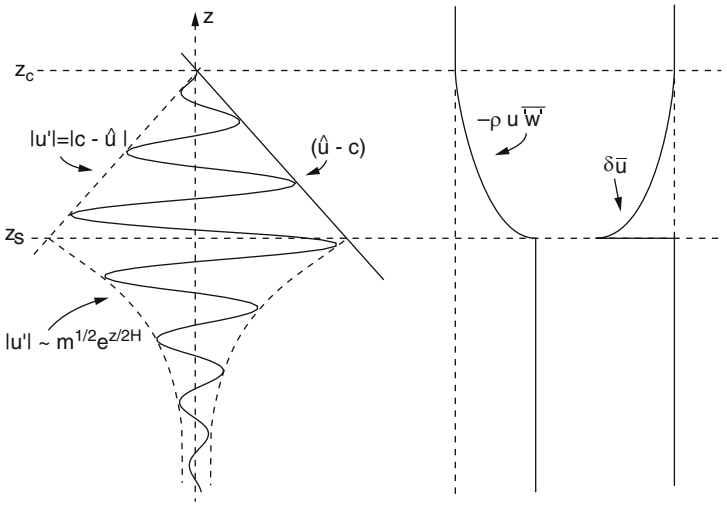



**Fig. 10.7** A schematic representation of growth and saturation of a vertically propagating gravity wave. After the saturation level $z_s$ has been reached, the wave amplitude decreases and is completely absorbed at $z_c$. The momentum flux (*shown on the right*) produces an acceleration of the zonal mean wind that makes it coincide with the phase velocity (Fritts 1984)

At this point, we can obtain the diffusion coefficient induced by wave breaking. The hypothesis is that once saturated, the energy of the wave is dissipated by a term proportional to the action itself. Based on Eq. 10.68, we can write

$$\nabla \cdot \left(\mathbf{u}_g A\right) = -2\delta A \tag{10.80}$$

where $\delta$ is the dissipation rate. The reason for the factor 2 is that the action is proportional to the square of quantities like the potential temperature for which the dissipative term is proportional simply to $\delta$. Substituting in Eq. 10.79 the values for the action and the group velocity, we obtain

$$\partial \left(c_{gz} A\right) / \partial z = -\left(c_{gz} A\right) / H = -2\delta A \Rightarrow \delta = c_{gz}/2H \tag{10.81}$$

where we have exploited the condition that saturation $A$ is constant except for the density factor that varies as $\rho = \rho_0 \exp\ (-z/H)$. At this point, substituting the group velocity in Eq. 10.81, the dissipation rate becomes

$$\delta = c_{gz}/2H = k(\overline{u} - c)^2/(2HN) \tag{10.82}$$

The diffusion coefficient is obtained assuming that above the saturation height the velocity is regulated by a diffusion equation

$$K_{zz}\frac{\partial^2 u'}{\partial z^2} = -m^2 K_{zz} u' = -\delta u' \tag{10.83}$$

And from this, we derive easily

$$K_{zz} = \frac{k(\overline{u} - c)^4}{2HN^3} \tag{10.84}$$

At this point, we can relax a bit our assumptions and allow at least the zonal velocity and $N$ to change slowly with altitude. The new expression for the dissipation rate may be obtained with the help of Eq. 10.79

$$\delta = \frac{k(\overline{u} - c)^2}{N}\left[\frac{1}{2H} - \frac{3}{2}\frac{\overline{u}_z}{(\overline{u} - c)} + \frac{1}{2N}\frac{\partial N}{\partial z}\right] \tag{10.85}$$

And dividing by $\mathrm{m}^2$, we obtain the diffusion coefficient. We can substitute in (10.85) typical values of the variables. For a horizontal wavelength of 200 km $\left(\mathrm{k} = 1.57\ 10^{-5}\mathrm{m}^{-1}\right)$ and an intrinsic velocity $(\overline{u} - c) = 30\ \mathrm{ms}^{-1}$, with $N = 0.02\ \mathrm{s}^{-1}$ and $H = 6$ we obtain $K_{zz} = 265\ \mathrm{m}^2\mathrm{s}^{-1}$. This value is much larger than the one found for the troposphere or in the boundary layer.

A typical example of what happens when a gravity wave propagates vertically is shown in Fig. 10.8 which is based on the results of a computer program reproduced in the Appendix. The cases shown in the figure refer to an atmosphere that has

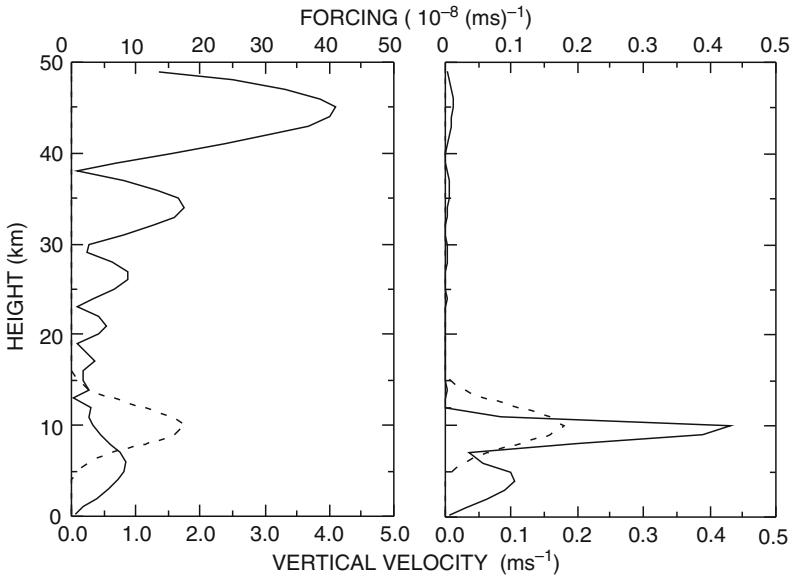



**Fig. 10.8** The vertical velocity as a function of height for a gravity wave without absorption (*left*) and with absorption (*right*). The critical level is at 12 km. The *dashed line* represents the forcing of the wave. Notice the different scales for the velocity and the forcing

a decreasing temperature from 280 K at the surface down to 200 K at 12 km constant up to 15 km and then increasing up to 270 K at 50 km. The zonal wind has a maximum of 40 ms$^{-1}$ at 12 km and then decreases until it becomes easterly $(-40\ \text{ms}^{-1})$ at 50 km. The wave is forced by a Gaussian centered at 10 km with a half width of 6 km.

With an assigned phase velocity of 50 ms$^{-1}$, the amplitude of the wave grows with altitude because the wind velocity never reaches the phase velocity. When the phase velocity is 40 ms$^{-1}$ the wave is absorbed at 12 km where the phase velocity is equal to the wind velocity.

We can see that compared to what is shown in Fig. 10.7, the decrease with altitude is much faster, although this could be a numerical effect. The wave amplitude below the critical level is not affected by the absorption. Details of this problem can be found in the Examples.

## E.10 Examples

### *E.10.1 Is the Phase Velocity a Vector?*

We discuss here a topic neglected in most elementary physics books and shows that what is called phase velocity does not have the characteristics of a vector. We refer to Fig. E.10.1 where we depict two wave crests for a two-dimensional wave

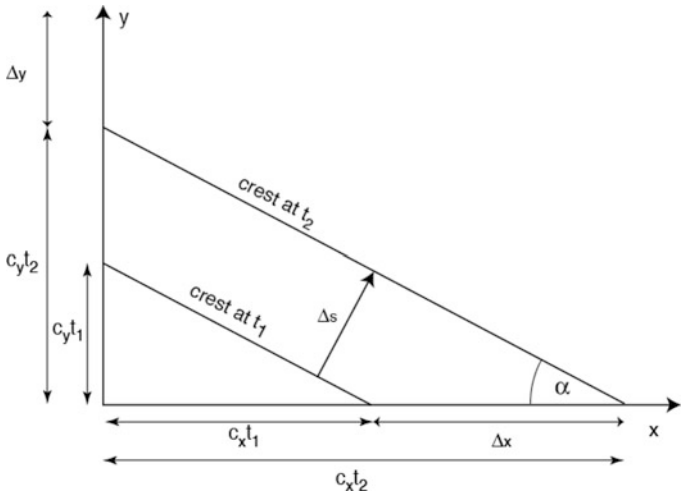



**Fig. E.10.1** Geometry for the propagation of a plane wave

propagating in x, y plane. We remember that the phase velocity along the x and y directions is simply

$$c_x = \omega/k;\ c_y = \omega/l$$

where $k$ and $l$ are the wave numbers along the x and y directions.

The wave propagates normally to the crest, and the distance $\Delta s$ may be found by observing that

$$\Delta s = \Delta x \sin\alpha = \Delta y \cos\alpha$$

so that

$$\frac{1}{\Delta s^2} = \frac{1}{\Delta x^2} + \frac{1}{\Delta y^2}$$

From the definition of the phase velocities, we have

$$\Delta x = c_x \Delta t = \omega \Delta t/k;\ \Delta y = c_y \Delta t = \omega \Delta t/l;$$

so that

$$\Delta s = \omega \Delta t \left(k^2 + l^2\right)^{-1/2}$$

and the propagation speed

$$c = \Delta s/\Delta t = \omega\left(k^2 + l^2\right)^{-1/2}$$

It can be easily verified that $c^2 \neq c_x^2 + c_y^2$ that is, the phase velocity, is not a physical vector.

It is interesting to show at this point that energy is transported by the group velocity. To achieve this, we write the flux of energy, $F$, integrated over depth

$$F = \int_{-h}^{0} p'u \, dz = \int_{-h}^{0} \rho g h' u \, dz \tag{E. 10.1}$$

where we have dropped the prime for the velocity. We assume a perturbation of the form

$$h' = A \cos(kx - \omega t) \tag{E.10.2}$$

The velocity $u$ can be obtained from the continuity equation

$$\frac{\partial h'}{\partial t} + h \frac{\partial u}{\partial x} = 0$$

Substituting for $h'$, we have

$$u = -\frac{\omega A}{kh} \cos(kx - \omega t) \tag{E.10.3}$$

so that the energy flux becomes

$$F = \int_{-h}^{0} \rho g h' u \, dz = \frac{\rho g \omega A^2}{kh} \int_{-h}^{0} \cos^2(kx - \omega t) \; dz = \frac{1}{2} \rho g A^2 \frac{\omega}{k} \tag{E.10.4}$$

The factor ½ is the average value for the cosine, while it can be easily shown that the quantity $\rho g A^2/2$ is just the average energy density of the wave. As a result, the energy flux is the product of the energy density by the group velocity $\omega/k$.

The energy density is the sum of the kinetic energy, $K$, and the potential energy, $U$

$$K = \tfrac{1}{2} \rho \left(u^2 + w^2\right) ; \; U = \int_{-h}^{0} \rho g z \, dz$$

We need to find the vertical component of the velocity

$$w = \frac{\partial h'}{\partial t} = -h \frac{\partial u}{\partial x} = -\omega A \sin(kx - \omega t)$$

we get then by summing

$$K = \frac{1}{4} \rho \omega^2 A^2 \left(1 + \frac{1}{(kh)^2}\right) \approx \frac{1}{4} \frac{\rho \omega^2 A^2}{(kh)^2}$$

where we have used the fact that $(kh)^{-2} \gg 1$. Integrating this over depth, we have

$$\langle K \rangle = \frac{1}{4}\frac{\rho\omega^2 A^2}{k^2 h} = \frac{1}{4}\frac{\rho c^2 A^2}{h} = \frac{1}{4}\frac{\rho g h A^2}{h} = \frac{1}{4}\rho g A^2$$

In the same manner, the potential energy integrated over depth becomes

$$\langle U \rangle = \frac{1}{2}\rho g h^2 = \frac{1}{2}\rho g A^2 \cos^2(kx - \omega t) = \frac{1}{4}\rho g A^2$$

And the total energy density is what we have used before

### *E.10.2 The Quasi-geostrophic Potential Vorticity in Log P Coordinates*

Taking into account the continuity, Eq. 10.32 becomes

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\zeta' + \upsilon'\frac{\partial f}{\partial y} - f\frac{1}{\rho_0}\frac{\partial}{\partial z^*}\left(\rho_0 w^*\right) \tag{E.10.5}$$

while Eq. 10.34 becomes

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\rho'}{\rho_0} - \frac{N_*^2 w^*}{g} = 0 \tag{E.10.6}$$

and the equation of hydrostatic equilibrium is

$$\frac{H}{p}\frac{\partial p'}{\partial z^*} + \frac{\rho'}{\rho_o}g = \frac{1}{\rho_o}\frac{\partial p'}{\partial z^*} + \frac{\rho'}{\rho_o}g = 0$$

As we did in Sect. 10.5, we can write the thermodynamic equation in the form

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{1}{\rho_0}\frac{\partial p'}{\partial z^*} + N_*^2 w^* = 0$$

And using $\psi = p'/f\rho_0$, we get

$$f\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\partial \psi}{\partial z^*} + N_*^2 w^* = 0 \tag{E.10.7}$$

From this equation, we can get $w^*$ which can be derived and substituted in Eq. 10.32 to obtain the quasi-geostrophic vorticity equation in log-pressure coordinates.

### *E.10.2 The Eliassen–Palm Flux Terms*

In Eq. 10.59, the values for the perturbation velocity and vorticity can be substituted to obtain

$$q'v' = \frac{\partial^2\psi}{\partial x^2}\frac{\partial\psi}{\partial x} + \frac{\partial^2\psi}{\partial y^2}\frac{\partial\psi}{\partial x} + \frac{1}{\rho_0}\frac{\partial\psi}{\partial x}\frac{\partial}{\partial z^*}\left(\frac{f^2\rho_0}{N_*^2}\frac{\partial\psi}{\partial z^*}\right) \tag{E.10.8}$$

The zonal mean of this equation can be obtained by manipulating the different terms. We can write

$$\begin{aligned}\rho_o q'v' = {} & \frac{\rho_o}{2}\frac{\partial}{\partial x}\left(\frac{\partial\psi}{\partial x}\right)^2 + \rho_o\frac{\partial}{\partial y}\left(\frac{\partial\psi}{\partial y}\frac{\partial\psi}{\partial x}\right) - \rho_o\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x\partial y} \\ & + \frac{\partial}{\partial z^*}\left[\frac{\partial\psi}{\partial x}\left(\frac{f^2\rho_o}{N_*^2}\frac{\partial\psi}{\partial z^*}\right)\right] - \frac{\partial^2\psi}{\partial z^*\partial x}\frac{\partial}{\partial z^*}\left(\frac{f^2\rho_o}{N_*^2}\frac{\partial\psi}{\partial z^*}\right)\end{aligned} \tag{E.10.9}$$

The zonal mean of the first term is zero, as are the third and the fifth. For example, the zonal mean for the third term is given by

$$\int_0^L \rho_0\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x\partial y}dx = \rho_0\frac{\partial\psi}{\partial y}\frac{\partial\psi}{\partial y}\Big|_0^L - \int_0^L \rho_0\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x\partial y}dx$$

which is obtained by integrating by parts. The result is that the integral and the mean are zero. The same thing is obtained for the last term.

### *E.10.3 Energy and EP Flux*

To obtain an equation that relates the EP flux to the energy of the wave, we multiply Eq. 10.68 by $\rho_0 u'$ to obtain (Holton 1975)

$$\rho_0\frac{1}{2}\left(\overline{u}\frac{\partial u'^2}{\partial t} + \frac{\partial u'^2}{\partial x}\right) + \frac{\partial\left(\rho_0\overline{u}u'w'\right)}{\partial z} - \overline{u}\frac{\partial\left(\rho_0 u'w'\right)}{\partial z} + \rho_0\frac{\partial\left(u'\Phi'\right)}{\partial x} - \rho_0\Phi'\frac{\partial u'}{\partial x} = 0$$

The last term of this equation can be substituted by Eq. 10.71 while the fourth from Eq. 10.69. After performing a zonal mean, we obtain

$$\rho_0\frac{1}{2}\overline{u}\frac{\partial u'^2}{\partial t} + \frac{\partial\left(\rho_0\overline{u}u'w'\right)}{\partial z} + \rho_0\overline{u}\frac{\partial\overline{u}}{\partial t} + \frac{\partial\left(\rho_0 w'\Phi'\right)}{\partial z} - \rho_0\Phi'_z w' = 0$$

Multiplying Eq. 10.70 by $\rho_0\Phi'_z/N^2$ it is possible to eliminate $\rho_0\Phi'_z w'$ to obtain after the zonal mean

$$\rho_0 \frac{1}{2}\frac{\overline{\partial u'^2}}{\partial t} + \frac{\rho_0}{N^2}\frac{1}{2}\frac{\overline{\partial \Phi_z'^2}}{\partial t} + \rho_0 \frac{1}{2}\frac{\partial \overline{u}^2}{\partial t} + \frac{d}{dz}\left(\rho_0 \overline{u}\overline{u'w'} + \rho_0 \overline{w'\Phi'}\right) = 0 \qquad \text{(E.10.10)}$$

Integrating over altitude, we have

$$\frac{1}{2}\frac{\partial}{\partial t}\int_0^\infty \rho_0 \left[\overline{u'^2} + \overline{\left(\Phi_z'/N^2\right)} + \overline{u}^2\right] dz = -\int_0^\infty \frac{d}{dz}\left[\rho_0 \overline{u}\overline{u'w'} + \overline{\rho_0 w'\Phi'}\right] dz$$

And using the condition that w′ = 0 for z = 0 and ∞, we get the conservation of total energy

$$\frac{1}{2}\frac{\partial}{\partial t}\int_0^\infty \rho_0 \left[\overline{u'^2} + \overline{\left(\Phi_z'^2/N^2\right)} + \overline{u}^2\right] dz = 0 \qquad \text{(E.10.11)}$$

This relation indicates that the total energy is in part the kinetic energy of the zonal current and the eddies and in part the oscillation energy $N^2\delta^{*2}/2$, where $\delta z^*$ is the amplitude of the oscillation. This can be easily obtained by considering that $\theta' = \overline{\theta}_z \delta z^*$, where $\theta'$ is related to the vertical acceleration of the parcel given by

$$\Phi_z' = -\theta' g/\overline{\theta} = \theta' N^2/\theta_z$$

The second term in Eq. C.7 is then obtained easily. Equation (10.71) can be rewritten in a different form considering that

$$\frac{1}{2}\frac{\partial}{\partial t}\overline{u}^2 = \overline{u}\frac{\partial \overline{u}}{\partial t} = -\frac{\overline{u}}{\rho_0}\frac{d}{dz}\left(\rho_0 u'w'\right)$$
$$\frac{1}{2}\frac{\partial}{\partial t}\overline{u}^2 = \overline{u}\frac{\partial \overline{u}}{\partial t} = -\frac{\overline{u}}{\rho_0}\frac{d}{dz}\left(\rho_0 u'w'\right)$$

And deriving the last term in Eq. 10.69, we have

$$\rho_0 \frac{1}{2}\frac{\overline{\partial u'^2}}{\partial t} + \frac{\rho_0}{N^2}\frac{1}{2}\frac{\overline{\partial \Phi_z'^2}}{\partial t} + \frac{d\overline{u}}{dz}\left(\rho_0 \overline{u'w'}\right) + \frac{d}{dz}\left(\rho_0 \overline{w'\Phi'}\right) \qquad \text{(E.10.12)}$$

This equation tell us that the change of energy per unit volume in the wave is in part due to the eddy transport term, similar to what we found in the boundary layer, and in part to the power flux divergence due to the pressure forces (last term). Equation E.10.12 can be written in the form that puts in evidence the energy conservation if we recur to Eq. 10.66 and assume that all the variables have a phase like $k\,(x - ct)$. We obtain

$$\overline{u'w'} = -\overline{\Phi' w'}/\left(\overline{u} - c\right)$$

that can be substituted in Eq. E.10.10 to give

$$\rho_0 \frac{1}{2}\frac{\partial \overline{u'^2}}{\partial t} + \frac{\rho_0}{N^2}\frac{1}{2}\frac{\partial \overline{\Phi'^2}}{\partial t} - \rho_0 \frac{d\overline{u}}{dz}\frac{\overline{w'\Phi'}}{\overline{u}-c} = -\frac{d}{dz}\left(\overline{\rho_0 w'\Phi'}\right)$$

The quantity $\rho_0 w'\Phi'$ can be regarded as power flux that can be written as the product of the energy density ($E$) for the vertical component of the group velocity. We obtain then

$$\frac{\partial E}{\partial t} - \frac{d\overline{u}}{dz}\frac{w_g E}{\overline{u}-c} = \frac{\partial E}{\partial t} - \frac{E}{\overline{u}-c}\frac{D_g}{Dt}(\overline{u}-c) = -\frac{d}{dz}\left(\overline{\rho_0 w'\Phi'}\right)$$

where $D_g$ indicates the total derivative where the velocity components are substituted by the group velocity.

### *E.10.4 The WKB Approximation*

This approximation is used to solve a differential equation of the form

$$W_{zz} + m^2 W = 0 \tag{E.10.13}$$

where $m$ is a function of $z$. If $m$ changes slowly, the local solution may be of the form $\exp(\pm imz)$. The extension of the solution to a greater domain was made by Liouville (and attributed to Wentzel, Kramers, and Brillouin 100 years later) by introducing the variables

$$\varphi = \int m\, dz\, \vartheta = m^{1/2} W \tag{E.10.14}$$

so that Eq. C.7 becomes

$$d^2\vartheta/d\varphi^2 + (1+\delta)\,\vartheta = 0\ ;\ \delta = m^{-3/2} d^2\left(m^{-1/2}\right)/dz^2 \tag{E.10.15}$$

and if $d << 1$ the approximate solution for Eq. C.9 is of the form $\vartheta = \exp(\pm i\phi)$ which is equivalent to a solution for Eq. 9.81 of the form

$$W = m^{-1/2} \exp\left(\pm i \int m\, dz\right) \tag{E.10.16}$$

If $m$ changes slowly, the solution becomes

$$W(z) \propto \exp(imz) \Rightarrow w' = A m^{-1/2} \exp\left[z/2H + i\,(kx + mz - ct)\right] \tag{E.10.17}$$

where $m = N/(\overline{u} - c)$

### *E.10.5 The Numerical Solution to the Wave Equation*

The starting point is an equation similar to Eq. 10.82 that contains also the nonlinear terms

$$W_{zz} + Q^2 W = 0 \tag{E.10.18}$$

where

$$Q^2 = \frac{N^2}{(c-\overline{u})^2} + \frac{\overline{u}_{zz} + \overline{u}/H}{(c-\overline{u})} - \frac{1}{4H^2} \tag{E.10.19}$$

In Eq. E.10.18, we put

$$W = W'^* e^{-z/2H}$$

where $W'^*$ is the perturbation to the vertical velocity. The perturbation may arise from a forcing $F(z)$ function of the altitude so that Eq. E.10.18 becomes

$$W_{zz} + Q^2 W = F(z) \tag{E.10.20}$$

The forcing is assumed to be an attenuated Gaussian

$$F(z) = \begin{cases} e^{-z/2} A \left( e^{-\left(z - z_f / z_w\right)^2} - e^{-4} \right) / H^2 & \left| z - z_f \right| < 2 z_w \\ 0 & \left| z - z_f \right| > 2 z_w \end{cases}$$

To solve Eq. E.10.20 numerically, we follow a method suggested by Lindzen so that we assume a vertical grid such that

$$z_k = k\Delta,\ k = 1,\ 2,\ ....,\ K$$

where the grid spacing is given by

$$\Delta = z_{top} / (K+1)$$

Equation C.3 is transformed in the finite difference form

$$W_{k+1} + \left(\Delta^2 Q_k^2 - 2\right) W_k + W_{k-1} = \Delta^2 F_k \tag{E.10.21}$$

for $k = 2, \ldots,\ K-1$. For $k = 1$ we have

$$W_2 + \left(\Delta^2 Q_1^2 - 2\right) W_1 = \Delta^2 F_1 - W_{bot} \tag{E.10.22}$$

where $W_{bot}$ is the condition on the velocity at the lower boundary. For the upper boundary $k = K$ we can have two conditions, the rigid lid

$$W_{k-1} + \left(\Delta^2 Q_k^2 - 2\right) W_k = \Delta^2 F_k \tag{E.10.23}$$

and the radiation condition

$$2W_{k+1} + \left(\Delta^2 Q_k^2 - 2i\Delta Q_k - 2\right) W_k = 0 \tag{E.10.24}$$

The first condition reflects the condition that the velocity is zero at the top while the second assumes a solution of the form

$$W\left(z_{top}\right) \propto \exp\left(-iQ\left(z_{top}\right) z\right)$$

The system that results from Eqs. (E.10.21), (E.10.22), (E.10.23), and (E.10.24) is tridiagonal and can be inverted with standard methods. The program included in the CD written by one of my students uses the method suggested by Lindzen and calculates also the geopotential and momentum fluxes.

### *E.10.6 A Few More Things About Mountain Waves*

We may consider following Durran (1990) the two-dimensional flow which satisfies the linearized governing equations

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} &= 0 \\ \frac{\partial w}{\partial t} + \frac{\partial p}{\partial z} &= b \\ \frac{\partial b}{\partial t} + N^2 w &= 0 \\ \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} &= 0 \end{aligned} \tag{E.10.25}$$

In these equations, b represents the buoyancy given by $b = g\left(\theta - \theta_0\right)/\theta_s$ and the pressure $p = c_p \theta_s \left(\pi - \pi_0\right)$ and $N^2 = \left(g/\theta_s\right) d\theta_0/dz$. All the values with the subscript (s) are the reference value, while $\pi = \left(p/p_s\right)^{R/c_p}$ and the quantities with subscript (0) are hydrostatically equilibrium values.

To obtain a single equation for w, we derive the first with respect to z and the second with respect to x to eliminate pressure

$$\frac{\partial}{\partial t}\left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right) = \frac{\partial b}{\partial x}$$

This equation is then derived with respect to t while the third in (E.10.25) with respect to x to eliminate b

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial u}{\partial z}-\frac{\partial w}{\partial x}\right)+N^2\frac{\partial w}{\partial x}=0$$

We derive this equation with respect to x and eliminate u with the continuity equation to get

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)w+N^2\frac{\partial^2 w}{\partial x^2}=0 \tag{E.10.26}$$

We look for the solution of the form

$$w=w_0\cos\left(kx+mz-\omega t\right) \tag{E.10.27}$$

And from the continuity equation, we have

$$u=-\frac{m}{k}w_0\cos\left(kx+mz-\omega t\right) \tag{E.10.28}$$

We can also easily get the buoyancy and the pressure

$$b=\frac{N^2}{\omega}w_0\sin\left(kx+mz-\omega t\right) \tag{E.10.29}$$

$$p=-\frac{\omega m}{k^2}w_0\cos\left(kx+mz-\omega t\right) \tag{E.10.30}$$

The dispersion relation can be found by forcing (E.10.27) in (E.10.26) to get

$$\omega^2=\frac{N^2k^2}{k^2+m^2}=N^2\cos^2\phi \tag{E.10.31}$$

This is the same as (10.16). Also, (E.10.27) and (E.10.28) imply that $u/w=-m/k$ meaning that air parcel motions are parallel to wave front.

## *E.10.7 Waves Forced by Sinusoidal Ridges*

Some other properties for mountain waves can be obtained assuming a forcing at the lower boundary given by a sinusoidal profile

$$h(x)=h_m\sin kx \tag{E.10.32}$$

and the condition of a steady state. In this case, all the time derivatives could be reduced to $\partial/\partial t \rightarrow u_0 \partial/\partial x$ where $u_0$ is the background wind along the $x$ direction. In this case, (E.10.26) reduces to

$$u_0^2 \frac{\partial^2}{\partial x^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right) w + N^2 \frac{\partial^2 w}{\partial x^2} = 0$$

that can be integrated twice with respect to x, giving

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial z^2}+\frac{N^2}{u_0^2} w = 0 \tag{E.10.33}$$

We can integrate this equation assuming a solution of the form

$$w\,(x,\ z) = w_1(z)\cos kx + w_2(z)\sin kx \tag{E.10.34}$$

that can be substituted into (E.10.33), giving the equations

$$\frac{\partial^2 w_i}{\partial z^2}+\left(\frac{N^2}{u_0^2}-k^2\right) w_i = 0 \tag{E.10.35}$$

where $w_i$ is either $w_1$ or $w_2$. We can put $m = \left(N^2/u_0^2 - k^2\right)^{1/2}$ and $\mu^2 = -m^2$ and the equation can have two possible solutions, one varying exponentially with altitude and the other freely propagating with height

$$w_i(z) = \begin{cases} A_i e^{\mu z} + B_i e^{-\mu z} & k > N/u_0 \\ C_i \cos mz + D_i \sin mz & k < N/u_0 \end{cases} \tag{E.10.36}$$

The first condition happens when the intrinsic frequency $Nk > u_0$ and of it we need to take only the solution decaying with altitude having the form

$$w\,(x,\ z) = Be^{-\mu z}\cos\ kx$$

And the constant is determined in such a way that

$$w\,(x,\ 0) = B\cos\ kx = u_0 \frac{\partial h}{\partial x} = u_0 h_m k \cos kx$$

so that the solution is given by

$$w\,(x,\ z) = u_0 h_m k e^{-\mu z}\cos\ kx \tag{E.10.37}$$

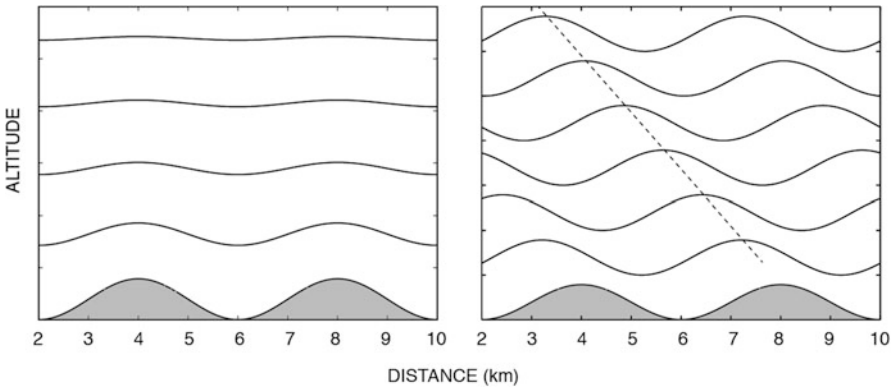



**Fig. E.10.2** Streamlines of stationary waves above the sinusoidal profile. On the *left*, the amplitude attenuates with height. On the *right*, freely propagating waves upward and westward. The constant phase is indicated by the *broken line* (Durran 1990)

The dispersion relation obtained forcing a solution of the form $\mathcal{R}\left(e^{i(kx+mz-\omega t)}\right)$ is

$$\omega = u_0 k \pm \frac{Nk}{(k^2+m^2)^{1/2}} \tag{E.10.38}$$

In order to have $\omega = 0$ we need to take into account only the negative roots, and the vertical group velocity is given by

$$\frac{\partial \omega}{\partial m} = \frac{Nkm}{(k^2+m^2)^{3/2}} \tag{E.10.39}$$

Vertical propagation of energy requires that k and m have the same sign. So for the second solution of (E.10.36), we should have

$$w\,(x,\ z) = u_0 h_m k \cos\,(kx+mz) \tag{E.10.40}$$

The first of the two solutions is shown on the left of Fig. E.10.2 where $u_0 = 10 \text{ms}^{-1}$ and $h_m = 500$ m. The propagating solution shown on the right has $u_0 = 5\ \text{ms}^{-1}$.

### What Happens Over an Isolated Mountain

We have treated a similar problem as an application of the conservation of potential vorticity. We go back on the same problem using the formalism we have developed for gravity waves. The idea is to reconstruct the isolated mountain in trends

of Fourier components so that even the vertical velocity we have found can be written as

$$w(x,\ z) = \mathrm{Re}\int_0^\infty \widehat{w}(k,\ z)\, e^{-ikz} dk \tag{E.10.41}$$

that can be substituted in (E.10.23) to give

$$\frac{\partial^2 \widehat{w}}{\partial z^2} + \left(l^2 - k^2\right)\widehat{w} = 0 \tag{E.10.42}$$

where we have put $l^2 = (N/u_0)^2$. The boundary condition becomes

$$\widehat{w}(k,\ 0) = u_0 ik\widehat{h}(k) \tag{E.10.43}$$

where $\hat{h}(k)$ is the Fourier transform of the mountain profile

$$\widehat{h}(k) = \frac{1}{\pi}\int_{-\infty}^{\infty} h(x) e^{-ikx} dx$$

Then the solutions to (E.10.42) will be for $k^2 > l^2$

$$\widehat{w}(k,\ x) = \widehat{w}(k,\ 0)\exp\left[-\left(k^2 - l^2\right)^{1/2} z\right] \tag{E.10.44}$$

while for $k^2 > l^2$

$$\widehat{w}(k,\ x) = \widehat{w}(k,\ 0)\exp\left[i\left(l^2 - k^2\right)^{1/2} z\right] \tag{E.10.45}$$

From the definition of streamfunction $\eta(x, z)$ we get

$$w/u_0 = \partial\eta/\partial x \tag{E.10.46}$$

And integrating with respect to $x$

$$\eta(x,\ z) = \frac{1}{u_0}\int_0^x w(\xi, z)\, d\zeta = \frac{1}{u_0}\int_0^x dx \mathrm{Re}\int_0^\infty w(k,\ z)\, e^{ikx} dk$$

so we get

$$\eta(x,\ z) = \frac{1}{iku_0}\mathrm{Re}\int_0^\infty w(k,\ z)\, e^{ikx} dk$$

We can substitute the values for $w(k, z)$ and obtain

$$\eta(x,z) = \frac{1}{iku_0}\mathrm{Re}\left\{\int_0^l \widehat{w}(k,0)\exp\left[i\left(l^2-k^2\right)^{1/2}z\right]e^{ikx}dk + \int_l^\infty \widehat{w}(k,0)\exp\left[-\left(k^2-l^2\right)^{1/2}z\right]e^{ikx}dk\right\}$$

Substituting (E.10.43), we have

$$\eta(x,\ z) = \mathrm{Re}\left\{\int_0^l \widehat{h}(k)\exp\left[i\left(l^2-k^2\right)^{1/2}z\right]e^{ikx}dx + \int_l^\infty \widehat{h}(k)\exp\left[-\left(k^2-l^2\right)^{1/2}z\right]e^{ikx}dx\right\} \tag{E.10.47}$$

Just to make an example, we use e simple profile

$$h(x) = h_m a^2/\left(x^2+a^2\right) \tag{E.10.48}$$

whose Fourier transform is given by

$$\widehat{h}(k) = h_m a e^{-ka} \tag{E.10.49}$$

We may consider two different cases to obtain an analytic solution which corresponds to the case $al \gg 1$ or $al \ll 1$. The parameter a gives an idea of the width of the ridge, while from the definition of $1(N/u_0)$ we notice that the quantity $al$ is the ratio between the time it takes for a particle to traverse the ridge and the period of an oscillation. So for strong wind and narrow ridge and weak stability $al \ll 1$ and the first integral becomes small so we have

$$\begin{aligned}\eta(x,z) &= \mathrm{Re}\left\{\int_0^\infty h_m a e^{-ka}e^{-kz}e^{ikx}dx\right\}\\ &= h_m a\mathrm{Re}\left\{\frac{1}{-(a+z)+ix}\right\} = \frac{h_m(a+z)}{(a+z)^2+x^2}\end{aligned} \tag{E.10.50}$$

Notice that when $z \to 0$, $\eta(x,\ z) \to h(x)$. In the opposite case, when $al \ll 1$ buoyancy effects dominate and mathematically, $\hat{h}(k)$ is small in the range $l < k < \infty$ so that the second integral in (E.10.47) makes no contribution and we have

$$\begin{aligned}\eta(x,z) &= \mathrm{Re}\left\{h_m a\int_0^\infty e^{-ka}e^{ilz}e^{ikz}dk\right\}\\ &= h_m a\mathrm{Re}\left\{\frac{-1}{-a+i(lz+kx)}\right\} = h_m a\frac{(a\cos lz - x\sin lz)}{a^2+x^2}\end{aligned} \tag{E.10.51}$$

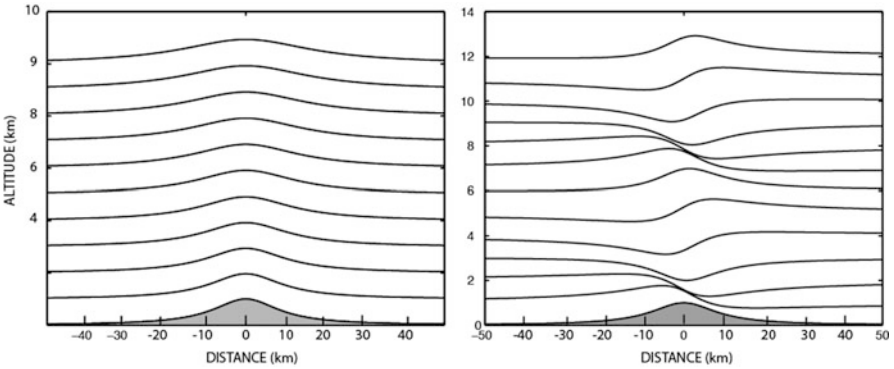



**Fig. E.10.3** Streamlines over an isolated obstacle. On the *left* is shown the approximation $al << 1$, while on the *right* is the case when $al >> 1$. Notice that the different vertical scale and the streamline are plotted at 1 km interval (Durran 1990)

Results in (E.10.50) and (E.10.51) are illustrated in Fig. E.10.3. In particular, the second case corresponds to vertical propagating nondispersive waves. The flow is periodic in the vertical so that at $z = \pi/l$ the streamline shape is the inverted $h(x)$ profile. The parameters for the figure on the left are $h_m = 1000m$, $a = 10km$. For the figure on the right, $h_m = 1000m$, $a = 10km$, $N = 0.01/s$, and $u_0 = 10m/s$. This implies a period $z = \pi/l = \pi\,(u_0/N) \approx 3146m$. In the figure, the ground profile is repeated every 6 km.

## References[1]

### *Books*

Andrews DG, Holton JR, Leovy CB (1987) Middle atmosphere dynamics. Academic, Orlando
Gill AE (1982) Atmosphere-ocean dynamics. Academic Press, New York
Holton JR (1975) The dynamic meteorology of the stratosphere and mesosphere. Meteorol Monogr 37. American Meteorological Society
Holton JR (1992) An introduction to dynamic meteorology. Academic Press, New York
Lindzen RS (1990) Dynamics in atmospheric physics. Cambridge University Press
Pedlosky J (1987) Geophysical fluid dynamics. Springer, New York

[1]This chapter draws heavily on four books (Andrews et al., Gill, Holton, Lindzen) and the papers by Bretherton, Fritts, and Lindzen. A large section on mountain waves has been added. These being another example of teaching with advanced science.

## *Articles*

Bretherton FP (1966) The propagation of groups of internal gravity waves in a shear flow. Q J Roy Meteorol Soc 92:466

Durran DR (1986) Mountain waves. In: Ray PS (ed) Mesoscale meteorology and forecasting. Springer

Durran DR (1990) Mountain waves and downslope winds. In: Atmospheric processes over complex terrain. Meteor Monogr 23. American Meteorological Society

Fritts DC (1984) Gravity waves saturation in the middle atmosphere: a review of theory and observations. Rev Geophys Space Phys 22:275

Lindzen RS (1970) TIdes and gravity waves in the upper atmosphere. In: Fiocco G (ed) Mesospheric models and related experiments. Reidel, Dordrecht, Holland

Lindzen RS (1981) Turbulence and stress owing to gravity wave and tidal breakdown. J Geophys Res 86:9707+

Smith RB (1979) The influence of the mountains on the atmosphere. Adv Geophys 21. Academic Press

# Chapter 11
# The Data on the Atmospheric Circulation

What we have learned so far can be applied to understanding the general circulation of the atmosphere. With this term, we mean all the large-scale motions that charaterize the atmosphere of the Earth. We already know the general features of motions, in the sense that we expect them to be geostrophic or quasi-geostrophic. Also we know that they are quasi-horizontal and that in general we can neglect vertical motions.

The theory of the general circulation of the atmosphere is not an academic matter because weather forecast and its improvement are based on it. At this point we do not want to make any attempt to review the ideas or the history on atmospheric circulation. The interested reader may look at the references given at the end of the book. We would rather start from the observed data and after a qualitative assessment of them we will present some of the theories in the next chapter.

## 11.1 The General Features

The main difference with respect to the old days is that today we have a wealth of data on atmospheric circulation. Data obtained through ground-based methods have been complemented with those obtained from satellite observations. A qualitative image of the circulation could then be that represented in Fig. 11.1. This figure refers to the situation found in the northern hemisphere in summer.

At the subsolar point located at about 20° latitude north, the intense heating produced by the absorption of solar radiation determines an almost convective circulation, with air rising to the top of the troposphere. This air has a considerable water vapor content that through condensation produces even more positive buoyancy. As in the case of a convective cell, the "hot plume" at the top of the troposphere separates in two branches directed to the north and south. If we follow the northern branch, the air, under the Coriolis force, will move toward east reaching a latitude



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_11

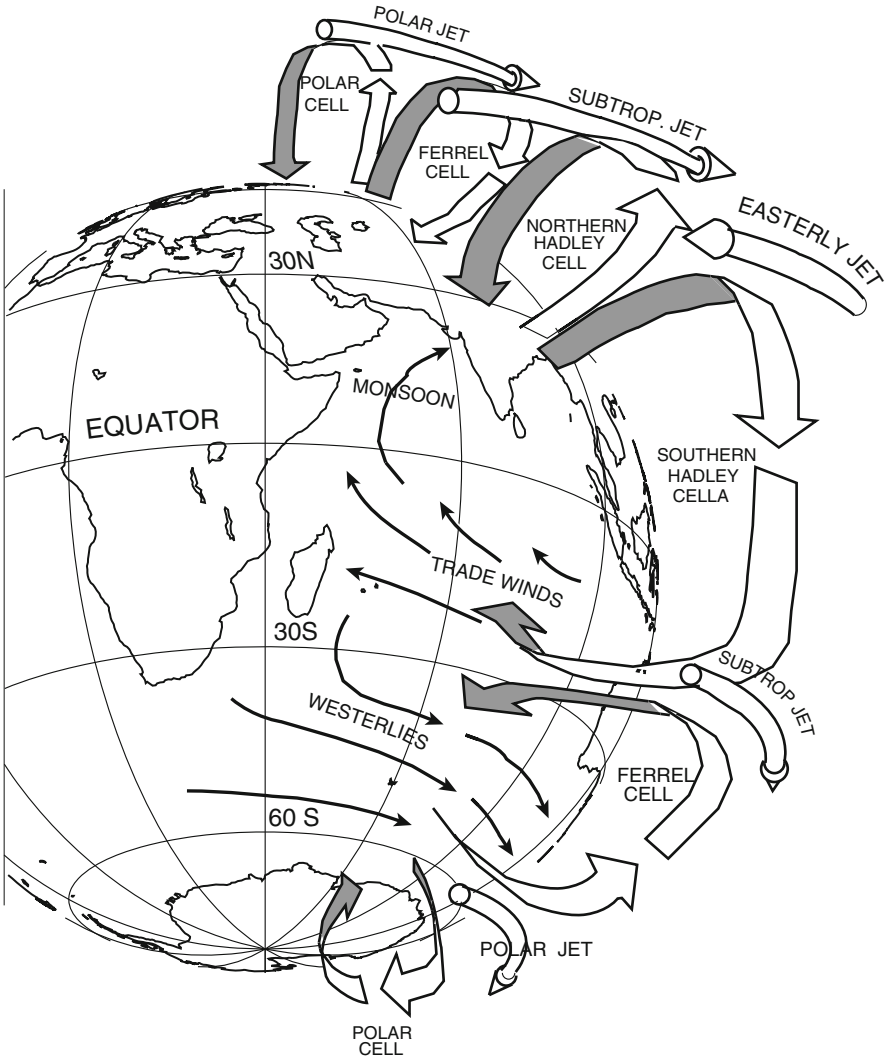



**Fig. 11.1** The general characteristics of the atmospheric circulation referred to the summer in the northern hemisphere. In this figure, the meridional character of the circulation in emphasized (Adapted from Washington and Parkinson 1986)

belt around 30–35°. At this point, the air has cooled to the point that it starts to subside along a "cold plume" and, closing the cell, moves near the ground toward west. The easterly component of this motion corresponds to the *trade winds*. One of the first questions we have to ask about the general circulation of the atmosphere is why these convective cells (known as *Hadley cells*) are limited in latitude and if there is any reason why they do not extend from the equator to the pole.

The figure shows how the circulation on the meridional plane, at middle latitudes, is dominated by a cell that is not produced by differential absorption and must have its origin in non-thermal processes. For this reason, it is called an *indirect* cell and it is also referred as a Ferrell Cell. One of the effects of these indirect cells is to produce the westerly winds that at middle latitudes drive the zonal circulation. The changes with altitude of these winds are due to the temperature latitudinal gradient (i.e. the thermal wind). In Fig. 11.1, "solid" arrows represent the zonal winds. As we will see shortly except for the tropical region (that is between the Hadley Cells), winds are always westerly although they reach a maximum in the high troposphere in the tropical region.

Labels attached to the different jets are meant to indicate the region and not a change in direction. In this figure we mainly show the meridional circulation, although there is also some indication about local or regional circulations like the monsoons.

Everything depicted in Fig. 11.1 actually can be represented by a meridional mass stream function as shown in Fig. 11.2. The units of mass/time refer to the atmospheric mass crossing a latitude circle at the altitude and latitude indicated in the figure. The polar fluxes are not represented because they are lower than $10^{10}$ kg $s^{-1}$. The data to which this figure refer represent a 10-year average and it clearly shows that there is no symmetry between the two hemispheres.

A quite general picture for the zonal winds is given in Fig. 11.3. As in the previous case, this is an annual mean and again the situation is not exactly symmetric between the two hemispheres. The structure of the trade winds is quite clear, as is the presence of the equatorial jet, as already shown in Fig. 11.1. A simple consideration can be made regarding the zonal winds and their contribution to the angular momentum of the Earth–atmosphere system, where the zonal winds are

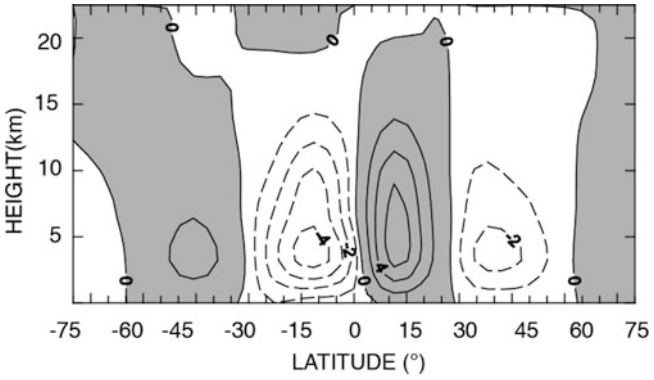



**Fig. 11.2** The mean annual meridional circulation of the atmosphere. The isopleths of the mass stream function are in units of $10^{10}$ kg $s^{-1}$ and are plotted at intervals of 1 unit for the anticlockwise circulation (*dashed lines*) and at interval of 2 for the clockwise cells (*shaded*). The flatness of the curves at the *top* is due to lack of data. The Hadley and Ferrell cells show up clearly. The figure is based on the data by Oort and Rasmussen. Southern latitudes are indicated with negative values

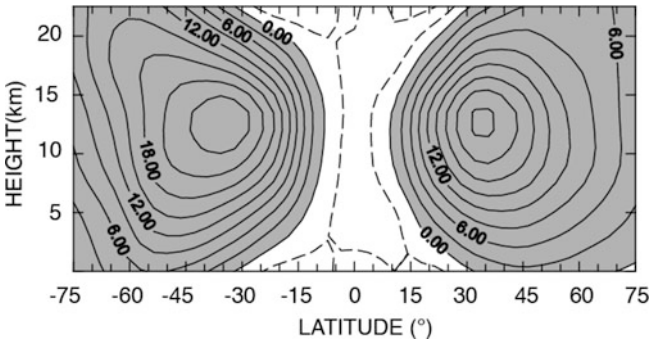



**Fig. 11.3** The average zonal winds over a period of 10 years. The isopleths are reported at intervals of 3 m $s^{-1}$. Although the maximum values are reached at the tropics, the zonal winds remain westerly even in the high altitude polar regions. The easterly winds (*dashed*) are the trade winds in the lower tropical troposphere. The westerly winds are *shaded* (All data are taken from Oort 1973; Peixoto and Oort 1992; Oort and Peixoto 1983)

westerly they "push" the Earth and the atmosphere transfers angular momentum to the solid Earth. On the other hand, when the winds are easterly (that is at the tropics) the atmosphere slows down the Earth and gains angular momentum. One would argue that the angular momentum lost by the solid planet at the tropics must be regained in other regions. However, this is a too generic argument to give useful results.

However the energy budget looks more promising from this point of view. We will start then to look at some energetics of the atmosphere: a very interesting and fundamental topic.

## 11.2 The Energy Budget of the Atmosphere

As we have seen since the first chapter, the main source of energy for the atmosphere is solar radiation. The power incident at one point of the Earth's surface depends on astronomical parameters (season and latitude). By contrast, the radiation absorbed depends mainly on the albedo and in that we include everything: the effects of the atmosphere, the clouds, the soil, etc. Nowadays satellite data can be used to make a quite accurate budget of the net zonally averaged solar radiation (that is incident minus reflected). In the same fashion, the emitted infrared radiation can also be measured. These two terms are plotted in Fig. 11.4, and they actually represent the net energy gain from the solar radiation and the net energy loss by the Earth due to planetary radiation. What is rather evident from this figure is that there is a net gain of energy in the tropical region and a net loss in the high latitude and polar regions. The reason for this behavior is because the solar radiation absorbed decreases toward the pole due to a geometrical factor but also to the increasing

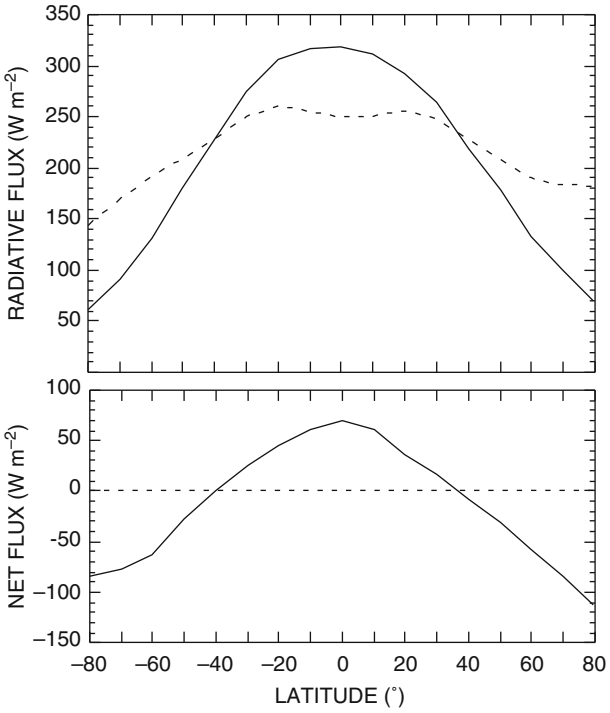



**Fig. 11.4** The *upper part* of the figure shows the zonally averaged absorbed solar radiation (*solid line*) and the emitted infrared radiation (*dashed line*). The *lower part* shows the net radiation absorbed that is the sum of the two previous terms

albedo of the polar regions. On the other hand, the infrared radiation is emitted mainly at the tropopause and depends on its temperature. This does not change very much with latitude because the tropopause height decreases going from the equator to the pole. The net radiation, that is the difference between the absorbed and emitted radiation, is positive approximately between 40 °N and 40 °S and negative at higher latitudes both north and south. The earth-atmosphere system must be in thermal equilibrium and this means that its temperature must be stationary, and consequently the integral of the net radiation on its surface must be zero. If $F_A$ is the net radiation we must have

$$\int_{-\pi/2}^{\pi/2} 2\pi a^2 F_A \cos\phi d\phi = 0 \tag{11.1}$$

where $a$ is the Earth's radius and φ the latitude. It is interesting then from the data of Fig. 11.4 to calculate the same integrals starting from the north pole to find the net radiation up to some latitude, as shown in Fig. 11.5.

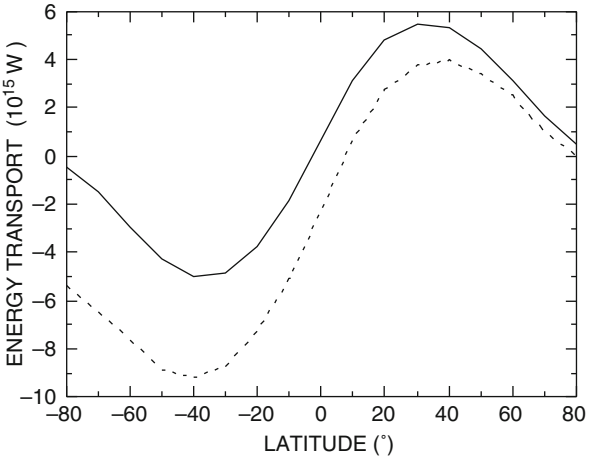



**Fig. 11.5** The net energy as a function of latitude calculated according to Eq. (11.1). The energy emitted at the north pole is assumed to be zero. See text for the meaning of the *dashed line*

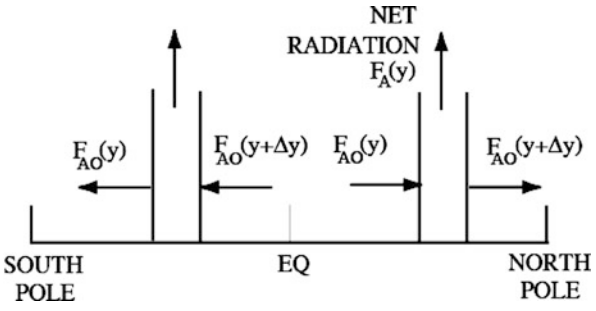



**Fig. 11.6** The energy budget in a semi-infinite box delimited by two walls at latitudes $y$ and $y + \Delta y$

However as a first result we found that the energy at the South Pole is not zero (dashed line). Several reasons can explain this result, although the question is really that the albedo of the Earth–atmosphere system is poorly known in detail. The albedo strongly influences the calculation of the absorbed and reflected radiation. A few tricks are then used to correct this result and the one adopted in this case is to multiply the absorbed radiation by 1.025, which is a very small factor indeed but enough to correct the calculation. Actually the curve shown is the integral multiplied by the cosine of the latitude.

The fact that the sign is positive in the northern hemisphere and negative in the other has no special significance except for the assumption made on the coordinate system. This is shown in Fig. 11.6, where the energy fluxes are represented, as if they were only a function of latitude. We can imagine two walls extending from the surface to the top of the atmosphere. The energy fluxes across the walls (due to the

atmosphere and ocean circulation, for example) are directed toward the poles. If the total flux is indicated with $F_{AO}$ then we have

$$F_{AO}(y + \Delta y) - F_{AO}(y) > 0$$

This means that in that particular latitude interval, the net radiation is absorbed so that at the equilibrium it must be

$$F_{AO}(y + \Delta y) - F_{AO}(y) = 2\pi a^2 F_A \cos\phi \Delta\phi$$

or equivalently

$$\Delta F_{AO}/\Delta y = 2\pi a F_A \cos\phi \tag{11.2}$$

It is clear then that the curve shown in Fig. 11.5 gives the change of the total flux with latitude. In the northern hemisphere, in regions with net absorption ($F_A > 0$), the total flux must increase with increasing latitude and vice versa for $F_A < 0$.

The calculation of $F_{AO}$ is simple. We have

$$F_{AO}(y) = \int_0^{\infty} dz \int_0^{2\pi a\cos\phi} a d\lambda F_D(y,\ \lambda,\ z) = 2\pi a \cos\phi \int_0^{\infty} \overline{F}_D(y, z) dz \tag{11.3}$$

where $\overline{F}_D(y,\ z)$ is the zonally averaged flux at altitude z and latitude y in W m$^{-2}$. If the result of the integral over altitude is $F_E$ we have

$$F_{AO}(y) = 2\pi a \cos\phi F_E$$

which substituted in Eq. (11.2) gives

$$\frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}(\cos\phi F_E) = F_A \tag{11.4}$$

This relation expresses the conservation of energy because it gives the divergence of the total energy flux equal to net radiative flux. The problem we have at this point is to see whether the observations can justify Eq. (11.4). We start by rewriting the same equation as

$$2\pi a^2 \cos\phi F_E = 2\pi a^2 \int_{-\pi/2}^{\phi} \cos\phi F_A d\phi \tag{11.5}$$

where the right-hand side is just the curve shown in Fig. 10.5. The relevance this may have on the study of the atmospheric circulation should be emphasized. Equation (11.5) gives the constraints that any theory on the mechanism for energy transport must obey.

### 11.2.1 Forms of Energy

The energy (per unit surface) for the atmosphere can be divided into internal energy $E_I$, potential $E_P$ and kinetic $E_K$. They are given by the following integrals

$$E_1 = \int_0^\infty \rho C_v T \, dz \tag{11.6a}$$

$$E_P = \int_0^\infty \rho g z \, dz = -E_1 = \int_0^\infty z \, dp \tag{11.6b}$$

$$E_k = \int_0^\infty \rho \, dz \left(u^2 + v^2\right)/2 \tag{11.6c}$$

In the last expression we have neglected the vertical velocity. The potential energy can be written in a slightly different form

$$-\int_{p_0}^0 z \, dp = \int_0^\infty p \, dz = R \int_0^\infty \rho T \, dz \tag{11.7}$$

which added to Eq. (11.6a) gives

$$E_1 + E_p = \int_0^\infty \rho C_v T \, dz + R \int_0^\infty \rho T \, dz = \int_0^\infty \rho C_p T \, dz = C_p E_1 / C_v \tag{11.8}$$

This is a rather interesting conclusion because it says that the sum of the internal and potential energy (also called *total potential energy*) is about 1.4 times the internal energy. Assuming a linear lapse rate $\Gamma$ for the temperature, we can evaluate the average value for Eq. (11.8)

$$E_1 = C_v T_0 p_0 / g \, (1 - R\Gamma / g)$$

which, with appropriate numerical values, gives an energy of about $5 \cdot 10^{14}$ J m$^{-2}$ corresponding to $6.7 \cdot 10^{32}$ J for the globe. Of all this energy, only a tiny amount is available for motion. Before trying to understand why, we like to consider the kinetic term. From the equations of motion, we can easily evaluate the rate of change of the kinetic energy. We start from the equations of motion in terms of the geopotential

$$Du/Dt - fv = -\partial \Phi / \partial x$$

$$Dv/Dt + fu = -\partial \Phi / \partial y$$

multiplying the first by $u$ and the second by $v$ and adding

$$\frac{DK}{Dt} = -\rho\left(u\frac{\partial\Phi}{\partial x} + v\frac{\partial\Phi}{\partial y}\right) \tag{11.9}$$

In this case $K$ is the kinetic energy per unit volume. In the same way we can find the rate for the potential energy per unit volume $P$

$$\frac{DP}{Dt} = \frac{D}{Dt}\left(\rho C_p T\right) = \rho\frac{dQ}{dt} + \frac{Dp}{Dt} = \rho\left(\frac{dQ}{dt} - wg\right) \tag{11.10}$$

The right-hand side of Eq. (11.9) can be written as

$$\left(u\frac{\partial\Phi}{\partial x} + v\frac{\partial\Phi}{\partial y}\right) = \nabla\cdot(\Phi\mathbf{V}) - w\frac{\partial\Phi}{\partial z} = \nabla\cdot(\Phi\mathbf{V}) - wg$$

And the rate of change of the total energy $E$ is given by

$$\frac{DE}{Dt} = \frac{DK}{Dt} + \frac{DP}{Dt} = -\rho\left(\nabla\cdot(\Phi\mathbf{V}) - \frac{dQ}{dt}\right) \tag{11.11}$$

This equation expresses the fact that the total energy increases with heating and is dissipated by the geopotential flux. The same can be cast in a Eulerian form much more simply. From the definition of total derivative, we have

$$\frac{\partial E}{\partial t} = -\nabla\cdot(K\mathbf{V}) - \rho\nabla\cdot\left[(gz\mathbf{V}) + \left(C_p T\mathbf{V}\right)\right] + \rho\frac{dQ}{dt} \tag{11.12}$$

The first term on the right-hand side represents the flux of kinetic energy which is usually negligible. In this way, Eq. (11.12) can be put in a conservative form

$$\frac{\partial E}{\partial t} + \nabla\cdot F_E = \rho\frac{dQ}{dt} \tag{11.13}$$

where

$$F_E = \rho\left(gz\mathbf{V}\right) + \rho\left(C_p T\mathbf{V}\right)$$

We can perform the divergence (neglecting the vertical terms) and make a time and a zonal average

$$\frac{\partial\left[\overline{E}\right]}{\partial t} = -\frac{\partial}{\partial y}\left(pg\left[\overline{zv}\right] + \rho C_p\left[\overline{vT}\right]\right) + \rho\frac{d\left[\overline{Q}\right]}{dt} \tag{11.14}$$

The zonal mean in this case is indicated with the brackets, while the time average is indicated with the overbar. This is the rate of change of energy per unit volume,

and if we want the total change it is necessary to integrate Eq. (11.14) with respect to volume. This operation is more convenient in spherical coordinates, so we get for the total change of energy

$$\int \overline{\frac{\partial [E]}{\partial t}} dV = 2\pi a \int_0^\infty dz \int_{-\pi/2}^{\phi} d\left[\rho \cos\phi \left(g\,[\overline{zv}] + C_p[\overline{vT}]\right)\right] + 2\pi a^2 \int_0^\infty dz \int_{-\pi/2}^{\phi} \rho \cos\phi \frac{d\,[\overline{Q}]}{dt} d\phi \tag{11.15}$$

The diabatic term includes both the radiative heating and the latent heat contribution. In stationary conditions the right-hand-side integral can be related to Eq. (11.5) if we put

$$F_E = \int_0^\infty \rho \left(g\,[\overline{zv}] + C_p\,[\overline{vT}]\right) dz; \qquad F_A = \int_0^\infty \rho \frac{d\,[\overline{Q}]}{dt} dz \tag{11.16}$$

Actually in Eq. (11.5), the quantity $F_A$ included only the radiative contribution which in this case is extended also to the latent heat. This distinction is not important in any case. We have expressed the simple fact that diabatic heating (radiative and latent heat) must be balanced by the sensible heat flux and the geopotential flux. This can be easily verified using data similar to those reported by Oort and Rasmussen. However, we have used at length concepts like zonal average, or time average, and we are at the point that we need to think about them a little.

### *11.2.2 Decomposition of Transport*

Equation (11.14) gives the zonally average flux of the sensible heat and the geopotential. Actually, the quantities appearing in that equation are quite variable in time and in space. It is then convenient to resort to averages and the relative deviations.

We consider, for example, a quantity $A$ and for it we can define the time average over the interval $\tau$

$$\overline{A} = \frac{1}{\tau} \int_0^\tau A\, dt \tag{11.17}$$

And the deviation from the mean $A'$ will be such that

$$A = \overline{A} + A' \tag{11.18}$$

We have talked a lot about zonal mean which we remember is given by

$$[A] = \frac{1}{2\pi}\int_0^{2\pi} A d\lambda \tag{11.19}$$

while the deviations $A^*$ will be such that

$$A = [A] + A^* \tag{11.20}$$

Knowing these simple rules we can proceed with more complex operations like

$$\overline{AB} = \overline{\left(\overline{A} + A'\right)\left(\overline{B} + B'\right)} = \overline{AB} + \overline{A'B'} \tag{11.21}$$

or

$$[AB] = [A][B] + \left[A^* B^*\right] \tag{11.22}$$

We are mainly interested in zonal averages of time averages like the quantities appearing in Eq. (11.14). To accomplish this we consider first the zonal mean

$$AB = \left([A] + A^*\right)\left([B] + B^*\right) = [A]\ [B] + [A]B^* + [B]A^* + A^* B^*$$

$$[AB] = [A][B] + \left[A^* B^*\right]$$

Then we know that

$$[A] = \overline{[A]} + [A]'; \qquad [B] = \overline{[B]} + [B]'$$

And substituting we have for the zonal mean of a product

$$[AB] = \overline{[A][B]} + \overline{[A]}[B]' + \overline{[B]}[A]' + [A]'[B]' + \left[A^* B^*\right]$$

while the time average

$$\overline{[AB]} = \overline{[A][B]} + \overline{[A]'[B]'} + \overline{[A^* B^*]} \tag{11.23}$$

It is instructive to evaluate the same quantity starting from the time average

$$\overline{AB} = \left(\overline{A} + A'\right)\left(\overline{B} + B'\right) = \overline{AB} + \overline{A'B'}$$

And considering that $\overline{A} = \left[\overline{A}\right] + \overline{A}^*$ we have

$$\overline{AB} = \left[\overline{A}\right]\left[\overline{B}\right] + \left[\overline{A}\right]\overline{B}^* + \left[\overline{B}\right]\overline{A}^* + \overline{A}^*\overline{B}^* + \overline{A'B'}$$

so that the zonal mean

$$\overline{[AB]} = \overline{[A]}\,\overline{[B]} + \left[\overline{A}^{*}\overline{B}^{*}\right] + \overline{[A'B']} \tag{11.24}$$

Equations (11.23) and (11.24) are the same so that also the right-hand side must be the same as can be easily shown.

A first conclusion is that any quantity can be written as a function of its spatial and time averages

$$\overline{A} = \left[\overline{A}\right] + \overline{A}^{*} + [A]' + A'^{*} \tag{11.25}$$

The first term ([A]) represents the zonal mean of the time-averaged value. For example in Fig. 10.1, this term could be appropriate for the trade winds or, even better, the westerlies shown in Fig. 10.3. The second term $\overline{A}^{*} = \overline{A} - [A]$ represents the time-averaged of the deviations from the zonal mean and so may refer to the asymmetric part of time average quantities, like the monsoon circulation. The third term $[A]' = [A] - [A]$ represents the temporal deviations from the zonal mean and then the fluctuations with respect to the zonal mean. Finally, the last quantity is the instantaneous value of the deviation from the zonal mean and consequently the lows and highs. The meaning of these terms will be more clear when we apply them to the energy budget of the Earth–atmosphere system.

### *11.2.3 The Details of the Energy Budget*

We apply immediately what we have illustrated in the previous section. The average quantities appearing in Eq. (11.14) can be written

$$\left[\overline{vT}\right] = \left[\overline{v}\right]\left[\overline{T}\right] + \left[\overline{v}^{*}\overline{T}^{*}\right] + \left[\overline{v'T'}\right] \tag{11.26}$$

This relation says that the total sensible heat transport is due to the mean zonal meridional circulation (first term), to the mean zonal stationary eddies (that is the time average of the deviations from the zonal mean) and to the mean zonal transient eddies (deviations from the time average). This last term takes care of the correlation between the time deviations. In the way we can also write the geopotential and the latent heat term

$$L\left[\overline{vq}\right] = L\left[\overline{v}\right]\,\left[\overline{q}\right] + L\left[\overline{v}^{*}\overline{q}^{*}\right] + L\left[\overline{v'q'}\right] \tag{11.27}$$

where $L$ is the latent heat of condensation and $q$ is the mass mixing ratio for water vapor. The data for calculating the quantities (11.26), (11.27) and (11.15) can be obtained from the Oort and Rasmussen statistics. One first result is shown in Fig. 11.7, where the different terms of the budget are shown as a function of

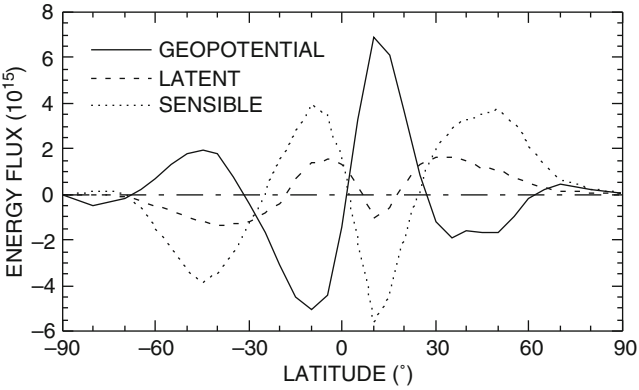



**Fig. 11.7** The flux of the different components that contribute to the energy flux. The values are referred to the flux crossing a circle of latitude

latitude. We can notice a number of interesting things. First of all, if we look at the behavior of the latent heat meridional flux, we may note a first positive maximum at about 10° south and a negative one at 10° north. This means that in the narrow equatorial region precipitation is much higher than evaporation so that the region is a sink for water vapor. The water vapor flux in this region in both hemispheres is toward the equator where it condenses and rains out. Tropics, on the other hand, are a source for water vapor because the flux is positive in the northern hemisphere (from the tropics to higher latitudes) and negative in the southern hemisphere (from the tropics toward the south pole). A rather similar situation is obtained for the sensible heat, for which the equatorial region is actually a sink. At first sight this may look contradictory because the equator should be a region of heat production. Actually, if we look at the meridional circulation (Fig. 11.2) below 5 km, there is a region where cold air in the Hadley cell returns from high latitude moving toward the equator. The region where heat is exported from the equator is located at high altitude, where density and temperature are low enough to have a negligible effect on the flux. At these latitudes, the mean circulation dominates transport. At middle latitudes transient eddies are dominating the transport, and the tropics and middle latitudes are the main sources. In the case of the geopotential flux the situation is the opposite because the source is located at the equator. The reason is that the values of the mean circulation are high but opposite in sign both at high and low altitude. The tropics in this case are a sink for the fluxes arriving from the high latitudes. An interesting point is the opposite sign of the geopotential and sensible heat. The detailed explanation for that will be found in the discussion of the energy cycles. However an intuitive explanation is found considering that the geopotential flux is related to the kinetic energy while the sensible heat flux is related to the change in potential energy.

It should be obvious that when the kinetic energy increases, the potential energy decreases and vice versa, while their sum must be equal to the net diabatic flux. This

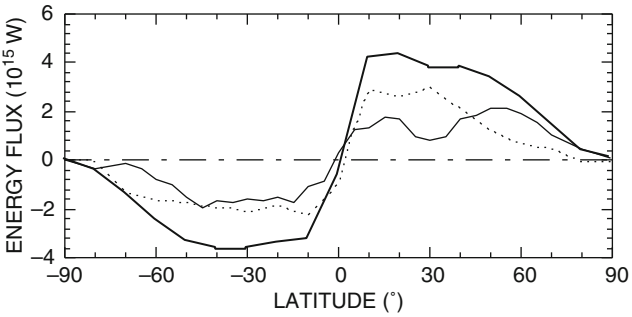



**Fig. 11.8** The budget between the diabatic (*thick line*) and dynamic terms. The atmospheric contribution is shown with the *thin line* while that of the ocean with the *dashed line*

sum is shown in Fig. 11.8, and we notice immediately a problem: apparently the two fluxes are not able to balance the diabatic term. We assume that the missing term is due to the oceanic heat flux which is obtained as a difference. Both the oceanic and atmospheric fluxes are directed toward the respective poles in the two hemispheres and are approximately of the same order of magnitude. Dealing here only with the atmospheric part we will require that any theory on the general circulation be constrained to explain the behavior of the energy fluxes with latitude and in particular the fact that they have a maximum around middle latitudes.

## 11.3 The Mean Zonal Circulation

The debate about what keeps the atmospheric circulation going can be reduced to the bone by saying that, while for the tropics there are few doubts about the engine (a kind of convective cell or direct cell), for the middle and high latitudes the eddies could be responsible.

The data in the previous paragraph are consistent with a simple theory (more like a diagnostic theory) taken from a series of lectures given by John Wallace in 1978 at the National Center for Atmospheric Research.

The starting point again is the local forcing due to the pressure gradients are balanced by the Coriolis acceleration.

$$\frac{D\overline{u}}{Dt} = \frac{\partial \overline{u}}{\partial t} + \overline{u}\frac{\partial \overline{u}}{\partial x} = -f\left(\overline{v} - \overline{v}_g\right) \tag{11.28}$$

Again we notice that the quantity with the overbar denotes time averages. In the Eq. (11.28) we have kept the most important terms. The zonal acceleration resulting from this equation can be easily calculated so that, because the advective term is about 30 m $s^{-1}$ $day^{-1}$, the difference between the meridional velocity and its geostrophic component is about 3 m $s^{-1}$. We are interested in the zonal mean values

and it may be useful to consider local equilibrium. We consider the deviation from the time average so that $u = \overline{u} + u'$ and, $v = \overline{v} + v'$. With these substitutions, Eq. (11.29) becomes, adding the meridional advective term,

$$\frac{\partial \overline{u}}{\partial t} = -\overline{u}\frac{\partial \overline{u}}{\partial x} - \overline{v}\frac{\partial \overline{u}}{\partial y} - \frac{\partial}{\partial x}\overline{u'^2} - \frac{\partial}{\partial y}\overline{u'v'} + f\left(\overline{v} - \overline{v}_g\right) = 0 \tag{11.29}$$

This is the equation that describes the equilibrium for the zonal wind. From the data, it is possible to establish that the transient terms (third and fourth on the right-hand side) are about three times smaller than the first so that Eq. (11.29) is valid locally. Then if Eq. (11.29) is zonally averaged we should write each term like $u = [\overline{u}] + u'$. We obtain from Eq. (11.29)

$$\frac{\partial [\overline{u}]}{\partial t} = -\frac{1}{2}\frac{\partial}{\partial x}\left[\overline{u}^2\right] - \overline{v}\frac{\partial [\overline{u}]}{\partial y} - \left[\overline{v}^*\frac{\partial \overline{u}^*}{\partial y}\right] - \frac{1}{2}\frac{\partial}{\partial x}\left[\overline{u'^2}\right]$$
$$-\frac{\partial}{\partial y}\left[\overline{u'v'}\right] + f\left([\overline{v}] - \left[\overline{v}_g\right]\right) = 0 \tag{11.30}$$

The terms that appear as derivative with respect to $x$ when averaged zonally disappear, as happens for, because it is proportional to the pressure gradient in the $x$ direction. The third term on the right can be written as

$$-\left[\overline{v}^*\frac{\partial \overline{u}^*}{\partial y}\right] = -\frac{\partial}{\partial y}\left[\overline{u}^*\overline{v}^*\right] + \left[\overline{u}^*\frac{\partial \overline{v}^*}{\partial y}\right]$$

Assuming that the wind has zero divergence we obtain

$$\frac{\partial \overline{u}^*}{\partial x} = -\frac{\partial \overline{v}^*}{\partial x} \Rightarrow \left[\overline{u}^*\frac{\partial \overline{v}^*}{\partial y}\right] = -\left[\overline{u}^*\frac{\partial \overline{u}^*}{\partial x}\right] = -\frac{1}{2}\frac{\partial\left[\overline{u}^{*2}\right]}{\partial x} = 0$$

Considering these simplifications Eq. (11.29) becomes

$$\frac{\partial [\overline{u}]}{\partial t} = [\overline{v}]\left(f - \frac{\partial [\overline{u}]}{\partial y}\right) - \frac{\partial}{\partial y}\left\{\left[\overline{u}^*\overline{v}^*\right] + \left[\overline{u'v'}\right]\right\} = 0 \tag{11.31}$$

In this equation we notice how the momentum budget is determined by both mean quantities, like the first term on the right-hand side which are the meridional vorticity net flux and eddy terms. These are of stationary types (due to deviations from zonal mean quantities) and transient (due to deviations from time mean quantities). In particular, the second term on the right represents the divergence of the momentum flux. If we indicate with $G$ the divergence of the eddy fluxes and with $F$ the force per unit mass due to friction we obtain

$$\frac{\partial [\overline{u}]}{\partial t} = [\overline{v}]\left(f - \frac{\partial [\overline{u}]}{\partial y}\right) + G + F = 0 \tag{11.32}$$

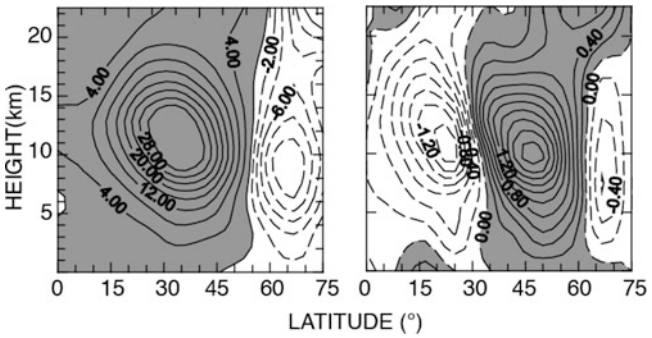



**Fig. 11.9** The *left panel* shows the annual mean values of the zonal momentum flux (units of $m^2\ s^{-2}$) while the *right panel* shows the divergence (units of $m\ s^{-1}\ day^{-1}$)

This relation can be used to establish the order of magnitude of the meridional circulation and its direction. In particular both $G$ and $F$ can be obtained from the data (i.e. Oort and Rasmusson). For example, the transport of the zonal momentum

$$\left\{[\overline{u}^*\overline{v}^*] + \left[\overline{u'v'}\right]\right\}$$

and its divergence ($G$). Both quantities are shown in Fig. 11.9. From this figure we notice that the momentum flux is always positive for latitudes lower than 50° and presents a maximum at about 30° at an altitude of about 12 km. Above 50° N the flux becomes negative, at least up to 75° N, so that for this latitude range the Earth is "pushed" by the wind.

The divergence of this flux actually represents the acceleration on the mean zonal wind which is then negative up to 30° N and then becomes positive. The maximum value of this acceleration (annual mean is of the order of 1.7 $m\ s^{-1} day^{-1}$) can reach even 5 $m\ s^{-1} day^{-1}$ in winter. These data show that there must exist a source of momentum at middle latitudes, as is shown schematically in Fig. 11.10. In this figure is also shown the position of the jet and the $F$ term. The friction is essentially confined to the planetary boundary layer and will accelerate the motion in the tropical region. In that region, the zonal wind is negative and the friction will give a positive acceleration. The middle and high latitudes are dominated by westerlies, and in this case the zonal wind is positive and the Earth is pushed by the winds. In the regions where it is possible to neglect friction (for example at the jet level) Eq. (11.32) has a very simple form that gives the order of magnitude for the meridional velocity

$$[\overline{v}] = -G/\left(f - \partial\left[\overline{u}\right]/\partial y\right) \tag{11.33}$$

For middle latitude $f = 10^{-4}\ s^{-1}$, while from the previous figure $G$ is about 5 $m\ s^{-1}$ $day^{-1}$ and $\partial[\overline{u}]/\partial y \approx 0.2\ 10^{-4}\ s^{-1}$. The upper branch of the Ferrel cell has then a velocity around 0.4 $m\ s^{-1}$ directed toward south because $G$ is positive.

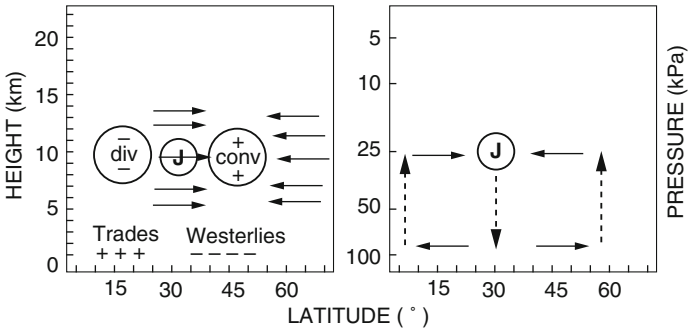



**Fig. 11.10** A schematic illustration of the meridional circulation (*right*). On the *left* are shown the regions of positive and negative divergence for the momentum flux. The positive and negative signs near the surface indicate the action of friction

The rate of change of temperature can be obtained directly from the thermodynamic equation so that

$$\frac{\dot{q}}{C_p} = \frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + w\left(\frac{\partial T}{\partial z} + \frac{g}{C_p}\right) \tag{11.34}$$

We can proceed in the same way as in the case of the acceleration of the mean flux to obtain

$$\frac{\partial \left[\overline{T}\right]}{\partial t} + (\Gamma_d - \Gamma)\left[\overline{w}\right] = \frac{\dot{q}}{C_p} - \frac{\partial}{\partial y}\left[\overline{v'T'}\right] \tag{11.35}$$

From this equation, we can see that the temperature change is determined not only by the diabatic term but also by the eddy heat flux divergence. For stationary conditions, we obtain for the vertical velocity

$$[w] = (P + Q) / (\Gamma_d - \Gamma) \tag{11.36}$$

where we have put

$$P = -\partial\left[\overline{v'T'}\right]/\partial y \qquad Q = \left[\overline{\dot{q}}\right]/C_p.$$

In Fig. 11.11, the term $P + Q$ of Eq. (11.36) is shown as the annual mean. The diabatic heating is maximum at the equator while for the eddy flux the divergence is negative below 30° N and positive at higher latitudes. This means that because the term $(\Gamma_d - \Gamma)$ is always positive, there will be upward motions in the equatorial region and downward motions near the tropics. The motion will be again positive upward around 60° N. These estimations are then consistent with the previous one under the provision of non-acceleration of the zonal mean flux.

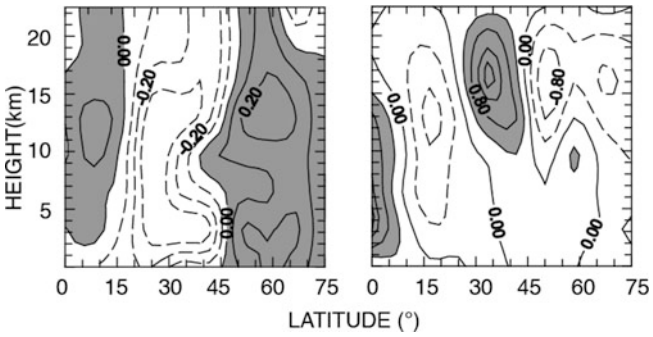



**Fig. 11.11** Annual mean values of the divergence of the eddy heat flux (*left*) in °C day$^{-1}$ and of the diabatic heating (*right*). Notice the prevalence of the diabatic heating near the equator and of the eddy term at middle and high latitudes

The conclusion we may drawn is that almost all the characteristics of the mean meridional circulation can be obtained from the condition of zero acceleration on the zonal wind and thermal equilibrium. From Fig. 11.9, we see that at 45° N the zonal wind acceleration due to the eddy flux is such that the zonal wind increases at high altitude more than at low altitude and consequently the tendency is to increase the wind shear. On the other hand, from Fig. 11.11, at the same latitude the eddies tend to decrease the latitudinal temperature gradient heating the atmosphere at high latitude and cooling it at low and middle latitudes, thus reducing the temperature latitudinal gradient. Reducing the temperature gradient will reduce the thermal wind.

The Ferrel cell in a sense is the consequence of these tendencies because, as it can be seen from Fig. 11.10, the meridional velocity is negative at the jet level and so slows down the zonal wind while the same meridional velocity at the lower level speeds up the zonal wind. Also the vertical branches of the cell are responsible for a cooling in the region of 60° N (adiabatic expansion of the ascending air) and a heating around 30° N (compression of descending air). The Ferrel cell is then generated by the eddies, although this qualitative approach needs to be justified by a mechanism that should be sound also from an energetic point of view.

The Hadley cell can be justified in the same way because from Fig. 11.9 we notice that in the equatorial region the eddies slow down the wind at high altitude (reducing the shear), but at the same time the eddies also increase the latitude temperature gradient. In reality, the Hadley cell is so direct that it does not need the presence of the eddies; rather as we will see in a moment, the eddies reduces the intensity of the zonal winds.

All these are diagnostic considerations. They simply tell us that the observations are consistent with the circulation, that is, the physics is sound. However physics is more than that and we should be able to find a theory that would explain why the direct cell breaks before reaching the pole, and why the eddies are responsible of the Ferrel cell and find a meaning for all the other data.

A possible idea is the following: The latitudinal thermal gradient it is established for geometrical reason and it is enough to produce a convective cell where the zonal wind increases with altitude to balance the pressure gradient that is a consequence of the thermal gradient. The zonal currents that arise from this situation are unstable in the sense that the minimum wiggle is enough to make them grow and eventually break. This instability generates eddies that can provide the engine for the circulation at middle latitude. We will start on this road with the Hadley circulation which for obvious reasons we will assume to be symmetric.

## E.11 Examples

### *E.11.1 Waves and Momentum Flux*

We will anticipate some of the theory on the general circulation that we will report on the next chapter but in simpler terms. This is based mainly on the lectures given by Held (2000). In particular we refer to the fact that poleward eddy momentum flux is the consequence of Rossby wave propagation toward the equator. Rossby waves can be thought of as being stirred at mid latitudes by baroclinic instability and then propagating to the subtropics. The waves may approach critical latitudes where their phase velocity equals the background zonal flows. At those latitudes waves may grow in amplitude and breaks irreversibly, resulting in the absorption of wave activity. Randel and Held (1991) introduced the concept of phase speed spectrum which decomposes eddy fluxes as a function of phase speed wavenumber.

We may start by noting that the thermal wind equation gives

$$f\frac{\partial U}{\partial z} = -g\frac{\partial T}{\partial y} \tag{E.11.1}$$

Because the gradient $\partial T/\partial y$ is always negative in the troposphere this implies that $U>0$ throughout the troposphere. At the surface, the situation could be a little more tricky. There the pressure gradient is constantly directed toward the tropics and the geostrophic balance can be achieved only with a zonal easterly wind.

Some interesting consideration can be made for the momentum flux. We start by writing the vorticity equation

$$\frac{\partial \zeta}{\partial t} = -u\frac{\partial \zeta}{\partial x} - v\frac{\partial \zeta}{\partial y} - \beta v \tag{E.11.2}$$

where $\zeta$ indicate the relative vorticity. By linearizing this equation ($u = \overline{U}, v = v'$) we have

$$\zeta' = \frac{\partial v'}{\partial x} - \frac{\partial U}{\partial y}; \quad \frac{\partial \zeta'}{\partial t} = -U\frac{\partial \zeta'}{\partial x} - \beta v' + \frac{\partial^2 U}{\partial y^2}v'; \quad \frac{\partial \zeta'}{\partial t} = -U\frac{\partial \zeta'}{\partial x} - \gamma v'$$

where

$$\gamma = \beta - \frac{\partial^2 U}{\partial y^2}$$

In case $U = \cos t$ $\gamma = \beta$ and we have the usual plane wave solution

$$\psi = A \cos (kx + ly - \omega t)$$

with the dispersion relation

$$\omega = ck = Uk - \frac{\beta k}{k^2 + l^2} \equiv \Omega (k, l) \tag{E.11.3}$$

We can evaluate the meridional transport of zonal momentum by eddies

$$\overline{u'v'} = -\overline{\frac{\partial \psi}{\partial y}\frac{\partial \psi}{\partial x}} = -A^2 \frac{kl}{2} \tag{E.11.4}$$

This shows that if $kl > 0$ we have southward momentum flux while if $kl < 0$ we have northward momentum flux. We can now evaluate the meridional group velocity

$$G_y = \frac{\partial \Omega}{\partial l} = \frac{2\beta kl}{(k^2 + l^2)^2} \tag{E.11.5}$$

If we compare (E.11.5) with (E.11.14), we find that meridional flux of zonal momentum is positive when $kl < 0$ and vice versa. We now consider a source and north of the source the meridional group velocity must be positive so we chose $kl > 0$ while south of the source we choose $kl < 0$ so that $G_y$ is negative.

Meaning that the meridional momentum flux is opposite to the group velocity. This argumentation can be seen also from another point. We assume that there is small amount of damping in the propagation and the addition implies an imaginary part to the frequency that is

$$\omega + i\kappa = \Omega (k, l)$$

We can compute $l$ by assuming that the new solution is close to the inviscid solution

$$\omega + i\kappa = \Omega (k, l_0) + \frac{\partial \Omega (k, l_0)}{\partial l} l_1$$

And this substituted in the original wave solution gives

$$l_1 = i\kappa \left( \frac{\partial \Omega (k, l_0)}{\partial l} \right)^{-1} = \frac{i\kappa}{G_y (k, l_0)}$$

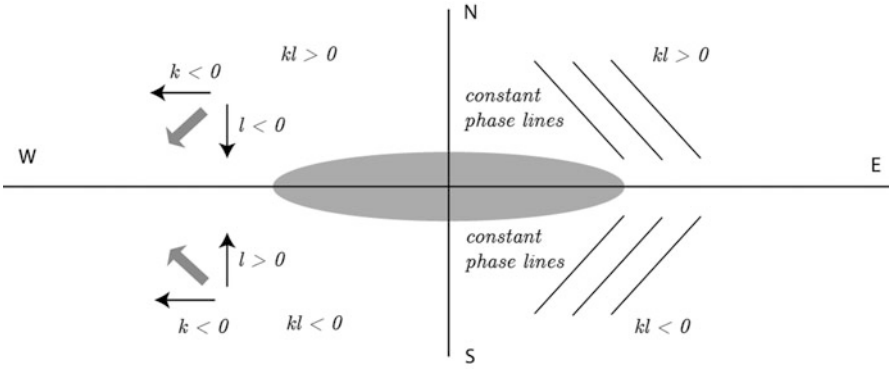



**Fig. E.11.1** Propagation of Rossby waves with respect to source shown as the elliptical shadow

Where we have assumed $l = l_0 + l_1$ and $\omega = \Omega\,(k, l_0)$. We have then

$$\tilde{\psi} = \mathrm{Re}\left[\exp\left(il_0 y - \frac{\kappa y}{G_y}\right)\right]$$

In the presence of damping, the solution must decay away from the source and this requires that $G_y$ to be positive north of the source.

Vice versa south of source $G_y$ must be negative.

The situation is illustrated in Fig. E.11.1, where the source is shown as and elliptical shadow. The constant phase lines to the north and south of the source are illustrated to the left of the figure. On the left, the grey arrows show the propagation of momentum which is coherent with the direction of the group velocity.

## *E.11.2 Waves and Vorticity Flux*

The angular momentum flux and the vorticity flux are related. Consider the meridional vorticity flux

$$\zeta' v' = \left(\frac{\partial v'}{\partial x} - \frac{\partial u'}{\partial y}\right) v' = \frac{1}{2}\frac{\partial v'^2}{\partial x} - \frac{\partial u'v'}{\partial y} + u'\frac{\partial v'}{\partial y} = \frac{\partial}{\partial x}\left[\frac{1}{2}\left(v'^2 - u'^2\right)\right] - \frac{\partial u'v'}{\partial y}$$

When averaged zonally (over $x$)

$$\overline{\zeta' v'} = -\frac{\partial \overline{(u'v')}}{\partial y} \tag{E.11.6}$$

so that the average vorticity flux is the divergence of the momentum flux. The equation of motion in the $x$ direction is given by

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = fv - \frac{1}{\rho}\frac{\partial p}{\partial x}$$

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} + fv - \frac{1}{\rho}\frac{\partial p}{\partial x} = fv - \frac{\partial\left(u^2\right)}{\partial x} - \frac{\partial\left(uv\right)}{\partial y}$$

Obtained using the continuity equation. Then zonally averaging we obtain

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial\left(\overline{u'v'}\right)}{\partial y} \tag{E.11.7}$$

Equating (E.11.6) and (E.11.7) we have

$$\frac{\partial \overline{u}}{\partial t} = \overline{v'\zeta'} \tag{E.11.8}$$

Relating the acceleration of the zonal flow to the vorticity eddy flux. We now consider a more general form of the vorticity equation

$$\frac{\partial \zeta}{\partial t} = -u\frac{\partial \zeta}{\partial x} - v\frac{\partial \zeta}{\partial y} - \beta v + S - D \tag{E.11.9}$$

where $S$ is the stirring and $D$ is some unspecified damping. If this equation is linearized about the zonal flow we have

$$\frac{\partial \zeta'}{\partial t} = -\overline{u}\frac{\partial \zeta'}{\partial x} - \gamma v' + S' - D' \tag{E.11.10}$$

where

$$\gamma = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} \tag{E.11.11}$$

We now multiply (E.11.10) by $\zeta'/\gamma$ and average over $x$. We obtain

$$\frac{\partial P}{\partial t} = -\overline{v'\zeta'} + \frac{\overline{S'\zeta'}}{\gamma} - \frac{\overline{D'\zeta'}}{\gamma} \tag{E.11.12}$$

where

$$P = \frac{1}{2\gamma}\overline{\zeta'^2} = \frac{\gamma}{2}\overline{\eta^2} \tag{E.11.13}$$

One can think of $\eta = \zeta'/\gamma$ being the ratio of vorticity (dimension $s^{-1}$) and $\gamma$ dimension ($m^{-1}s^{-1}$) as the meridional parcel displacement that would create the perturbation vorticity. $P$ is referred as *pseudomomentum* density of the waves. Based on (E.11.12) and (E.11.8) we can write

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial P}{\partial t}$$

$P$ is always defined positive so that in a region $P$ increase with time in the same region the mean zonal velocity will decrease with time. $P$ is then a measure of wave amplitude.

### *E.11.3 More on Pseudomomentum*

There is another way to understand $P$. Suppose we consider a fluid parcel at latitude $y_0$ at time $t$. Consider now a previous time $t = 0$ and locate the same particle which trace a curve $y - y_0 = \eta(x)$. If this curve behaves normally and that is a single value curve in $x$. The vorticity flux through this latitude can be calculated as

$$\overline{\int_{y_0}^{y_0+\eta} [\zeta(0) + \gamma y] dy} = \frac{\gamma}{2}\overline{\eta^2} \tag{E.11.14}$$

If we assume

$$u = -\partial\psi/\partial y \quad v = \partial\psi/\partial x$$
$$\psi' = A\cos(kx + ly - \omega t) \Rightarrow \zeta' \approx -\left(k^2 + l^2\right)\psi'$$

An from the definition of pesudomomentum and averaging over cosine we get

$$P = \frac{1}{4\gamma}\left(k^2 + l^2\right)^2 A^2 \tag{E.11.15}$$

At this point we have from (E.11.5)

$$G_y P = \beta k l A^2/2\gamma$$

In our case $\beta = \gamma$ and from (E.11.4) we get

$$G_y P = -\overline{u'v'} \tag{E.11.16}$$

In general in presence of forcing and dissipation we have

$$\frac{\partial P}{\partial t} = -\partial\left(-\overline{u'v'}\right)/\partial y + \text{sources} - \text{sinks}$$

The conclusion is that in statistically steady state the eddy momentum flux convergence $\partial\left(-\overline{u'v'}\right)/\partial y$ is determined by the sources and sinks of pseudomomentum.

It is interesting to note that eddy energy is not conserved in this simple model but the waves can exchange energy with the mean flow. If we use (E.11.7) and multiply by $u$ and then integrate over $y$

$$\frac{\partial}{\partial t}\int \frac{1}{2}\overline{u}^2 dy = \int \overline{u'v'}\frac{\partial \overline{u}}{\partial y}dy$$

The term on the left is the change in kinetic energy of the flow that must be conserved, so changes in zonal kinetic energy must be compensated by changes in eddy energy

$$\frac{\partial}{\partial t}\int \frac{1}{2}\overline{u'^2 + v'^2}dy = -\int \overline{u'v'}\frac{\partial \overline{u}}{\partial y}dy$$

The global integrated eddy kinetic energy decays as eddies propagates from regions of large zonal flow to region of weaker flow.

## References[1]

### *Books*

Oort AH (1973) Global atmospheric circulation statistics, 1958–1973, NOAA Professional paper 14. U.S. Printing Office

Oort AH, Peixoto JP (1983) Global angular momentum and energy balance requirements from observations. Adv Geophys 25:355

Peixoto JP, Oort AH (1992) Physics of climate. American Institute of Physics, New York

[1]Here we are in the terrain of Peixoto and Oort's book and their paper on *Advances in Geophysics*. Data probably can be updated on the Web with respect to Oort. The example are drawn from Isaac Held lectures.

## *Articles*

Held IM. The zonal mean climate in the troposphere. Available at (if it is still possible) http://www.gfdl.noaa.gov/ih

Randel WJ, Held IM (1991) Phase speed spectra of transient eddy fluxes and critical layer absorption. J Atmos Sci 48:688–697

Wallace JM (1978) Maintenance of the zonally averaged circulation: a eulerian perspective, in the general circulation: theory, modeling and observations. *NCAR Technical Note, NCAR/CQ-6+1978-ASP*

Washington WM, Parkinson CL (1986) An introduction to three dimensional climate modeling. University Science Books, Oxford University Press, Sausalito, CA, USA

# Chapter 12
# Theories on the General Circulation of the Atmosphere

It would have been too much to have introduced in the previous chapter the theory (or the theories) on the general circulation of the atmosphere. In this chapter, we will try to understand the mechanisms that are important to general circulation. We will start with a rather ambiguous division between tropical and extra-tropical circulation, but this may be justified due to the fact that the physical processes that determine the circulation are quite different and can thus be treated differently. One thing to notice is that such theories, discussed since the seventeenth century, are far from being definitive, but as usual our purpose is to hope that the reader become so interested and involved as to try to formulate a theory by himself (i.e., a very arduous proposition).

## 12.1 The Equatorial Circulation

With this name, we will indicate all the circulations that may explain the Hadley cell, and the simplest example is that of a circulation that is perfectly symmetric with respect to the equator. This example is from A. Gill and although, as we will see, it does not have many points in common with reality, it may be helpful to understand the point.

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_12) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_12

### 12.1.1 Gill's Symmetric Circulation

We will start by using the motion and thermodynamic equations. The heating produced by the solar radiation is assumed to be symmetric with respect to the equator. We will write the thermodynamic equation including the diabatic heating term which is expressed in $s^{-1}$ because it has been normalized by dividing for the average temperature:

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\frac{\rho'}{\overline{\rho}}-\frac{N^2}{g}w'+q'=0 \tag{12.1}$$

The heating term $q'$ represents the net heating, that is, the sum of the solar $q$ and infrared components. The planetary component is approximated by a relaxation term $-\alpha T'/T$ where $T'$ represents the temperature perturbation. We can obviously write $T'/T=-\rho'/\rho$ and the net heating becomes

$$q'=q+\alpha\rho'/\overline{\rho}.$$

In this specific case, we will assume that the variables are independent of longitude (i.e., $x$) so that Eq. (12.1) becomes

$$\left(\frac{\partial}{\partial t}+\alpha\right)\frac{\rho'}{\overline{\rho}}-\frac{N^2}{g}w'+q=0 \tag{12.2}$$

When writing the equations of motion, we should introduce also a viscous term proportional to the velocity. It is to note that such a term has nothing to do with the friction between the atmosphere and the surface but is rather a manifestation of the viscous stress we treated in Chap. 8. The reason for this term (also called *Rayleigh friction*) is that without it we could not have stationary solutions. Indicating with $r$ the friction coefficient, we have for the equations of motion

$$\left(\frac{\partial}{\partial t}+r\right)u-fv=0;\qquad\left(\frac{\partial}{\partial t}+r\right)v+fu=-\frac{1}{\overline{\rho}}\frac{\partial p}{\partial y} \tag{12.3}$$

A stationary solution is possible only in the case when the temperature meridional gradient is balanced by a meridional circulation (i.e., $v$ and $w$), and this is possible only if the equilibrium is established between the Coriolis force and the friction, as is clearly shown in the first part of Eq. (12.3).

To eliminate pressure between Eqs. (12.2) and (12.3), we recur to the hydrostatic equilibrium and obtain

$$\frac{N^2}{g}\frac{\partial w}{\partial y}-\frac{\alpha}{rg}\left(f^2+r^2\right)\frac{\partial v}{\partial z}-\frac{\partial q}{\partial y} \tag{12.4}$$

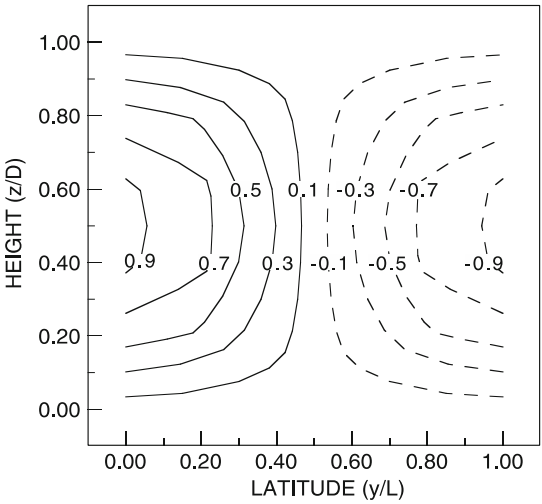



**Fig. 12.1** The normalized heating function given by Eq. (12.6). Positive values indicate heating

where we have assumed the Coriolis parameter to be constant with latitude. We can now introduce a streamfunction $\Psi$ on the $y$, $z$ plane such that

$$v = -\frac{\partial \psi}{\partial z}; \quad w = \frac{\partial \psi}{\partial y}$$

Substituting in Eq. (12.5), we have

$$\frac{N^2}{g}\frac{\partial^2 \psi}{\partial y^2} - \frac{\alpha}{rg}\left(f^2 + r^2\right)\frac{\partial^2 \psi}{\partial z^2} - \frac{\partial q}{\partial y} = 0 \tag{12.5}$$

This equation is solved in a domain defined in latitude by the equator ($y = 0$) and $y = L$. In altitude, it goes from the surface $z = 0$ to $z = D$, $y = L$. In this region, the heating is described as

$$q = q_0 \cos(ly) \sin(mz) \tag{12.6}$$

where $l = \pi/L$ and $m = \pi/D$. Figure 12.1 shows this function and the heating effect near the equator and the cooling at the tropics are evident.

The solution for the streamfunction can be guessed by the boundary conditions. At the top ($z = D$), the vertical velocity must be zero (rigid lid condition), and this implies $\Psi$ to be constant with latitude. We then assume that for both the top and the surface $\Psi = 0$. We look for a solution of the form

$$\psi = \psi_0 \sin(ly) \sin(mz) \tag{12.7}$$

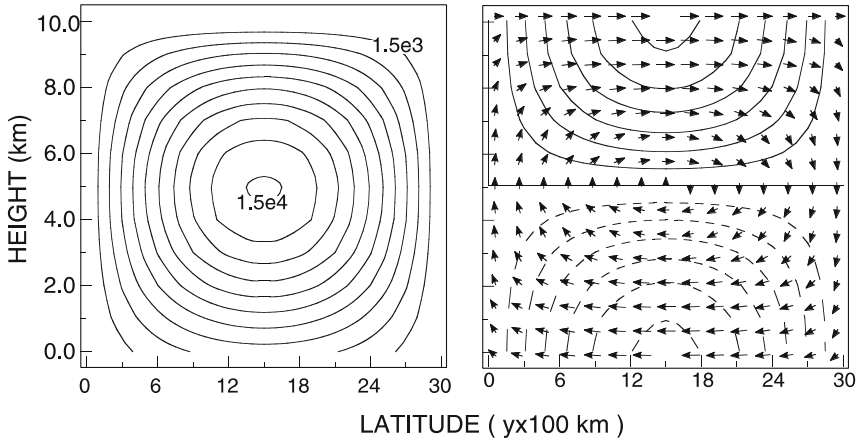



**Fig. 12.2** The streamfunction (in $m^2\ s^{-1}$) for the symmetric Gill circulation (*left*). Contours are at intervals $1.5 \times 10^3$. The latitude scale in km corresponds to amplitude from 0° to 30°. Notice that the condition of a vertical velocity at the lower boundary implies a discontinuity in the bottom contours. (*right*) The circulation. The *solid and dashed lines* represent the easterlies and westerlies, respectively. Notice that the contours are traced at intervals of 30 m $s^{-1}$ so that the maximum velocity is 180 m $s^{-1}$. *Arrows* (not in scale) represent the meridional circulation

that substituted in Eq. (12.5) gives for the amplitude:

$$\psi_0 = \frac{lrgq_0}{[N^2l^2r + \alpha m^2 (f^2 + r^2)]} \tag{12.8}$$

It is very instructive to see how the velocity field behaves. We have

$$v = -m\psi_0 \sin(ly) \cos(mz); \quad w = l\psi_0 \cos(ly) \sin(mz) \tag{12.9}$$

From the stationary condition of the first part of Eq. (12.3), we can obtain the zonal velocity $u = fv/r$ which gives the zonal wind proportional to the meridional wind. This means, as is shown in Fig. 12.2, a positive meridional velocity (northward) for $D/2 < z < D$ and a negative (southward) one in the half of the region near the equator. As a further consequence, the zonal velocity will be positive in the high troposphere (westerly) and negative in the lower troposphere (trade winds).

The solution seems to work at least qualitatively but can be further implemented. The condition of zero vertical velocity at the ground is not realistic. As we have seen in Chap. 8, the velocity at the top of the boundary layer is given by

$$w = \zeta (K/2f)^{1/2}$$

This condition changes the boundary conditions and we now look for a different solution:

$$\psi = \psi_0 \sin(ly) Z(z) \tag{12.10}$$

that substituted in Eq. (12.5) gives a vertical structure equation for $Z$:

$$Z''(z) - \lambda^2 Z = \mu^2 \sin(mz) \tag{12.11}$$

where

$$\lambda^2 = N^2 l^2 r / \left[\alpha \left(f^2 + r^2\right)\right]; \qquad \mu^2 = -lrgq_0 / \left[\psi_0 \alpha \left(f^2 + r^2\right)\right] \tag{12.12}$$

The solution of Eq. (12.11) has the form

$$Z = A \sinh(\lambda z + a) + B \sin(mz)$$

where $A$ and $B$ are constants to be determined and the streamfunction

$$\psi = \psi_0 \sin(ly) \left[C \sinh(\lambda z + a) + \sin(mz)\right] \tag{12.13}$$

the condition that for $z = D$, $Y = 0$ gives $a = -\lambda D$ at the lower boundary:

$$\frac{\partial \psi}{\partial y} = -\frac{\partial u_g}{\partial y}\left(\frac{K}{2f}\right)^{1/2} \Rightarrow \psi = -\gamma u_g; \qquad \gamma = \left(\frac{K}{2f}\right)^{1/2}$$

We remember that

$$u_g = -(1/\rho f)(\partial p / \partial y)$$

and, expressing everything as a function of $v$, we have

$$-(1/\rho f)(\partial p / \partial y) = \left(r^2 + f^2\right) v / r$$

so that we have for the boundary condition

$$\psi|_{z=0} = -\gamma \frac{\left(r^2 + f^2\right) v}{fr}$$

This condition gives immediately the constant $C$ in Eq. (12.13) so that we have for the streamfunction

$$\psi = \psi_0 \sin(ly) \left[\sin(mz) + C \sinh \lambda (D - z)\right] \tag{12.14}$$

where

$$C = \frac{m}{r(2/Kf)^{1/2} \sinh \lambda D + \lambda \cosh \lambda D} \tag{12.15}$$

We can see now if this simple theory has some relation with reality. We can specify the domain to be $L = 3000$ km (i.e., about 30°) and $D = 10$ km and assume

for $f$ an average latitude such that $f \sim 2 \cdot 10^{-5}$. The coefficients $\alpha$ and $r$ are assumed of the order of 1/20 day (d) $\cong 5 \cdot 10^{-7}$ s$^{-1}$ and the calculated value for $\lambda$ results in $2.2 \cdot 10^{-2}$ m$^{-1}$ so that $\lambda D \cong 5$ and C $\cong 3.5 \cdot 10^{-3}$. The amplitude $\Psi_0$ can be obtained through the boundary conditions and assuming a vertical velocity upward from the boundary layer of a few mms$^{-1}$. With 1 mm s$^{-1}$, we obtain $\Psi_0 \cong 1.5 \cdot 10^4$ m$^2$s$^{-1}$, and with this value, we can proceed to the calculation of the zonal and meridional velocities. We have $v = m\ \Psi_0 \cong 4.5$ m s$^{-1}$ and $u = fv/r \cong 180$ m s$^{-1}$. These velocities are really too high and of course we could try to tune the parameters to get closer to the truth, but the game is not worth the effort. As a matter of fact, we see a number of problems with this solution. Zonal winds reach values up to 180 m s$^{-1}$ although they show the right sign: westerlies at altitude and easterlies near the surface. As shown by Eq. (12.3), the equilibrium in the zonal direction is achieved between the Coriolis force and friction. At the surface, the condition on the vertical velocity is such that $w$ is positive (upward) near the equator and negative on the opposite side, and this again is consistent with the zonal velocity gradient.

Although we have not obtained a realistic result, the example has been useful to work out a direct circulation, which is not exactly a convective circulation. We will be a little closer to reality in the next section.

### *12.1.2 The Nonlinear Symmetric Circulation*

We will now report a simple model for the axisymmetric circulation as published a few years ago by Adam Sobel and Tapio Schneider. The model calculates the circulation in a thin layer of thickness $\delta$ at the top of the troposphere assumed to be of constant height $H$. We start out by writing down the thermal wind equation

$$-f\frac{\partial u}{\partial z} = \frac{g}{T}\frac{\partial T}{\partial y}$$

Integrating from the surface to altitude $H$, we get

$$-fu = \frac{gH}{T_0}\frac{\partial T}{\partial y} \tag{12.16}$$

where $u$ is the velocity in the top layer and the temperature gradient is the average in the troposphere. Assuming the beta plane approximation, this equation can be written as

$$-\beta y u = \frac{gH}{T_0}\frac{\partial T}{\partial y} \tag{12.17}$$

We then write the momentum equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = fv - \frac{1}{\rho}\frac{\partial p}{\partial x}$$

The zonal average cancels the x-derivatives, and at the steady state, we have

$$v\left(\frac{\partial u}{\partial y} - f\right) = 0 \tag{12.18}$$

The last equation we need is the thermodynamic equation

$$\frac{D\theta}{Dt} = \frac{\theta_E - \theta}{\tau}$$

That is a "Newtonian cooling" approximation, with the potential temperature $\theta$ relaxing to the equilibrium value $\theta_E$ with characteristic time $\tau$. When the Lagrangian derivative is written explicitly, we get

$$\frac{\delta \Delta z}{H}\frac{\partial v}{\partial y} = \frac{\theta_E - \theta}{\tau} \tag{12.19}$$

In this formulation, we assume that heat is transported in a top layer of thickness $\delta$ in the upper troposphere. Here the temperature $T$ and the potential temperature $\theta$ are related by the usual formula:

$$T = \theta(p_t/p_s)^{R/C_p}$$

where $p_t$ and $p_s$ are the pressure of the tropopause and the surface, respectively, $\Delta_z$ is the potential temperature difference between surface and the top of the layer, and $T_0$ is the reference temperature at the surface.

Equation (12.18) can be integrated immediately:

$$u = \frac{1}{2}\beta y^2 + \text{cost} \tag{12.20}$$

It can be recognized that this is simply the conservation of angular momentum when we use $\beta = 2\Omega\cos\varphi/a \sim 2\Omega/a$ and y $\sim a$ sin $\varphi$. We get easily

$$u = \frac{1}{2}\frac{2\Omega}{a}(a\sin\varphi)^2 = \Omega a \sin^2\varphi \tag{12.21}$$

This is very similar to the requirement of angular momentum conservation:

$$\Omega a^2 = a\cos\varphi\,(\Omega a\cos\varphi + u)$$

from which the zonal wind can be obtained:

$$u = \Omega a \sin^2\varphi/\cos\varphi \tag{12.22}$$

When $\varphi$ is small, (12.22) coincides with (12.21).

We now assume that the equilibrium potential temperature is prescribed, and following the original formulation of Held and Hou (1980), we assume a dependence of the form

$$\frac{\theta_E(\varphi, z)}{\theta_0} = 1 - \frac{2}{3}\Delta_y P_2(\sin\phi) + \Delta_z\left(\frac{z}{H} - \frac{1}{2}\right) \tag{12.23}$$

The meaning of this expression is very simple. It can be easily shown that

$$P_2(\sin\varphi) = \left(3\sin^2\varphi - 1\right)/2$$

For $z = H/2$ and small angle approximation $y = a\varphi$, we get

$$\frac{\theta_E}{\theta_0} = 1 + \frac{2}{3}\Delta_y - 2\Delta_y\left(\frac{y}{a}\right)^2$$

And when vertically averaged,

$$\frac{\theta_E}{\theta_0} = 1 + \frac{1}{3}\Delta_y - \Delta_y\left(\frac{y}{a}\right)^2 \tag{12.24}$$

The difference in potential temperature between the equator ($y = 0$) and the pole ($y = a\pi/2$ is simply $\theta_0\Delta_y(\pi/2)^2$, while for fixed latitude, the difference between altitude $H$ and the surface is $\theta_0\Delta_z$. The solution (12.24) can also be written in the form

$$\theta_E = \theta_{E0} - \Delta\theta\left(\frac{y}{a}\right)^2 \quad \theta_{E0} = \theta_0\left(1 + \Delta_y/3\right) \quad \Delta\theta = \theta_0\Delta_y \tag{12.25}$$

We now assume that beyond some latitude $y_H$, there is no meridional circulation which corresponds to $v = 0$. For such latitude according to (12.19), $\theta = \theta_E$. Now we can integrate (12.17), and substituting for u the expression (12.20), we obtain

$$-\frac{1}{2}\beta^2 y^3 = \frac{gH}{T_0}\frac{\partial T}{\partial y} \quad \Rightarrow \frac{\partial T}{\partial y} = -\frac{T_0}{2gH}\beta^2 y^3 = -\frac{T_0}{2L_\rho^4}y^3$$

where $L_\rho = \left(\sqrt{gH}/\beta\right)^{1/2}$ is the equatorial Rossby radius. The integration over y gives

$$T = C' - \frac{T_0}{8L_\rho^4}y^4$$

Substituting for the potential temperature, we have

$$\theta = C - \left(\frac{p_s}{p_t}\right)^{R/C_p}\frac{T_0}{8L_\rho^4}y^4 = C - Dy^4$$

where

$$D = \left(\frac{p_s}{p_t}\right)^{R/C_p}\frac{T_0}{8L_\rho^4} \tag{12.26}$$

Notice that the definition for $L_\rho$ is a bit different from the previous $L_\rho = \sqrt{gH}/f$ because in this case the equator $f$ is zero.

The determination of $C$ is straightforward because for $y = y_H$, $\theta = \theta_E$. So we have

$$\theta = \theta_E(y_H) + D\left(y_H^4 - y^4\right) = \theta_{E0} - \Delta\theta\left(\frac{y_H}{a}\right)^2 + D\left(y_H^4 - y^4\right) \tag{12.27}$$

We can then use this expression with (12.19) and (12.24) to find the dependence of the meridional velocity on latitude:

$$v = \frac{H}{\delta\Delta_z\tau}\left[\frac{\Delta\theta}{a^2}\left(y_H^2 y - \frac{y^3}{3}\right) + D\left(\frac{y^5}{5} - y_H^4 y\right)\right] \tag{12.28}$$

The constant of integration has been set to zero because symmetry requires $v = 0$ at the equator. To determine the boundary of the Hadley cell, $y_H$, we require that $v = 0$ for $y = y_H$ so we get

$$y_H = \left(\frac{5\Delta_\theta}{6a^2 D}\right)^{1/2} \tag{12.29}$$

Equation (12.27) can be put in another form that can be handled more easily:

$$v = \frac{H}{\delta\Delta_z\tau}\frac{y_H^5 D}{5}\left[\left(\frac{y}{y_H}\right) - 2\left(\frac{y}{y_H}\right)^3 + \left(\frac{y}{y_H}\right)^5\right] \tag{12.30}$$

Beyond the latitude of the Hadley cell, we may assume that the zonal velocity corresponds to equilibrium temperature $\theta_E$ so that, from (12.17) and (12.24), we have

$$u_E = \frac{gH\Delta_y}{\Omega\theta_0 a} \tag{12.31}$$

Beyond $y_H$, because of (12.19), we assume $v = 0$ and $u = u_E$.

At this point, it is interesting to proceed with some arithmetic and calculate the values of the different quantities. From the paper of Sobel and Schneider, we get the following numbers:

$$\begin{array}{llll} \tau = 37\text{d}; & H = 16\,\text{km}; & \delta = 4\,\text{km}; & \mathrm{T}_0 = 300\text{K}; \\ \Delta_z = 60\,\text{K}; & \Delta_y = 100\,\text{K}; & \theta_0 = 330\text{K} & \end{array}$$

And we have then

$$L_R^4 \approx 4 \cdot 10^{26}\,\text{m}^4; \quad D = 1.3 \cdot 10^{-25}\,\text{K m}^{-4}; \quad y_H = 2690\text{km}$$

The latitude of the Hadley cell is roughly 24° in the "small angle" approximation. This is quite important in evaluating, for example, (2.31) because we can easily neglect terms with power larger than 1. We can then assume a linear behavior of the meridional velocity with attitude.

We can now evaluate the momentum transport $vu$ as preliminary to evaluate the zonal velocity at the surface. Using (12.20) and (12.27), we obtain

$$uv = \frac{5H^2\Delta_y^2 y_H^3 \beta}{72\Delta_z \tau y_1^4 D}\left[\left(\frac{y}{y_H}\right)^3 - 2\left(\frac{y}{y_H}\right)^5 + \left(\frac{y}{y_H}\right)^7\right] \tag{12.32}$$

This actually represents the meridional transport of zonal momentum. To obtain the zonal velocity at the surface, we must write the complete zonal momentum equation (see E.12 examples) that we rewrite for convenience:

$$\frac{Du}{Dt} - \frac{uv\tan\varphi}{a} - fv = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{\partial}{\partial z}\left(K\frac{\partial u}{\partial z}\right) \tag{12.33}$$

When averaged zonally, the pressure term disappears and the equation can be written in a different form:

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial y} - \frac{uv\tan\varphi}{a} - fv = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{\partial}{\partial z}\left(K\frac{\partial u}{\partial z}\right)$$

where again the zonal average has been applied. The transport of zonal momentum can be outlined by writing this equation in the form (with $\nabla\cdot\mathbf{u} = 0$)

$$\nabla\cdot(\mathbf{u}u) - \frac{uv\tan\varphi}{a} - fv - \frac{\partial}{\partial z}\left(K\frac{\partial u}{\partial z}\right) = 0 \tag{12.34}$$

At the surface ($z = 0$), the condition for the stress becomes $K\partial u/\partial z = Cu$ where $C$ is the appropriate constant. With the definition of angular momentum,

$$M = \Omega a^2\cos^2\varphi + ua\cos\varphi \tag{12.35}$$

Equation (12.32) can be written as (see E.12)

$$\nabla\cdot(\mathbf{u}M) = K\frac{\partial^2 M}{\partial z^2} \tag{12.36}$$

We have worked up to this moment with vertical integrated quantities so that the last equation one integrated vertically becomes (**U** is now the integrated velocity)

$$\nabla\cdot(\mathbf{U}M) = K\frac{\partial M}{\partial z} \tag{12.37}$$

Nevertheless Eq. (12.34) is useful to introduce what is known as Hide's theorem (from Raymond Hide), and in this, we will follow the argument given by Lindzen (1990). Suppose that $M$ has a local maximum in the fluid. Around this maximum, a contour can be found where $M$ is constant. Integration of (12.34) gives the left-hand side always zero (the integral can be split in one containing the divergence of $u$ and the other containing the gradient of $M$), while the right-hand side gives a negative contribution (the viscous flux is down gradient). This situation is not consistent and $M$ cannot have a maximum at the interior. With a similar argument, we can rule out the presence of a maximum at the surface. The only possibility is that the surface wind is easterly, and only in this case, the contribution of the right-hand term can be zero (see E.12). An upper bound for M is when at the equator the zonal wind is zero and in this case

$$M_{\max} < \Omega a^2$$

To obtain the zonal velocity at the surface, we still use the Eq. (12.35) in the small angle approximation which becomes

$$\frac{\partial}{\partial y}(vM) \approx \frac{\partial}{\partial y}(uva) = K\frac{\partial M}{\partial z} = K\frac{\partial ua}{\partial z} = -aCu\,(0)$$

where we have used the boundary condition at the surface and at the top, respectively (zero zonal velocity). We can then obtain the velocity at the surface by deriving Eq. (12.30), that is,

$$u(0) = -\frac{5}{24}\frac{H^2\Delta_y^2 y_H^2\beta}{\Delta_z\tau y_1^4 DC}\left[\left(\frac{y}{y_H}\right)^2 - \frac{10}{3}\left(\frac{y}{y_H}\right)^4 + \frac{7}{3}\left(\frac{y}{y_H}\right)^6\right] \tag{12.38}$$

This expression shows that the zonal velocity corresponds to easterlies for latitude below about 15° and to westerlies above that limit.

The last quantity to evaluate is the meridional heat flux. This can be accomplished by integrating Eq. (12.19) using (12.26) for the potential temperature. We get simply

$$\frac{\delta\Delta_z}{H}v = \frac{5}{36}\frac{H\Delta_y^2 y_H}{\tau y_1^4 D}\left[\left(\frac{y}{y_H}\right) - 2\left(\frac{y}{y_H}\right)^3 + \left(\frac{y}{y_H}\right)^5\right] \tag{12.39}$$

It is convenient when plotting Eqs. (12.38), (12.37), (12.31), (12.29), and (12.26) to normalize with an appropriate factor. The zonal velocity is normalized with the $u_E$ value given by (12.30). At the surface, the zonal velocity is normalized by the factor

$$5\,H^2\Delta_y^2 y_H^2\beta/\left(24\Delta_z\tau y_1^4 DC\right)$$

The momentum transport is normalized with the factor

$$5H^2\Delta_y^2 y_H^3\beta/\left(72\Delta_z\tau y_1^4 D\right)$$

and finally the heat transport by the factor

$$5H\Delta_y^2 y_H/\left(36\tau y_1^4 D\right)$$

The theory we have illustrated is quite simple and does not take into account any kind of viscosity, and this can be done much more easily if we look at the problem from the vorticity point of view.

### *12.1.3 The Vorticity Equation and Viscosity*

We can use the expression (12.36) together with the hypothesis that the vertical flux of angular momentum is simply proportional to the velocity:

$$K\partial M/\partial z = kaU\cos\varphi$$

where $k$ is related to viscosity. Equation (12.36) then becomes (Fig. 12.3)

$$\frac{V}{a^2\cos\varphi}\frac{\partial M}{\partial\varphi} = kU \tag{12.40}$$

where $V$ and $U$ are the vertical integrated velocities (meridional and zonal). It is easy to show that

$$\frac{1}{a^2\cos\varphi}\frac{\partial M}{\partial\varphi} = \frac{1}{a}\frac{\partial u}{\partial\varphi} - \frac{u\tan\varphi}{a} - f$$

If near the equator the Coriolis term is redefined as to include the centrifugal term, we obtain $f = f + u\tan\varphi/a$, and considering that all the quantities depend only on $\varphi$, we have for the relative vorticity $\zeta = -(1/a)(\partial u/\partial\varphi)$ so that (12.40) becomes

$$kU = -(\varsigma + f)V \tag{12.41}$$

This equation has very important consequences, and if we derive it with respect to y, we obtain a corresponding equation for the vorticity:

$$-k\zeta = \frac{\partial}{\partial y}[(\zeta + f)V] \tag{12.42}$$

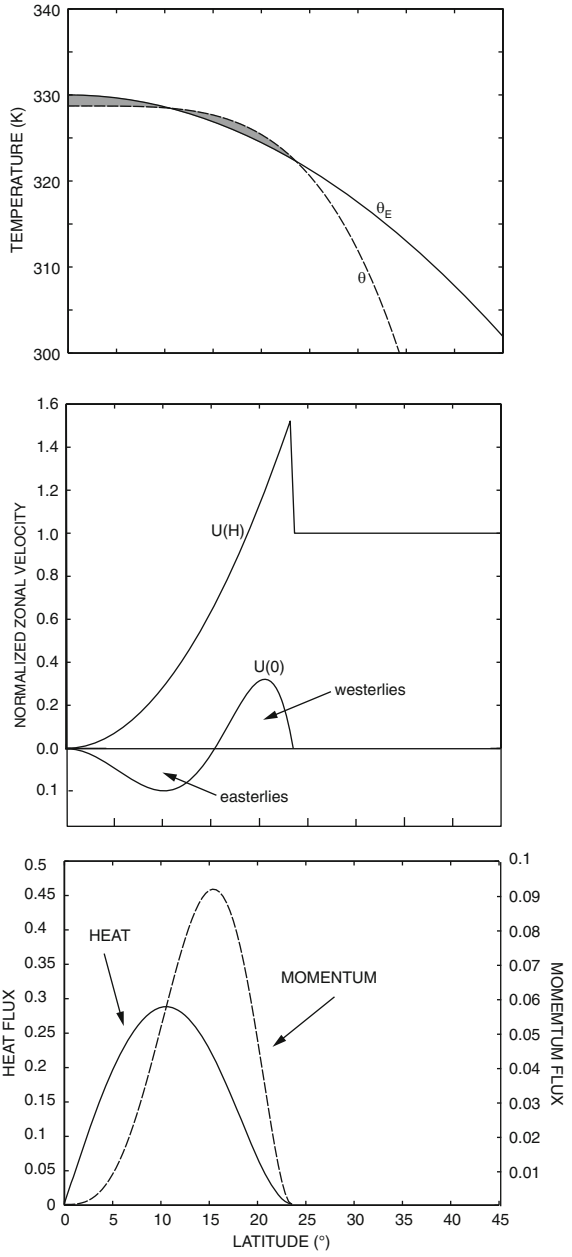



**Fig. 12.3** The latitudinal behavior of the temperature (*top*), the zonal wind (*middle*), and the heat and momentum flux (*bottom*). The *shaded area* in the *top panel* are equal and give a criteria to obtain the limit latitude for the Hadley cell. Notice the different scale (even if normalized) for the zonal wind at the surface (Sobel and Schneider 2009)

The same equation can be obtained following a paper by Polvani and Sobel (2002), starting from primitive equations. In this particular form, these equations are derived in the examples and can be written as

$$\begin{aligned}
&\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial y} - v = -\alpha u \\
&\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial y} + u = -\frac{\partial h}{\partial y} - \alpha v \\
&\frac{\partial h}{\partial t} + v\frac{\partial h}{\partial y} + h\frac{\partial v}{\partial y} = Q
\end{aligned} \tag{12.43}$$

Here h is the depth of the atmosphere, $\alpha$ is the internal friction, and $Q$ is the heating rate or thermal forcing that is parameterized according to

$$Q = \frac{\eta_E - \eta}{\tau} \tag{12.44}$$

where $\eta$ is a related to h by the relation $h = 1 + \eta$ and this is an indication of "temperature," while $\tau$ is a time constant. In the steady-state condition $\partial/\partial t = 0$, we derive the first of (12.42) with respect to y and the second with respect to x and then subtract to obtain

$$\frac{\partial \zeta}{\partial t} + \frac{\partial}{\partial y}\left[v\left(\zeta + 1\right)\right] = -\alpha\zeta$$

That is the same as (12.41). The system (12.42) can be solved numerically (see the examples) and the results are shown in Fig. 12.4. The main effect of the friction introduction is to smooth out the discontinuity in the different parameters, and consequently, the latitudinal delimitation of the Hadley cell is not so well defined. Again this is an exercise which has the function to have further insight in the physical aspect of the problem.

## 12.2 The Middle Latitude Circulation

The Hadley circulation, as we have shown, is unable to transport heat beyond a certain latitude. The zonal currents created by this direct cell however are unstable with respect to small perturbations. This characteristic allows, through a completely different mechanism, the heat to be transported up to the middle and high latitudes. In this section, we will illustrate the mechanism at first qualitatively and then in a more accurate and quantitative way. The dynamic instability (called *baroclinic*) resulting from the amplification of small perturbations of the zonal current is not easy to visualize because it is intrinsically three dimensional. But again, as always

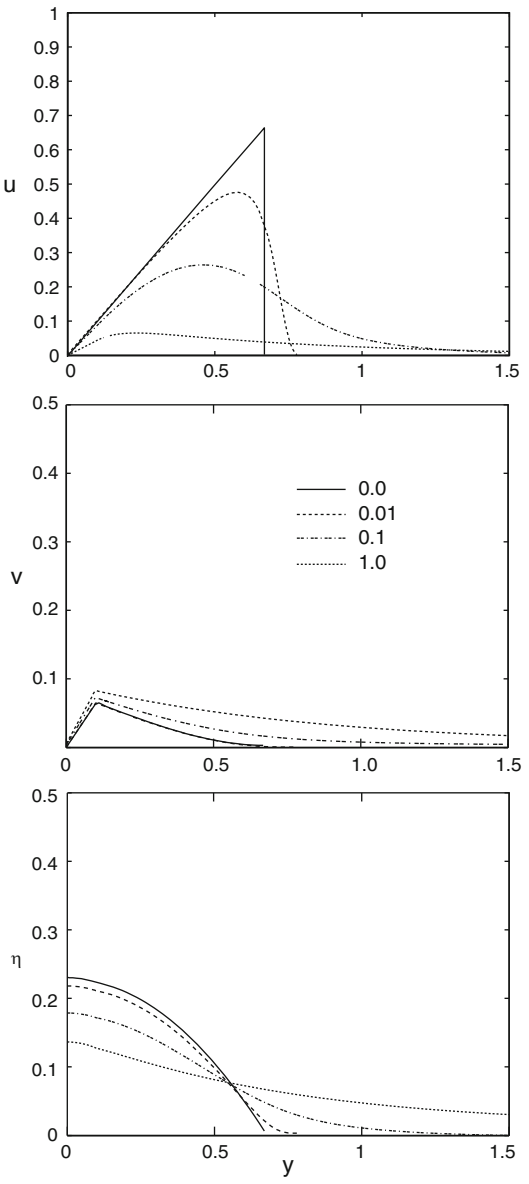



**Fig. 12.4** The numerical solution for the system (12.42). The *solid curve* shows the solution for the inviscid case for the zonal velocity (*top*), the meridional velocity (*middle*), and the parameter η. The *dotted lines* show the same solutions for different values of friction α (Polvani and Sobel 2002)

in physics, the mathematics does not give any particular insight if first we do not try to understand the physical process. We will follow once more the beautiful textbook by Wallace and Hobbs and then we will go to something that looks more original and that will already give quantitative indications.

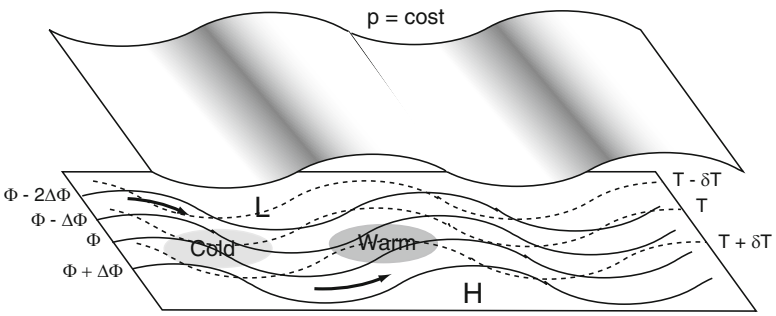



**Fig. 12.5** The geopotential lines (*solid*) and the isotherms (*dashed*) in a growing baroclinic wave on a constant pressure surface (shown above). L and H indicates low and high pressure

### *12.2.1 The Baroclinic Instability: Qualitative Treatment*

We consider a westerly current that is slightly perturbed as shown in Fig. 12.5. Such a current implies that the geopotential decreases slowly northward. For sake of simplicity, we assume that the phase velocity of the wave is the same as the wind in such a way that an observer fixed with the wave sees only the meridional movements. As shown in the figure, the wind taken parallel to the geopotential brings cold air to low latitudes and warm air to high latitudes (area *C* and *W*). Area *A* will tend to cool and area *W* to warm. The net result is that the isotherms will assume a wave character similar to the geopotential although the temperature wave will be shifted by one fourth of the wavelength. This is because the meridional velocity will be maximum at the points of inflection of the geopotential wave. The condition for the wave to grow further is not only the increase in the temperature perturbation but also the increase in the kinetic energy. This can be obtained only if the cold air that moves from north to south sinks during its meridional motion. The potential energy lost in this way is replenished if the air that moves northward (half a wave downwind) rises at the same time.

This mechanism for the amplification of kinetic energy can be understood a little better if we refer to Fig. 12.6 which shows the trajectories of an air parcel on the meridional plane and compares their slopes with that of the isentropes. If we indicate with $\alpha = dz/dy$ the slope of the isentropes with respect to the horizontal plane and noting that $d\theta/dy < 0$ and $d\theta/dz > 0$ with $\theta$ potential temperature,

$$\alpha = -(d\theta/dy) / (d\theta/dz) \tag{12.45}$$

On the other hand, if we denote with $\gamma$ the slope of the trajectories and with $w'$ and $v'$ the vertical and meridional components of the velocity of the air parcel, we should have

$$w'/v' = \tan\gamma \tag{12.46}$$

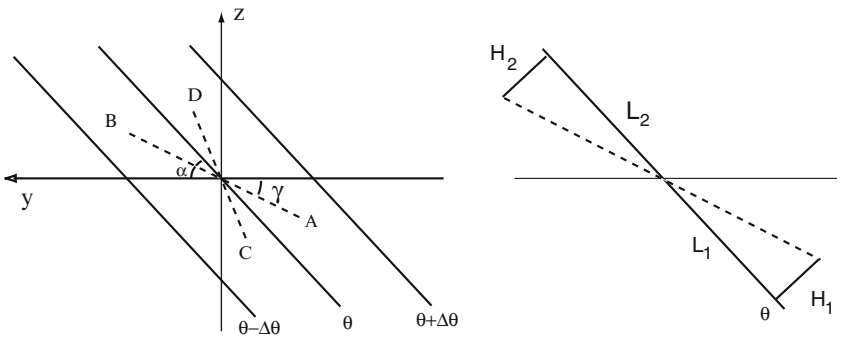



**Fig. 12.6** Possible trajectories of an air parcel with respect to the isentropes on the meridional plane. The slope of the trajectories is γ, while that of the isentropes is α

It is quite evident then that if $\theta'$ indicates the deviation from the mean of the potential temperature, we could have the following cases:

$$\begin{array}{llll} \gamma > \alpha & w' > 0 & \theta' < 0 & w'\theta' < 0 \\ \gamma < \alpha & w' > 0 & \theta' > 0 & w'\theta' > 0 \end{array} \tag{12.47}$$

The first of these conditions refers, for example, to a mass that moves from *C* to *D* and, being adiabatic, the motion conserves the potential temperature. The second condition corresponds to the case in which the parcel moves from *A* to *B*. Similar conclusions can be reached for parcels moving southward. From these considerations, we understand that the amplification is possible only in the second case, when the slope of the trajectory is smaller than the slope of the isentropes (Fig. 12.6).

We can now find which angles maximize the instability. Again with reference to Fig. 12.6, the deviation in the potential temperature can be expressed as

$$\theta' = \theta_1 - \theta_2 = \theta + H_1 \partial\theta/\partial n - (\theta - H_2 \partial\theta/\partial n) \tag{12.48}$$

where with $n$ we have indicated the direction normal to the isentrope $\theta$. Because the angles $\alpha$ and $\gamma$ are small, we can write

$$\partial\theta/\partial n \cong \partial\theta/\partial z$$

so that

$$\theta' \cong (H_1 + H_2)\, \partial\theta/\partial z = (L_1 + L_2) \tan(\alpha - \gamma)\, \partial\theta/\partial z \tag{12.49}$$

and the heat transport along the vertical

$$w'\theta' = \gamma v' (L_1 + L_2)(\alpha - \gamma)\, \partial\theta/\partial z \tag{12.50}$$

which is maximum when $\gamma = \alpha/2$. At this point, we can use Eq. (12.48) to relate the angles with the average values of the variables and establish under what conditions the instability may arise. Based on the thermal wind equation, we write

$$f\frac{\partial u}{\partial z} = -\frac{R}{H}\frac{\partial T}{\partial y} \cong -\frac{R}{H}\left(\frac{T}{\theta}\frac{\partial \theta}{\partial y} - \frac{RT}{C_P}\frac{\partial p}{\partial y}\right) \cong -\frac{g}{\theta}\frac{\partial \theta}{\partial y}$$

And with $H$ the scale height, we have

$$f\frac{u}{H} \cong -\frac{g}{\theta}\frac{\partial \theta}{\partial y} \Rightarrow u \approx -\frac{gH}{f\theta}\frac{\partial \theta}{\partial y} \tag{12.51}$$

This rough evaluation for $u$ is necessary to express the condition for instability:

$$w'/v' \approx w'/u' = \gamma < \alpha \tag{12.52}$$

At this point, we need to express the vertical velocity as a function of $u$. This can be done by recurring to a scale analysis of the vorticity equation

$$\frac{D}{Dt}(\zeta + f) = -(\zeta + f)\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = (\zeta + f)\frac{\partial w'}{\partial z}$$

Thus we have immediately

$$\frac{D}{Dt}(\zeta + f) = O\left(\frac{U}{L}\left(\frac{U}{L}\right)\right); \qquad f\left(\frac{\zeta}{f} + 1\right)\frac{\partial w'}{\partial z} \approx f\frac{\partial w'}{\partial z}$$

because

$$\zeta/f = O(\zeta + f) = O << 1$$

so that we can write

$$\frac{\partial w}{\partial z} = O\left(\frac{1}{f}\frac{U^2}{L^2}\right) = O\left(R_0\frac{U}{L}\right) \Rightarrow w = O\left(\frac{R_0 H u}{L}\right) \tag{12.53}$$

By combining Eqs. 12.50, 12.51, and 12.52, we obtain the condition for instability:

$$\frac{H}{fL^2}\left(-\frac{gH}{f\theta}\frac{\partial \theta}{\partial y}\right) < -\frac{d\theta/dy}{d\theta/dz}$$

from which we have the typical horizontal length for the instability to occur:

$$L_R = \frac{H}{f}\left(-\frac{g}{\theta}\frac{\partial \theta}{\partial z}\right)^{1/2} = \frac{H}{f}N = \frac{H}{f}\left(\frac{g}{H}\right)^{1/2} = \frac{\sqrt{gH}}{f} \tag{12.54}$$

with $N$ as the Brunt–Väisälä frequency. We already know this number as the *Rossby deformation radius* which with the typical values for $N$, $f$, and $H$ are of the order of 3000 km. The baroclinic instability is then possible only for wavelengths greater than $L_R$. At this point, intuition and order of magnitude calculation cannot give us much more, so we have to go to a more analytical approach if we want to find more precise numbers.

### 12.2.2 The Baroclinic Instability: The Eady Problem

The most important developments on the theory of general circulation at middle latitudes were reached around the time of the Second World War. This was probably stimulated by the need to provide the allied bombers with more precise weather forecast in their raids over Germany. In the years following the war, between 1945 and 1950, this progress was made firm with the formal contributions of J. Charney in 1947 and E.T. Eady in 1949.

Eady solved the problem by considering the vorticity equation in the $z^*$ coordinate system introduced in Chap. 10. In his approach, the Coriolis parameter was considered constant with latitude ($\beta = 0$). The quasi-geostrophic potential vorticity equation is then written as

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\left(\nabla^2\psi + \frac{f_0^2}{N^2}\frac{\partial^2\psi}{\partial z^{*2}}\right) = 0 \tag{12.55}$$

and the thermodynamic equation can be written in a more general form as

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\theta' + v'\frac{\partial\overline{\theta}}{\partial y} + w'\frac{\partial\overline{\theta}}{\partial z^*} = 0 \tag{12.56}$$

In (E.12.2), we show that this can be written as

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\partial\psi}{\partial z^*} - \frac{\partial\psi}{\partial x}\frac{\partial\overline{u}}{\partial z^*} + w'\frac{N_*^2}{f} = 0 \tag{12.57}$$

In order to solve Eq. (12.55), we consider the following conditions, which are the same that were used by Eady. We assume the atmosphere is confined between two surfaces at $z^* = \pm H$ and a constant vertical shear for the zonal velocity such that

$$\overline{u} = z^*\partial\overline{u}/\partial z^*$$

On the two rigid lids that limit the atmosphere, we assume $w' = 0$ so that Eq. (12.70) becomes

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\partial\psi}{\partial z^*} - \frac{\partial\psi}{\partial x}\frac{\partial\overline{u}}{\partial z^*} = 0$$

We look now for a solution of the form

$$\psi\left(x, y, z^*, t\right) = \Psi\left(z^*\right) \cos ly \exp\left[ik\left(x - ct\right)\right]$$

that can be substituted in Eq. (12.54) to give

$$d^2\Psi/dz^{*^2} - \alpha^2\Psi = 0 \tag{12.58}$$

where

$$\alpha^2 = N^2\left(k^2 + l^2\right)/f^2$$

We notice that we have omitted the asterisk from the Brunt–Väisälä frequency. We can obtain another equation for $\Psi$ using Eq. (12.56) written for the boundaries at $z^* = \pm H$:

$$\left(\overline{u} - c\right) d\Psi/dz^* - \left(d\overline{u}/dz^*\right)\Psi = 0 \tag{12.59}$$

The solutions for (12.57) must satisfy Eq. (12.58) at the boundaries so that the solution for $\Psi$ has the form

$$\Psi\left(z^*\right) = A \sinh \alpha z^* + B \cosh \alpha z^*$$

We will have two equations in the coefficients $A$ and $B$

$$\begin{aligned} cA - d\overline{u}/dz^* \left(\tanh\left(\alpha H\right) - 1/\alpha H\right) B = 0 \\ cB - d\overline{u}/dz^* \left(\coth\left(\alpha H\right) - 1/\alpha H\right) A = 0 \end{aligned} \tag{12.60}$$

First of all, we could find the phase velocity:

$$c = \pm d\overline{u}/dz^* H[\left(\tanh\left(\alpha H\right) - 1/\alpha H\right)\left(\coth\left(\alpha H\right) - 1/\alpha H\right)]^{1/2} \tag{12.61}$$

We find that the phase velocity has an imaginary component $c_\mathrm{i}$ that determines the growth of the wave amplitude with time with a time constant $(kc_\mathrm{i})^{-1}$. Equation (12.60) shows that we will have instability when

$$\tanh\left(\alpha H\right) < 1/\alpha H,$$

that is, when $\alpha H < 1.1997$. Figure 12.7 shows the growth rate as a function of $\alpha H$, and we notice that the maximum (the minimum time constant) happens for $\alpha H = 0.8031$ so that

$$kc_i = 0.3098\left(f/N\right) d\overline{u}/dz^*$$

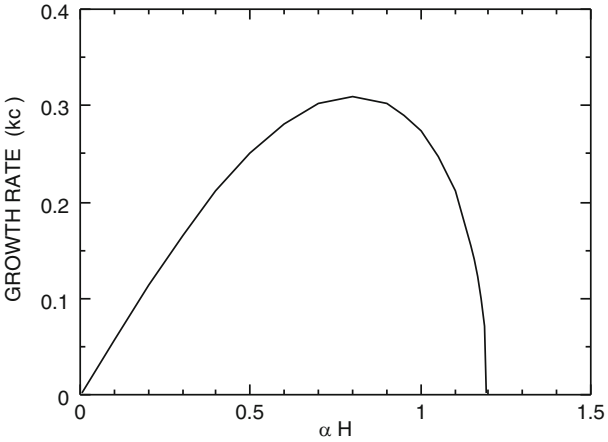



**Fig. 12.7** The growth rate of a baroclinic wave in the Eady problem as a function of a quantity proportional to the wave number

The conditions expressed for realistic data give for $l = 0$, $kNH/f < 1.1997$ which correspond to a length of 5200 km. We can also find the ratio *A/B* from Eq. (12.60):

$$\frac{A}{B} = \left( \frac{1/\alpha H - \tanh \alpha H}{\coth \alpha H - 1/\alpha H} \right)^{1/2} \tag{12.62}$$

For the value $\alpha H$ that maximizes the growth rate, this ratio is 1. 502.

Based on this, we can now calculate the geopotential, the temperature, and the vertical velocity which are useful to understand the instability mechanism. The geopotential can be easily calculated because it is proportional to the solution of Eq. (12.57) so that substituting for the values of $A$ and $B$ and taking the real part

$$\Phi' = \cos kx \frac{\sinh(\alpha z^*)}{\sinh \alpha H} + \sin kx \frac{\cosh(\alpha z^*)}{\cosh \alpha H} \tag{12.63}$$

This is the part of the solution that depends on altitude and longitude and it is only the perturbed part. To obtain the total for the geopotential at each altitude, we need to add the part proportional to the zonal current $-\overline{uy}$. The temperature is obtained through (12.55):

$$T = \frac{H\alpha}{R} \left( \cos kx \frac{\cosh(\alpha z^*)}{\sinh \alpha H} + \sin kx \frac{\sinh(\alpha z^*)}{\cosh \alpha H} \right) \tag{12.64}$$

The vertical velocity is more complicated to obtain, although we should use Eq. (12.66). In practice, after substituting $\Phi'$ and obtaining $w'$ to solve for the streamfunction, we integrate with respect to $x$:

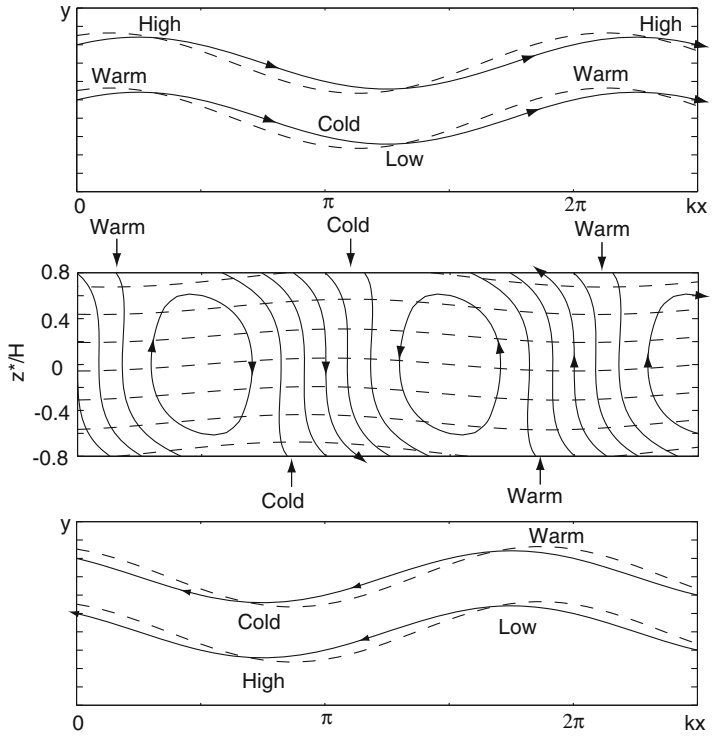



**Fig. 12.8** The *upper* and *lower parts* of the figure show the geopotential (*solid line*) and isotherms (*dashed lines*) at the upper and lower boundary, respectively. The figure part in the *middle* shows the streamfunction in the vertical plane (Calculation and figure based on Gill 1982)

$$\psi' = fN^{-2}k\left(d\bar{u}/dz^*\right)\left[\left(C' - C\right)\sin kx + \left(D - D'z\right)\cos kx\right]$$
$$C = \sinh\left(\alpha z^*\right)/\sinh\alpha H; \qquad D = \cosh\left(\alpha z^*\right)/\cosh\alpha H$$
$$C' = \alpha\cosh\left(\alpha z^*\right)/\sinh\alpha H; \qquad D' = \alpha\sinh\left(\alpha z^*\right)/\cosh\alpha H \tag{12.65}$$

In Fig. 12.8, some of these quantities are shown to emphasize the phase relationship. The upper part of the figure shows the behavior with latitude and longitude of the geopotential and the temperature at the upper boundary. The same quantities for the lower boundary are shown in the bottom part of the figure. Notice the phase shift of $\pi/2$ so that the lines connecting the high and the low are tilted westward about 48°. The temperature has a phase shift of about 48° eastward between the upper and lower level. The figure part in the middle shows the streamfunction on the $x$-$z$ plane, and from it, we see that warm air rises east of the low as the surface slopes slightly toward west to subside finally east of the high pressure.

In the same figure, the isentropes are shown (dashed lines) which give the warm and cold regions. Notice that the potential temperature increases with height. This figure shows the fastest-growing mode. A careful study of the sign of the meridional

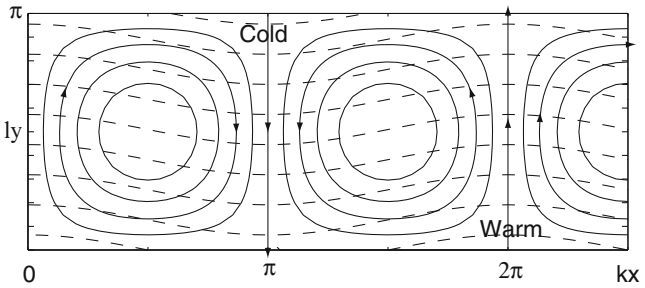



**Fig. 12.9** Geopotential lines (*solid*) and temperature (*dashed*) for an Eady wave at the steering level. Notice the two currents, one from north to south and downward and the other from south to north and upward. The net result is a northward transfer of heat and transformation of potential in kinetic energy (Redrawn from Gill 1982)

velocity (obtained differentiating the geopotential with respect to $x$) gives an idea of the direction of the current which is mainly easterly. The most interesting results on the $x$-$z$ plane are a number of "convective cells" sloping slightly to the west that bring warm air from the lower to the upper level. This scheme is then very similar to that illustrated qualitatively at the beginning of this section. Another interesting aspect is the advection at the *steering level* ($z^* = 0$). This can be seen from Fig. 12.9 where we show the geopotential together with the temperature lines. These actually are the intersections of the isentropes with the horizontal surface. Currents that move mainly in the north–south direction at the same time move upward and downward. The result is the net transfer of heat to the north according to what we have seen in the latest section. In particular, we can establish that the heat flux is proportional to the quantity:

$$\overline{v\theta} = \frac{1}{f}\overline{\frac{\partial \Phi\prime}{\partial x}\frac{\partial \Phi\prime}{\partial z^*}}$$

which can be shown to be always a positive quantity by using Eqs. (12.66) and (12.67).

The previous analysis, although it has some limitations, is useful to understand the structure of a growing baroclinic wave. The main limitation of the Eady solution is that it neglects that the Coriolis parameter is a function of latitude. These limitations are removed in the Charney theory.

### *12.2.3 The Baroclinic Instability: The Charney Problem*

We can have an idea of the $\beta$ effect looking first of all at the vertical velocity which can be obtained from Eq. (12.62). It can be shown that the vertical velocity decreases with increasing wavelength (i.e., with decreasing $k$). This explains the existence of a minimum wavelength for the instability because for $k \to 0$ then $\lambda \to \infty$ and the

vertical velocity goes to zero, that is, the motion is horizontal. We can make a scale analysis of the vorticity equation

$$D\zeta/Dt + v\beta \approx fdw/dz \tag{12.66}$$

that becomes

$$O\left(U^2/L^2\right) + \beta O(U) \approx fO\left(w/H\right)$$

As a consequence of $L \to \infty$, the advection of planetary vorticity balances the stretching of the vortex tubes. The fact that the motion is quasi-horizontal means that the term on the right-hand side can be assumed to be zero so that using $\zeta = \nabla^2\psi$ we have as a first approximation

$$(\overline{u} - c)\,\alpha^2 = \beta$$

where $\alpha^2 = k^2 + l^2$. For stationary waves ($c = 0$), we have for the zonal velocity

$$\overline{u} = \beta/\alpha^2 \approx Hdu/dz$$

so we get

$$du/dz \approx \beta L^2/H \tag{12.67}$$

This quantity is shown in Fig. 12.10 together with the other limit conditions obtained for short wavelengths. It should be noted that the maximum instability in

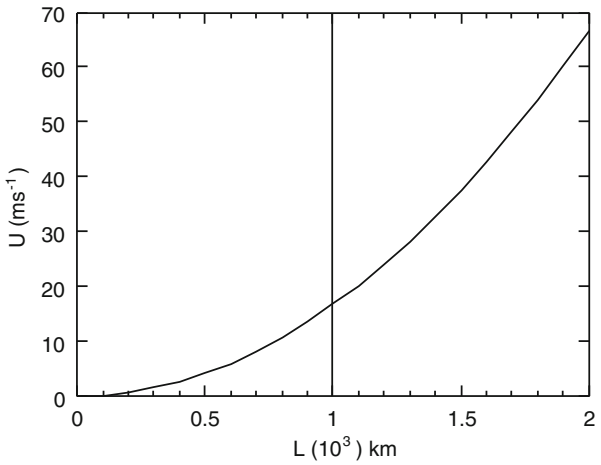



**Fig. 12.10** The relation between the vertical shear and the most unstable wavelength for the baroclinic instability. The *vertical line* refers to the wavelength given by Eq. (12.41), while the curve is the effect of a variable Coriolis parameter

this case is obtained for a wavelength reduced by a quantity $2^{1/2}$. The figure is drawn for $\beta = 1.67 \cdot 10^{-11}$ m$^{-1}$ s$^{-1}$ and an atmospheric thickness of 10 km.

What is known as the *Charney effect* includes the $\beta$ effect, and in this case, it is most appropriate to use the quasi-geostrophic potential vorticity of Eq. (9.58) in which we neglect the horizontal shear. We obtain

$$\overline{q}_y = \beta + \frac{\varepsilon}{H}\frac{\partial u}{\partial z}$$

with $\varepsilon = f^2/N^2$. In case the solution is of the same kind as in Eq. (12.65), we have the vertical structure equation

$$(\overline{u} - c)\left[(\Psi_{zz} - \Psi_z/H) - \alpha^2\Psi\right] + \overline{q}_y\Psi = 0 \tag{12.68}$$

This equation cannot be solved in a simple way so that in a sense the most straightforward way to arrive at a solution is through a numerical method. The effect of $\beta$ gives us the possibility to consider only long waves and in particular Rossby waves for which it is not necessary to consider a lid, as in the case of the Eady problem. After Charney solved the problem analytically, Roger Phillips worked out a solution using a two-level model that gives a lot of physical insight into the problem.

### *12.2.4 The Baroclinic Instability: Two-Level Model*

This model is based on the vorticity and thermodynamic equations written in pressure coordinates. The first should be manipulated a bit to make it easier to use. The first step is to divide the wind vector into geostrophic and ageostrophic components. The relative vorticity is due only to the geostrophic component, while the divergence is due to the ageostrophic component so that we have

$$\zeta \approx \zeta_g; \qquad \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = \left(\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y}\right) = -\frac{\partial w}{\partial p} \tag{12.69}$$

Also we may use Eqs. (7.28) and (7.29) and neglect the relative vorticity with respect to planetary vorticity and the equation becomes

$$\frac{\partial \zeta_g}{\partial t} + \mathbf{V}_g \cdot \nabla\left(\zeta_g + f\right) = f_0\frac{\partial \omega}{\partial p} \tag{12.70}$$

where we have set $f = f_0 + \beta y$. This is also called (we have already seen this) quasi-geostrophic vorticity equation because it considers geostrophy only in the advection term. The thermodynamic equation is written in the well-known form

$$\frac{\partial T}{\partial t} + \mathbf{V}_g \cdot \nabla T = \left(\frac{\sigma p}{R}\right)\omega \tag{12.71}$$

where

$$\sigma = -RT_0 p^{-1} d \ln \theta_0 / dp$$

is the static stability as shown in Chap. 5.

Equations (12.70) and (12.71) are the basis of our model, and they can be connected if we introduce a streamfunction such that $\psi = \Phi / f_0$ so that

$$\mathbf{V}_g = \mathbf{k} \times \nabla \psi \qquad \zeta_g = \nabla^2 \psi;$$

And from the definition of geopotential,

$$T = -\frac{f_0}{R} p \frac{\partial \psi}{\partial p} \tag{12.72}$$

And considering all the derivatives are calculated at constant pressure, we have

$$\frac{\partial \nabla^2 \psi}{\partial t} + \mathbf{V}_g \cdot \nabla \left( \nabla^2 \psi + f \right) = f_0 \frac{\partial \omega}{\partial p} \tag{12.72a}$$

$$\frac{\partial}{\partial t} \left( \frac{\partial \psi}{\partial p} \right) + \mathbf{V}_g \cdot \nabla \left( \frac{\partial \psi}{\partial p} \right) = - \left( \frac{\sigma}{f_0} \right) \omega \tag{12.72b}$$

These two equations are coupled through $\omega$ and can be solved using a model that assumes the atmosphere to be a two-layer channel. The Coriolis parameter in the channel changes linearly. The model is sketched in Fig. 12.11.

The vorticity equation is solved at levels 1 and 3, while the thermodynamic equation is solved only at the separation plane 2. The prediction of the vertical velocity is made only at this level, while rigid lids are assumed at levels 0 and at the surface, meaning $\omega = 0$.

We again assume that the streamfunction describes a basic state (a zonal current $U$ function of pressure) perturbed by a time-dependent term:

$$\psi = -U(p)y + \psi'(x, p, t) \tag{12.73}$$

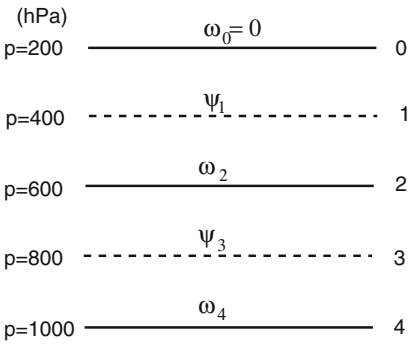



**Fig. 12.11** The general layout of the pressure levels and the variables in the two-layer model

Equations (12.72a) and (12.72) become

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial x}\right)\frac{\partial^2\psi}{\partial x^2}+\beta\frac{\partial\psi}{\partial x}-f_0\frac{\partial w'}{\partial p}= \tag{12.74}$$

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial x}\right)\frac{\partial\psi}{\partial p}-\frac{\partial U}{\partial p}\frac{\partial\psi}{\partial x}+\left(\frac{\sigma}{f_0}\right)\omega'=0 \tag{12.75}$$

At this point, these equations are converted to finite difference equations so that we can write, referring to Fig. 12.11,

$$\left.\frac{\partial\omega'}{\partial p}\right|_1\approx\frac{\omega_2'-\omega_0'}{p_2-p_0}\approx\frac{\omega_2'}{\Delta p};\qquad \left.\frac{\partial\omega'}{\partial p}\right|_3\approx\frac{\omega_4'-\omega_2'}{p_4-p_2}\approx-\frac{\omega_2'}{\Delta p};$$

$$\left.\frac{\partial\psi}{\partial p}\right|_2\approx\frac{\psi_3-\psi_1}{\Delta p};\quad \left.\frac{\partial U}{\partial p}\right|_2\approx\frac{U_3-U_1}{\Delta p};\quad \psi_2=\frac{\psi_3+\psi_1}{2};\quad U_2=\frac{U_3+U_1}{2}$$

and the finite difference equations corresponding to Eq. (12.73) become

$$\begin{aligned}\left(\frac{\partial}{\partial t}+U_1\frac{\partial}{\partial x}\right)\frac{\partial^2\psi_1'}{\partial x^2}+\beta\frac{\partial\psi_1'}{\partial x}-f_0\frac{\omega_2'}{\Delta p}=0\\ \left(\frac{\partial}{\partial t}+U_3\frac{\partial}{\partial x}\right)\frac{\partial^2\psi_3'}{\partial x^2}+\beta\frac{\partial\psi_3'}{\partial x}+f_0\frac{\omega_2'}{\Delta p}=0\end{aligned} \tag{12.76}$$

while the thermodynamic equation is

$$\left(\frac{\partial}{\partial t}+U_m\frac{\partial}{\partial x}\right)\left(\psi_1'-\psi_3'\right)-U_T\left(\frac{\partial\psi_1'}{\partial x}+\frac{\partial\psi_3'}{\partial x}\right)-\frac{\Delta p\sigma}{f_0}\omega_2'=0 \tag{12.77}$$

$$U_m=\left(U_1+U_3\right)/2;\quad U_T=\left(U_1-U_3\right)/2$$

The solution of this system follows a standard procedure that consists in eliminating $\omega'_2$ between the two equations and obtaining two new equations in the variables $\psi_1'+\psi_3'=\psi_m';\quad \psi_1'-\psi_3'=\psi_T'$

The first of these quantities is proportional to the temperature at level 2, while the second is proportional to the geopotential (Fig. 12.12). Through a number of manipulations, it is possible to arrive at two different equations for the above quantities:

$$\left(\frac{\partial}{\partial t}+U_{m1}\frac{\partial}{\partial x}\right)\frac{\partial^2\psi_m}{\partial x^2}+\beta\frac{\partial\psi_m}{\partial x}+U_T\frac{\partial}{\partial x}\left(\frac{\partial^2\psi_T}{\partial x^2}\right)=0 \tag{12.78}$$

$$\left(\frac{\partial}{\partial t}+U_m\frac{\partial}{\partial x}\right)\left(\frac{\partial^2\psi_T}{\partial x^2}-2\lambda^2\psi_T\right)+\beta\frac{\partial\psi_T}{\partial x}+U_T\frac{\partial}{\partial x}\left(\frac{\partial^2\psi_m}{\partial x^2}+2\lambda^2\psi_m\right)=0 \tag{12.79}$$

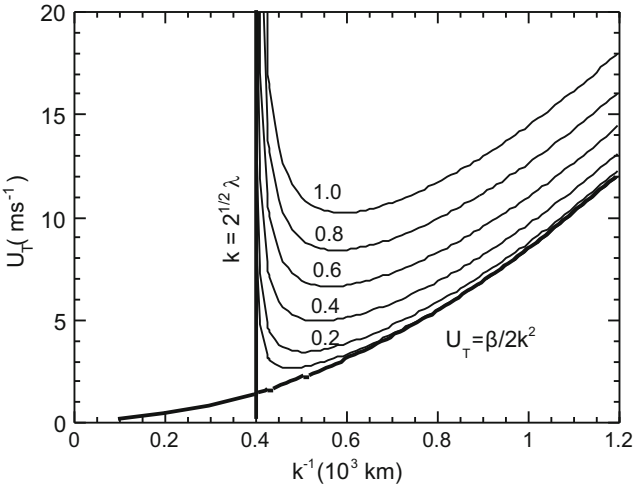



**Fig. 12.12** The wind shear as a function of the reciprocal of the wavelength. The vertical asymptote corresponds to the minimum wavelength for instability. The curves are given for different growth rates from $n = 1$ to 0. Units are $10^{-5}$ s$^{-1}$

where

$$\lambda^2 = f^2/\left(\sigma\Delta p^2\right)$$

We can now assume the two functions to be of the form

$$\psi_m = Ae^{ik(x-ct)}; \qquad \psi_T = Be^{ik(x-ct)} \tag{12.80}$$

Substituting these expressions in Eqs. (12.78) and (12.79), we have an algebraic linear system in the two unknowns $A$ and $B$:

$$\begin{aligned} U_T\left(k^2-2\lambda^2\right)A - \left[(c-U_m)\left(k^2+2\lambda^2\right)+\beta\right]B &= 0 \\ \left[(c-U_m)\,k^2+\beta\right]A - \qquad\qquad k^2U_TB &= 0 \end{aligned} \tag{12.81}$$

that has solutions when the determinant of the coefficient is zero. This condition gives a quadratic equation in the phase velocity $c$ that has the solution

$$c = U_m - \frac{\beta\left(k^2+\lambda^2\right)}{k^2\left(k^2+2\lambda^2\right)} \pm \left[\frac{\beta^2\lambda^4 - U_T^2k^4\left(4\lambda^4-k^4\right)}{k^4(k^2+2\lambda^2)^2}\right]^{1/2} \tag{12.82}$$

The phase velocity has an imaginary part, that is, the quantity between parentheses is negative. When this happens, the solution to Eq. (12.80) is unstable, and the neutral condition (i.e., the quantity is zero) is given by

$$U_T^2 = \beta^2 \lambda^4 / k^4 \left(4\lambda^4 - k^4\right)$$

This equation has an asymptote that gives the lower limit for the wavelength of instability:

$$L_c = \sqrt{2}\pi/\lambda = \pi \Delta p \sqrt{2\sigma}/f_0 \tag{12.83}$$

At this point, we can define a growth rate for the wave given as a product of the wave number and the imaginary part of the phase velocity, $n = kc_i$. The imaginary part is given by the square root of the quantity in parentheses in Eq. (12.67) which is the inverse of the time constant for the wave growth. This time is shown in Fig. 12.14 as a function of the shear and the reciprocal of the wave number. This figure refers to $\beta = 1.67 \cdot 10^{-11}$ m$^{-1}$ s$^{-1}$ and $\lambda = 1.76 \cdot 10^{-6}$ m$^{-1}$.

We notice that the neutrality curves (the vertical asymptote and the curve) are similar to those shown in Fig. 12.10, obtained on the basis of very simple considerations. These curves delimitate the stability from the instability area.

It is rather instructive to examine at this point these results in terms of the air parcel's motion on the vertical plane ($p$–$x$). The first step is to show that the streamfunctions at the two levels also give information on the middle-level temperature. Relation between geopotential and streamfunction gives $fd\psi = d\Phi = -RTdp/p$

Integrating between levels 1 and 3 and assuming that the temperature at level 2 is the average between levels 1 and 3, we obtain

$$T_2' = f\left(\psi_1' - \psi_3'\right) / \left[R \ln\left(p_3/p_1\right)\right] \tag{12.84}$$

In this way, Eq. (12.81) has a very simple meaning because

$$\left(\frac{\partial}{\partial t} + U_m \frac{\partial}{\partial x}\right) T_2' \left[R \ln\left(p_3/p_1\right)/f\right] - 2U_T v_2' - \left(\frac{\Delta p \sigma}{f_0}\right) \omega_2' = 0 \tag{12.85}$$

If we zonally average and use stationary conditions, we get an omega equation reduced to the bone: this is a diagnostic equation for the vertical velocity

$$\omega_2' \propto -U_T v_2' = v_2' \partial \overline{T} / \partial y \tag{12.86}$$

where we have used the relation between the thermal wind and the meridional temperature gradient $U_T \propto (\partial T/\partial y)$. Equation (12.86) tells us that the motion is downward ($\omega'_2 > 0$) when we have cold advection ($v'_2 \partial T/\partial y > 0$) and vice versa. If we refer to Fig. 12.13, this means having downward motion in correspondence to the negative temperature perturbation and upward motion for the warm perturbation. Also the maximum of the southward component of the velocity corresponds to the inflection point of the streamfunction at the upper level. The opposite can be said for the $\psi_3$ function at the lower level.

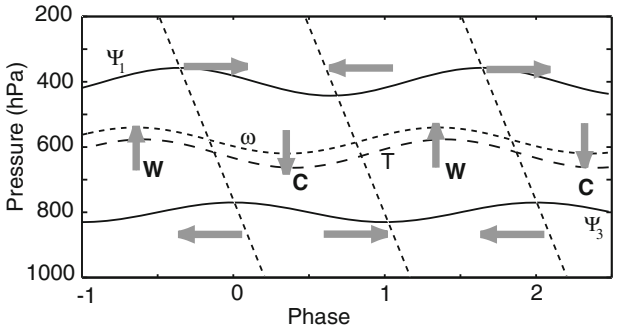



**Fig. 12.13** Phase relationship and circulation on the $p$–$x$ plane in the two-level model. The streamfunctions at the two levels are out of phase by about 64° and the constant phase lines are tilted westward. Vertical velocity has a phase of about 116° with respect to the streamfunction at level 3. The temperature has a phase of 108°. *C* and *W* indicate cold and warm sectors, respectively. *Upward* motions correspond to the maximum of the vertical velocity (Haltiner and Williams 1980)

Another point deals with the vorticity. Actually the divergent motion shown in Fig. 12.13 is distributed between the horizontal planes at 400 and 800 hPa. The convergence appearing at 400 hPa in the downward columns is such that the vorticity increases at those points while decreasing at the divergent points at the lowest level. This can also be obtained from Eq. (12.80) where at level 1 the vorticity tendency, without advection terms, is proportional to $\omega'_2$, while the opposite happens at level 3. This circumstance keeps the perturbation growing. Figure 12.13, redrawn from the Haltiner and Williams textbook, shows the solutions for a two-level model with the following characteristics: $p_0 = 200$ hPa, $\Delta p = 400$ hPa, and $\beta = 1.67 \cdot 10^{-11}$ $\mathrm{m^{-1}\ s^{-1}}$.

The analytical solutions are, with $c_\mathrm{r}$ as the real component of the phase velocity,

$$\begin{aligned} \psi'_1 &= (14.2/k)\cos\left[k\,(x - c_r t) + 64°\right] e^{nt} \\ \psi'_3 &= (10/k)\cos\left[k\,(x - c_r t)\right] e^{nt} \end{aligned} \tag{12.87}$$

so that the vertical velocity and the temperature are

$$\begin{aligned} \omega'_2 &= -1,36 \cdot 10^{-3} \cos\left[k\,(x - c_r t) + 116°\right] e^{nt} \\ T'_2 &= 4,35 \cos\left[k\,(x - c_r t) + 108°\right] e^{nt} \end{aligned} \tag{12.88}$$

The units are $\mathrm{m^2\ s^{-1}}$ for $\psi$ and hPa $\mathrm{s^{-1}}$ for the vertical velocity, and the temperature is given in C. Notice that the phases of $\omega'_2$ and $T'_2$ are almost the same with the temperature lagging slightly. The barotropic component is given by

$$\psi'_1 + \psi'_3 = (20.6/k)\ \cos\left[k\,(x - c_r t) + 38°\right] e^{nt}$$

so that the vorticity advection $U_T$ $(\partial\zeta_3/\partial x)$ is proportional to the quantity $-\sin[k(x-c_r t)+38°]$ and the effect of the vertical motion on the vorticity is not exactly in phase with $\omega'_2$; it can be shown that this effect is due to $\beta$.

We have now talked enough about the baroclinic instability, and it is now time to go back to the question that brought us here in the first place. The Hadley circulation cannot export heat beyond the tropical latitudes, while the eddies generated by the baroclinic instability should be able to do that. We can now go back to our energetic consideration that we introduced on Chap. 11.

## 12.3 Energetics of the Baroclinic Waves

As we have seen in Chap. 11, the total energy of the atmosphere is of the order of $3\ 10^9$ j m$^{-2}$. The mass per unit surface is of the order of $10^4$ kg m$^{-2}$; if only 10 % of this energy could be available for the atmospheric motions, we would have an average wind of 200 m s$^{-1}$ which is very unrealistic. In order to make available all the potential energy, the temperature of the atmosphere should be at absolute zero and the atmosphere itself should be squeezed very near the ground. It is much more reasonable to think about the existence of a minimum energy state (corresponding to a stratified atmosphere at rest) and refer to it when calculating the energy available for the motion. This energy is called the *available potential energy* (APE).

To have a more precise idea of what we are talking about, we refer to Fig. 12.14, taken from an example reported by Gill. We can imagine a fluid divided by a vertical sector that separates two fluids at density $\rho_1$ and $\rho_2$. If the separation is removed, the two fluids will stratify with the one with higher density at the bottom.

The potential energy of the initial configuration can be written as

$$E_p = \int_0^H \rho g z dz = \int_0^H (\rho_1 + \rho_2)\, g z dz = (\rho_1 + \rho_2)\, g H^2/4$$

and similarly the final energy as

$$E_p = (3\rho_1 + \rho_2)\, g H^2/8$$

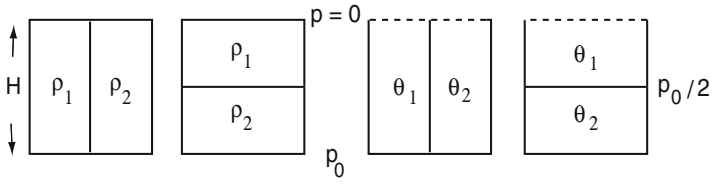



**Fig. 12.14** A simple scheme for the calculation of available potential energy (APE): at the *left* two fluids of different density and to the *right* with different potential temperatures

The available energy is then given by the difference:

$$APE = (\rho_2 - \rho_1)\, gH^2/8 \tag{12.89}$$

If all the APEs should transform into kinetic energy, we would have an average velocity for the fluid:

$$U^2 = (\rho_2 - \rho_1)\, gH/\left[2\,(\rho_2 + \rho_1)\right]$$

A very similar calculation can be done in the case that the conservative quantity is the potential temperature (in the previous case, it was the mass). Then we must assume that the upper surface extends to zero pressure. From the definition of total energy, we obtain

$$E_p + E_I = C_p \int_0^\infty \rho T dz = \frac{C_p}{g} \int_0^{p_r} T dp = \frac{C_p \theta}{g\,(k+1)\, p_0^k} p_0^{k+1} \tag{12.90}$$

where with $p_0$ we intend the maximum pressure and $k = R/C_p$. Using this equation for the initial and final energy, we obtain for APE

$$APE = \frac{C_p P_r\,(\theta_1 - \theta_2)}{2g\,(1+k)} \left[1 - \left(\frac{1}{2}\right)^k\right] \tag{12.91}$$

Actually the atmosphere is much nearer to this second case because it can be assimilated to a number of layers, each with its own potential temperature. The total energy can then be written as (Fig. 12.14)

$$E_p + E_I = C_p \int \left(p/p_0\right)^k \theta dM \tag{12.92}$$

where $dM = -dpdxdydz/g$. When the atmosphere reaches the status of minimum energy, the mass of the single layer is conserved and so also the average pressure $p$, because this depends only on the mass above that layer.

The *total potential energy* (TPE) per unit surface is then given by

$$TPE = \frac{C_p p_0^{-k}}{g} \int \theta p^k dp \tag{12.93}$$

Integrating by parts between the surface pressure ($p_0$) and zero, we obtain

$$TPE = \frac{C_p p_0^{-k}}{g\,(1+k)} \int p^{k+1} d\theta \tag{12.94}$$

The available potential energy becomes then

$$APE = \frac{C_p p_0^{-k}}{g\,(1+k)} \int \left(p^{k+1} - \overline{p}^{k+1}\right) d\theta \tag{12.95}$$

This equation can be put in a simpler form if we assume as usual that $p = \overline{p} + p'$ so that we have

$$\left(\overline{p} + p'\right)^{k+1} = \overline{p}^{k+1} + (k+1)\,\overline{p}^{k} p' + k\,(k+1)\,\overline{p}^{k-1} p'^2/2 + \cdots$$

When we average, the second term goes to zero and the APE becomes

$$APE = \frac{R}{2gp_0} \int \overline{p}^{k+1} \overline{(p'/\overline{p})^2} d\theta \tag{12.96}$$

This same relation can be expressed with an integral over pressure:

$$APE = \frac{1}{2} \int_0^{p_0} \frac{\overline{T}}{\Gamma_d - \overline{\Gamma}} \overline{\left(\frac{T'}{\overline{T}}\right)^2} dp \tag{12.97}$$

Based on this equation, we can calculate the order of magnitude for the ratio between TPE and APE. Considering $\overline{T}'^2 = (15C)^2$ and comparing with Eq. (12.107), we have

$$APE/TPE \approx \left(T'/T\right)^2 \approx (15/200)^2 \approx 1/200$$

The ratio between average kinetic energy and TPE can be found by putting

$$K = \frac{1}{2} \int V^2 dp \qquad TPE = \frac{c_v}{gR} \int c^2 dp$$

where

$$c^2 = C_p RT/C_v$$

is the square root of the sound velocity so that because $V/c \cong 0.05$ we have

$$K/TPE \approx 1/2000; \qquad K/APE \approx 1/10$$

With all these nice things in mind, we are ready to see how energy changes and transforms in the baroclonic instability.

### 12.3.1 Energy in the Two-Level Model

We start by writing the perturbation kinetic energy per unit mass $K'$. In the two-level model, this is given by

$$K' = \frac{1}{2}\left[\left(\frac{\partial \psi_1'}{\partial x}\right)^2\right] + \frac{1}{2}\left[\left(\frac{\partial \psi_3'}{\partial x}\right)^2\right] \tag{12.98}$$

where, in this case, the brackets indicate the zonal mean. The perturbation APE can be calculated by using the definition of the previous paragraph with Eq. (12.87). We assume that the temperature perturbation coincides with $T'_2$; then the integral can be evaluated with the condition to change it as energy per unit mass. We obtain

$$P' = \frac{g\overline{T'^2_2}}{2\left(\Gamma_d - \Gamma\right)\overline{T}} \tag{12.99}$$

From the definition of $\sigma$, we obtain easily

$$\sigma = -\frac{RT}{p}\frac{\partial \ln \theta}{\partial p} = \left(\frac{R}{p}\right)^2 \frac{T}{g}\left(\Gamma_d - \Gamma\right) = \left(\frac{f}{\lambda \Delta p}\right)^2 \tag{12.100}$$

We now put $\ln(p_3/p_1) \approx \Delta p/p$ in Eq. (12.90) and substitute for $T'_2$ in Eq. (12.104). We get

$$P' = \lambda^2\left[\left(\psi_1' - \psi_3'\right)^2\right]/2 \tag{12.101}$$

We multiply now the first part of Eq. (12.75) by $-\psi'_1$ and the second by $-\psi'_3$ and then sum and average zonally. We have

$$dK'/dt = -f\left[\omega_2'\left(\psi_1' - \psi_3'\right)\right]/\Delta p = -\left(2f/\Delta p\right)\left[\omega_2'\psi_T'\right] \tag{12.102}$$

From Eq. (12.76), if we multiply through by $-(\psi'_1 - \psi'_3)$ and averaging again zonally,

$$\begin{aligned} dP'/dt &= \lambda^2 U_T\left[\left(\psi_1' - \psi_3'\right)\partial\left(\psi_1' + \psi_3'\right)/\partial x\right] + f\left[\omega_2'\left(\psi_1' - \psi_3'\right)\right]/\Delta p \\ &= 4\lambda^2 U_T\left[\psi_T'\partial\psi_m'/\partial x\right] + \left(2f/\Delta p\right)\left[\omega_2'\psi_T\right] \end{aligned} \tag{12.103}$$

This equation shows first of all how the change in the perturbation APE is partly due to the change of the perturbation kinetic energy. The conversion is negative (i.e., the kinetic energy increases at the expense of the potential energy) for upward motion ($\omega'_2 < 0$) and the perturbation thickness is positive ($(\psi'_1 - \psi'_3) > 0$). In this case, warm air rises from low levels to replace cold air at the upper level, a situation

in which the potential energy decreases. The first term to the right of Eq. (12.103) is the net generation of the total energy of the perturbation as can be seen by summing Eqs. (12.102) and (12.103):

$$\left(dP' + dK'\right)/dt = 4\lambda^2 U_T \left[\psi_T \left(\partial \psi_m / \partial x\right)\right] \tag{12.104}$$

The growth of the total energy depends on the phase relationship between $\psi_T$ and $\psi_m$. It can be shown that if $kx_0$ is the phase difference between the two streamfunctions, then

$$\left(dP' + dK'\right)/dt \propto \sin\left(kx_0\right)$$

This means that the phase difference between temperature and geopotential must be $0 < kx < \pi$ and the maximum generation of energy happens when the phase difference is just $\pi/2$. This conclusion is the same to the one we have introduced in a qualitative way in Sect. 12.1, and it is also clear at this point why the ridge and through axis must slope westward.

At this point, our understanding of the circulation at middle latitudes is much better. The average potential energy $[P]$ generated by the temperature differences (as given by Eq. (12.97)) created by the differential heating is converted into perturbation potential energy $P'$ as shown by the term in Eq. (12.104). This energy is then converted in kinetic energy $K'$ through the lowering of the center of mass, as implicitly expressed in Eq. (12.102). This continuous process would deplete completely the APE of the mean flux if this were not replenished through the absorption of the solar radiation. A very simplified scheme is presented in Fig. 12.15, as taken from Wallace and Hobbs. What drives the system is the absorption of solar and planetary radiation which heats the lower troposphere and cools the upper troposphere, which is "adjusted" by the convective instability through the development of small-scale turbulence that also dissipates energy. The differential heating generates a latitudinal temperature gradient that is responsible for the buildup of APE. The Hadley circulation is a typical direct circulation (warm air rises and cold air sinks) that transforms APE into kinetic energy of large-scale motions. At middle latitude, responsible for this transformation is the baroclinic instability through a rather complex mechanism. In both cases, the planetary boundary layer (PBL) is where these motions are dissipated, while the final sink is the small-scale turbulence that through a cascade of energy goes down to the molecular scale.

It is not strictly necessary that part of the kinetic energy is converted back in APE (actually fluxes are rather small) because the energy from the sun is, for all effects, infinite. The atmosphere is like an engine that transforms thermal energy into motions that are dissipated by "friction" processes. The efficiency of such an engine can be easily evaluated as the ratio between the rate of change of kinetic energy and the energy received by the sun:

$$\left(dK'/dt\right)/Q\left(1-\alpha\right) = 2.2/240 = 0.009$$

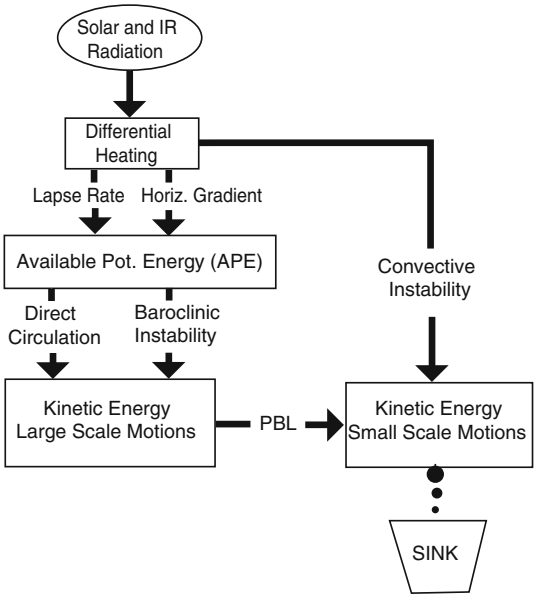



**Fig. 12.15** A simplified scheme of the energy cycle of the atmosphere. The sink is the final dissipation of energy (Adapted from Wallace and Hobbs 1977)

which is less than 1 %. On the other hand, the same order of magnitude for the efficiency can be obtained as the ratio between the temperature difference between pole and equator and the average temperature of the troposphere (250 K).

We can now make some considerations as to what we learned in Sect. 11.2.3 about geopotential and sensible heat flows. Equation (12.118) gives as the meridional flux of sensible heat simply the term $(dP' + dK')/dt$ and the changes of kinetic and potential energy are of opposite sign. At this point, a good conclusion for this chapter could be an introduction to the transport problem. This is not a very popular topic for the troposphere, while it is quite hot for the stratosphere. In this case, for transport, we mean all the processes that contribute to transport, mainly across latitude, heat, chemical, and scalar tracers like the potential temperature.

### *12.3.2 The Parameterization of Transport*

The considerable amount of work on the baroclinic instability should give us some indication about the transport mechanisms at middle latitude. Green, in the early 1970s, tried to exploit the theory to calculate large-scale transport. He started from the theory of the mixing length so that a generic term for the eddy transport like $u's'$ (with $s$ some generic quantity) can be written as

$$u's' = -u'l\partial\overline{s}/\partial x = -K\partial\overline{s}/\partial x \tag{12.105}$$

where $K$ is a diffusion coefficient and $\bar{s}$ is the zonal mean value of the quantity being transported. If we consider only transport in the meridional plane, we have

$$\overline{v's'} = -K_{vy}\partial\bar{s}/\partial y - K_{vz}\partial\bar{s}/\partial z \tag{12.106}$$

where $\bar{s}$ is a function of $y$ and $z$, $K_{vy} = \overline{v'l_y}$, $K_{vz} = \overline{v'l_z}$. Similarly, for the vertical, we can write

$$\overline{w's'} = -K_{wy}\partial\bar{s}/\partial y - K_{wz}\partial\bar{s}/\partial z; \quad K_{wy} = w'l_y; \quad K_{wz} = w'l_z \tag{12.107}$$

If we assume that the baroclinic instability is responsible for the eddies, then

$$\frac{K_{vz}}{K_{vy}} \approx \frac{\overline{v'l_z}}{\overline{v'l_y}} \approx \frac{l_z}{l_y} \approx \gamma \approx \frac{1}{2}\alpha \approx -\frac{1}{2}\frac{\partial\theta/\partial y}{\partial\theta/\partial z} \tag{12.108}$$

and the heat flux can be written as

$$\overline{v'\theta'} = -K_{vy}\left(\frac{\partial\overline{\theta}}{\partial y} - \frac{1}{2}\frac{\partial\theta/\partial y}{\partial\theta/\partial z}\frac{\partial\overline{\theta}}{\partial z}\right) = -\frac{1}{2}K_{vy}\frac{\partial\overline{\theta}}{\partial y} \tag{12.109}$$

Again we have a closure problem because we need to express the diffusion coefficient as a function of average quantities. Some help can be found in the discussion about energy and baroclinic instability. The perturbation kinetic energy and APE can now be written as

$$\begin{aligned} K' &= \frac{1}{2}\int dydxdzp\left(u'^2 + v'^2\right) \\ APE &= \frac{1}{2}\int dydxdp\frac{T'^2}{\overline{T}\partial\overline{\theta}/\partial z} \end{aligned} \tag{12.110}$$

Notice that the first part of the equation expresses the total kinetic energy, while in the second ones, we have assumed $\theta = T$. Assuming average values for the velocities $V = |v'|$ and $U = |u'|$ and putting

$$M = \int \frac{dxdydp}{g} \qquad T' \approx \frac{L}{2}\frac{\partial\overline{\theta}}{\partial y}$$

with $L$ some characteristic length, we obtain

$$K' = MV^2 \qquad APE = \frac{1}{2}\frac{M}{\overline{T}}\frac{L^2}{4}\frac{\left(\partial\overline{\theta}/\partial y\right)^2}{\left(\partial\overline{\theta}/\partial z\right)} \tag{12.111}$$

If all the APEs are transformed into kinetic energy, from this equality, we can get the velocity

$$V^2 = \frac{g}{\overline{T}} \frac{L^2}{8} \frac{\left(\partial\overline{\theta}/\partial y\right)^2}{\left(\partial\overline{\theta}/\partial z\right)}$$

and using the definition for the Rossby deformation radius, we get

$$V = uL/\left(2\sqrt{2}L_R\right) \tag{12.112}$$

where $u$ is the thermal wind given by Eq. (12.52). We can now assume that the characteristic length $L$ is the one for which we have the fastest growth rate and this is about five times the Rossby radius so that at least for the sake of order of magnitude, we write $K_{vy} \approx VL_R$ and we have

$$K_{vy} \approx -\frac{g}{\overline{T}} \left(\frac{H}{f}\right)^2 \left(\frac{\partial\overline{\theta}}{\partial y}\right) \left(\frac{g}{\theta} \frac{\partial\overline{\theta}}{\partial z}\right)^{1/2} \tag{12.113}$$

and the heat fluxes are then

$$\begin{aligned} \overline{v'\theta'} &\approx \frac{1}{2} \frac{gH^2}{\theta f^2} \left(\frac{\partial\overline{\theta}}{\partial y}\right)^2 \left(\frac{g}{\theta} \frac{\partial\overline{\theta}}{\partial z}\right)^{1/2}; \\ \overline{w'\theta'} &\approx -\frac{1}{4} \left(\frac{gH}{\theta f}\right)^2 \left(\frac{\partial\overline{\theta}}{\partial y}\right)^3 \left(\frac{g}{\theta} \frac{\partial\overline{\theta}}{\partial z}\right)^{-1/2} \end{aligned} \tag{12.114}$$

These equations may be validated by experimental data. However, a very interesting thing is that the meridional heat flux is not a simple diffusion process because it is not proportional to the meridional temperature gradient. Also from Eq. (12.114), we can see the existence of some feedback mechanism. For example, if the temperature meridional gradient should increase for some reason, then the heat flux would increase even faster thus increasing the heat exchange along latitudes with the result of bringing the temperature gradient back to the unperturbed value. Then we are in the presence of a negative feedback. The same increase would lead to an increase of the vertical flux (remember that $\partial\theta/\partial y$ is negative) so the static stability would also increase and again this would contribute to bringing the meridional gradient back to its initial value.

These considerations are quite important because based on them we can simplify greatly the heat transport in two-dimensional climate models.

## 12.4 The General Circulation: A Reductionist Approach

We cannot embark here on a contribution to the reductionism debate, although in the past there have been reductionist temptations even in atmospheric physics. At the time (we were in the 1970s), it was called the similitude theory. It was originally formulated by G. Golitsyn (1970) and exposed in a slightly different form by Conway Leovy, in the preprint that was never published anywhere but continues to inspire us. The basic idea is to find some general law to which all the planet atmospheric circulations could be reduced. We have attempted something like that for the thermal structure, and we will try to do something like that now for the circulation.

A good starting point is to consider the potential energy for unit mass in the atmosphere, given by $gH$, as shown already in Chap. 11. If this energy would be emitted as radiation, then the average emitted power would be

$$gH/\tau_r \tag{12.115}$$

where $\tau_r$ is the radiative time constant given by (see Chap. 1)

$$\tau_r = p_e C_p / \varepsilon g \sigma T_e^3$$

with $T_e$ and $p_e$ as values of temperature and pressure at the emission level and $\varepsilon$ as the emissivity. Equation (12.115) can be made nondimensional using as normalization constants the Earth's angular velocity and the radius $a$ so that

$$A = \frac{1}{\Omega \tau_r} \frac{gH}{(\Omega a)^2} \tag{12.116}$$

Consider now the Hadley cell as shown in Fig. 12.16, with the assumption that the heat transported by the circulation in the upper branch is dissipated as radiation. We will have at the upper branch

$$\rho_e C_p w (T_1 - T_2) \approx \varepsilon \sigma T_e^4$$

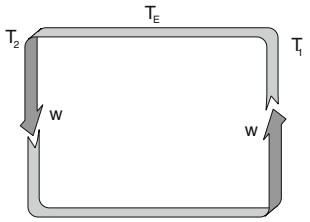



**Fig. 12.16** Heat transport in the Hadley cell

which is equivalent to

$$\rho_e w \delta T / T \approx \varepsilon \sigma T_e^3 / C_p$$

Comparing this equation with the definition of radiative relaxation time, we have

$$w \delta T / T \approx R T_e / (g \varepsilon \tau_r) \approx H / (\varepsilon \tau_r) \tag{12.117}$$

This is a very important link between the temperature difference across the Hadley cell, the vertical velocity, and the dissipated power.

Another important relation can be found directly from Eq. (12.22) which gives the zonal velocity for the constant angular momentum. However, we use the Rossby number, $Ro = u/(\Omega a)$, to write

$$Ro = \alpha \sin^2 \varphi_s / \cos \varphi_s \tag{12.118}$$

In this case, $\varphi_s$ is the Hadley cell's northern limit where air starts to subside. The coefficient $\alpha$ is just an adjustment term that takes into account the non-perfect conservation of the angular momentum. Using the notation

$$r_s^2 = \sin^2 \varphi_s / \cos \varphi_s$$

we can transform Eq. (12.17) in a slightly different form:

$$\frac{\delta T}{T} = \frac{u \Omega a}{2gH} r_s^2 = \frac{Ro (\Omega a)^2}{2gH} r_s^2 \tag{12.119}$$

Equations 12.116,12.117,12.118, and 12.119 should allow us in principle to evaluate the Hadley cell extension (Fig. 12.16); however we need an additional relation that links the vertical velocity to the Rossby number. The vorticity conservation can be written as

$$\Omega v / a \approx f w / H$$

and considering that $f = 2\Omega \sin \phi$, we have $v/a \approx (w/H) \sin \phi \approx (w/H)\, r_s$ so that

$$v / (\Omega a) \approx r_s \, (w / \Omega H)$$

At this point, we can assume the ratio between meridional and zonal velocity to be of the order of the *local Rossby number*

$$Ro_s \approx u / (a \Omega \sin \varphi) \approx Ro / r_s$$

With the approximation that $r_s \approx \sin\varphi$, that is, $v = u Ro_s$, we have

$$v / (\Omega a) \approx r_s \, (w / \Omega H) \approx Ro_s R_o \approx Ro^2 / r_s$$

and taking into account Eq. (12.47)

$$w/\Omega H \approx \alpha Ro \tag{12.120}$$

We can now introduce the normalized quantities

$$W = w/\Omega H; \quad G = \left[gH/(\Omega a)^2\right](\delta T/T) \tag{12.121}$$

so as to define all the relevant quantities for the circulation to two constants, $A$ (already defined) and an adjustable constant $\gamma$:

$$\begin{aligned} Ro &= (\gamma A)^{1/3}; & r_s &= \alpha^{-1/2}(\gamma A)^{1/6} \\ G &= (2\alpha)^{-1}(\gamma A)^{1/2}; & v/(\Omega a) &= r_s W = \alpha^{1/2}(\gamma A)^{1/2} \end{aligned} \tag{12.122}$$

These equations determine the main characteristics of the circulation on each planet, and they can be completed by another equation concerning the fractional change of the potential temperature across a scale height, $\delta\theta/(H\theta)$.

To do this simple thing, we need to introduce the inertial instability, the first dynamical instability in a very long line.

### *12.4.1 The Inertial Instability*

We will repeat here the details about the inertial instability given in Chap. 4. We proved that the acceleration of the parcel displaced by its initial position is

$$Dv/Dt = D^2(\delta y)/Dt^2 = -f\left(f - \partial u_g/\partial y\right)\delta y \tag{12.123}$$

This has the form of an "oscillator" equation and will give different motions according to the sign of the terms in parentheses:

$$\left(f - \frac{\partial u_g}{\partial y}\right)\begin{cases} > 0 & \text{stable} \\ = 0 & \text{neutral} \\ < 0 & \text{unstable} \end{cases} \tag{12.124}$$

When the third condition is satisfied, a parcel displaced normally to the zonal current will move farther away from its initial position. The condition for instability can also take into account also the stratification of the fluid beside the latitudinal shear of the wind. In this case, it is necessary to make the derivative along isentropic surfaces, that is, the condition to be satisfied is the following:

$$f - (\partial u/\partial y)_\theta < 0.$$

And if we apply the rules learned in Chaps. 5 and 7, we have

$$\left(\frac{\partial u}{\partial y}\right)_\theta = \left(\frac{\partial u}{\partial z}\right)_y \left(\frac{\partial z}{\partial y}\right)_\theta + \left(\frac{\partial u}{\partial y}\right)_z$$

so that the condition for instability becomes

$$f - \frac{\partial u}{\partial y} + \left(\frac{\partial u}{\partial z}\right) \frac{(\partial\theta/\partial y)}{(\partial\theta/\partial z)} < 0 \tag{12.125}$$

From the thermal wind equation, we have

$$f\partial u/\partial z = -(g/\theta)\,\partial\theta/\partial y$$

which can be substituted in Eq. (12.125) for the meridional gradient of $\theta$:

$$f - \frac{\partial u}{\partial y} - \frac{1}{g}\left(\frac{\partial u}{\partial z}\right)^2 \frac{f\theta}{(\partial\theta/\partial z)} < 0$$

From the definition of Eq. 8.28 of the Richardson number, we have

$$1 - \frac{1}{f}\frac{\partial u}{\partial y} - \frac{1}{Ri} < 0 \tag{12.126}$$

For the stability, we should then have

$$1 - Ro > 1/Ri$$

so that in the case $Ro < 1$ it must be $Ri > 1$. Again, using the definition of the Richardson number, we have for the stability condition

$$\frac{g}{\theta}\frac{\partial\theta}{\partial z}\left(\frac{\partial u}{\partial z}\right)^{-2} \approx \frac{g}{H}\frac{\delta\theta}{\theta}\frac{H^2}{u^2} \approx \frac{1}{Ro^2}\frac{gH}{(a\Omega)^2}\frac{\delta\theta}{\theta} > 1 \tag{12.127}$$

And we get

$$\frac{1}{(\gamma A)^{2/3}}\frac{gH}{(a\Omega)^2}\frac{\delta\theta}{\theta} > 1 \tag{12.128}$$

### *12.4.2 A Comparison Among the Planets*

An application of the above equations can be made on the Hadley cells of the different planets. In Fig. 12.17, the relationship between the northern limit of the

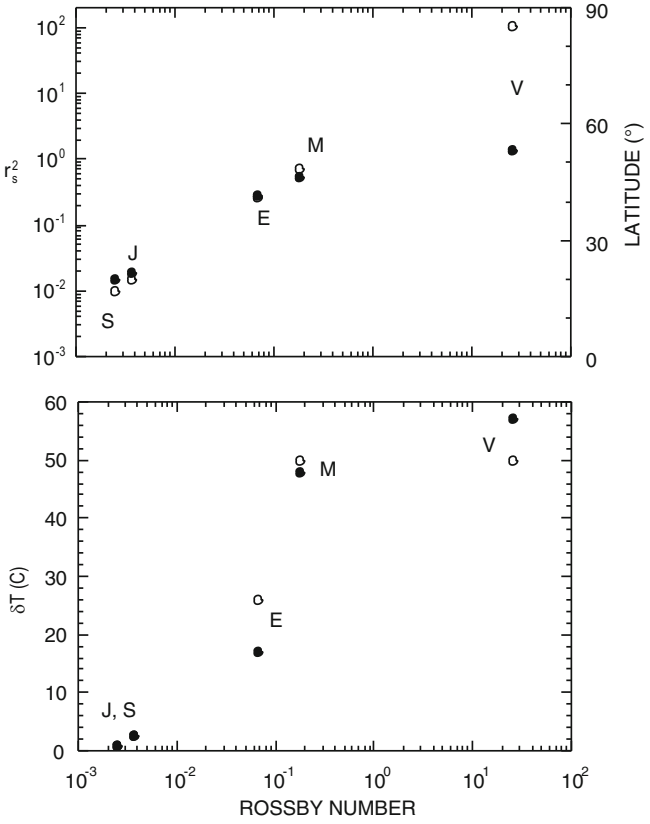



**Fig. 12.17** The relation between the Rossby number and the latitudinal extension of the Hadley cells for a few planets of the solar system (*top*): *V* Venus, *E* Earth, *M* Mars, *J* Jupiter, *S* Saturn. At the *left* the parameter $r_s$ is given. For the same planets, the *bottom figure* gives the temperature change across the scale height. In both figures, the *black circle*s are the observed data and the *white circles* the calculated values

Hadley cell and the Rossby number is shown. We notice that, except for Venus, the agreement with the real data is quite good. We have not attempted to evaluate the meridional velocity because this is a quite complicated quantity to measure. For Venus, the poor agreement may be due to the high Rossby number. The extension of the cell up to polar latitude may be the reason for some of the instability. The good agreement for the other planets is a temptation to "push the envelope" of the theory. Mars, for example, is sometime shrouded with planetary dust storms that change the emissivity of the atmosphere. Reasonable values could be 0.15 for a normal situation to 0.5 in case of a storm so that the relaxation time reduces from 9 to 1.5 days. This implies a change in the zonal wind velocity on the planet that may go from about 60 m $s^{-1}$ to 100 m $s^{-1}$ and a corresponding increase in the amplitude of the Hadley cell. This behavior is what is observed during dust storms on Mars.

The predictions of the theory about static stability are given in the lower part of Fig. 12.17, and we notice immediately how the agreement is good in this case. In the case of Jupiter, the data and theory in practice overlap.

The most interesting result of this theory is that knowledge of the Rossby number does not imply a linear relation between this same number and the zonal wind.

The most obvious approach would be to assume that the potential energy ($gH$), transformed in available energy through the efficiency ($\delta T/T$), was proportional to the square of the zonal velocity. However, the temperature difference across the Hadley cell depends itself on the zonal velocity and then on the Rossby number.

## E.12 Examples

### *E.12.1 The Thermodynamic Equation*

We start from (12.70)

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\theta\prime+v\prime\frac{\partial\overline{\theta}}{\partial y}+w\prime\frac{\partial\overline{\theta}}{\partial z^*}=0$$

And using the thermal wind relation

$$\frac{\partial\overline{u}}{\partial z^*}=-\frac{g}{f}\frac{1}{\overline{\theta}}\frac{\partial\overline{\theta}}{\partial y}$$

The equation becomes

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\frac{\theta\prime}{\overline{\theta}}-v\prime\frac{f}{g}\frac{\partial\overline{u}}{\partial z^*}+\frac{w\prime}{\overline{\theta}}\frac{\partial\overline{\theta}}{\partial z^*}=0 \tag{E.12.1}$$

Introducing the streamfunction

$$\psi=\frac{\phi}{f} \tag{E.12.2}$$

where $f$ is the geopotential, we have

$$\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\frac{\theta'}{\overline{\theta}}-\frac{f}{g}\frac{\partial\psi}{\partial x}\frac{\partial\overline{u}}{\partial z^*}+\frac{w'}{g}N_*^2=0 \tag{E.12.3}$$

where the Brunt–Väisälä frequency is defined as

$$N_*^2=\frac{g}{\overline{\theta}}\frac{\partial\overline{\theta}}{\partial z^*}$$

Using the relation between streamfunction and geopotential, we have

$$T' = \frac{H}{R}\frac{\partial \Phi}{\partial z^*} = f\frac{H}{R}\frac{\partial \psi}{\partial z^*} \tag{E.12.4}$$

And then we may assume

$$\frac{\theta'}{\overline{\theta}} \simeq \frac{T'}{\overline{T}} = \frac{f}{\overline{T}}\frac{H}{R}\frac{\partial \psi}{\partial z^*} = \frac{f}{g}\frac{\partial \psi}{\partial z^*}$$

Substituting in (E.12.17), we have

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\frac{\partial \psi}{\partial z^*} - \frac{\partial \psi}{\partial x}\frac{\partial \overline{u}}{\partial z^*} + \frac{w'}{f}N_*^2 = 0$$

which is the form we need.

### *E.12.2 The Hadley Circulation as a Shallow Water Case*

The equations of a shallow water system can be written as a function of the perturbation height $h$:

$$\begin{aligned} \partial u/\partial t + u\partial u/\partial x + v\partial u/\partial y - fv &= -g\partial h/\partial x \\ \partial v/\partial t + u\partial v/\partial x + v\partial v/\partial y + fu &= -g\partial h/\partial y \end{aligned} \tag{E.12.5}$$

When averaged zonally and adding the friction term, they become

$$\begin{aligned} \partial u/\partial t + v\partial u/\partial y - fv &= -\alpha u \\ \partial v/\partial t + v\partial v/\partial y + fu &= -g\partial h/\partial y - \alpha v \end{aligned}$$

These equations can now be normalized using the Rossby radius $((gH)^{1/2}/f$ as length scale and the Coriolis parameter as the inverse of the time scale. The result is simply

$$\begin{aligned} \partial u/\partial t + v\partial u/\partial y - v &= -\alpha u \\ \partial v/\partial t + v\partial v/\partial y + u &= -\partial h/\partial y - \alpha v \end{aligned} \tag{E.12.6}$$

where now all the variables are nondimensional. To write the continuity equation in the same quantities, we need to start from the full continuity:

$$\frac{1}{H}\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}\left[\left(1 + \frac{h}{H}\right)u\right] + \frac{\partial}{\partial y}\left[\left(1 + \frac{h}{H}\right)v\right] = 0 \tag{E.12.7}$$

We can then introduce a normalized thickness $\eta = h/H$ and write the equation

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x}\left[(1+\eta)\, u\right] + \frac{\partial}{\partial y}\left[(1+\eta)\, v\right] = 0 \tag{E.12.8}$$

The thickness $h$ can be interpreted as geopotential thickness:

$$h + H = RT/g = (RT_0/g)\,(1 + T/T_0) \tag{E.12.9}$$

so that

$$1 + h/H = 1 + T/T_0 \tag{E.12.10}$$

In this way, Eq. (E.12.7) can be interpreted as describing the temperature of the fluid, and if we include the forcing and average zonally, we get

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial y}\left[(1+\eta)\, v\right] = Q \tag{E.12.11}$$

where the forcing may be represented as a Newtonian cooling term

$$Q = \frac{\eta_E - \eta}{\tau} \tag{E.12.12}$$

where $\eta_E$ is an assigned equilibrium temperature. Equations E.12.6, E.12.7, E.12.8, E.12.9, E.12.10, and E.12.11 can be solved numerically.

### *E.12.3 The Hadley Circulation: Numerical Solution*

We consider the time-independent solution so that the system to be solved is given by

$$\begin{aligned} &v\partial u/\partial y - v = -\alpha u \\ &v\partial v/\partial y + u = -\partial h/\partial y - \alpha v \\ &\frac{\partial}{\partial y}\left[(1+\eta)\, v\right] = Q \end{aligned} \tag{E.12.13}$$

where $h = 1 + \eta$. To simplify things further, the quantity $h_E$ is specified as follows:

$$\eta_E(y) = \begin{cases} H_E & \text{for} \quad |y| < y_E \\ 0 & \text{for} \quad |y| > y_E \end{cases} \tag{E.12.14}$$

The first of Eq. E.12.9 is simply

$$\frac{\partial u}{\partial y} = 1 - \alpha \frac{u}{v} \tag{E.12.15}$$

The second and third equations can be written as

$$\begin{aligned} v\partial v/\partial y + u &= -\partial \eta/\partial y - \alpha v \\ v\frac{\partial \eta}{\partial y} + (1+\eta)\frac{\partial v}{\partial y} &= Q \end{aligned} \tag{E.12.16}$$

This system can be solved for the derivatives $\partial\eta/\partial y$ and $\partial v/\partial y$ so that

$$\frac{\partial v}{\partial y} = -\left[uv + \alpha v^2 + Q\right]\left[v^2 - (1+\eta)\right]^{-1} \tag{E.12.17}$$

$$\frac{\partial \eta}{\partial y} = \left[(1+\eta)(\alpha v + u) + Qv\right]\left[v^2 - (1+\eta)\right]^{-1} \tag{E.12.18}$$

The system (E.12.11), (E.12.13), and (E.12.14) can be solved using MATLAB, for example; the appropriate boundary conditions and the results are given in Fig. 12.14. A possible MATLAB script follows:

```
function xdot=hadley(t,x);
xdot=zeros(3,1);
alfa=1.0;
xdot(1)=1-alfa*x(1)/x(2);
if(t<0.1)
    q=1-x(3);
else
    q=-x(3)
end
aux=x(2)*x(2)-(1+x(3));
aux1=1/aux;
xdot(2)=-(x(1)*x(2)+alfa*x(2)*x(2)+q)*aux1;
xdot(3)=((1+x(3))*(alfa*x(2)+x(1))+q*x(2))*aux1;
The following instructions run the program and plot the results
[t,x]=ode45('hadley',[0 0.95],[0,0.001,0.17]);
plot(t,x(:,1),t,x(:,2),t,x(:,3)), axis([0 1 0 1])
```

### *E.12.4 The Hadley Circulation on Slow-Rotating Planet?*

An approach similar to the Held and Hou model can be applied to a slow-rotating planet like Venus. We can simplify the expression for the equilibrium temperature:

$$\theta_E = \theta_{E0} - \frac{2}{3}\Delta\theta P_2(\sin\varphi) \approx \theta_{E0} - \frac{\Delta\theta}{a^2}y^2 \tag{E.12.19}$$

where a is the radius of the planet and $\sin\varphi \approx y/a$. The zonal velocity which conserves the angular momentum is given by

$$u_M = \frac{\Omega a^2 \sin^2\varphi}{a\cos\varphi} \approx \frac{\Omega y^2}{a}$$

The average wind shear is obtained simply by dividing by the scale height and then the cyclostrophic thermal wind equation.

Now the same approach could be used for a slow-rotating planet where the thermal wind equation is now substituted by the cyclostrophic relation:

$$\frac{u^2\tan\varphi}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial y} \approx -R\frac{\partial T}{\partial y}$$

so that

$$\frac{\partial\left(u^2\right)}{\partial z} \approx \frac{u_M^2}{H} = \frac{Ra}{Hy}\frac{\partial T}{\partial y} = \frac{ga}{Ty}\frac{\partial T}{\partial y} = \frac{ga^2}{y\theta_0}\frac{\partial\theta_M}{\partial y} \tag{E.12.20}$$

We can obtain $\partial\theta/\partial y$ from this equation

$$\frac{\partial\theta_M}{\partial y} = -\frac{u_M^2 y\theta_0}{Hga^2} = -\frac{\Omega^2 y^5\theta_0}{Hga^4}$$

that can be easily integrated to give

$$\theta_M = \theta_{M0} - \frac{\Omega^2\theta_0}{a^4 gH}\frac{y^6}{6} \tag{E.12.21}$$

Calling $\pm Y_H$ as the limits of the Hadley cell, the condition for the net heating reads as

$$\int_o^{Y_H}\theta_M dy = \int_o^{Y_H}\theta_E dy$$

A second equation is obtained by imposing that

$$\theta_M(Y_H) = \theta_E(Y_H)$$

The condition on the integration gives

$$\theta_{M0} - \frac{\Omega^2\theta_0}{a^4 gH}\frac{Y_H^6}{42} = \theta_{E0} - \frac{\Delta\theta_0}{a^2}\frac{Y_H^2}{3} \tag{E.12.22}$$

while the equivalence of the temperature as

$$\theta_{M0} - \frac{\Omega^2\theta_0}{a^4 gH}\frac{Y_H^6}{6} = \theta_{E0} - \frac{\Delta\theta_0}{a^2}Y_H^2 \tag{E.12.23}$$

Equating the difference $\theta_{E0} - \theta_{M0}$, we obtain

$$Y_H = \left(\frac{14gHa^2\Delta\theta}{3\Omega^2\theta_0}\right)^{\frac{1}{4}} \tag{E.12.24}$$

To obtain the difference $\theta_{E0} - \theta_{M0}$, we subtract (E.12.22) and (E.12.23) to get

$$\theta_{E0} - \theta_{M0} = \frac{2}{9}\left(\frac{14gH\Delta\theta^3}{3\Omega^2 a^2\theta_0}\right)^{\frac{1}{2}}$$

For Venus, we may use $H = 15.9\text{km};\ \Omega = 2\pi/224.7\text{d}^{-1}; a = 6051\text{km}$ g $= 8.87\text{ms}^{-2}$ and get easily $YH > a$, that is, the Hadley cell would be larger than the hemisphere.

### *E.12.4 Transport by Eddies (James 1994, p. 125)*

Consider a disturbance geostrophically and hydrostatically balanced as shown in Fig. E.12.1 where three isobaric surfaces are shown at $p + \Delta p$, $p$, and $p - \Delta p$. As we have seen before, the streamfunction is related to the geopotential height $Z'$ by

$$\psi = \frac{g}{f}Z' \tag{E.12.25}$$

The wind component is given by

$$u' = -\frac{\partial\psi}{\partial y} = -\frac{g}{f}\frac{\partial Z'}{\partial y}; \quad v' = \frac{\partial\psi}{\partial x} = \frac{g}{f}\frac{\partial Z'}{\partial x}; \tag{E.12.26}$$

The temperature of the middle level (at pressure $p$) is given by

$$T_0 = \frac{p_o g}{R}\frac{Z'_{p+\Delta p} - Z'_{p-\Delta p}}{2\Delta p} \tag{E.12.27}$$

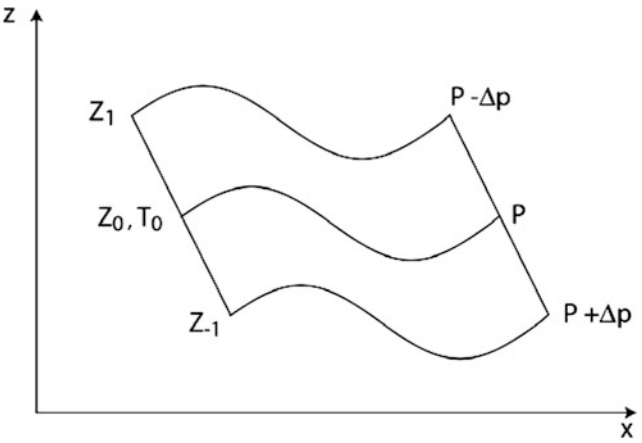



**Fig. E.12.1** Schematic illustration of a wavelike disturbance with westward phase tilt with altitude (James 1994)

We assume that the disturbance is sinusoidal and has a constant amplitude with altitude but has a phase shift with height:

$$Z_p = Z_{Rp} + A \sin(kx + i\delta), \quad i = -1, 0, 1 \tag{E.12.28}$$

The difference in geopotential height is given by

$$Z'_{p+\Delta p} - Z'_{p-\Delta p} = A\left[(\sin kx + \delta) - (\sin kx - \delta)\right] = A\cos(kx)\sin(\delta)$$

so that the temperature

$$T'_0 = \frac{p_o g}{\Delta p R} A \cos(kx) \sin(\delta) \tag{E.12.29}$$

There is no eddy fluctuation of temperature without phase shift. According to (E.12.20), the poleward wind is given by

$$v' = \frac{kg}{f} A \cos(kx) \tag{E.12.30}$$

and the poleward temperature flux is given by

$$\left[\overline{v'T'}\right] = \frac{1}{2} \frac{p_0 g^2 A^2 k}{fR\Delta p} \sin\delta \tag{E.12.31}$$

Inspection of Fig. E.12.1 makes this result obvious. The westward phase tilt means that thickness, and then temperature, is a maximum where the poleward wind is a maximum. Thus there must be a poleward temperature flux.

## References[1]

### *Books*

Gill AE (1982) Atmosphere-ocean dynamics. Academic, New York
Haltiner GJ, Williams RT (1980) Numerical prediction and dynamic meteorology. Wiley, New York
Holton JR (1992) An introduction to dynamic meteorology. Academic, New York
James IN (1994) Introduction to circulating atmosphere. Cambridge University Press, Cambridge
Lindzen RS (1990) Dynamics in atmospheric physics. Cambridge University Press, Cambridge/New York
Lindzen RS, Lorenz EN, Platzman GW (eds) (1990) The atmosphere- a challenge: the science of Jule Gregory Charney. American Meteorological Society, Boston
Pedlosky J (1987) Geophysical fluid dynamics. Springer, New York
Wallace JM, Hobbs PV (1977) Atmospheric science: an introductory survey. Academic, New York

### *Articles*

Eady ET (1949) Long waves and cyclone waves. Tellus 1(3):156
Golitsyn GS (1970) A similarity approach to the general circulation of planetary atmospheres. Icarus 13:1
Green JSA (1960) A problem in baroclinic stability. Q J Roy Meteorol Soc 86:237
Green JSA (1970) Transfer properties of the large-scale eddies and the general circulation of the atmosphere. Q J Roy Meteorol Soc 96:157
Held IM, Hou AY (1980) Nonlinear axially symmetric circulations in nearly inviscid atmosphere. J Atmos Sci 37:515
Held IM, Hoskins BJ (1985) Large scale eddies and the general circulation of the troposphere. Adv Geophys 28:3
Leovy CB Circulation in planetary atmospheres. Unpublished manuscript
Polvani LM, Sobel AH (2002) The Hadley circulation and the weak temperature gradient approximation. J Atmos Sci 59:1744
Schneider EK (1983) Martian great dust storms: interpretative axially symmetric models. Icarus 55:302
Sobel AH, Schneider T (2009) Single-layer axisymmetric model for a Hadley circulation with parameterized eddy momentum forcing. J Adv M Earth Sys 1:1
Stone PH (1972) A simplified radiative-dynamical model for the static stability of rotating atmospheres. J Atmos Sci 29:405

[1]Here the basic textbooks are Gill, Haltiner and Williams, and Lindzen. But the real sources are the papers listed. Some of them may look historical but they still are very instructive to read. The Hadley section has been completely rewritten.

# Chapter 13
# Radiation for Different Uses

In Chap. 2, we introduced the problem of radiative transfer in the atmosphere. After that, we were able to make some simple calculations. However, in order to go on, we need to introduce some more sophisticated algorithms. For example, for the solar radiation, the method we suggested is not fast enough to do practical calculations. Also the treatment for the diffuse radiation developed for the molecular component (Rayleigh scattering) must be extended also to aerosol particles. This, for example, will enable us to calculate how much radiation goes through a very thick cloud (optical thickness larger than 10).

Regarding infrared radiation, we need to learn how a single gas with specific spectroscopic characteristics contributes to the emission and absorption of radiation. Also, in this case, we may use a method that takes hours on a computer or one that can be solved in a few seconds. The latter may look more like engineering solutions, but as always, we prefer to allow the reader to practice because we are convinced that a theory can be corrected and improved only through exercise.

We will start with the solar radiation and its interaction with aerosols. In the process, we will make a pass at the Mie theory on light scattering by spherical particles. We will end up with the treatment of the IR radiation based on the band approach developed by Robert Cess.

## 13.1 Parameterization of Gaseous Absorption

We have seen that for ozone there are no major problems in calculating the absorption if the cross sections are known. However, if we consider water vapor, we will start to have some major problems because in the near infrared, this gas has

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_13) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_13

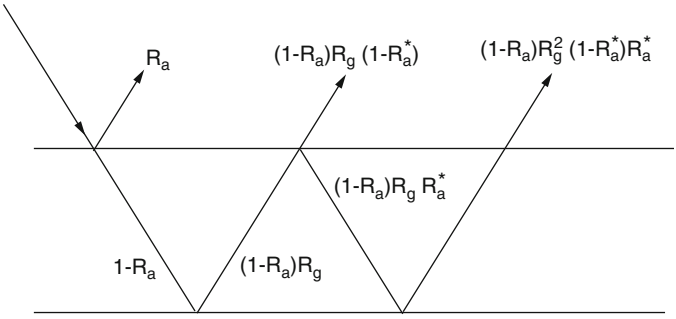



**Fig. 13.1** The scheme to evaluate the combined reflectivity of the atmosphere and the surface

lines that are not resolved. Also the absorption coefficient may depend on pressure.

In this first part, we will limit the discussion to gas absorption although we know there are a few objections, because in the case of water vapor, it is not easy to separate the gas from the liquid phase, that is, without clouds.

We can start by calculating the albedo which results as a combination between the effect of the atmosphere and the surface. We assume that the atmosphere is a single reflecting layer over a surface that has a reflectivity $R_g$. The reflectivity of the atmosphere depends on the zenith angle of the sun and is given by $R_a(\mu_0)$ where $\mu_0 = \cos\theta$. We also assume the atmosphere has an internal reflectivity $R^*_a$ that coincides with the average of the reflectivity over all the zenith angles (Fig. 13.1):

$$R_a^* = 2\int_0^1 R_a\ (\mu_0)\ \mu_0\, d\mu_0$$

The reason that justifies this approach is that the radiation reflected by the surface is completely diffuse so that the internal reflectivity may well be the average of $R_a(\mu_0)$. Referring to the figure, we can easily calculate the total reflectivity:

$$\begin{aligned} R(\mu_0) &= R_a(\mu_0) + [1 - R_a(\mu_0)]\left(1 - R_a^*\right)\left[1 + R_g R_a^* + R_g^2 R_a^{*2} + \ldots\right] \\ &= R_a(\mu_0) + [1 - R_a(\mu_0)]\left(1 - R_a^*\right) R_g / \left(1 - R_g R_a^*\right) \end{aligned} \tag{13.1}$$

As we will see in a while, the value of the atmospheric reflectivity is a function of the wavelength because it is determined by the Rayleigh scattering, meaning that it will be comparatively higher in the ultraviolet than in the near infrared. It can be shown that in the region where the ozone absorption is important, the reflectivity is given by

$$R_a(\mu_0) = 0.219/(1 + 0.816\mu_0) \quad \Rightarrow \quad R_a^* = 0.144 \tag{13.2}$$

with typical values of the order of 10–15 %.

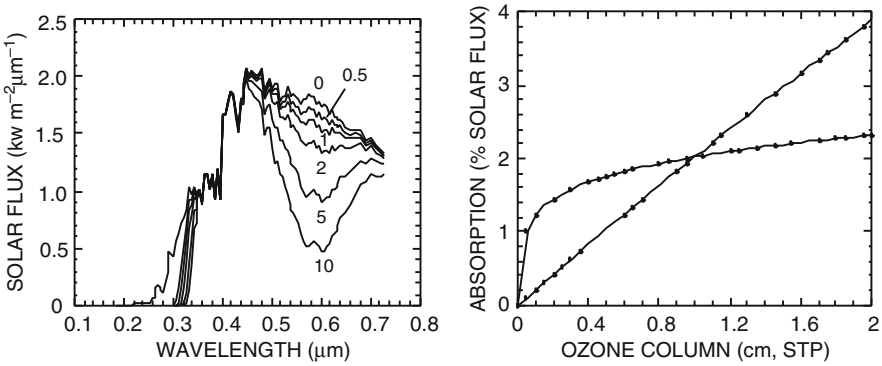



**Fig. 13.2** The ozone absorption as a function of the column density of the same gas. At the *left*, the solar spectrum (in kW m$^{-2}$ s$^{-1}$μm$^{-1}$) is shown after passing through increasing amounts of gas, indicated in cm SPT. At the *right*, the percentage absorption of the solar flux is shown (Lacis and Hansen 1974)

We will find the explanation for these relations only later. For the time being, we will act like those strategists who isolate pockets of strenuous resistance, having in mind only the final target which in our case is to learn how the atmosphere heats up when absorbing the solar radiation.

### *13.1.1 The Ozone Absorption*

With the method introduced in Chap. 2, we can divide the spectral region in two parts (ultraviolet and visible) and for each of them establish an empirical formula that gives the absorption as a function of the ozone total amount.

As shown in Fig. 13.2, this method is easily justified because the two spectral regions are very well separated and the modalities of absorption are rather different. In this figure, the ozone total content is expressed in cm at standard temperature and pressure conditions (STP) and this requires a little explanation. As already said for the columnar density, the units are molecules cm$^{-2}$. The unit of cm is easily connected to that because it represents the height of the gas column at STP. Knowing that a mole of gas occupies 22.42 l, it is easy to find that 1 cm is equivalent to a number of molecules given by

$$1\ \text{cm} = 6.023 \times 10^{23}/22.42 \times 10^{3} = 2.687 \times 10^{19}\text{molecules} \tag{13.3}$$

The same unit multiplied by 1000 has the name of *Dobson unit* (DU). In the case of the Earth's atmosphere, the total ozone amount is around 0.3–0.4 cm, that is, 300–400 DU. The amounts shown in the figure are justified because the ozone traversed by the solar radiation depends on the sun's zenith angle. At sunrise and sunset, the solar radiation may easily go through 10 cm SPT of ozone, not only in

the UV but also in the visible in the spectral region known as the Chappuis band. This considerable absorption takes place in the yellow region of the spectrum and explains the red color of the sky at sunset and sunrise. On the right of the same figure, the ozone absorption (expressed as percentage) is shown in the two different spectral regions as a function of the total ozone amount. The absorption in the visible is almost linear because the cross sections are small, while in the ultraviolet there are clear signs of saturation.

Approximating these curves with analytical functions can parameterize the absorption. In this way, it is possible to avoid the long calculation over the cross sections. A. Lacis and J. Hansen, in a rather famous paper of 1974, gave the following parameterization for the visible:

$$A_{oz}^{vis}(x) = \frac{0.02118x}{1 + 0.042x + 0.000323x^2} \tag{13.4}$$

where $A_{oz}^{vis}(x)$ is the fraction of the incident solar flux absorbed by an amount $x$ of ozone given in cm. This relation is valid for an ozone amount that goes from $10^{-4}$ up to 10 cm of ozone. For the ultraviolet, the upper limit is 1 cm and the absorption is given by

$$A_{oz}^{uv}(x) = \frac{1.082x}{(1 + 138.6x)^{0.805}} + \frac{0.0658x}{1 + (103.6x)^3} \tag{13.5}$$

The total absorption is given by the sum of Eqs. (13.4) and (13.5). As shown in Fig. 13.2, the absorption is both of the direct and the diffuse radiation. The absorption path in the two cases is different because for the direct beam, it is enough to divide the geometrical path by the cosine of the zenith angle, while in the case of the diffuse radiation, both refraction and diffusion tend to increase the geometrical path. If we imagine dividing the atmosphere in a number of layers and $u_l$ is the total ozone in the $l^{\text{th}}$ layer at the angle $\mu_0$, the effective ozone amount is given by

$$x_l = u_l M \tag{13.6}$$

where

$$M = 35/\left(1224\mu_0^2 + 1\right)^{1/2} \tag{13.7}$$

is an effective amplification factor that takes into account the effects mentioned above. The direct radiation goes straight to the ground where it is reflected and scattered back to the $l^{\text{th}}$ layer this time from below. The total amount of ozone traversed is then the direct and the diffuse path:

$$x_l^* = u_t M + \overline{M}\,(u_t - u_l) \tag{13.8}$$

where $u_t$ is the total ozone and $\overline{M}$ is an amplification factor taking into account the longer path for the upward diffuse radiation; we assume $\overline{M} \approx 1.9$. If we consider two adjacent layers $l$ and $l+1$, the fraction of the total flux absorbed between the two layers is given by the difference between the direct and diffuse absorption:

$$A_{l,oz} = \mu_0 \left\{ \left[ A_{0z}\left(x_{l+1}\right) - R\left(\mu_0\right) A_{0z}\left(x^*_{l+1}\right) \right] - \left[ A_{0z}\left(x_l\right) - R\left(\mu_0\right) A_{0z}\left(x^*_l\right) \right] \right\} \quad (13.9)$$

Here the reflectivity is given by Eq. (13.3). This equation gives the net flux absorbed between two contiguous layers separated by a height interval $\Delta z$ and a pressure difference $\Delta p$. The net heating is then given by

$$\frac{\Delta T}{\Delta t} = \frac{F_0 g A_l}{C_p \Delta p} \quad (13.10)$$

where $F_0$ is the solar flux at the top of the atmosphere.

### 13.1.2 The Water Vapor Absorption

The parameterization of the absorption by water vapor is much more complicated than that of ozone. The absorption coefficient for water vapor is highly dependent on frequency and values with enough resolution are not available. The absorption coefficient also depends on pressure. The water vapor absorption is important in the lower troposphere where diffusion and absorption have the same weight so that it is difficult to separate the two effects as in the case of ozone. We will deal here only with gaseous absorption that refers to clear sky conditions and will postpone the treatment of the cloudy sky case until after we introduce the diffusion in the clouds.

We use a similar approach for parameterization as in the case of ozone. The data we use are shown in Fig. 13.3. They show the absorptivity (what an ugly word!) as a function of a quantity that is related to the total vapor amount. The absorptivity in this case is the fraction of the total solar flux absorbed and can be parameterized in terms of the precipitable water. We can have a parameterization for each fitting formula, and two examples are those suggested by G. Yamamoto

$$A_{wv}\left(y\right) = 2.9y / \left[ \left(1 + 141.5y\right)^{0.635} + 5.925y \right] \quad (13.11)$$

and the older one suggested by F. Fowler

$$A_{wv}\left(y\right) = 0.0946 y^{0.303} \quad (13.12)$$

In both cases, $y$ indicates the *precipitable water*, defined as follows. We need to remember the definition of specific humidity given in Chap. 2, $w = \rho_w / \rho$, where $\rho_w$ is the water vapor density. Then, if we want to evaluate the total content of water vapor in a column (mass per unit surface), we calculate the integral

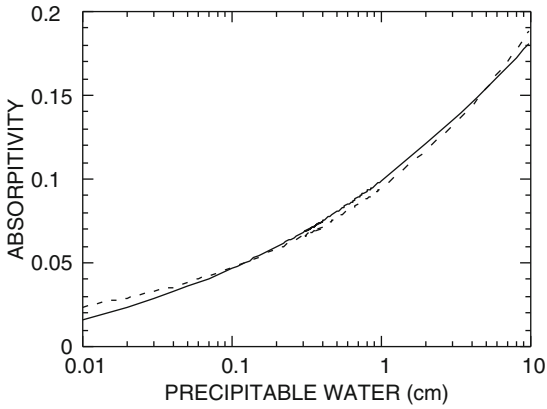



**Fig. 13.3** Absorptivity of water vapor as a function of the precipitable water (cm). The two curves refer to the different parameterizations given by Eqs. (13.11) and (13.12). *Dashed line* refers to the Yamamoto parameterization

$$y = \int_0^z \rho_w dz = \frac{1}{g}\int_p^{p_0} \frac{\rho_w}{\rho} dp = \frac{1}{g}\int_p^{p_0} w dp \tag{13.13}$$

where we have used the hydrostatic equilibrium approximation. The two curves corresponding to Eqs. (13.11) and (13.12) are presented in Fig. 13.3, and we can see there is a sensible difference between the two. Notice that the precipitable water is now expressed in cm, and this is because of the obvious fact that the 1 g cm$^{-2}$ corresponds to 1 cm of water. The relation of Eqs. (13.11) and (13.12) refers to standard conditions for pressure and temperature, so the extension to the different conditions found in the troposphere requires some correction or the introduction of an equivalent value for $y$ given by

$$y^{eff} = y\left(\frac{p}{p_0}\right)^n \left(\frac{T_0}{T}\right)^{\frac{1}{2}} \tag{13.14}$$

where $p_0$ and $T_0$ are reference values and $n$ is an exponent that may vary between 0 and 1. From the definition of Eq. (13.13), we can find the effective water vapor traversed by the solar radiation up to the $l^{\text{th}}$ layer:

$$y_l = \frac{M}{g}\int_0^{p_l} w\left(\frac{p}{p_0}\right)^n \left(\frac{T_0}{T}\right)^{1/2} dp \tag{13.15}$$

where $M$ is given by Eq. (13.7). For the radiation reflected from the surface, we use an expression similar to Eq. (13.8) for the effective water vapor quantity:

$$y_l^* = y_t + \frac{5}{3M}(y_t - y_l) \tag{13.16}$$

where we assume the amplification factor for the diffuse radiation to be 5/3. The fraction of the flux absorbed by the $l^{\text{th}}$ layer will be

$$A_{l,wv} = A_{wv}\left(y_{l+1}\right) - A_{wv}\left(y_l\right) + R_g\left[A_{wv}\left(y_l^*\right) - A_{wv}\left(y_{l+1}^*\right)\right] \tag{13.17}$$

It is interesting to note how in this relation the contribution of the diffuse radiation is simply multiplied by the reflectivity of the surface and not the combined reflectivity of the atmosphere and surface as in Eq. (13.9). This is because in the spectral region where the water vapor absorption is significant (i.e., in the near infrared), the effect of Rayleigh diffusion is negligible. It is instructive at this point to evaluate how much radiation reaches the ground. If we assume that the atmospheric absorption is due only to ozone and water vapor, then we can divide all the energy flux that reaches the ground in two regions, one containing 64.7 % of the radiation (the visible and the ultraviolet) and the remaining the region significant to the water vapor absorption. We will start from the last one because we can neglect diffusion. We have then

$$A_{g,wv} = \mu_0\left[0.353 - A_{wv}\left(y_t\right)\right]\left(1 - R_g\right) \tag{13.18}$$

where $y_t$ is the total content. In practice, from all the available flux, we subtract the portion absorbed along the atmosphere and that reflected by the surface.

For the region, where the ozone absorption is important, we have a more complicated relation:

$$A_{g,oz} = \mu_0\left[0.647 - R_r\left(\mu_0\right) - A_{oz}\left(Mu_t\right)\right]\left(1 - R_g\right) / \left(1 - R_a^* R_g\right) \tag{13.19}$$

In this case, we subtract the ozone absorption, the part reflected on the whole atmosphere, while the factor in the denominator takes into account all the multiple reflections between the ground and the atmosphere. As a consequence, $R_r(\mu_0)$ in this case is given by

$$R_r\left(\mu_0\right) = 0.28/\left(1 + 6.43\mu_0\right) \tag{13.20}$$

with an average value around 6 %. Referring to Figs. 13.3 and 13.4, we see that for a value of $y \sim 2$, we have an absorptivity of $\approx 0.12$ so that water vapor is actually responsible for most of the solar radiation absorption.

This absorption produces, as shown in Fig. 13.4, a diurnally averaged heating rate of about 0.5 °C day$^{-1}$. This calculation refers to middle altitude average conditions, while the maximum heating rate for $H_2O$ may even reach 1 °C day$^{-1}$.

We are now in a condition to evaluate the heating rate in the entire atmospheres However, although we have taken into account molecular scattering, we still have neglected the effects of aerosols (or more generally atmospheric particulates) on the solar radiation. We need to take these into account now.

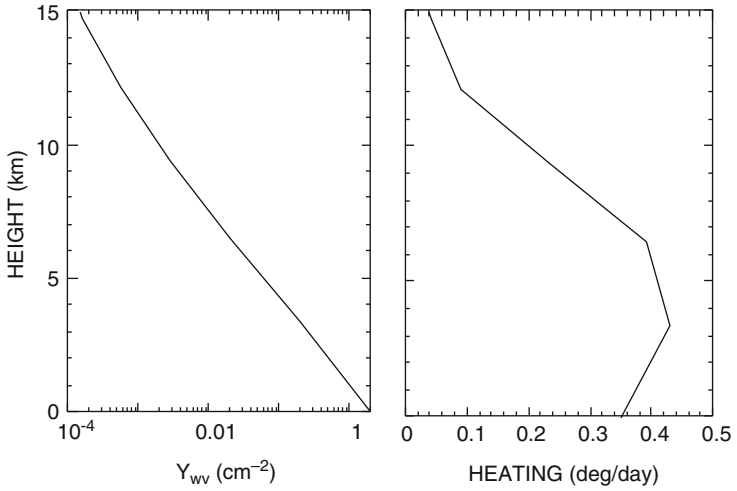



**Fig. 13.4** The equivalent water vapor amount (g $cm^{-2}$) above a certain altitude is shown at the *left*. This refers to the average conditions at midlatitude. At the *right*, we represent the corresponding diurnal averaged heating rate

## 13.2 The Interaction of Solar Radiation with Particulates in the Atmosphere

Since Chap. 3, we have considered the absorption by molecules and atoms. We dedicated the entire Chap. 9 to the atmospheric particulates and know that solar radiation interacts strongly with the cloud particles, for example. Anybody that has taken a decent general physics course must remember that a beautiful example of such an interaction is the rainbow which can be treated either with the geometrical or wave optics approximation.

In this case, the raindrop dimensions are large with respect to wavelength. When the particle dimensions become comparable to those of the wavelength, then it is necessary to recur to another classical treatment of radiation scattering. This one was introduced by a German physicist, Gustav Mie, who wrote, among other things, a seminal book on this topic. Curiously, this is another example of a problem that was well ahead of the possibility to solve it. As a matter of fact, computers fast enough to make all the calculations were made available only at the time of Mie's death (in 1957). There are different "regimes" where different approximations for treating the scattering of radiation must be applied, and these are illustrated in Fig. 13.5, inspired by the Wallace and Hobbs' textbook. The quantity that separates these regimes is related to the ratio between the wavelength and the radius and is called the size parameter $\chi$:

$$\chi = 2\pi r/\lambda \tag{13.21}$$

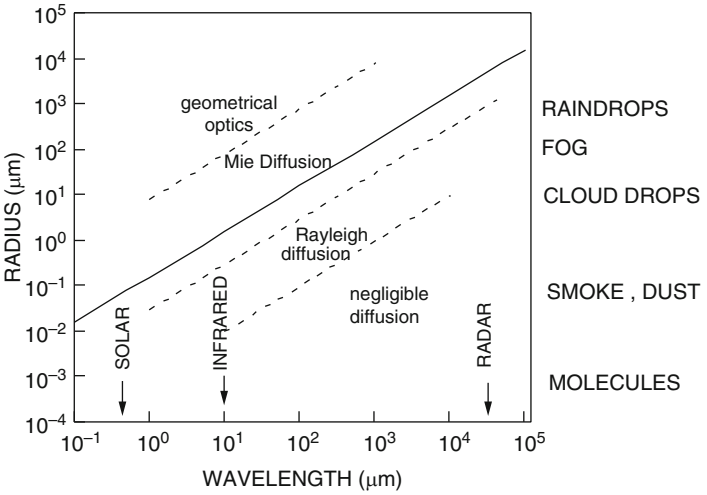



**Fig. 13.5** The different regimes applicable to radiation scattering fixed by the ratio between particle dimensions and wavelength. As a reference, typical wavelengths (solar, IR, etc.) are indicated; particle dimensions are to the *right* (from Wallace and Hobbs 1977)

where the meaning of the symbols is clear. The Mie scattering is predominant when this ratio is roughly one. This means, for example, that smoke or dust particles with the radius of about 0.1 $\mu$m must be treated within the Mie theory, while for very large values of $\chi$ ($\sim$50), the geometrical optics approximation becomes important. Figure 13.5 shows that raindrops can be treated in the geometrical optics approximation, while for the IR even relatively small particles (1 $\mu$m) may fall in the Mie regime. These dimensions are typical of the volcanic stratospheric particles and of the now famous polar stratospheric particles. Diffusion becomes negligible when size parameters are very small so that in this case neither Rayleigh nor Mie scattering is important nor everything reduces to pure absorption.

The fact that the interaction of radiation depends on the particle radius may complicate many calculations, considering, for example, that, within a cloud, particles have different radii. We have treated extensively the subject of size distribution in Chap. 9, and we will assume that everyone has read and learned something from that chapter.

### *13.2.1 Optical Properties of the Particles*

We cannot undertake a complete theory of radiation scattering from small particles, so we will assume that the reader already knows a few things and others will be justified on the basis of intuition. We need to clarify in any case the problem on which we are supposed to be working by referring to Fig. 13.6.

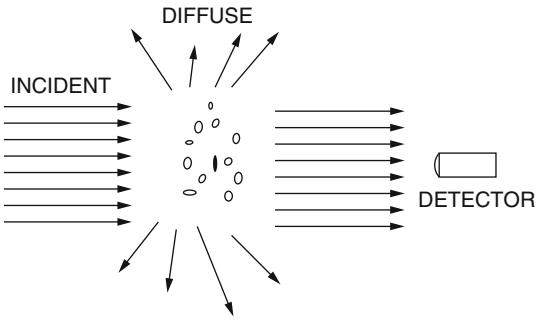



**Fig. 13.6** The diffusion and absorption processes for a beam of light incident on a cloud of particles

We have a parallel beam of light of intensity $I_i$ falling on a medium that contains a number of particles that scatter and absorb radiation. Even if absorption is absent (like in the case of Rayleigh scattering), the detector will measure an energy or power that is less than the incident radiation. The incident power in this case would be redistributed in space and the detector will capture only part of it. The power absorbed and scattered is, respectively,

$$\begin{aligned} W_{\text{abs}} &= I_i C_{\text{abs}} \\ W_{\text{sca}} &= I_i C_{\text{sca}} \end{aligned} \tag{13.22}$$

where $C_{\text{abs}}$ and $C_{\text{sca}}$ are the cross sections for absorption and scattering that have the dimensions of an area and are usually normalized to the geometrical cross section, so that we can define an *absorption* ($Q_{\text{abs}}$) and *scattering* ($Q_{\text{sca}}$) *efficiency*:

$$\begin{aligned} Q_{\text{abs}} &= C_{\text{abs}}/\pi r^2 \\ Q_{\text{sca}} &= C_{\text{sca}}/\pi r^2 \end{aligned} \tag{13.23}$$

These quantities depend on the size parameter defined by Eq. (13.21) and the refractive index that may be complex $m = n + ik$. The sum of the two efficiencies is called the *extinction efficiency* $Q_{\text{ext}} = Q_{\text{abs}} + Q_{\text{sca}}$. Figure 13.7 shows an example of extinction efficiency for a raindrop as a function of $\chi$.

Although we still do not know how to calculate this efficiency, a first question to ask is whether changing $\chi$ is equivalent to changing the wavelength or radius. Figure 13.7 is obtained by keeping the wavelength constant, and this is because in this way we use always the same refractive index and change only the radius. When the wavelength is changed keeping the radius constant, then at some point the particle would become completely opaque at the radiation, and $Q_{\text{ext}}$ would remain fixed at 2. We can see that this limit in Fig. 13.7 is reached asymptotically. This looks like a paradox because the cross section tends to be twice the geometrical cross section. However, we will see later how this fact could be explained on the basis of wave optics.

A good starting point is to consider the interaction of an electromagnetic wave with a dielectric sphere with a relative dielectric constant $\varepsilon_{\text{r}}$. The sphere is immersed

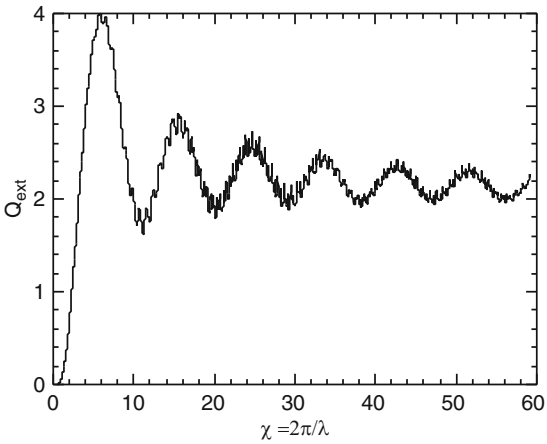



**Fig. 13.7** The extinction efficiency as a function of the size parameter for a raindrop

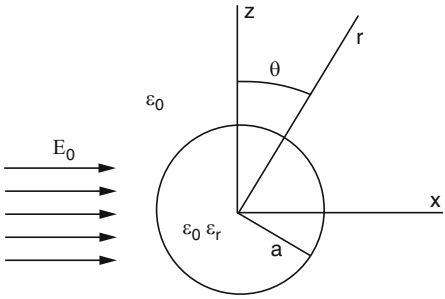



**Fig. 13.8** A dielectric sphere of radius $a$ in the presence of a polarized electromagnetic wave traveling in the $z$ direction

in a uniform electric field of intensity $E_0$, as shown in Fig. 13.8. This is a rather classical problem in electrostatics, and it can be shown that the resulting external field is given by the sum of the preexisting field plus a dipole field oriented in the same direction as the field $E_0$. The intensity of the dipole moment is

$$\mathbf{p} = \varepsilon_0 \alpha \mathbf{E}_0 \tag{13.24}$$

where the polarizability $\alpha$ is given by

$$\alpha = 4\pi a^3 \left(\varepsilon_r - 1\right) / \left(\varepsilon_r + 1\right) \tag{13.25}$$

The dipole is oriented along $x$ and the angle shown in the figure is complementary to the one used normally for the dipole. The situation shown corresponds to an incident wave that travels in the $z$ direction polarized in the $x$ direction, that is, $\mathbf{i}E_0$ exp $i(kz - \omega t)$. The dipole will oscillate with the same frequency of the incident field, that is,

$$\mathbf{p} = \mathbf{i}\varepsilon_0 \alpha E_0 \ \exp\left(-i\omega t\right),$$

and will emit electromagnetic radiation according to

$$W_{\text{sca}}(t) = \frac{p^2\omega^4}{6\pi\varepsilon_0 c^3} e^{2i\omega t} = \frac{p^2 k^3 \omega}{6\pi\varepsilon_0} e^{2i\omega t} \tag{13.26}$$

Keeping in mind the value for $p$ and that of the Poynting vector $S = c\varepsilon_0 {E_0}^2$,

$$W_{\text{sca}}(t) = k^4 S|\alpha|^2/6\pi \tag{13.27}$$

And according to the definition for the scattering cross section, we have

$$C_{\text{sca}} = k^4|\alpha|^2/6\pi \tag{13.28}$$

Until now, the approximation for small spheres, or that comparable with the wavelength, has been used implicitly when we have assumed the sphere to be uniformly polarized. In practice, Eq. (13.33) is similar to Eq. (3.21), obtained with the same method for the Rayleigh scattering. In Eq. (13.33), the polarizability appears squared because it could be a complex quantity, which means we could also have some absorption effect. However, to obtain the value for the absorption, we need to recur to a more sophisticated procedure based on what happens to the scattered component of the electric field, as shown in Fig. 13.9. The electric field vector is decomposed in two components:

$$E^i_{//} \quad E^i_{\perp}; \tag{13.29}$$

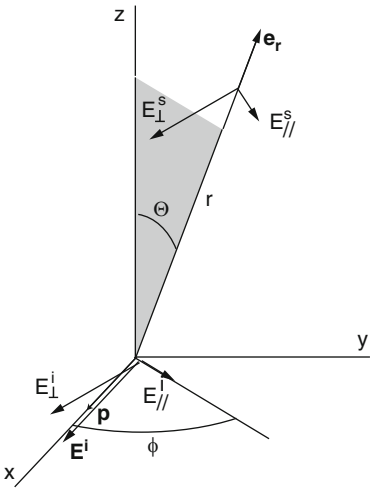



**Fig. 13.9** The scattering geometry with the polarized wave traveling along $z$. The scattering plane is *shadowed*

parallel and perpendicular to the scattering plane. This is defined as the plane which contains the propagation and the $r$ directions. At large distances ($kr >> 1$) from the origin, where the sphere is located, the electric field is given by

$$\mathbf{E}_s = -\frac{k^2}{4\pi\varepsilon_0}\frac{e^{ikr}}{r}\mathbf{e_r} \times (\mathbf{e_r} \times \mathbf{p}) \tag{13.30}$$

where $\mathbf{e}_\mathrm{r}$ is the unit vector in the $r$ direction. From Eq. (13.30), we can find the single components parallel and normal to the scattering plane:

$$E^s_{//} = -\frac{e^{-ikr+ikz}}{4\pi r}k^2\alpha\ \cos\theta; \qquad E^s_{\perp} = -\frac{e^{-ikr+ikz}}{4\pi r}k^2\alpha \tag{13.31}$$

while the component of the incident field can be written as

$$E^i_{//} = E_0\ \cos\varphi\ e^{-ikz}; \qquad E^i_{//} = E_0\ \sin\varphi\ e^{-ikz} \tag{13.32}$$

The relation between the two fields is written in more synthetic form:

$$\begin{bmatrix} E^s_{//} \\ E^s_{\perp} \end{bmatrix} = \frac{e^{-ikr+ikz}}{ikr}\begin{bmatrix} S_2(\theta) & 0 \\ 0 & S_1(\theta) \end{bmatrix}\begin{bmatrix} E^i_{//} \\ E^i_{\perp} \end{bmatrix} \tag{13.33}$$

where

$$S_1(\theta) = -\frac{ik^3\alpha}{4\pi}; \qquad S_2(\theta) = -\frac{ik^3\alpha}{4\pi}\ \cos\theta \tag{13.34}$$

Equation (13.37) is quite important for the scattering theory. The particular diagonal form of the matrix is valid only in the case of the homogeneous sphere, and the two functions $S_1$ and $S_2$ give simply the angular dependence of the scattered field with respect to the incident one. In order to find the cross section, we consider the $z$ direction where the field is given by the superposition of the incident and the scattered fields. To simplify further the situation, we assume that the incident field is polarized in the normal direction and use the far field approximation ($z >> x, y$). We have then

$$r = \left(x^2 + y^2 + z^2\right)^{1/2} \approx z + \left(x^2 + y^2\right)/2z$$

and also

$$E^s_{\perp} = \frac{e^{-ik(r-z)}}{ikr}S_1(\theta)\,E^i_{\perp} \tag{13.35}$$

so that in the direction $z$, $\theta \sim 0$ the resultant field

$$E_{\perp}^{i} + E_{\perp}^{s} \approx E_{\perp}^{i} \left\{ 1 + \frac{S_1(0)}{ikz} e^{-ik\left(x^2+y^2\right)/2z} \right\} \tag{13.36}$$

The square of this quantity is proportional to the Poynting vector so that retaining terms of the order of $1/z$

$$\left|E_{\perp}^{i} + E_{\perp}^{s}\right|^2 \approx \left|E_{\perp}^{i}\right|^2 \left\{ 1 + \frac{2}{kz} \mathrm{Re}\left[ \frac{S_1(0)}{i} e^{-ik\left(x^2+y^2\right)/2z} \right] \right\} \tag{13.37}$$

To evaluate the absorbed power and then $C_{\mathrm{ext}}$, we normalize Eq. (13.37) to the incident amplitude and integrate over a sphere of cross section $\pi a^2$ to get

$$\frac{1}{\left|E_{\perp}^{i}\right|^2} \int \left|E_{\perp}^{i} + E_{\perp}^{s}\right|^2 dxdy = \pi a^2 + C_{\mathrm{ext}} \tag{13.38}$$

The integral appearing in Eq. (13.38) is explained, for example, in Feynman's

$$\int e^{-ik\left(x^2+y^2\right)/2z} dxdy = 2\pi z/ik$$

with the result

$$C_{\mathrm{ext}} = \left(4\pi/k^2\right) \mathrm{Re}\left[S_1(0)\right] = k\mathrm{Im}\left[\alpha\right] \tag{13.39}$$

We have found a relation between the extinction and the polarizability so that if this does not have complex components, there is no absorption either. The theory we have illustrated is however valid only in the case when the diffusion from the particles is negligible and then the absorption coincides with the extinction. We are now in the condition to explain the paradox shown in Fig. 13.7 which for large dimensions with respect to the wavelength, $C_{\mathrm{ext}}$ becomes twice the geometrical cross section. We consider the dielectric sphere as in Fig. 13.10 with refractive index $n$. The optical path difference between a ray passing at distance $a\cos\psi$ from

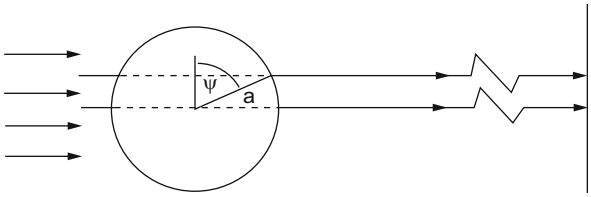



**Fig. 13.10** The geometry of anomalous diffraction

the center and one that passes outside is given by $2a(n-1)\sin\psi$ and the phase difference $4\pi a(n-1)\sin\psi/\lambda$. The amplitude of the wave on the screen on the right side of the figure at a large distance is

$$E_q = E_0 e^{-i\rho \sin\psi} \tag{13.40}$$

where we have put $\rho = 4\pi(\mathrm{n}-1)/\lambda$. The amplitude of the diffuse ray will be given by the difference between $E_0$ and $E_\mathrm{q}$ so that

$$E_{sca} = E_0\left(1 - e^{-i\rho \sin\psi}\right) \tag{13.41}$$

Operating in the same way as in Eq. (13.37), we have

$$S(0) = \frac{k^2}{2\pi}\int \left(1 - e^{-i\rho \sin\psi}\right) dxdz \tag{13.42}$$

The elementary area on which to perform the integral is given by the circular shadow, that is, $2\pi a^2 \cos\psi \sin\psi \, d\psi$, and the integral is then

$$S(0) = k^2a^2\int_0^{\pi/2} \left(1 - e^{-i\rho\sin\psi}\right) \cos\psi \sin\psi \, d\psi = k^2a^2K(i\rho) \tag{13.43}$$

$$K(i\rho) = \frac{1}{2} + \frac{e^{-i\rho}}{i\rho} - \frac{e^{-i\rho}-1}{\rho^2} \tag{13.44}$$

with the final result

$$C_{\mathrm{ext}} = \frac{4\pi a^2}{x^2}\mathrm{Re}\left[S(0)\right] = 4\pi a^2 \mathrm{Re}\left[K(i\rho)\right] \tag{13.45}$$

This result attributes the extinction mainly to a diffraction phenomenon and, as shown in Fig. 13.11, is not a bad approximation when compared with the exact calculation already shown in Fig. 13.7. The case is again a water sphere with a real refractive index of 1.33. We can see that both the phase and the amplitude are well reproduced except for small values of the size parameter where the error is not negligible. Notice that the scale is not exactly the same as the one used in Fig. 13.7 because it is also a function of $\rho$.

The paradox can now be explained in the following way: the geometric cross section removes part of the light beam and a similar quantity is removed by the diffraction that fills out the "void" left by the geometrical shadow of the sphere. This cannot be applied to the shadow of your geranium pot on the window (as van de Hulst explained) because the angles must be small. So the most obvious application is when you look at asteroids through a telescope or, most simply, when you see dust grains illuminated by the sun.

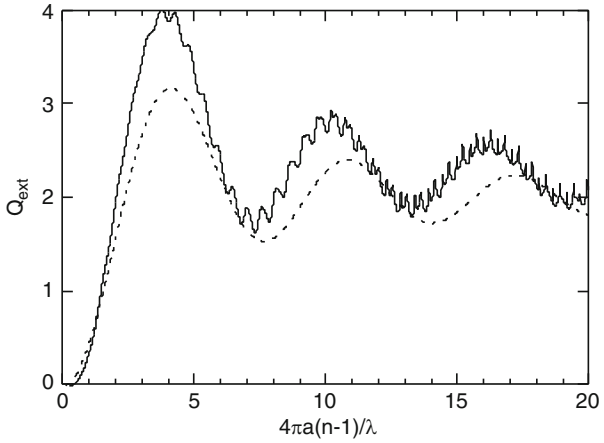



**Fig. 13.11** The cross section for extinction as a function of a quantity proportional to the size parameter. The *solid line* is the same shown in Fig. 13.7, while the *dashed line* is the approximation of Eq. (13.45)

### *13.2.2 Phase Functions and Mie Scattering*

We have introduced the phase function when we dealt with the Rayleigh scattering, and on that occasion, we saw that it was rather isotropic. If we consider a large number of molecules, the isotropy increases because the scattering among molecules is incoherent.

The forward scattering direction is rather special because, as shown in Fig. 13.9, if we indicate with $\mathbf{k}$ and $\mathbf{e}_r$ the unit vectors for the incident and scattering directions, respectively, while $\mathbf{l}$ is the separation between the two scattering elements, then the forward phase difference is given by

$$\Delta\varphi = 2\pi \left[\mathbf{l} \cdot (\mathbf{k} - \mathbf{e_r})\right] / \lambda$$

In this way, when the incident and the scattering directions coincide, we have a reinforcement of the intensity. This is quite evident in Fig. 13.12, where the phase functions are shown together with the distribution of the diffuse radiation on the scattering plane (and sometimes on the one normal to it) for water spheres at a wavelength of 0.55 $\mu$m.

This figure shows a few interesting things. For small particles ($\chi \sim 1$), the radiation is scattered almost isotropically around the axis normal to the direction of incidence so that it looks like Rayleigh scattering. As the dimension of the particle increases, the forward scattering becomes increasingly important, and for large particles, the forward-scattered radiation is about a thousand times the radiation scattered backward. This effect seems to contradict what you see (or do not see) when driving in the rain or fog. When you drive in the fog, knowing that most of

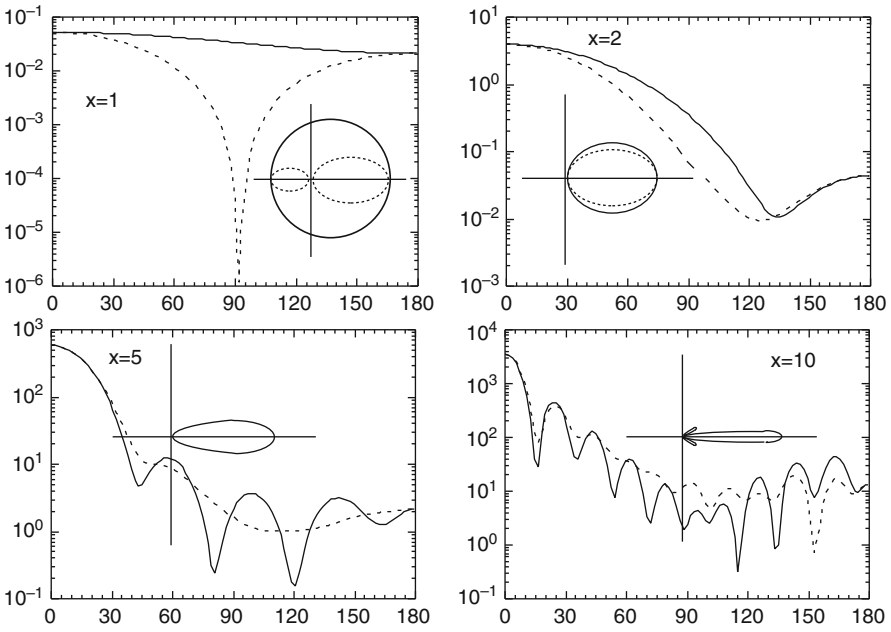



**Fig. 13.12** The phase function vs. the scattering angle and the angular distribution of the radiation (*inserts*). The *solid line* is the phase function for radiation polarized normally to the scattering plane, while the *dashed line* refers to parallel polarization. For each figure, the size parameter is also shown

the radiation is scattered forward does not help you to find the way. On the contrary, even the small amount that is scattered backward gives you a lot of trouble.

## 13.3 Radiative Transfer in the Presence of Scattering

When scattering is present, the equation for radiative transfer gets more complicated and, among other things, must be considered in three dimensions. We refer to Fig. 13.13 and consider that the radiation beam arrives from some direction $\mathbf{\Omega}_0$ which is characterized by the azimuth angle $\Phi_0$ and zenith angle $\pi + \Theta_0$, where the factor $\pi$ is due to the downward direction of the beam. Such radiation interacts with an atmospheric portion of thickness $dz$ and follows different paths. If we consider just the direction $\mathbf{\Omega}$, we can see that the radiation is simply attenuated:

$$dI\,(z, \mathbf{\Omega}) = C_{\text{ext}} N I\,(z, \mathbf{\Omega})\, dz/\cos\theta \tag{13.46}$$

where $N$ is the number of particles per unit volume. The quantity $C_{\text{ext}}$ that appears in this relation must be intended as an average over the different sizes of the particles.

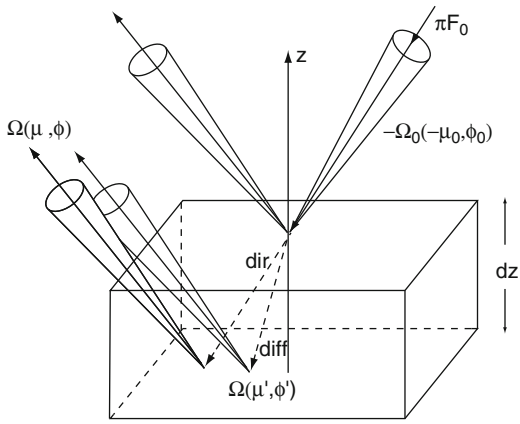



**Fig. 13.13** Radiative transfer in an infinitesimal part of the atmosphere. The $z$ axis is the local vertical

In the same direction, we will find the contribution of the diffuse radiation in the direction $\mathbf{\Omega}'$ which following another scattering event has been diffused in the direction $\mathbf{\Omega}$. In order to consider this effect, we must carry out an integration on all the events of this sort, that is, $\mathbf{\Omega}'$ is completely generic. We obtain

$$dI\left(z,\mathbf{\Omega}\right)=C_{\text{sca}}Ndz/\cos\theta\int_{4\pi}I\left(z\mathbf{\Omega}'\right)\frac{P\left(\mathbf{\Omega},\mathbf{\Omega}'\right)}{4\pi}d\mathbf{\Omega}' \tag{13.47}$$

Notice that $P(\mathbf{\Omega},\ \mathbf{\Omega}')$ is the phase function that establishes which portion of the intensity $I(z,\ \mathbf{\Omega})$ goes in the direction $\mathbf{\Omega}$. Finally, another contribution may arise from the radiation attenuated down to the altitude $z$ and diffused in the direction $\mathbf{\Omega}$:

$$dI\left(z,\mathbf{\Omega}\right)=\frac{C_{\text{sca}}Ndz}{\cos\theta}F(z)\frac{P\left(\mathbf{\Omega}-\mathbf{\Omega}_0\right)}{4\pi} \tag{13.48}$$

where the flux at altitude $z$ is given by

$$F(z)=\pi F_0\exp\left\{-\frac{1}{\cos\theta_0}\int_z^{\infty}C_{ext}N\left(z'\right)dz'\right\} \tag{13.49}$$

and the sum of all contributions gives

$$\begin{aligned}-\mu\frac{dI\left(z,\mathbf{\Omega}\right)}{d\tau}&=I\left(z,\mathbf{\Omega}\right)-\frac{\omega}{4\pi}\int_{4\pi}I\left(z,\mathbf{\Omega}'\right)\frac{P\left(\mathbf{\Omega},\mathbf{\Omega}'\right)}{4\pi}d\mathbf{\Omega}'\\&-\tfrac{\omega}{4\pi}\pi F_0P\left(\mathbf{\Omega},\mathbf{\Omega}_0\right)e^{-\tau/\mu_0}\end{aligned} \tag{13.50}$$

where the quantity

$$\omega=C_{\text{sca}}/C_{\text{ext}} \tag{13.51}$$

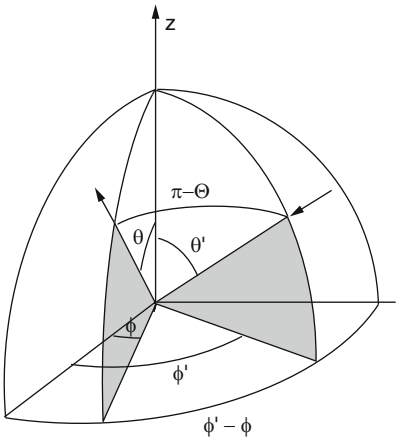



**Fig. 13.14** The relation between incident and diffuse radiation. The incidence and scattering planes are *shadowed* (Liou 1980)

is called the *single scattering albedo* and has the value of 1 for conservative scattering (i.e., no absorption). Also we have for the optical thickness

$$\tau(z) = \int_z^\infty C_{\text{ext}} N\left(z'\right) dz' \tag{13.52}$$

Equation (13.50) is not really easy to handle so we need to make it simpler. The first step in this direction can be made by referring to Fig. 13.14 to find a relation between the scattering angle and the azimuth and zenith angles. Using the rule of spherical geometry and referring to the triangle determined by the local zenith and the incident and exit directions, we have

$$\cos\Theta = \mu\mu' + \left(1-\mu^2\right)^{1/2}\left(1-\mu'^2\right)^{1/2}\cos\left(\varphi-\varphi'\right)$$

scattering. Substituting Eq. (13.55) in Eq. (13.50) and integrating over $\phi$, we obtain

$$\mu\frac{d\left(I_0+\mu I_1\right)}{d\tau} = -\left(I_0+\mu I_1\right)+\omega\left(I_0+\mu g I_1\right)+\frac{1}{4}\omega F_0 e^{-\tau/\mu_0}\left(1+3g\mu_0\mu\right) \tag{13.53}$$

With the same method used in Chap. 3, we can find the other two equations from Eq. (13.53), integrating first with respect to $\mu$ and second by multiplying through $\mu$ and integrating. In this way, we can separate the two quantities $I_0$ and $I_1$.

The phase function and the intensity of radiation are then functions of the zenith and azimuth angles. A first approximation to solve Eq. (13.50) is to assume that the intensity is only a function of the zenith angle. However, this time, unlike what we used in Chap. 3 (double stream), the series expansion stops at the first order in the zenith angle:

$$I(\tau,\mu) = I_0(\tau)+\mu I_1(\tau) \tag{13.54}$$

Similarly, the phase function can be developed in terms of Legendre polynomials:

$$P(\Theta) = 1 + \varpi(\tau)\cos\Theta \tag{13.55}$$

At this point, the integral that appears in Eq. (13.50) can be drastically simplified:

$$\int_0^{2\pi} d\varphi' \int_{-1}^{1} P(\mu, \varphi; \mu', \varphi')\, I(\mu', \varphi')\, d\mu' = 4\pi (I_0 + \mu g I_1) \tag{13.56}$$

where a new quantity appears known as *asymmetry factor* $g$:

$$g = \frac{\varpi}{3} = \frac{1}{2}\int_{-1}^{1} P(\Theta)\cos\Theta d(\cos\Theta) \tag{13.57}$$

This factor gives an idea of how symmetric the scattering is with respect to the scattering plane. It ranges then between 0 for Rayleigh scattering and 1 for total forward scattering:

$$\begin{aligned} \frac{dI_1}{d\tau} &= -3[1-\omega(\tau)]I_0 + \tfrac{3}{4}\omega(\tau)F_0 e^{-\tau/\mu_0} \\ \frac{dI_0}{d\tau} &= -[1-\omega(\tau)g(\tau)]I_1 + \tfrac{3}{4}\omega(\tau)g(\tau)\mu_0 F_0 e^{-\tau/\mu_0} \end{aligned} \tag{13.58a, b}$$

From these two equations, we can determine the intensities and then the fluxes:

$$F(\tau) = 2\pi \int_0^{\pm 1} \left(I_0 + \mu I_1\right)\mu d\mu = \pi\left[I_0(\tau) \pm \frac{2}{3} I_1(\tau)\right] \tag{13.59}$$

where $\mu > 0$ corresponds to $F^{\downarrow}(\tau)$ and $\mu < 0$ to $F^{\uparrow}(\tau)$. Again such fluxes are only for the diffuse radiation so that the total downward flux will be

$$S = \mu_0 \pi F_0 e^{-\tau/\mu_0} + F^{\downarrow} \tag{13.60}$$

The net flux at some height or at some optical thickness will be

$$F = \mu_0 \pi F_0 e^{-\tau/\mu_0} + F^{\downarrow} - F^{\uparrow} = \mu_0 \pi F_0 e^{-\tau/\mu_0} + 2H \tag{13.61}$$

where we have put $G = \pi I_0$ and $H = 2\pi I_1/3$. It is interesting to obtain from Eq. (13.60) a relation for the absorbed flux:

$$\frac{dF}{d\tau} = -\pi F_0 e^{-\tau/\mu_0} + 2\frac{dH}{d\tau} = -(1-\omega)\left(\pi F_0 e^{-\tau/\mu_0} + 4G\right) \tag{13.62}$$

from which we see that in the case $\omega = 1$ there is no energy absorption. As we will see, (13.62) is useful when we work with discrete layers because it is possible to avoid numerical derivatives which always have some degree of approximation.

The absorbed flux is proportional to the derivative of the flux with respect to height which is related to the optical thickness by

$$d\tau = -\rho \sum_i \chi_i K_i dz$$

where $\chi_i$ and $K_i$ are the mixing ratio and the absorption coefficient of the i$^{\text{h}}$ component. To obtain the derivative with respect to height, we need to multiply Eq. (13.62) by $d\tau/dz$.

Equation (13.58a, b) constitutes a first-order system that requires boundary conditions. The first is to have the diffuse flux at the top of the atmosphere to be zero:

$F^{\downarrow}(0) = 0$ at $\tau = 0$

The second condition has to do with the reflectivity of the surface to which we assign an albedo $A$. The upward diffuse flux at the surface is then given by the total downward flux there multiplied by the albedo. The two conditions translate as

$$\begin{aligned} I_0(0) + 2I_1(0)/3 &= 0 \\ I_0(\tau^*) - 2I_1(\tau^*)/3 &= A\left[I_0(\tau^*) + 2I_1(\tau^*)/3 + \mu_0 F_0 e^{-\tau^*/\mu_0}\right] \end{aligned} \tag{13.63}$$

where $\tau^*$ is the total thickness of the atmosphere.

The system of Eq. (13.58a, b) does not have a simple solution if the single scattering albedo and the asymmetry factor are a function of height. This is what happens in the real case so that the method to solve Eq. (13.58a, b) is to divide the atmosphere in $N$ homogeneous layers where $\omega$ and $g$ have constant values $\omega_{\text{i}}$ and $g_{\text{i}}$ for the $i$th layer. For each layer, we find the solutions

$$\begin{aligned} I_0(\tau) &= I_0^i(\tau) = C_1^i e^{-k_i\tau} + C_2^i e^{+k_i\tau} - \alpha_i e^{-\tau/\mu_0} \\ I_1(\tau) &= I_1^i(\tau) = P_i\left(C_1^i e^{-k_i\tau} - C_2^i e^{+k_i\tau}\right) - \beta_i e^{-\tau/\mu_0} \end{aligned} \tag{13.64}$$

that are valid for the interval $\tau_{i-1} < \tau < \tau_i$ and where the coefficients are

$$\begin{aligned} k_i &= [3(1-\omega_i)(1-\omega_i g_i)]^{1/2}; \quad P_i = [3(1-\omega_i)/(1-\omega_i g_i)]^{1/2} \\ \alpha_i &= 3\omega_i F_0 \mu_0^2 [1 + g_i(1-\omega_i)]/4\left(1 - k_i{}^2\mu_0^2\right) \\ \beta_i &= 3\omega_i F_0 \mu_0 \left[1 + 3g_i(1-\omega_i)\mu_0^2\right]/4\left(1 - k_i{}^2\mu_0^2\right) \end{aligned} \tag{13.65}$$

At this point, we have two unknowns for each layer ($C_1^i$, $C_2^i$). Imposing the continuity between adjacent layers, we find $2N-2$:

$$I_0^i(\tau_i) = I_0^{i+1}(\tau_i) \quad ; \quad I_1^i(\tau_i) = I_1^{i+1}(\tau_i), \quad i = 1, 2 \ldots, N-1 \tag{13.66}$$

The remaining two equations are provided by the conditions at the top and at the surface. This way to proceed is called the Eddington approximation, and there is

also a slight variant that takes into account a pronounced forward scattering. In this case, the phase function is approximated by a Dirac $\delta$:

$$\widetilde{P}\left(\mu,\mu'\right)=2f\delta\left(\mu-\mu'\right)+(1-f)\,P\left(\mu,\mu'\right) \tag{13.67}$$

It can be shown that this function enables us to obtain an equation similar to Eq. (13.57) where $\tau$, $\omega$, and $g$ are scaled according to

$$\omega=\left(1-g^{*2}\right)\omega^{*}/\left(1-\omega^{*}g^{*}\right)\quad;\quad g=g^{*}/\left(1+g^{*}\right)\quad;\quad\tau=\left(1-\omega^{*}g^{*2}\right)\tau^{*} \tag{13.68}$$

where $\omega^{*}$, $g^{*}$, and $\tau^{*}$ are the real values. We have also assumed $f=g^{2}$.

What we have sketched here is a rather simple and elegant theory that has the advantage of being easily adapted to computers because it reduces a continuum to a discrete system and then to a system of linear equations. It is worth then to look at some solutions just to appreciate its power. By the way, when the scaling is used, the approximation becomes the $\delta$-Eddington approximation.

### 13.3.1 Few Simple Applications of the δ-Eddington Approximation

At this point, we can finally justify a number of things that simply have been taken for granted starting from the reflectivity of the atmosphere. If we reduce the atmosphere to a single layer with optical thickness $\tau^{*}$, then its reflectivity is

$$R\left(\mu_0,\tau^{*}\right)=F^{\uparrow}(0)/\mu_0\pi F_0 \tag{13.69}$$

This calculation is particularly simple in the case of conservative scattering, that is, when $\omega=1$. In this case, Eq. (13.58a, b) simplifies further and the solution is

$$I_0\left(\tau\right)=B_1-\frac{3}{4}\mu_0{}^{2}F_0e^{-\tau/\mu_0}-B_2T\left(\tau\right)\quad;\quad I_1\left(\tau\right)=B_2-\frac{3}{4}\mu_0F_0e^{-\tau/\mu_0} \tag{13.70}$$

$$T\left(\tau\right)=\int_0^{\tau}\left[1-g\left(\tau'\right)\right]d\tau' \tag{13.71}$$

has the role of an equivalent optical thickness. The constants $B_1$ and $B_2$ can be easily found:

$$\begin{aligned}B_2&=3\mu_0F_0\left(1-A\right)\left[2+3\mu_0+\left(2-3\mu_0\right)e^{-\tau^{*}/\mu_0}\right]\left\{4\left[4+3\left(1-A\right)T\left(\tau^{*}\right)\right]\right\}^{-1}\\B_1&=\left(3\mu_0{}^{2}/4+\mu_0/2\right)F_0-2B_2/3\end{aligned} \tag{13.72}$$

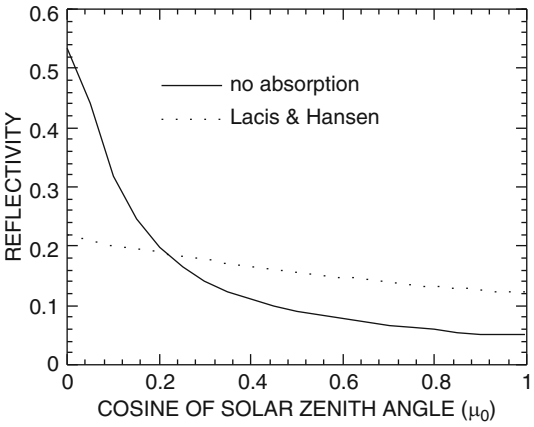



**Fig. 13.15** The reflectivity of a Rayleigh atmosphere (*solid line*) compared to the Lacis and Hansen (1974) approximation, Eq. (13.2)

And the reflectivity is given by

$$R\left(\mu_0, \tau^*\right) = 1 - (4B_2/3\mu_0 F_0) \tag{13.73}$$

In the case of a Rayleigh atmosphere, we have $g = 0$ so that the function in Eq. (13.71) coincides with the optical thickness. We can then build reflectivity curves like those shown in Fig. 13.15, where Eq. (13.73) is compared with the reflectivity given by Eq. (13.2), indicated as in Lacis and Hansen. We notice that the behavior with the zenith angle is quite different especially for large angles. The reason is that we have neglected the absorption. Equation (13.2) has been obtained by fitting a function to data that take into account the atmospheric absorption. This may explain why the reflectivity decreases for large solar zenith angles when the absorption is greater. Equation (13.73) can also be interpreted as the reflectivity of a cloud with conservative scattering (something not at all impossible). Then we can use larger values for optical thickness and not isotropic scattering. In this case, it is rather interesting to calculate the reflectivity as a function of $T$ rather than $\tau^*$. We obtain then what is shown in Fig. 13.16, where the reflectivity is plotted as a function of $T$. This means that reflectivity measurements do not give information about $\tau$ when we do not have information about $g$. Something we are learning is that, at least for conservative scattering, the larger the optical thickness, the larger the reflectivity. For example, if we refer to Fig. 13.16, we find that for $\tau^* = 16$ the reflectivity for $g = 0.7$ is about 73 %, which means a transmission of 27 %. (Keep in mind that for conservative scattering the sum of the reflectivity and the transmission must be one.) This is rather surprising because the direct transmission would be $\exp(-\tau/\mu_0) \approx 1.2 \cdot 10^{-7}$, so that the effect of the diffuse radiation is very important. This is the reason why we have enough light even when it is very cloudy. Thunderstorm clouds (for which the conservative scattering is not a good approximation) have a thickness of the order of 10 and so reduce the direct radiation practically to zero.

A further consideration can be made for the albedo in the presence of a cloudy sky. In this case, we see that for conservative scattering the reflectivity depends only

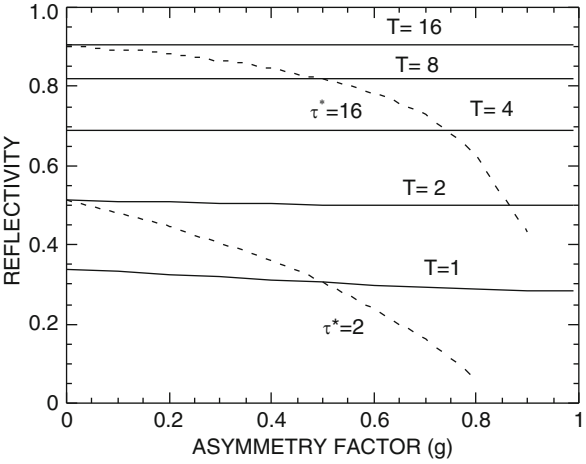



**Fig. 13.16** The reflectivity as a function of the asymmetry factor for different values of the equivalent optical thickness $T$. Also included are two curves for the real values of 2 and 16 (Shettle and Weinman 1970)

on the equivalent optical thickness. Curves shown in Fig. 13.16 can be fitted by a function like

$$aT\left(\tau^*\right) / \left[b + cT\left(\tau^*\right)\right]$$

The coefficients could be found, for example, by averaging over the zenith angle. However, in their paper, Lacis and Hansen have $a = c = 0.13$, $b = 1$.

We have frequently mentioned conservative scattering that does not correspond to practical cases. In general, nonconservative scattering is also important in the slicing of the atmosphere mentioned before. In this process, we can easily find layers that are heterogeneous in composition, containing absorbing gases and particles. Suppose, for example, that in a particular layer there are particulates with optical thickness $\Delta\tau_{\mathrm{p}}$ with albedo and asymmetry factor $\omega_{\mathrm{p}}$ and $g_{\mathrm{p}}$, respectively. The same layer may contain an absorbing gas of thickness $\Delta\tau_{\mathrm{a}}$, and everything will be in an atmosphere of optical thickness $\Delta\tau_R$ due to Rayleigh extinction. We can introduce equivalent values for the optical thickness, single scattering albedo, and asymmetry factor defined as follows:

$$\begin{aligned}
\Delta\tau &= \Delta\tau_p + \Delta\tau_a + \Delta\tau_R \\
\omega &= \left(\omega_p \Delta\tau_p + \Delta\tau_R\right) / \Delta\tau \\
g &= \omega_p \Delta\tau_p g / \Delta\tau
\end{aligned} \tag{13.74}$$

With these parameters, the procedure illustrated above can be repeated.

The influence of the nonconservative scattering is in any case clear because it tends to decrease the reflectivity with respect to the conservative case for the same value of the optical thickness. We now may understand why cloudy planets have such large albedo values, because their clouds are pretty good mirrors.

We are in good shape for what concerns solar radiation, but to go on, we must make some progress also on the IR side.

## 13.4 The Transfer of Infrared Radiation

Formally the computation with the IR radiation may look simpler. Among other things, wavelengths are such that in many case they are larger than the dimension of the particles and scattering can be neglected. The complication arises from the fact that the absorption and emission terms are due to a very large number of spectral lines. We are again facing the problem that, because our models have to treat many other processes, we cannot afford to reserve so much computer time only for the IR. A similar problem is also present for the solar UV, for so-called Schumann–Runge bands, although this is a quite limited problem with respect to the infrared.

In the following paragraphs, we will illustrate the formal solution to the problem of radiative transfer in the infrared together with very simple shortcuts. We will use a few spectroscopic notions just to show that in geophysics nothing is enough.

### *13.4.1 The Formal Solution*

The starting point could be Eq. (3.32), whose solution can be written formally:

$$I_\nu(\tau,\mu) = I_\nu(\tau_1,\mu)\, e^{-(\tau_1-\tau)/\mu} + \int_\tau^{\tau_1} B_\nu(\tau')\, e^{-(\tau'-\tau)/\mu} \frac{d\tau'}{\mu} \tag{13.75}$$

where the significance of the various terms is obvious. This same equation can be separated in two parts referring to the upper ($\mu > 1$) and bottom ($\mu < 1$) hemispheres, respectively (Fig. 13.17):

$$\begin{aligned} I_\nu(\tau,\mu) &= I_\nu(\tau_1,\mu)\, e^{-(\tau_1-\tau)/\mu} + \int_\tau^{\tau_1} \pi B_\nu(\tau')\, e^{-(\tau'-\tau)/\mu} \frac{d\tau'}{\mu} \\ I_\nu(\tau,-\mu) &= I_\nu(0,-\mu)\, e^{-\tau/\mu} + \int_0^{\tau_1} \pi B_\nu(\tau')\, e^{-(\tau-\tau')/\mu} \frac{d\tau'}{\mu} \end{aligned} \tag{13.76}$$

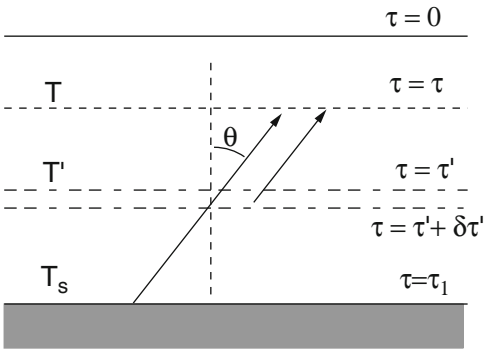



**Fig. 13.17** The scheme to calculate monochromatic intensity

The monochromatic Planck function that appears in these equations must be calculated for the temperature corresponding to the respective optical thickness. These equations have a very simple meaning. The first part of Eq. (13.76) simply says that the intensity of radiation at some level $\tau$ is given by the intensity at the surface (where optical thickness is $\tau_1$) attenuated for the absorption term and increased by the contribution of each layer of thickness $\tau'$ comprised between $\tau$ and $\tau_1$. This process is shown schematically in Fig. 13.18. An integration of Eq. (13.76) over the angle gives the corresponding monochromatic fluxes measured as power per unit surface. We have

$$F^{\uparrow}(\tau) = 2\pi\, B_{\nu}(T_s) \int_0^1 e^{-(\tau_1-\tau)/\mu}\mu\, d\mu + 2\int_0^1 d\mu \int_{\tau}^{\tau_1} \pi B_{\nu}(\tau')\, e^{-(\tau'-\tau)/\mu} d\tau'$$
$$F^{\downarrow}(\tau) = 2\int_0^1 d\mu \int_0^{\tau} \pi B_{\nu}(\tau') e^{-(\tau-\tau')/\mu} d\tau' \tag{13.77}$$

The simplification of the second part of Eq. (13.77) is due to the fact that at the top of the atmosphere ($t = 0$) we have $I_{\nu}(0, -\mu) = 0$ because above that level there are no sources of infrared radiation.

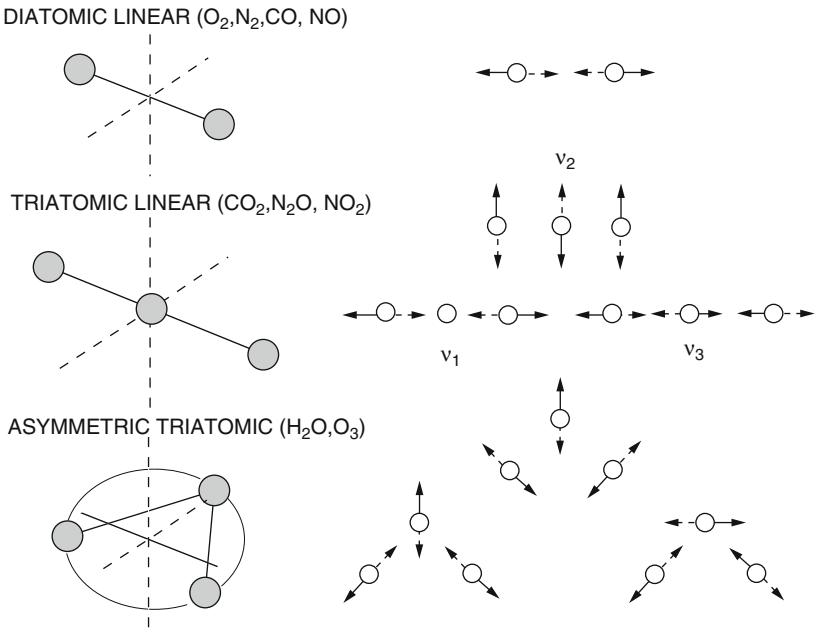



**Fig. 13.18** A qualitative scheme of the degree of freedom for the vibrational motion of a few molecules of atmospheric interest. At the *left*, the geometrical arrangement is shown, while at the *right* the vibrational motions

The formal solutions we have found are not very useful in practice because the absorption lines there are made up of thousands of lines that contribute to the optical thickness. This means that it is very time consuming to integrate line by line although every once in a while for reference this is done. There are other solutions that may approximate the effects of the line structure with appropriate averages. The most efficient of these methods is based on the parameterization of the absorption in terms of the total gas content. We need however to talk about the molecular spectra or more generally about infrared spectroscopy.

## 13.5 Molecular Spectra

The molecules we will deal with are mostly biatomic and linear like $O_2$ (oxygen), $N_2$ (nitrogen), $NO$ (nitrogen oxide), and $CO$ (carbon monoxide). In this case, the molecules are symmetric and have two principal momenta of inertia. The same is true for the linear triatomic molecules like carbon dioxide ($CO_2$), nitrous oxide ($N_2O$), and nitrogen dioxide ($NO_2$). Finally, we have asymmetric molecules like $H_2O$ and ozone ($O_3$) with three distinct principal momenta of inertia.

Some of these molecules have zero dipole moment (those with perfect symmetry), and this is important in the interaction with the external electromagnetic field.

The total energy of the molecules is given by the sum of the electronic state energy $E_e$, the vibrational energy $E_v$, the rotational energy $E_r$, and the one due to the translational motion:

$$E = E_e + E_v + E_r + E_t \tag{13.78}$$

Except for the last term, all the other energies are quantized thus giving rise to transitions and spectral lines that are classified based on their energy or frequency:

$$\begin{aligned}
&\text{Electronic transition} \approx 10.000 -> \text{cm}^{-1}\ (\text{E} > 1.24\ \text{eV})\ \ \text{visible, UV}\\
&\text{Vibrational transition} \approx 500 - 2000\ \text{cm}^{-1}\ (0.06 < \text{E} < 0.25\ \text{eV})\ \ \text{IR}\\
&\text{Rotational transition} \approx 1 - 500\ \text{cm}^{-1}\ (0.0001 < \text{E} < 0.06\ \text{eV})\ \ \text{microwave}
\end{aligned}$$

It is interesting to note that the energy range for the last two transitions is of the order of the thermal energy $kT \approx 0.025$ eV so that both the vibrational and rotational energy are compatible with the temperature that we find in the atmosphere.

Let us consider what happens for the carbon dioxide molecules. In Fig. 13.18, we can see that these molecules, beside the rotational motion around the axis, have also three vibration modes at the frequencies $\nu_2 = 667\ \text{cm}^{-1}$ ($15\,\mu$m) and $\nu_3 = 2349\ \text{cm}^{-1}$ ($4.3\,\mu$m). The symmetric vibration mode has a zero net dipole moment so that it does not interact, at least at the first order, with the external electromagnetic field. The vibrational energy in the different modes is quantized according to the relation

$$E_v = \sum_{i=1,3} (\upsilon_i + 1/2)\, h\nu_i \tag{13.79}$$

where $\upsilon_i$ and $\nu_i$ are the vibrational quantum number and frequency, respectively. For the rotational transition, the quantum number is given by

$$I\omega = \hbar\sqrt{J\left(J+1\right)} \tag{13.80}$$

and the energy of rotation

$$E_r = I\omega^2/2 = \hbar^2 J\left(J+1\right)/2I = hcBJ\left(J+1\right) \tag{13.81}$$

where $B = h/8\pi^2 cI$. The transition rules are $\Delta J = \pm 1$ so that for $\Delta J = 1$, the absorption lines are equally spaced at $2B$. In practice, the knowledge of the spectrum gives us the dependence of the absorption coefficient on frequency.

The absorption coefficient $k_\nu$ itself is determined by two factors, the line strength $S_\mathrm{i}$ and the shape as a function of frequency $f\left(\nu, \nu_i\right)$:

$$k_\nu = S_i f\left(\nu, \nu_i\right) \tag{13.82}$$

If the absorption cross section is $\sigma$, the line strength is given by

$$S_i = \left[\sigma_i\left(1 - e^{-h(\upsilon_u - \upsilon_l)/kT}\right)e^{-h\nu_l/kT}\right]/Z\left(T\right) \tag{13.83}$$

where $\nu_u$ and $\nu_\mathrm{l}$ are the frequencies of the upper and lower state, respectively, while

$$Z(T) = \sum_n e^{-E_n/kT} \tag{13.84}$$

is the factor that gives the Boltzmann distribution. This means that in writing Eq. (13.84), we have assumed the thermodynamic equilibrium.

### *13.5.1 Spectral Line Shape*

The theory of the spectral line shape is a quantum theory and not simple at all. We will try to follow a classical approach that assumes the molecule to be a harmonic oscillator of frequency $\omega_0$ that emits electromagnetic radiation. The same molecule collides with other molecules and between collisions travels in straight lines. In the absence of other processes besides the finite lifetime for emission, the shape of the line is given by (see again Feynman)

$$f\left(\nu - \nu_0\right) = \frac{1}{\pi}\frac{\alpha_N}{\left(\nu - \nu_0\right)^2 + {\alpha_N}^2} \tag{13.85}$$

where $\alpha_N = 1/4\pi\tau$ with $\tau$ as the lifetime of the transition. For a typical transition in the infrared, the lifetime is of the order of 0.1 s so that $\alpha_N \approx 1\ s^{-1}$ or in units of wave number $3 \cdot 10^{-11}\ cm^{-1}$ which is a very small width when compared with a typical IR wavelength.

We now take into account when we consider the Doppler shift due to the molecular motion. If the frequency at rest is $\nu_0$, for a molecule moving at speed $u$, the Doppler shift will be

$$\nu - \nu_0 = u\nu_0/c$$

and for a Maxwellian velocity distribution

$$\exp\left(-mu^2/kT\right)$$

the line width will be proportional to

$$\exp\left[-\frac{mc^2(\nu-\nu_0)^2}{2kT{\nu_0}^2}\right] = \exp\left[-(\nu-\nu_0)^2/\alpha_D^2\right] \tag{13.86}$$

where

$$\alpha_D^2 = 2kT{\nu_0}^2/mc^2 = v_0^2 v_{\text{rms}}^2/c^2$$

is the half width of the line and $v_{\text{rms}}^2$ is the mean square velocity. However, because it must hold the condition

$$\int_{-\infty}^{+\infty} f(\nu)\, d\nu = 1,$$

Eq. (13.86) can be normalized as

$$f(\nu-\nu_0) = \frac{1}{\sqrt{\pi}\alpha_D}\exp\left[(\nu-\nu_0)^2/\alpha_D^2\right] \tag{13.87}$$

It can be easily calculated that for carbon dioxide the line width at 250 K is $\alpha_D \approx 7 \cdot 10^{-4}\ cm^{-1}$, much higher than the natural lifetime.

The line shape can be also determined by the collision and in this case is called Lorentzian. Here we consider that the oscillator may change phase at each collision. We also assume that such collisions have a probability distribution of the type

$P(t) = \exp(-t/\tau)$ where $\tau$ is the interval between two collisions and is given by the mean free path divided by the mean square velocity. In practice, the incoherent sum of all the phase shifts implies that the time for the emission is just $\tau$ so that the line shape is similar to the natural one:

$$f(\nu-\nu_0) = \frac{1}{\pi}\frac{\alpha_L}{(\nu-\nu_0)+{\alpha_L}^2} \tag{13.88}$$

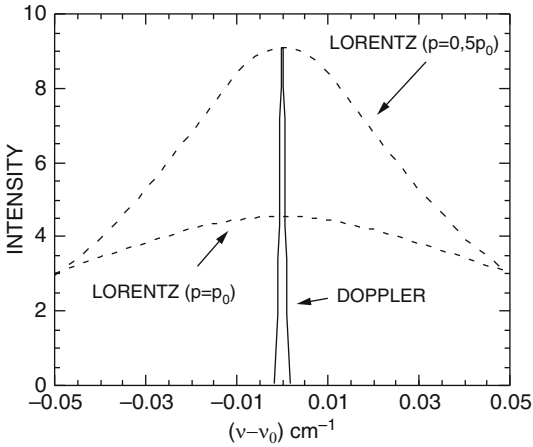



**Fig. 13.19** The comparison between a Doppler and a Lorentz line for different values of pressure

where $\alpha_L$ is now the Lorentzian half width given by

$$\alpha_L = p\frac{\sigma}{c}\left(\frac{1}{mkT}\right)^{1/2} \tag{13.89}$$

The half width depends now on pressure and temperature (and on the cross section $\sigma$). Again, for a gas like carbon dioxide, at a temperature of 300 K and at pressure of 100 hPa, $\alpha_L \approx 0.07$ cm$^{-1}$ which is about 100 times larger than the Doppler width. In Fig. 13.19, a comparison between the Lorentz and Doppler shapes is shown for a temperature of 300 K, while the two Lorentz lines refer to two different pressure values. We notice that the pressure broadening is much larger than that of the Doppler. However, the former scales linearly with pressure, and so we may argue that the Doppler broadening becomes important for pressure lower than 10 hPa which is for altitudes higher than 30 km.

## 13.6 Models for the Line Absorption

As shown by Eq. (13.79), what interests us most in the radiative transport of IR radiation is not the monochromatic absorption, but rather the integrated value over a certain frequency domain. If $u$ indicates the total amount of a gas along the absorption path, we can define the total absorption due to a single line as

$$A(u) = \int_{\nu_1}^{\nu_2}\left(1 - e^{-k_\nu u}\right)d\nu \tag{13.90}$$

This quantity is sometimes indicated through an equivalent width $W(u)$, meaning a rectangular absorption line with the absorption given by Eq. (13.90). We may have

a regime of *linear absorption* when the exponent in Eq. (13.90) is much less than 1. In this case, we have, keeping in mind Eq. (13.83),

$$A(u) = Su\int_{\nu_1}^{\nu_2} f(\nu)\, d\nu = Su \tag{13.91}$$

We may also have a *strong absorption* regime when, for example, for a Lorentzian line we have $|\nu - \nu_0| >> \alpha_L$, so that the absorption coefficient becomes

$$k_\nu \approx S\alpha_L/\pi(\nu - \nu_0)^2$$

and the absorption

$$A(u) = \int_{\nu_1}^{\nu_2} \left\{1 - \exp\left[-Su\alpha_L/\pi(\nu - \nu_0)^2\right]\right\} d\nu \tag{13.92}$$

This approximation is possible on every frequency range if the amount of absorbing gas makes the exponential small enough. The integral in Eq. (13.92) can be solved using the substitution:

$$y = Su\alpha_L/\pi(\nu - \nu_0)^2$$

so that

$$A(u) = 2(Su\alpha_L/\pi)^{1/2}\int_0^\infty y^{-1/2}e^{-y}dy = 2(Su\alpha_L/\pi)^{1/2}\Gamma(1/2) = 2(Su\alpha_L)^{1/2} \tag{13.93}$$

The conclusion is that in the linear regime, the absorption is independent from the line shape and is simply proportional to the amount of the absorbing gas. In the strong absorption regime, the absorption itself is proportional to the square root of the gas amount but depends on the line shape (i.e., the square root of its half width). The half width however depends on pressure so that the strong regime becomes less important at high altitudes.

At this point, the concept of equivalent width becomes useful. As shown in Fig. 13.20, we notice that the absorption given by Eq. (13.93) is equivalent to that of a rectangular line shape with a width given by Eq. (13.93). The same concept of equivalent can be introduced for the Doppler lines. It can be easily shown that, in the strong limit corresponding to the conditions,

$k_\nu u >> 1$ and $|\nu - \nu_0| \ll \alpha_D[\log(Su/\alpha_D)]^{1/2}$,

we obtain an equivalent width

$$W(u) \approx \alpha_D[\log(Su/\alpha_D)]^{1/2}$$

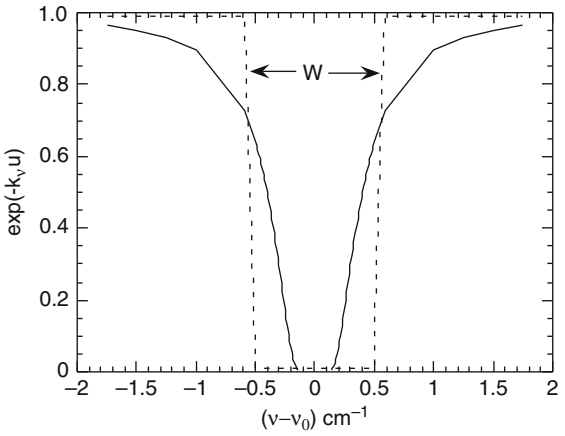



**Fig. 13.20** The equivalent width *W*. The absorption of the line is equivalent to that of a rectangular line of appropriate width

We have now clarified (at least we hope) many things about spectral lines in the infrared, while the problem of the line-by-line integration is still there. We will treat this problem directly by an application to the calculation of the infrared fluxes that are relevant to the climate.

### *13.6.1 A Formulation of the Infrared Flux*

A drastic simplification for the infrared flux was obtained by Robert Cess in 1971, and then implemented by V. Ramanathan, by assuming that the main contribution to the IR radiative flux is due to the vibrational and rotational bands of carbon dioxide, water vapor, and ozone. Also, this approach substituted the line-by-line calculation with a fitting of analytical functions to the total absorption of IR radiation by the different gases.

The starting point again is Eq. (13.78) which can be reformulated in the following way. First of all, we define a *transmissivity*

$$T\left(z, z'\right) = 2\int_0^1 e^{-(\tau'-\tau)/\mu}\mu \, d\mu \tag{13.94}$$

between altitudes $z$ e $z'$ so that the fluxes can be written as

$$\begin{aligned} F^{\uparrow}(z) &= \pi B_i(0) T_i\left(z, 0\right) + \int_0^z \pi B_i\left(z'\right) dT_i\left(z, z'\right) \\ F^{\downarrow}(z) &= \int_z^{\infty} \pi B_i\left(z'\right) dT_i\left(z, z'\right) \end{aligned} \tag{13.95}$$

where the optical thickness has been substituted with height and the integration has been performed on the frequency in such a way

$$\pi B_i(z) = \frac{1}{\Delta\nu_i}\int_{\Delta\nu_i} \pi B_\nu\left(z\right)d\nu$$

for the appropriate band

$$F_i^{\downarrow}(z) = \pi B_i\left(z_0\right)\left[T_i\left(z, z_0\right) - T_i\left(z, \infty\right)\right] + \int_{z_0}^{z} \pi B_i\left(z'\right) dT\left(z, z'\right) \tag{13.96}$$

With the altitude $z_0$, we indicate the base of an isothermal region that is at the top of our atmospheric model so that $B_i(z_0) = B_{\mathrm{i}}(\infty)$. We then introduce the definition for the absorptivity or absorptance:

$$A_i\left(z, z'\right) = 1 - T_i\left(z, z'\right) \tag{13.97}$$

so that the expression for the fluxes becomes

$$\begin{aligned} F_i^{\uparrow}(z) &= \pi B_i(0) + \pi \int_0^z A_i\left(z, z'\right) dB_i\left(z'\right) \\ F_i^{\downarrow}(z) &= \pi B_i\left(z_0\right) A_i\left(z, \infty\right) + \pi \int_{z_0}^{z} A_i\left(z, z'\right) dB_i\left(z'\right) \end{aligned} \tag{13.98}$$

These equations can be expressed as a function of the emissivity defined for a thin layer:

$$\varepsilon\left(z, z'\right) = \sum_i A_i\left(z, z'\right) B_i\left(z'\right)/B\left(z'\right), \quad \tilde{\varepsilon}\left(z, z'\right) = \sum_i A_i\left(z, z'\right) dB_i\left(z'\right)/dB\left(z'\right)$$

so that Eq. (13.98) becomes

$$\begin{aligned} F_i^{\uparrow}(z) &= \pi B_i(0) + \pi \int_o \tilde{\varepsilon}\left(z, z'\right) dB_i\left(z'\right) \\ F_i^{\downarrow}(z) &= \pi B_i\left(z_0\right) \varepsilon\left(z, \infty\right) + \pi \int_{z_0}^{z} \tilde{\varepsilon}\left(z, z'\right) dB_i\left(z'\right) \end{aligned} \tag{13.99}$$

These relations could simplify the computation of the infrared flux because for each band (e.g., the $CO_2$ 15 $\mu$m) the absorptance can be expressed with empirical functions and the integrals in Eq. (13.106) rapidly computed. When many years ago I discovered this method, I was amazed: it was so simple!

### 13.6.2 The Band Absorptivities According to Cess and Ramanathan

For carbon dioxide, we take into account the 15 $\mu$m band, including all the isotopes. To arrive at a simple formulation for the absorptance, we must consider the two regimes for the weak and strong line limit. Although the two cases were introduced for single lines, their extension to the $a$ band should be considered a linear dependence with the absorbed amount in the weak limit and a square root dependence in the strong limit. The simplification to be adopted is then

$$\begin{aligned} A(U) &= A_0 U = Su; \quad U << 1 \\ A(U) &= 2A_0\sqrt{\beta U}; \quad \beta << 1, \quad \beta u << 1, \quad u/\beta >> 1 \\ A(U) &= 2A_0 \ln U; \quad U >> 1 \end{aligned} \tag{13.100}$$

The first of the equations is obvious and contains again the parameter $A_0$ which is called the band parameter ($\mathrm{cm}^{-1}$) and is related to the band strength $S$ by the formula

$$A_0 = Su/U$$

Considering the dimension for $S$ ($\mathrm{cm\ g}^{-1}$) and $u$ ($\mathrm{g\ cm}^{-2}$), it is clear that $U$ must be a normalized optical path. The reason for introducing this quantity is that the atmosphere is not homogeneous, so that the optical path for radiation is weighted with the pressure if $U$ is defined as

$$U = \int_z^\infty \left(Sp_c/A_0\right)dz \tag{13.101}$$

where $p_c$ is the partial pressure of the gas in consideration.

The second equation of Eq. (13.100) introduces a new parameter $\beta$, called the line width parameter, defined as

$$\beta = 4\tilde{\alpha}/\delta$$

where $\tilde{\alpha}$ has the function of a mean line width. If $\beta << 1$, then the lines within the band are well spaced (i.e., do not overlap) so that the absorptivity of the single lines can be added. This condition is essential because it can be shown that even for the strong line we have, in case $Su >> 1$,

$$T = 1 - A(u) = \exp\left[-(Su\pi\alpha)^{1/2}/\delta\right] \qquad \Rightarrow \quad A(u) \approx (Su\pi\alpha)^{1/2}/\delta$$

and the last approximation can only be true if $\alpha << \delta$. The $\beta$ parameter must also be weighted so that

$$\beta = \frac{4\alpha_0}{U\delta}\int_{U_1}^{U_2} P dU \tag{13.102}$$

where $\alpha_0$ is the half width of the line at some reference pressure.

## 13.7 $\delta$-Eddington in the Infrared

We have developed the $\delta$-Eddington approximation in connection with the scattering of the ultraviolet and visible radiation. We have mentioned that the IR scattering is not important especially in the presence of particulates. However, the processes of absorption and emission from the particulates are important. In such cases, it may be useful to evacuate the emissivity and the absorptivity of the particles.

The way we proceed is to consider a single layer of optical thickness $\tau^*$ at temperature $T$ emitting as a blackbody. Fluxes in the two hemispheres can be evaluated using Eqs. (13.61) and (13.66):

$$\begin{aligned} F^{\uparrow}(\tau) &= \pi\left[C_1\left(1-\tfrac{2}{3}P\right)e^{-k\tau} + C_2\left(1+\tfrac{2}{3}P\right)e^{k\tau} + B\right] \\ F^{\downarrow}(\tau) &= \pi\left[C_1\left(1+\tfrac{2}{3}P\right)e^{-k\tau} + C_2\left(1-\tfrac{2}{3}P\right)e^{k\tau} + B\right] \end{aligned} \tag{13.103}$$

Constants may be determined by supposing that the diffuse flux is zero at the layer boundaries $F^{\uparrow}(\tau^*) = F^{\downarrow}(0) = 0$, so that we have

$$C_1 = B\frac{b_1 e^{k\tau^*} - b_2}{{b_2}^2 e^{-k\tau^*} - {b_1}^2 e^{k\tau^*}} \quad ; \qquad C_2 = B\frac{b_1 - b_2 e^{-k\tau^*}}{{b_2}^2 e^{-k\tau^*} - {b_1}^2 e^{k\tau^*}} \tag{13.104}$$

where $b_1 = 1 + 2P/3$ and $b_2 = 1 - 2P/3$. The net flux is then

$$F_e(\tau^*) = \pi\left(C_1 b_2 + C_2 b_1 + B\right) \tag{13.105}$$

For each wavelength, the emitted flux can be calculated as a function of the optical properties of the particles ($g$, $\omega$, $Q_{\text{ext}}$), so that the emissivity is given by

$$\varepsilon(\tau) = \left(\sigma T^4\right)^{-1}\int_{\Delta\lambda} F_e(\tau^*)\, d\lambda \tag{13.106}$$

The contributions of the particles to the fluxes at the height $z$ are given by

$$F^{\uparrow}(z) = \int_0^z \sigma T\left(z'\right)^4 d\varepsilon\left(z, z'\right) + [1 - \varepsilon(z, 0)]\sigma T_s^4$$
$$F^{\downarrow}(z) = \int_z^{\infty} \sigma T\left(z'\right)^4 d\varepsilon(z, z') \tag{13.107}$$

These equations need to be made discrete and we shall see that we have some applications for stratospheric aerosols.

## E.13 Examples

### *E.13.1 Color for Nonabsorbing Spheres*

If we look at Fig. 13.11, we notice that up to $\chi \approx 5$ the extinction cross section increases with the radius parameter. Then if we assume a constant refractive index with wavelength increasing $\chi$ correspond to decrease wavelength and this means that for a fixed radius particle population the shorter wavelength will be attenuated more than the longer ones. This is what is called reddening and this phenomenon is responsible for red color at sunset. On the other hand, the same figure shows that for $\chi$ between 5 and 10, the opposite happens and in this case we talk about blueing. This is a very rare phenomenon and we can detail the reasons on the following way. If we plot the extinction efficiency as a function of wavelength as in Fig. E.13.1, we see that for particle dimension between 0.1 and 0.3 μm (typical for small haze droplets), the extinction is a strong function of wavelength with shorter wavelength (i.e., blue) being extinguished far more than longer ones. This is again the reddening observed when haze is present. For larger particle (radius 1 μm), the behavior is more complex. In this, the near-infrared wavelength is strongly absorbed, while between 0.5 and 0.6 μm, there is a minimum which may hint to the appearance of green light in some severe thunderstorm. Finally, for much larger particles (10 μm), there is no pronounced wavelength dependence of the absorption, and this may explain that the sunlight does not experience any change in color when passing through thin clouds.

### *E.13.2 A Simple Model for Scattering*

We borrow this example from Andrews and we consider a plane-parallel atmosphere simplifying the two stream approximations with the assumption that of the down-

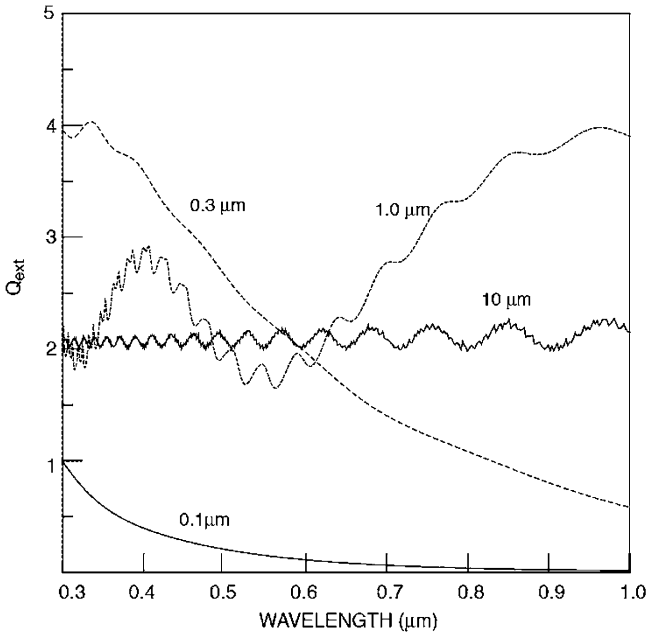



**Fig. E.13.1** The extinction efficiency as a function of wavelength for water droplets of assigned size (Petty 2006)

ward incident radiation, a fraction f is scattered forward, while a fraction $1-\mathrm{f}$ is scattered backward. These fractions result from the integration of all hemispheres.

We start by considering the downward penetration of the radiative flux irradiance $F_\nu^{\downarrow}$ at the frequency $\nu$. We could write the change of the flux with altitude as

$$\frac{dF_\nu^{\downarrow}}{d(-z)} = -a_\nu \rho F_\nu^{\downarrow} - s_\nu \rho (1-f) F_\nu^{\downarrow} + s_\nu \rho (1-f) F_\nu^{\uparrow} \tag{E.13.1}$$

This equation says that the downward flux decreases for absorption (the first term) and for the backscattered radiation (second term) and increases for the backscattering of the upward flux (last term). The terms $a_\nu$ and $s_\nu$ are the absorption and scatter coefficients. The extinction can be defined as $k_\nu = a_\nu + s_\nu$ so that the optical depth is given by $d\tau_\nu = -k_\nu \rho dz$ with the albedo for single scattering

$$\omega_\nu = s_\nu / (a_\nu + s_\nu) = s_\nu / k_\nu$$

And making the substitution, we obtain

$$\frac{dF_\nu^{\downarrow}}{d\tau} + F_\nu^{\downarrow} = \omega \left[ f F_\nu^{\downarrow} + (1-f) F_\nu^{\uparrow} \right] \tag{E.13.2a}$$

while for the upward flux

$$-\frac{dF^{\uparrow}}{d\tau}+F^{\uparrow}=\omega\left[fF^{\uparrow}+(1-f)\,F^{\downarrow}\right] \tag{E.13.2b}$$

The simplest case is if $\omega = 1$, that is, no absorption but scattering, then the two equations become

$$\begin{aligned}\frac{dF^{\downarrow}}{d\tau}+(1-f)\left(F^{\downarrow}-F^{\uparrow}\right)&=0\\ -\frac{dF^{\uparrow}}{d\tau}+(1-f)\left(F^{\uparrow}-F^{\downarrow}\right)&=0\end{aligned} \tag{E.13.3}$$

Summing the two equations, we obtain that the net flux is constant:

$$-F_z=F^{\downarrow}-F^{\uparrow}=\text{constant}$$

Now we can establish some boundary condition. We indicate with $F_{\text{inc}}$ the incident flux at $\tau = 0$ and assume that at the bottom of the layer ($\tau = \tau^*$), the downward flux is the transmitted flux $F_{\text{trans}}$. In summary

$$F^{\downarrow}(0)=F_{\text{inc}}\quad F^{\uparrow}(0)=F_{\text{refl}}\quad F^{\downarrow}\left(\tau^*\right)=F_{\text{trans}}\quad F^{\uparrow}\left(\tau^*\right)=0$$

We define a transmission coefficient $\mathrm{T} = F_{\text{trans}}/F_{\text{inc}}$ and a reflection coefficient $\mathrm{R} = F_{\text{refl}}/F_{\text{inc}}$. The result is $-F_z = F_{\text{inc}} - F_{\text{refl}} = (1-\mathrm{R})\,F_{\text{inc}} = F_{\text{trans}} - 0 = \mathrm{T}F_{\text{inc}}$ where we have used the constancy of $F_z$ which at the bottom is just $F_{\text{trans}} - F_{\text{refl}}$. We then have

$$\mathrm{R}+\mathrm{T}=\frac{F_{\text{refl}}}{F_{\text{inc}}}+\frac{F_{\text{trans}}}{F_{\text{inc}}}=\frac{F_{\text{inc}}}{F_{\text{inc}}}=1$$

meaning the conservation of irradiance.

We now can solve the first of the Eq. (E.13.3) by using

$$F^{\downarrow}-F^{\uparrow}=-F_z=\mathrm{T}F_{\text{inc}} \tag{E.13.4}$$

We have

$$\frac{dF^{\downarrow}}{d\tau}+(1-f)\,\mathrm{T}\,F_{\text{inc}}=0\quad\Rightarrow\quad F^{\downarrow}=-(1-f)\,\mathrm{T}\,F_{inc}\tau+\cos t$$

and the constant can be determined imposing that for $\tau = 0$, $F^{\downarrow} = F_{\text{inc}}$

$$F^{\downarrow}=F_{\text{inc}}\left[1-(1-f)\,\mathrm{T}\,\tau\right] \tag{E.13.5}$$

while using (E.13.4) we have

$$F^{\uparrow} = F_{\text{inc}} \left[\mathrm{R} - (1-f)\,\mathrm{T}\,\tau\right] \tag{E.13.6}$$

At this point, the reflection and transmission coefficients can be found. First we use

$$F^{\uparrow}\left(\tau^*\right) = 0 \quad \Rightarrow \quad \mathrm{R} = (1-f)\,\mathrm{T}\tau^*$$

Then, from $\mathrm{R} + \mathrm{T} = 1$, we get

$$\mathrm{T} = \frac{1}{1 + (1-f)\,\tau^*} \qquad \mathrm{R} = \frac{(1-f)\,\tau^*}{1 + (1-f)\,\tau^*} \tag{E.13.7}$$

We can see that for $\tau^* \to \infty$ reflection coefficient goes to 1 (perfect reflector) and transmission goes to 0.

### *E.13.3 Reflectivity and Transmission from Nonconservative Scattering*

Consider now the case in which $\omega < 1$. Then Eqs. (E.13.2a, and E.13.2b) becomes

$$\begin{aligned} \frac{dF^{\downarrow}}{d\tau} + F^{\downarrow} &= \omega\left[fF^{\downarrow} + (1-f)\,F^{\uparrow}\right] \\ -\frac{dF^{\uparrow}}{d\tau} + F^{\uparrow} &= \omega\left[fF^{\uparrow} + (1-f)\,F^{\downarrow}\right] \end{aligned} \tag{E.13.8}$$

We now put $\omega\,(1-f) = b$ and $1 - \omega f = a$ and then the equations become

$$\left(\frac{d}{d\tau} + a\right) F^{\downarrow} = bF^{\uparrow} \qquad \left(\frac{d}{d\tau} - a\right) F^{\uparrow} = -bF^{\downarrow} \tag{E.13.9}$$

If we derive the second equation with respect to $\tau$ and eliminate $F^{\downarrow}$,we obtain

$$\left(\frac{d^2}{d\tau^2} - \alpha^2\right) F^{\uparrow} = 0$$

where $\alpha^2 = a^2 - b^2$

The solution of the differential equation is of the form

$$F^{\uparrow} = A\sinh\alpha\tau + B\cosh\alpha\tau$$

Subject to the boundary condition $F^{\uparrow}(\tau^*) = 0$, the solution becomes

$$\begin{aligned} F^{\uparrow} &= A \sinh \alpha\tau - A \frac{\sinh \alpha\tau^*}{\cosh \alpha\tau^*} \cosh \alpha\tau \\ &= \frac{A}{\cosh \alpha\tau^*} \left( \sinh \alpha\tau \cosh \alpha\tau^* - \cosh \alpha\tau \sinh \alpha\tau^* \right) \end{aligned}$$

And absorbing all in the new constant

$$F^{\uparrow} = A \sinh \alpha \left( \tau^* - \tau \right)$$

This equation can be now substituted in second of (E.13.9) to have

$$A \left( \frac{d}{d\tau} - a \right) \sinh \alpha \left( \tau^* - \tau \right) = -bF^{\downarrow}$$

with the downward flux

$$F^{\downarrow} = \frac{A}{b} \left( \alpha \cosh \left( \tau^* - \tau \right) + a \sinh \alpha \left( \tau^* - \tau \right) \right)$$

We can obtain easily

$$F^{\downarrow}(0) = F_{\text{inc}} = \frac{A}{b} \left( \alpha \cos \alpha\tau^* + a \sinh \alpha\tau^* \right)$$

and also

$$F^{\uparrow}(0) = F_{\text{refl}} = A \sinh \alpha\tau^* \qquad F^{\downarrow}\left(\tau^*\right) = F_{\text{trans}} = \frac{A\alpha}{b}$$

and from the definition of coefficients

$$\mathrm{R} = \frac{F_{\text{refl}}}{F_{\text{inc}}} = \frac{b \sinh \alpha\tau^*}{\alpha \cosh \alpha\tau^* + a \sinh \alpha\tau^*} \qquad \mathrm{T} = \frac{F_{\text{trans}}}{F_{\text{inc}}} = \frac{\alpha}{\alpha \cosh \alpha\tau^* + a \sinh \alpha\tau^*}$$

For very large value of the optical thickness (Fig. E.13.2),

$$\mathrm{R} \to \frac{b}{\left(a^2 - b^2\right)^{1/2} + a} \qquad \mathrm{T} \to 0$$

Note also that in this case $\mathrm{R} + \mathrm{T} < 1$.

### *E.13.4 A MATLAB Program for the Delta-Eddington*

We present an example for a single layer for the delta-Eddington approximation. We just use the simple formula seen in the chapter and calculate the albedo as a

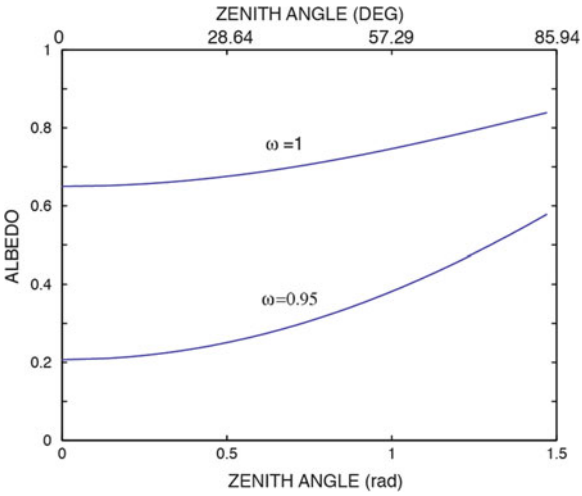



**Fig. E.13.2** Albedo as a function of the incident sun angle for two different values of the albedo of single scattering and g = 0.786

function of the incident zenith angle for a fixed and large optical thickness. The calculations refer to a case of conservative scattering ($\omega = 1$) and nonconservative scattering ($\omega = 0.95$). We can see the albedo increases with the zenith angle and the conservative case shows a much larger albedo:

MATLAB listing

```
mu=0.1:0.01:1.0
for i=1:length(mu)
xmu=mu(i)
omega = 0.95
g=0.786
tau=16.0
alb=0.0
xk=sqrt((3*(1-omega)*(1-omega*g)))
alfa=3*omega*xmu*xmu*(1+g*(1-omega))/(4*(1-(xk*xmu)^2))
beta=(3*omega*xmu*(1+3*g*(1-omega)*xmu*xmu))
  /(4*(1-(xk*xmu)^2))
p=sqrt((3*(1-omega)/(1-omega*g)))
a11=1+(2/3.)*p
a12=1-(2/3.)*p
a21=(1-alb-(2/3.)*p*(1+alb))*exp(-tau*xk)
a22=(1-alb+(2/3.)*p*(1+alb))*exp(+tau*xk)
c1=alfa+2*beta/3
c2=(alfa*(1-alb)-(2*beta/3)*(1+alb)+alb*xmu)*exp(-tau/xmu)
A=[a11 a12;a21 a22]
b=[c1;c2]
x=inv(A)*b
I0=x(1)*exp(-xk*tau)+x(2)*exp(xk*tau-alfa*exp(-tau/xmu)
```

```
   I1=(x(1)*exp(-xk*tau)-x(2)*exp(xk*tau))
   *p-beta*exp(-tau/xmu) I00=x(1)+x(2)-alfa
I10=(x(1)-x(2))*p-beta
fd=pi*(I0+2*I1/3.)
fu=alb*(fd+xmu*pi*exp(-tau/xmu))
Fu0=pi*(I00-2*I10/3.)
R(i)=Fu0/(xmu*pi)
fdir=xmu*pi*exp(-tau/xmu)
photo=fdir/xmu+fu+fd
t=(1-g)*tau
work1=3*xmu*(1-alb)*(2+3*xmu+(2-3*xmu)*exp(-tau/xmu))
work2=4*(4+3*(1-alb)*t) b2r=work1/work2 b1r
  =(3*xmu*xmu/4+0.5*xmu)-2*b2r/3.
rr(i)=1-4*b2r/(3*xmu)
teta(i)=acos(xmu)
end
plot(teta,R), axis([0 1.5 0 1])
hold
plot(teta,rr),axis([0 1.5 0 1])
```

### *E.13.5 Infrared Flux from Methane*

It is interesting to understand the power of the absorptance method by giving an example for methane. This gas has an absorption band at 1306 $cm^{-1}$ (7.65 $\mu$m) and the parameters for this band are

$$\begin{aligned} S &= 185\,(273/T) && \text{cm}^{-1}(\text{cm atm})^{-1} \\ A_0 &= 52(T/300)^{1/2} && \text{cm}^{-1} \\ \beta &= 0.17\,(p/p_0)\,(300/T) \end{aligned}$$

We can notice that these parameters depend on both pressure and temperature. The band strength, but in practice the unit of $(\text{cm atm})^{-1}$, substitutes for the $\text{cm}^2$ $\text{g}^{-1}$. At this point, we could easily evacuate the integral, but we are interested only in averaging quantities and order of magnitude. Most methane resides in the troposphere so that we can adopt a pressure-weighted average for the temperature around 263 K and an average value for pressure $p/p_0 = 0.514$. The amount of gas in the requested unit can be easily calculated through the molecules per $\text{cm}^2$:

$$N_{CH_4} = \frac{fM_{CH_4}\int_0^\infty \rho dz}{m_{CH_4}M}$$

where the $M$s are the molecular mass, while $m$ is the mass of a single molecule. To obtain the requested unit, this number must be divided by the Loschmidt number $2.687 \times 10^{19}$. The total methane amount is 1.4 cm atm, so that the normalized optical path is given by

$$U = Su/A_0 = 5.40$$

while the total absorptivity

$$A = 2A_0\sqrt{\beta U} \approx 69\ \mathrm{cm}^{-1}$$

And with the data, we can evaluate how much flux is absorbed by methane:

$$\Delta F^{\uparrow}_{CH_4}(\infty) = -A_{CH_4}(0,\infty)\,\pi\,[B_{CH_4}(\infty) - B_{CH_4}(0)]$$

The Planck function must be evaluated at the center of the band and its units will be W $\mathrm{m}^{-2}/\mathrm{cm}^{-1}$. These units are very strange in a sense because usually the Planck function is expressed as W $\mathrm{m}^{-2}$ $\mathrm{cm}^{-1}$. To obtain the units that we require (power per unit wave number), it is necessary to multiply the Planck function for the square of the wavelength. After that, we find that our Planck function for methane is given by

$$\pi B_\nu = 83.25/\,[\exp(-1878/T) - 1] \quad \mathrm{W\ m^{-2}cm}$$

We can now substitute for $T(0) = 290$ K and $T(\infty) = 250$ K so that

$$\Delta F^{\uparrow}_{CH_4}(\infty) = -69 \cdot 0.0825 \approx 5.7\ \mathrm{W\ m^{-2}}$$

This result, although we made a number of approximations, is quite good and represents the outgoing infrared flux absorbed by the methane

We were then right when we thought that these are back-of-the envelope calculations. When we need to compute the heating rate or when we have a gas whose mixing ratio changes with altitude, we need to go to something more complex.

## References[1]

### *Books*

Bohren CF, Huffman DR (1983) Absorption and scattering of light by small particles. Wiley, New York

Goody RM, Young YL (1989) Atmospheric radiation, theoretical basis. Oxford University Press, New York/Oxford

[1] In this chapter, the most important texts are those by Bohren and Goody and Young and the very useful Liou text. Again single papers are very useful like Lacis and Hansen and Ramanathan.

Liou KN (1980) An introduction to atmospheric radiation. Academic, Amsterdam/Boston
McCartney EJ (1976) Optics of the atmosphere: scattering by molecules and particles. Wiley, New York
McCartney EJ (1983) Absorption and emission by atmospheric gases: the physical processes. Wiley, New York
Petty GW (2006) A first course in atmospheric radiation. Sundog Publishing, Madison
Wallace JM, Hobbs PV (1977) Atmospheric science: an introductory survey. Academic, New York

## *Articles*

Lacis AA, Hansen JE (1974) A parameterization for the absorption of solar radiation in the Earth atmosphere. J Atmos Sci 31:118
Ramanathan V (1976) Radiative transfer within the earth troposphere and stratosphere: a simplified radiative convective model. J Atmos Sci 33:1330
Shettle EP, Weinman JA (1970) The transfer of solar irradiance through inhomogeneous turbid atmospheres evaluated by Eddington's approximation. J Atmos Sci 27:1048

# Chapter 14
# Simple Climate Models

So far we have introduced the different ways in which solar radiation and planetary radiation interact with the atmosphere and the surface. We have not fixed any particular number, and also we have been rather vague on how the vertical or latitudinal distribution of temperature is determined. We have the means now to do that and will start by looking at what we know about the radiative balance of the Earth–atmosphere system. We can refer to a classic figure like Fig. 14.1 which appears in different disguises in most of textbooks.

## 14.1 Energy Budget

We notice a few things from this globally averaged picture. First of all the solar radiation is depleted by almost 50 % when traversing the atmosphere by processes like scattering and absorption from both gases and clouds. As far as the infrared radiation is concerned, we can see that the "gray" approximation is really not valid because there are gases that absorb selectively and the same gases also contribute to emission. To complicate things, further, the planetary radiation also interacts with the clouds. In Fig. 14.1 there are two terms that we have especially mentioned in the chapter on the boundary layer, the latent and sensible heat. The surface could hardly be in radiative equilibrium and the excess of absorption at the surface is compensated by the presence of such fluxes.

All the processes shown in the figure are responsible for the temperature structure of the atmosphere and also determine the temperature at the surface. At a textbook level we must limit ourselves to simple things like asking the question how

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_14) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_14

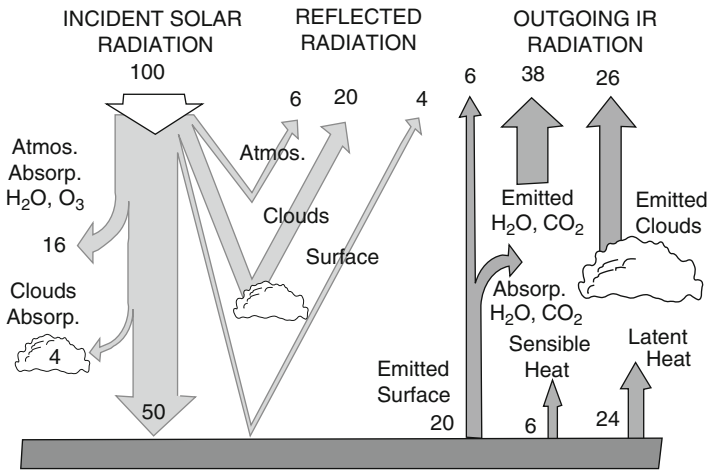



**Fig. 14.1** The interaction of solar and infrared radiation with the Earth–atmosphere system. The incident solar radiation outside the atmosphere is normalized to 100 units

altering one of those processes could change, for example, the globally averaged temperature. The answer to this question may be attempted at different levels of complexity. We are smart enough now to know that the atmosphere is rather complex and the dynamic processes should also be taken into account. We will show that the first level answer could be based on simple climate models that consider just the average properties of a planet or the parameterization of its atmospheric circulation. These models have different names like *radiative–convective models* or *energy balance climate models*. They do not give you any details on how climate may change in your hometown (who does?), but are important to understand the basic processes that determine the climate. For example, they may give you a good understanding of most of the feedback mechanisms like ice–albedo or water vapor feedback. To do that we need to use everything we have learned up to Chap. 13, including of course, the beautiful simplifications introduced by Cess and Ramanathan. We will start out at the zero dimension and go up just one step to one-dimensional energy balance models and then illustrate the radiative–convective models.

## 14.2 Zero-Dimensional Models and Feedback

We have encountered this kind of model already in Chap. 1 when we simply calculated the global average temperature of the Earth by equating the absorbed solar radiation and the infrared emitted radiation. The first realistic change we will introduce deals with the fact that the Earth–atmosphere system does not emit like a blackbody, using a parameterization based on experimental data. We can then assume that the emitted power in the infrared is linear with the temperature according to the relation

$$I = A + BT \tag{14.1}$$

where $I$ is the emitted power in W m$^{-2}$, $A$ and $B$ are constants, and $T$ is the temperature expressed in degrees Celsius. This relation was found by M. Budyko well before the satellite data could confirm it. It is interesting to note how Eq. 14.1 can be obtained by linearization of the blackbody emission law:

$$\sigma(273,15+T)^4 = \sigma(273,15)^4 + 4\sigma(273,15)^3 T$$

where $A = \sigma\ (273.15) = 315.58$ W m$^{-2}$ and $B = 4\sigma(273.15)^2 = 4.62$ W m$^{-2}$ C$^{-1}$. Actually, the values of the same coefficients found for Eq. 14.1 are $A = 203.3$ Wm$^{-2}$ and $B = 2.09$ Wm$^{-2}$ C$^{-1}$, and the reason is that, for the same surface temperature, because of the greenhouse effect, the outgoing flux is less than the corresponding blackbody flux. However the empirical relation like Eq. 14.1 takes into account a number of effects including cloudiness. Using Eq. 14.1 we can write the thermal equilibrium of the Earth in a slightly different way from what we saw in Chap. 1, that is,

$$A + BT_0 = \frac{S_0\left(1-\alpha_p\right)}{4} \tag{14.2}$$

where $T_0$ is the average global temperature, $S_0 = 1360$ W m$^{-2}$ and is the solar constant, and $\alpha_\mathrm{p}$ is the average planetary albedo. We obtain from this expression $T_0 = 14.9$ °C which is the global average temperature that takes into account the greenhouse effect.

We can now evaluate the climate sensitivity which we define as the global temperature change for a 1 % change of the solar constant. We have then

$$4\sigma T^3 \Delta T = \frac{\Delta S_0\left(1-\alpha_p\right)}{4}$$

so that

$$\beta_0 = \frac{S_0}{100}\frac{dT_0}{dS_0} \tag{14.3}$$

In this relation $\beta_0$ is the sensitivity parameter that corresponds to the previous definition. For a blackbody we find easily

$$\beta_0 = \frac{T}{400} = 0.63\,^\circ\mathrm{C}$$

If we use for the IR radiation the approximation Eq. 14.1, we obtain

$$\beta_0 = \frac{(A+BT_0)}{100B} = 1.12\,^\circ\mathrm{C}$$

The conclusion is that the greenhouse effect increases the climate sensitivity, something that we had seen already in Chap. 1.

The formulation we have adopted is useful to study one of the most important feedback mechanisms, referred to as *ice–albedo feedback*. This is based on a very simple consideration. A lower global temperature implies an increase in ice- or snow-covered areas and consequently albedo increases. Any perturbation that produces an initial decrease in temperature will be amplified by this mechanism and vice versa for a temperature increase. It is reasonable, based on this feedback, to assume that the albedo is a function of temperature. We may adopt a very simple dependence between two extremes: an ice-covered Earth with albedo 0.62 and an Earth free of ice with albedo 0.3. To these albedo values correspond, according to Eq. 14.2, the temperatures of −36.4 °C and 15 °C, respectively. We need to also make an assumption about the temperature dependence of the geographical extension of ice. Denoting with $x_s$ the sine of the lowest latitude reached by the ice sheets, we assume the following dependence:

$$x_x = \begin{cases} x_s = 1 & T_0 > 15\,^\circ\mathrm{C} \\ x_s = 0 & T_0 < 15\,^\circ\mathrm{C} \\ 1 + \dfrac{(T_0 - 15)}{30} & -15\,^\circ\mathrm{C} < T_0 < 15\,^\circ\mathrm{C} \end{cases} \tag{14.4}$$

In order to determine the temperature dependence of the albedo we have to use a more general definition than the one given in Chap. 1. The definition was based on the assumption that the distribution of the solar radiation with latitude was determined only by the geometrical factor. Actually we saw in Chap. 3 that the distribution is given by some function $S(x)$ (with $x$ the sine of the latitude) so that the albedo must be defined:

$$\alpha_p = \frac{1}{2}\int_{-1}^{1} dx\, S(x)\, \alpha(x) \tag{14.5}$$

where the function $S$ can be expanded in a series of Legendre polynomials

$$S(x) = 1 + S_2 P_2(x)$$

where $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial with $S_2 = -0.477$. The planetary albedo is then

$$\alpha_p = \int_0^{x_s} [1 + S_2 P_2(x)]\, \alpha_f dx + \int_{x_s}^{1} [1 + S_2 P_2(x)]\, \alpha_i dx \tag{14.6}$$

In Eq. 14.5 we have assumed that in the ice-free region $\alpha(x) = \alpha_\mathrm{f} = 0.3$, while in the region covered with ice $\alpha(x) = \alpha_\mathrm{i} = 0.62$. We have then

$$\overline{a} = 1 - \alpha_p = a_i + (a_f - a_i)\left[x_s + S_2 \frac{(x_s - x_s^3)}{2}\right] = H_0\,[x_s\,(T_o)] \tag{14.7}$$

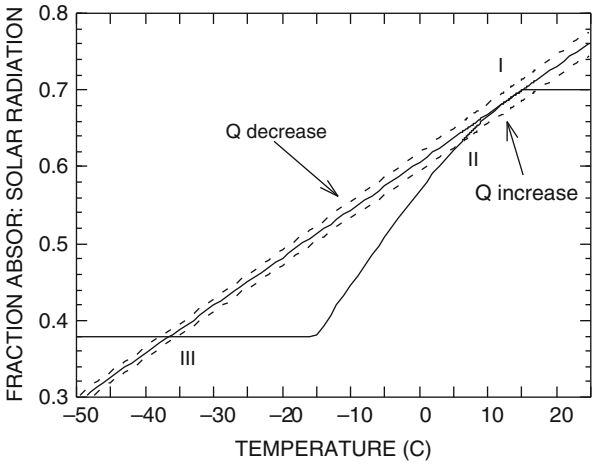



**Fig. 14.2** The solutions for Eq. (14.8) are shown. The three *straight lines* represent the function $4(A + BT)/S_0$: the *solid line* is for a solar constant of 1.36 kW $m^{-2}$ and the others are for a $\pm 2\,\%$ change. Solutions are indicated with I, II, and III

where with $\bar{a}$ we have indicated the co-albedo so that $a_i = 1 - \alpha_I$, $a_f = 1 - \alpha_f$ . Now the solution for the temperature $T_0$ is still given by Eq. 14.2 with the right-hand side substituted by $H_0[x_s\,(T_0)]$ so that

$$A + BT_0 = H_0\left[x_s\,(T_0)\right]\frac{S_0}{4} \tag{14.8}$$

The solution of this equation can be found graphically as shown in Fig. 14.2.

Surprisingly enough it looks like the climatic system we have invented has three different solutions, one around 15 °C corresponding to the present climate with a planet almost free of polar ice. The second solution is around 0 °C and the third one has a temperature of about −35 °C with a planet almost completely covered with ice. It is quite interesting to consider what happens by changing the solar constant as is shown qualitatively in the same figure for a variation of 2 %.

When the solar constant decreases, the temperature interval between the different solutions is also reduced. The opposite happens if the solar constant increases. This means that the solution marked as II is unstable. If for some reason (a perturbation) climate II should warm, then in order for that state to be a new equilibrium, the solar constant should decrease. However if the solar constant has remained constant, the climate will move further away from its initial state. The opposite will happen if the perturbation initially cools the planet. In this case the climate will move toward solution III and the Earth will be completely covered with ice. On the other hand points I and III represent stable solutions. If we consider the present climate and a perturbation that warms it, then in order to find a new equilibrium, the solar constant should increase. Remaining at a constant value, the planet will cool back to its initial state.

This rather complicated argument may be better understood if we refer to Fig. 14.3 where the global temperature is plotted as a function of the ratio between the solar constant $q$ and its reference value $q_0$. We notice how the curve is actually composed of two lines, an upper one near the solution I for an ice-free Earth and the other for solution III that corresponds to an ice-covered Earth. Along these

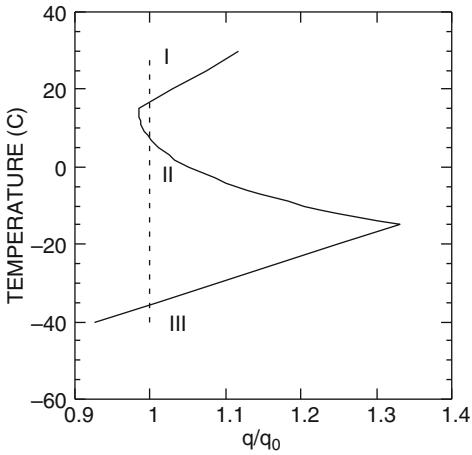



**Fig. 14.3** The global temperature in the zero-dimensional model as a function of the ratio between the solar constant and the reference value. The *dashed line* refers to the solution given in Fig. 14.2

two lines, because the albedo is constant, the temperature is linear with the solar constant. The middle tract is where the albedo is a function of the temperature. Starting from a value greater than one for the ratio $q/q_0$, the temperature decreases with the solar constant until the ratio becomes a little less than one. At this point the temperature would decrease even if the solar constant should increase; that is a clear sign for instability. The concept of climate sensitivity becomes rather useless in the presence of discontinuity in the temperature, so we need to further discuss this aspect. We consider now a time-dependent equation to the thermal equilibrium, Eq. 14.8, which is given by (assuming a unit thermal capacity)

$$\partial T_0/\partial t = S_0\left(1-\alpha_p\right)/4 - \left(A + BT_0\right) \tag{14.9}$$

From this equation we notice that the infrared term has the function of a relaxation term with a time constant equal to the inverse of $B$. We now apply the familiar perturbation method to study what happens around equilibrium points by assuming a temperature perturbation such that

$$T_0(t) = T_0^0 + T'(t)$$

so that Eq. 14.9 becomes

$$\partial T'/\partial t + BT' = q\left(dH_0/dT_0\right)T' \tag{14.10}$$

where we have put for simplicity

$$q = S_0\left(1-\alpha_p\right)/4$$

and used the condition for the solution

$$A + BT_0^0 = qH_0\left(T_0\right)$$

Deriving this equation with respect to $q$

$$B\frac{\partial T_0^0}{\partial q} = H_0 + q\left(\frac{\partial H_0}{\partial T_0^0}\right)\frac{\partial T_0^0}{\partial q}$$

Using this equation $B$ can be eliminated from Eq. 14.10 to obtain (see E.14.1)

$$\frac{\partial T'}{\partial t} = -H_0\left(\frac{\partial T_0^0}{\partial q}\right)^{-1} T' \tag{14.11}$$

This equation shows that in order to have stability it must be

$$\frac{\partial T_0^0}{\partial q > 0}$$

and vice versa. This criterion can be used to analyze Fig. 14.3.

This stability criterion can also be put in more familiar terms using the potential concept which is used in mechanics. We consider the function

$$F\left(T_0\right) = AT_0 + \frac{1}{2}BT_0^2 - q\int_0^{T_0} H_0\left(T_0'\right) dT_0' \tag{14.12}$$

(where we have substituted the double zero with a simpler notation) that is shown in Fig. 14.4. The minima and maxima of this function are calculated with the condition $dF/dT_0 = 0$ and are just the solutions I, II, and III that we found previously. We can also establish through Eq. 14.9 that

$$\dot{T}_0 = -dF/dT_0$$

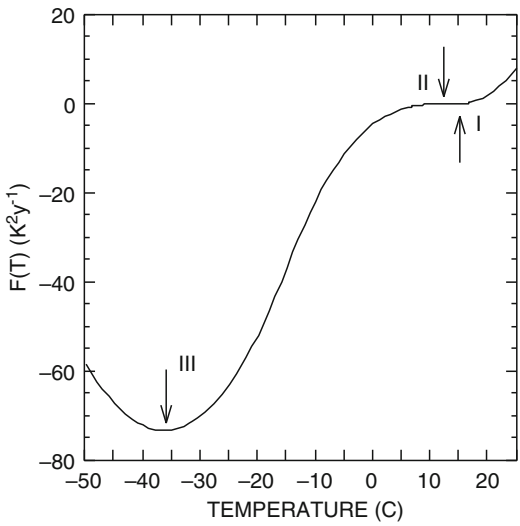



**Fig. 14.4** The potential function for the zero-dimensional model. The solutions found before are marked

so we have

$$\frac{dF}{dt} = \left(\frac{dF}{dT_0}\right)\dot{T}_0 = -\left(\dot{T}_0\right)^2 \leq 0 \tag{14.13}$$

This means that $F$ always decreases with time while the derivative of $T_0$ is proportional to the slope of $F$ at the point $T_0$. This again shows that solutions I and III are stable while solution II is unstable. We will see in the next Chap. 15 how these considerations, which may look like exam traps for the student, have very important implications because the climate seems to have two stable states, one corresponding to a planet covered with ice and the other to the present status.

It would seem possible that a small temperature perturbation could bring the system from one state to the other. As we will see it is not necessary that the perturbation be of the order of $-20$ °C, because even the classical theory of Brownian motion gives a finite probability for the escape from a potential well. The corresponding exit time will be a function of the height of the barrier. The theory of the ice age could be a nice application for this idea considering that the main period for it is of the order of 100,000 years.

## 14.3 One-Dimensional Energy Balance Climate Models

The next step in the improvement of the energy balance model is the introduction of the temperature dependence on the latitude. In this case the budget is established on a box delimited by two latitude bands as in Fig. 14.5. This box exchanges energy

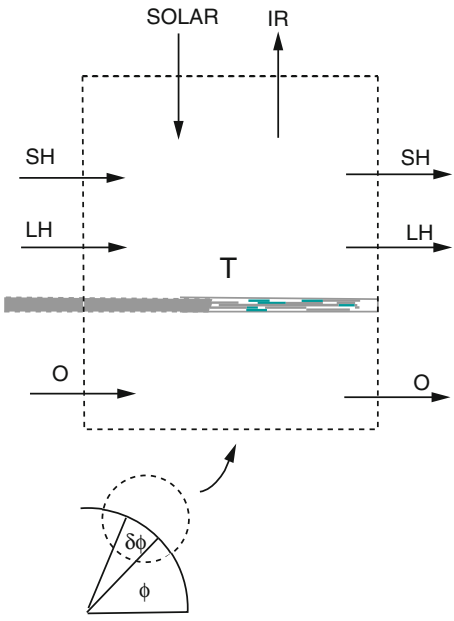



**Fig. 14.5** The energy budget within a box delimited by two latitude bands. $T$ is the temperature of the surface, *SH* and *LH* the sensible and latent heat fluxes, respectively, and $O$ the oceanic flux

through the atmospheric latent heat flux (LH) and sensible heat flux (SH) and from the ocean (O). Also the box absorbs and emits radiation. All these fluxes can be parameterized as we have seen in the previous chapters.

At the end of the 1960s almost at the same time, W. Sellers in the US and M. Budyko in the URSS proposed different parameterizations for the fluxes along the latitude. Budyko decided to assume that all the fluxes could be treated as a diffusive process, while Sellers calculated diligently all the different components appearing in Fig. 14.5. The first approach is much simpler especially after G. North found a very elegant way to solve the model.

We will report here both methods; in the Appendix the appropriate FORTRAN programs are listed. This could be an amusement, but as in the other cases we will show that models are useful to understand processes and the sensitivity of the system without going to the extreme and becoming convinced that they are more real than nature.

### *14.3.1 North's Model*

The simplest hypothesis that we can adopt is to assume that the latitudinal heat flux is proportional to the temperature gradient. This corresponds to assuming that the diffusion coefficients we found in Chap. 12 are constant. Indicating with $x$ the sine of latitude, we have for the flux in spherical coordinates

$$-D\left(1-x^2\right)^{1/2}\frac{dT}{dx}$$

However, looking at Fig. 14.5, we can see that the temperature change is instead proportional to the flux difference, that is, to the flux divergence:

$$-\frac{d}{dx}D\left(1-x^2\right)\frac{dT}{dx} \tag{14.14}$$

As for the radiative term we need to establish a parameterization for the albedo that will be a function of temperature and then latitude. We can assume that the albedo will be at a constant value 0.62 for a temperature lower than $-10$ °C and this will correspond to the ice-covered portion of the Earth for latitude higher than $x_s$. For the ice-free portion, albedo will be an assigned function of latitude:

$$a\left(x,x_s\right)=1-\alpha=\begin{cases} b_0 & x>x_s \\ a_0+a_2P_2(x) & x<x_s \end{cases} \tag{14.15}$$

where $b_0=0.38$, $a_0=0.697$, and $a_2=-0.0779$. On the other hand if we use the parameterization for the IR radiation $I(x)=A+BT(x)$, we obtain at the latitude $x_s$ $I_s=I(x_s)=182.4$ W m$^{-2}$, and the thermal equilibrium gives

$$I(x)-\frac{d}{dx}D\left(1-x^2\right)\frac{dT(x)}{dx}=qS(x)a\left(x,x_s\right) \tag{14.16}$$

From this relation the temperature can be eliminated using the fact that it is proportional to $I(x)$ and defining $D$ to include also $B$. We have

$$I(x) - \frac{d}{dx} D\left(1 - x^2\right) \frac{dI(x)}{dx} = qS(x)a\left(x, x_s\right) \tag{14.17}$$

The solution to this equation can be obtained in a very elegant way (due to G. North) using the fact that the Legendre polynomials are an eigenfunction of the diffusion spherical operator

$$\frac{d}{dx}\left(1 - x^2\right) \frac{dP_n(x)}{dx} = -n\left(n + 1\right) P_n(x) \tag{14.18}$$

Some restrictions can be used on terms like $P_n(x)$, because if we consider an annual mean and perfectly symmetrical hemispheres, then the solutions should not change when x changes sign. This means that each variable developed in a series of Legendre polynomials should contain only even terms. For example, for the IR

$$I(x) = \sum_{n \ \ pari} I_n P_n(x) \tag{14.19}$$

Substituting in Eq. 14.17 we have

$$\sum_{n \ \ pari} \left\{ I_n P_n(x) - I_n \frac{d}{dx} D\left(1 - x^2\right) \frac{dP_n(x)}{dx} \right\} = qS(x)a\left(x, x_s\right) \tag{14.20}$$

which is equivalent to

$$\sum_{n \ \ pari} \left\{ I_n P_n(x) \left[1 + n\left(n + 1\right) D\right] \right\} = qS(x)a\left(x, x_s\right) \tag{14.21}$$

We can now use the orthogonality and normalization properties of the polynomials:

$$\int_0^1 P_n(x) P_m(x) dx = \frac{\delta_{mn}}{(1 + 2n)} \tag{14.22}$$

Then if we multiply Eq. 14.21 by $P_m(x)$ and integrate, we will be left only with those terms for which $m = n$. We have

$$I_n\left[1 + Dn\left(n + 1\right)\right] = (1 + 2n)\, q \int_0^1 P_n(x) S(x) a\left(x, x_s\right) dx \tag{14.23}$$

Putting

$$H_n\left(x_s\right) = (1 + 2n) \int_0^1 P_n(x) S(x) a\left(x, x_s\right) dx \tag{14.24}$$

we can write

$$I_n \left[1 + Dn\,(n+1)\right] = qH_n\,(x_s) \tag{14.25}$$

Let us try to write this relation for the first few values of $n$:

$$\begin{aligned} I_0 &= qH_0\,(x_s) = q\int_0^1 S(x)a\,(x, x_s)\,dx \qquad n = 0 \\ I_2\,(1 + 6D) &= qH_2\,(x_s) \qquad\qquad\qquad\qquad n = 2 \end{aligned} \tag{14.26}$$

The first relation is just the one that gives the global average temperature. The second equation gives an indication of the temperature difference between the pole and the equator and may depend on the diffusion coefficient. Through this equation we can determine the value of $D$. The procedure is to calibrate the model with the present climatic conditions so that we put in Eq. 14.26 $x_s = 0.95$ which corresponds to a latitude of about 72°. We also have $H_0(0.95) = 0.698$, and with $q = 340$ W m$^{-2}$, we obtain $I_0 = 237.3$ W m$^{-2}$. Because the temperature at $x = x_s$ is −10 °C, we have an emitted power of $I(x_s) = I_0 + I_2P_2(x_s) = A - 10B = 182.4$ W m$^{-2}$, from which we can get $I_2 = -54.8$ W m$^{-2}$. This value, inserted in the second part of Eq. 14.26, gives the value for $D = 0.65$. We now have

$$T_0 = \frac{(I_0 - A)}{B} \qquad T_2 = \frac{I_2}{B}$$

and the temperature as a function of latitude can be written as

$$T = T_0 + T_2P_2(x) \tag{14.27}$$

The behavior of temperature obtained in this way can be compared with the experimental data in Fig. 14.6 which shows a satisfactory agreement for such a simple model.

It is possible now to test how sensitive is the ice line (i.e. $x_s$) to the value of the solar constant. The radiation emitted at the ice line can be expanded

$$I_s = \sum_{n\ \ pari} I_nP_n\,(x_s) \tag{14.28}$$

and the coefficients $I_n$ can be obtained from Eq. 14.25

$$I_n = qH_n \frac{(x_s)}{[1 + Dn\,(n+1)]}$$

So that Eq. 14.28 can be written as

$$I_s = \frac{q\sum_{n\ \ pari} H_n\,(x_s)\,P_n\,(x_s)}{[1 + Dn\,(n+1)]} \tag{14.29}$$

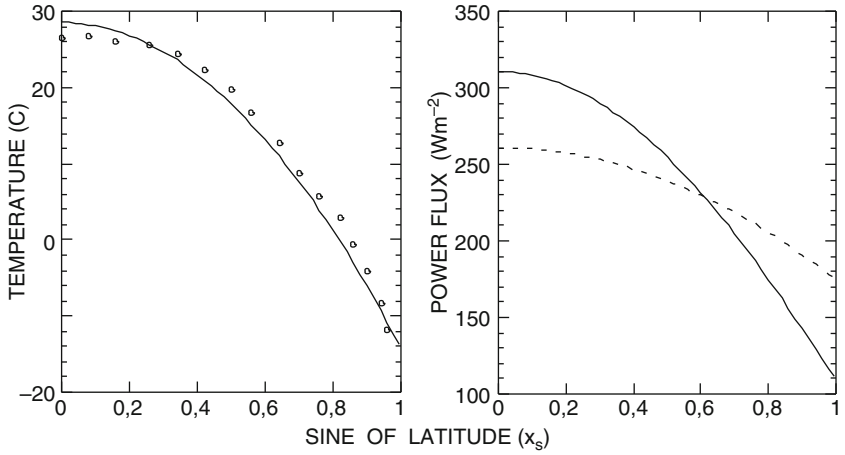



**Fig. 14.6** The temperature as a function of latitude at the *left* is compared with annual mean values (*circles*) for the northern hemisphere. At the *right* are shown the absorbed solar radiation (*solid line*) and the IR emitted radiation (*dashed*). This figure may be compared with Fig. 11.4 which shows the experimental data

This equation can be read in a slightly different way. Because $I_s$ is constant, then changing the solar constant implies the change of latitude $x_s$. This means that $q$ may be intended as a function of $x_s$ according to the relation

$$q(x_s) = I_s \left[ \sum_{n \ pari} \frac{H_n(x_s) P_n(x_s)}{1 + Dn(n+1)} \right]^{-1} \tag{14.30}$$

This relation gives a value of the solar constant that is consistent with the assigned latitude of the ice line, and in Fig. 14.7 this function is shown including three terms of the summation. We notice that, starting from the present conditions and decreasing the solar constant, the latitude of the ice line decreases as expected. However beyond the "knee" of the curve even an increase of the solar constant brings about a decrease of the latitude $x_s$ and this is a clear sign of instability. Also in this case, as in the zero-dimensional models, we have a transition from a warm Earth to a planet covered with ice. The same figure shows that to return to the present conditions the solar constant must increase more that 40 %. The reason is that when the Earth is ice covered, its albedo is quite high (0.7) so that even a solar constant of $1.42 q_0$ will produce a global mean temperature of $-30$ °C and a temperature at the equator of $-10$ °C.

This model has three different solutions; like the zero-dimensional model, two of them are stable (ice-covered Earth and the present climate) and one is unstable. The figure also shows the dependence of the solution on the adopted diffusion coefficient. We should expect that a higher diffusion coefficient would reduce the

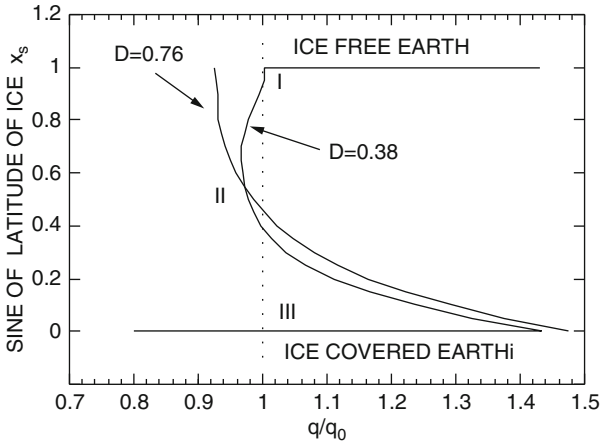



**Fig. 14.7** The latitude of the ice line as a function of the fractional change of the solar constant. Notice the stability of the present climate for a change in the solar constant up to 4 % of the present value. Also shown is a curve for a different value of the diffusion coefficient

temperature gradient along the latitude and then increase instability. When the solar constant decreases in such conditions, the global temperature is lowered and because of the reduced gradient, the temperature will change little with latitude and ice could form everywhere. The diffusion coefficient however cannot be changed at will because it must be constrained by the conditions observed for the present Earth.

The stability of the solution for the one-dimensional model can be carried out with the same approach used in the previous section. However in this case the analytical treatment is more complex, because the minima have to be found not with respect to a single variable but to a distribution of that variable with latitude. We will need to make another little detour, regarding mathematical analysis.

### *14.3.2 The Stability of the One-Dimensional Model*

We would like to extend here the considerations on the potential introduced for the zero-dimensional model. In that case we need to find minima and maxima of a function for certain values of the temperature. In this case the temperature is a function of latitude so that our process of minimization should give us a function.

The potential function is then a function of a function and in mathematical terms this is called a functional. A function may be regarded as an operation that establishes a correspondence between a certain value of a variable $x$ with the result of the operation $y(x)$. The variable can be constituted by a series of values (like a vector $x$) so that the function becomes $y(x)$. If this vector becomes of infinite dimension (i.e., a function), then we have a *functional*:

$$z = z\,[y(x)]$$

In our case the function is just $T(x)$. In the theory of the functionals, as we do for functions, we can find maxima and minima, which being functions are called *stationary functions*. The condition $dy(x)/dx = 0$ is given by an equation from Euler. We define a functional of the form

$$F[y(x)] = \int_a^b f\left(x, y, y'\right) dx \tag{14.31}$$

where $y'$ is indicative of a derivation with respect to $x$; then the Euler condition is

$$\frac{d}{dx}\left[\frac{\partial f\left(x, y, y'\right)}{\partial y'}\right] - \frac{\partial f\left(x, y, y'\right)}{\partial y} = 0 \tag{14.32}$$

Doing a simple substitution $x \to x,\ y \to T,\ y' \to \partial T/\partial x$, we obtain with $T_x = \partial T/\partial x$

$$\frac{d}{dx}\left[\frac{\partial f\left(x, T, T_x\right)}{\partial T_x}\right] - \frac{\partial f\left(x, T, T_x\right)}{\partial T} = 0$$

By comparing this with Eq. 14.18 a good candidate for our functional should satisfy the condition

$$\frac{\partial f\left(x, T, T_x\right)}{\partial T_x} = D\left(1 - x^2\right) T_x \qquad ; \frac{\partial f\left(x, T, T_x\right)}{\partial T} = I(x) - qS(x)a\left(x, x_s\right)$$

so that we can argue that the required functional will be

$$F[T(x)] = \int dx \left[\frac{1}{2} D\left(1 - x^2\right) T_x^2 + R(T) - qSA(T)\right] \tag{14.33}$$

where $R(T)$

$$R(T) = \int^T I\left(T'\right) dT' \qquad A(T) = \int^T a\left(T'\right) dT' \tag{14.34}$$

The manipulation of these equations is quite complicated but can be simplified in the case that we consider a spectral solution with only two modes $T = T_0 + T_2 P_2(x)$. We can then evaluate the integral of Eq. 14.33 so that we have

$$\int dx T_2^2 \left[\frac{1}{2} D\left(1 - x^2\right) \left(dP_2/dx\right)^2\right] + \int dx \int \left(A + BT'\right) dT' - q \int dx \int S(x) a\left(x, x_s\right) dT'$$

The first and the second term give

$$\int dx T_2^2 \left[\frac{1}{2} D\left(1 - x^2\right) \frac{dP_2}{dx} \frac{dP_2}{dx}\right] + \int dx \left(AT + BT^2/2\right)$$

The first integral can be solved by parts

$$\frac{1}{2}T_2^2 D\left(1-x^2\right)\frac{dP_2}{dx}P_2\Big|_0^1 - \int dx T_2^2\left[\frac{1}{2}P_{2_2}\frac{d}{dx}D\left(1-x^2\right)\frac{dP_2}{dx}\right]$$

At this point it is easy to show that using Eqs. 14.18 and 14.22 and using in the second integral the spectral expression for the temperature

$$F\left(T_0, T_2\right) = AT_0 + \frac{B}{2}{T_0}^2 + \frac{6D+B}{10}{T_2}^2 - M\left(T_0, T_2\right) \tag{14.35}$$

where

$$M\left(T_0, T_2\right) = q\int dT' \int S(x)a\Big(x, x_s\Big)dx \tag{14.36}$$

This integral is not simple at all because $x_s$ is a function of temperature. Instead of using Eq. 14.15 we may use a simplification based on the original Budyko idea which is to write the absorption as a step function

$$a(T) = a_0\theta\left(T_s - T\right) + a_1\theta\left(T - T_s\right) \tag{14.37}$$

where $\theta(Z)$ is a step function that is zero if $Z<0$ and is unity if $Z \geq 0$. In this way the ice-covered fraction of the Earth's surface has absorption $a_0$ and the ice-free one has $a_1$. With this simplification Eq. 14.36 becomes

$$M\left(T_0, T_2\right) = q\int_0^1 S(x)\Big(T - T_s\Big)\left[a_0\theta\left(T_s - T\right) + a_1\theta\left(T - T_s\right)\right]dx \tag{14.38}$$

The stationary values of the function $T$ are obtained by putting to zero the derivative $\partial F/\partial T_0, \partial F/\partial T_2$ resulting in the condition

$$\begin{aligned} A + BT_0 &= q\int_0^1 S(x)\left[a_0\theta\left(T_s - T\right) + a_1\theta\left(T - T_s\right)\right]dx \\ (6D+B)\,T_2 &= 5q\int_0^1 S(x)P_2(x)\left[a_0\theta\left(T_s - T\right) + a_1\theta\left(T - T_s\right)\right]dx \end{aligned} \tag{14.39}$$

which is identical to Eq. 14.26

The functional found is actually a surface above the plane $T_0$, $T_2$ and is represented in Fig. 14.8. We notice the points of the minima (solutions I and III) separated by a saddle point corresponding to the unstable solution II. We can understand a little better the nature of this instability if we write the rate of change of temperature:

$$C\frac{\partial T_n}{\partial t} = -\frac{\partial F}{\partial T_n}$$

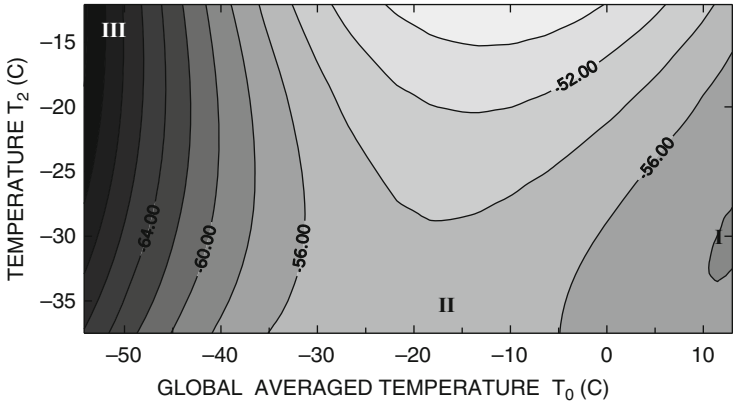



**Fig. 14.8** The potential function as a function of the global mean temperature $T_0$ and the second mode amplitude $T_2$. The values on the isopleths are normalized. Notice the minima, solutions I (*lower right*) and III (*upper left*), and the saddle point corresponding to solution II (*center*)

We have also

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial T_n}\frac{\partial T_n}{\partial t} = -C\left(\frac{\partial T_n}{\partial t}\right)^2 \tag{14.40}$$

As in the case of the zero-dimensional model, the rate of change for $F$ is always negative, so that a perturbation moving the system away from its equilibrium value will cause the system itself to go to lower values of $F$ until it reaches a new extreme. From this point of view the saddle point is clearly unstable.

### *14.3.3 The Sellers Model*

The solution obtained by North for the energy balance model is very elegant but has limitations because it includes all the fluxes in one single component. The model proposed by Sellers on the other hand separates the fluxes in to the different components. Again referring to Fig. 14.5, we obtain for the balance equation:

$$R_s = L\Delta c + \Delta C + \Delta F \tag{14.41}$$

where $R_s$ is the net radiative flux, $L$ the latent heat, $\Delta c$ the net flux of water vapor, and $\Delta C$ and $\Delta F$ the net flux of sensible heat and of the oceans, respectively. If $l_0$ indicates the length of the northernmost latitude circle and $l_1$ the same length for the southernmost boundary, the quantities appearing in Eq. 14.41 can be written as

$$\Delta F^i = \left(l_0 {F_0}^i - l_1 {F_1}^i\right)/A_0 \tag{14.42}$$

where $F^i$ indicates the generic flux and $A_0$ the surface between two latitude circles. The parameterization of the radiative term is rather simple because

$$R_s = q\,(1 - \alpha_s) - I_s \tag{14.43}$$

where the albedo is a function of the surface temperature

$$\begin{aligned} \alpha_s &= b - 0.009T_g \qquad T_g < 10C \\ \alpha_s &= b - 2.548 \qquad\quad T_g > 10C \end{aligned} \tag{14.44}$$

where $T_g$ is the surface temperature and $b$ is a parameter that is given in the program listed in the Appendix. The surface temperature is not the same as that of the corresponding latitude belt but takes into account the average elevation $Z$ of the land of that belt according to

$$T_g = T_0 - 0.0065Z \tag{14.45}$$

The parameter $b$ is chosen in such a way as to satisfy the experimental data. The infrared radiation on the other hand is a function of the temperature with a more complicated expression than the linear one by Budyko:

$$I_s = \sigma T_0^4 \left[1 - m \tanh\left(1.9 \cdot 10^{-16} T_0^6\right)\right] \tag{14.46}$$

where $\sigma$ is the Stefan–Boltzmann constant and $m$ is a quantity that takes into account the atmospheric transparency, equal to 0.5 for the present climate.

The flux parameterization takes into account the diffusion processes (due to the eddies) and the advection processes so that

$$\begin{aligned} c &= (vq - K_w \Delta q/\Delta y)\,\Delta p/g \\ C &= (vT_0 - K_h \Delta T/\Delta y)\left(\Delta p C_p\right)/g \end{aligned} \tag{14.47}$$

The quantity $\Delta p/g$ is an indication of the thickness of the atmosphere, while the velocities are parameterized based on the observed data:

$$\begin{aligned} v &= -a\left(\Delta T + \overline{\Delta T}\right) \qquad \text{north of } 5^\circ N \\ v &= -a\left(\Delta T - \Delta \overline{T}\right) \qquad \text{south of } 5^\circ S \end{aligned}$$

where the average value of the temperature difference is defined as

$$\overline{\Delta T} = \sum_{l=1}^{l=17} \left[\left(l_i \Delta T + l_{i+1} \Delta T\right) / \left(l_i + l_{i+1}\right)\right] \tag{14.48}$$

The quantity $q$ is the average specific humidity at sea level $q = \varepsilon e/p$ where $\varepsilon = 0.622$ and $p = 1000$ hPa, so that the transport term for the latent heat can be expressed directly through the Clausius–Clapeyron equation:

$$\Delta q = \frac{\varepsilon^2 L e_0 \Delta T}{p R_d {T_0}^2} \tag{14.49}$$

with $e_0$ the saturation pressure at the sea level

Finally we can express the sensible heat flux from the ocean given by

$$F = -K_0 \frac{\Delta z \Delta T}{\Delta y} \frac{l'}{l_1} \tag{14.50}$$

The ratio $l'/l_1$ is the fraction of the latitude circle covered by the ocean and $\Delta z$ is the average depth, while $\Delta y = 1.11 \cdot 10^8$ m. All the data that appear in the model formulation are given in the table at the Appendix together with the program listing.

The results of the Sellers model are given in Figs. 14.9 and 14.10. In the first one we notice that a difference with North's model is the asymmetry of the two hemispheres with respect to the albedo and temperature values. This asymmetry could be conserved also in North's model keeping the odd terms in Legendre polynomials. One of the characteristics of the Sellers model is to use a constant albedo when it becomes greater than 0.85. This happens in the southern hemisphere where the albedo in any case has larger values than in the northern hemisphere. This is an effect of the Antarctic continent having a larger extension than the Arctic ice. Figure 14.10 shows the sensible, latent, and oceanic heat fluxes. Although the behavior is rather similar to the data, the model underestimates the ocean flux and

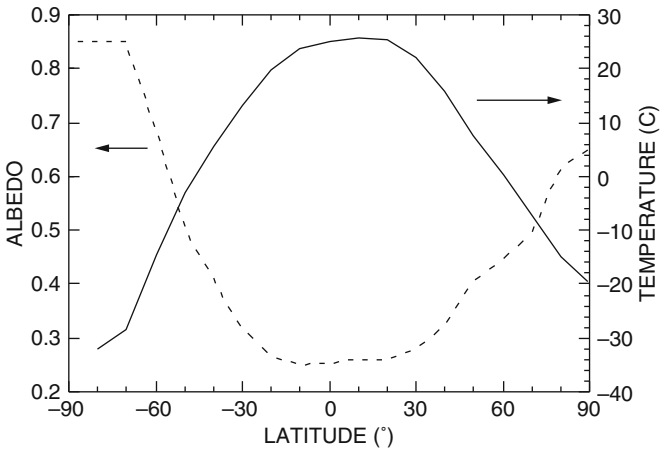



**Fig. 14.9** Albedo (*dashed line*) and temperature (*solid line*) as a function of latitude in the Sellers model. The negative value of the latitudes is for the southern hemisphere

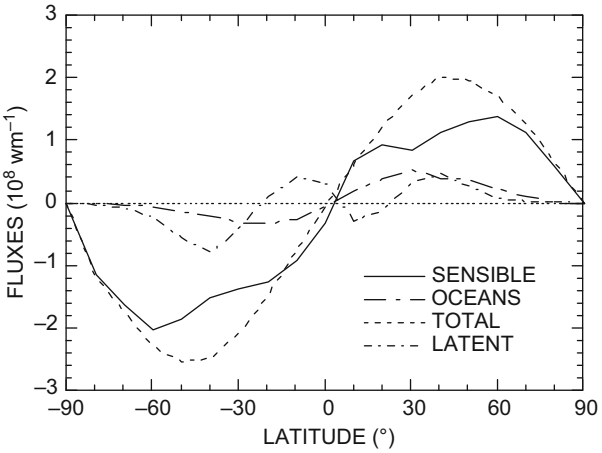



**Fig. 14.10** The sensible, latent, and oceanic fluxes as a result of the Sellers model. This figure can be compared with Fig. 11.8, although the units are not the same

this can be easily seen though a comparison with Fig. 11.8. This underestimation can be somewhat adjusted by changing the model parameters, but in any case it points out the intrinsic limitation of these models. Also, with the Sellers model, it is possible to estimate the sensitivity to changes in the solar constant, although in a way that is much less elegant than in North's model.

### *14.3.4 The Time Dependence of EBM*

If one maintains the Eq. 14.9 in its original form, we have

$$\partial T/\partial t = S_0\left(1-\alpha_p\right)/4 - (A+BT) \tag{14.51}$$

Redefining the temperature in degrees Kelvin and calling the solar term simply $Q$

$$CdT/dt = Q - [A + B\,(T-273)] \tag{14.51a}$$

where $C$ is the thermal capacity per unit area of the system. If we assume the model over land $C$ could be assumed to coincide with that of the atmospheric column $C \simeq \rho C_p H \simeq 1\cdot 10^3 \cdot 10^4 \simeq 10^7 JK^{-1}m^{-2}$. Over the ocean we must consider a much larger value. If we assume the ocean to be 100 m deep we have the corresponding capacity $C \simeq \rho C_0 100 \simeq 10^3 \cdot 4.18 \cdot 10^3 \cdot 100 \simeq 4.18 \cdot 10^8 JK^{-1}m^{-2}$ that is about 40 times as large. From Eq. 14.51 the equilibrium temperature is simply

$$T_{eq} = 273 + (Q-A)/B$$

If we assume that at time $t = 0$ the temperature differs from the equilibrium *T(0)* the solution for a generic time t is given by

$$T(t) = T_{eq} + \left[T(0) - T_{eq}\right] e^{t/\tau_0} \tag{14.52}$$

The constant $\tau_0$ is the ratio *C*/*B* and for land would be $\tau_0 \simeq 10^7/2.1 \simeq 5 \cdot 10^6 s$ that is about 55 days. If the initial temperature is warmer than the equilibrium, then the system will emit more radiation than absorbed from the sun $Q < [A + B(T\text{-}273)]$, and the system will cool.

Equation (14.51a) could be put in a quite different form if we assume the temperature to be given by the equilibrium value plus the small deviation T. Then the equation can be put in the form

$$C\frac{dT}{dt} = Q_{\text{net}} + BT$$

where $Q_{\text{net}}$ is just the difference between the absorbed flux and the emitted flux. In this relation the time constant would be the ratio C/B, and if we use the value for the ocean, the constant would be very long. We prefer then to use a value that corresponds to the one observed $\lambda = 15$ w m$^{-2}$K-1 and rewrite the equation as

$$C_O\frac{dT}{dt} - \lambda T = Q_{\text{net}} \tag{14.53}$$

where $C_O$ is the ocean heat capacity for unit area. This equation has a solution similar to (14.52) where the time constant is given by $\tau_0 = C_O/\lambda$ which is of the order of 300 d $\approx$ 10 months. A rather interesting application of (14.53) is to evaluate the frequency response of the climatic system. The idea is to compare the ocean response to short-term fluctuations of the atmosphere. We may imagine that those fluctuations are provided by the weather and assume that $Q_{\text{net}}$ in (14.53) is of the form

$$Q_{\text{net}} = \text{Re}\widehat{Q}_\omega e^{i\omega t}$$

where $\widehat{Q}_\omega$ is the amplitude of the stochastic component of the energy flux at frequency $\omega$ associated with the atmosphere (weather). We further assume that $\widehat{Q}_\omega$ is independent of frequency so that can be assimilated to a "white noise" when the amplitude is the same at all the frequencies. To solve the equation we assume a solution of the form $T = \text{Re}\widehat{T}_\omega e^{i\omega t}$ so that we have

$$C_O\widehat{T}_\omega i\omega - \lambda\widehat{T}_\omega = \widehat{Q}_\omega \quad \Rightarrow \quad \widehat{T}_\omega = \frac{\widehat{Q}_\omega}{C_O i\omega - \lambda}$$

and the solution gives the variance of the temperature.

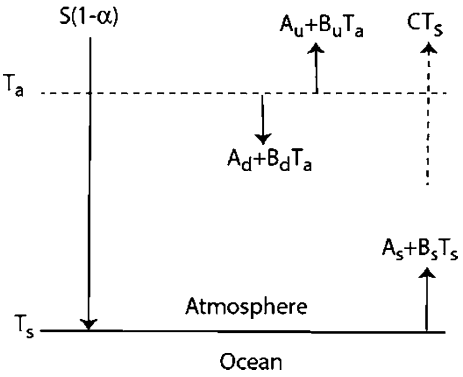



**Fig. 14.11** Scheme for a simple ocean-atmosphere energy balance model (From Vallis 2010)

$$\widehat{T}_\omega^2 = \left(\frac{\widehat{Q}_\omega}{C_O}\right)^2 \frac{1}{\omega^2 + (\lambda/C_O)^2} \tag{14.54}$$

The quantity $(\lambda/C_O)$ has the dimension of a frequency and the value of the constant is about $15/1.2 \cdot 10^7 = 0.1$ days$^{-1}$. We may consider the same model from another point of view following Vallis (2010). This mimics an energy balance model we saw as an example of the greenhouse effect and illustrated schematically in Fig. 14.11. We can write down the equation for the time dependence of the surface temperature and the atmospheric temperature

$$\begin{aligned} C_s \frac{\partial T_s}{\partial t} &= S\,(1-\alpha) + (A_d + B_d T_a) - (A_s + B_s T_s) \\ C_a \frac{\partial T_a}{\partial t} &= A_s + (B_s - C)\,T_s - (A_d + B_d T_a) - (A_u + B_u T_a) \end{aligned} \tag{14.55}$$

In this equations $C_a$ and $C_s$ are the heat capacities of the atmosphere and the surface, while constants A's and B's are somewhat determined empirically. If we assume an ocean mixed layer of 60 m depth, its $C_s \approx 25\,C_a$. The temperatures can be grouped to obtain

$$\begin{aligned} C_s \frac{\partial T_s}{\partial t} &= A_1 + B_d T_a - B_s T_s \\ C_a \frac{\partial T_a}{\partial t} &= A_2 + B_2 T_s - B_3 T_a \end{aligned} \tag{14.56}$$

where $A_1 = S(1-a) + A_d - A_s$, $A_2 = A_s - A_d - A_u$, $B_2 = B_s - C$ and $\mathrm{B}_3 = B_d + B_u$.

If we take into account the fluctuations, we can write $T_s = \overline{T}_s + T_s'$ and $T_a = \overline{T}_a + T_a'$ where $\overline{T}_s$ and $\overline{T}_a$ are the stationary solutions and we introduce a random forcing $\sigma \dot{W}$ on the right-hand side of the second of (14.56) representing the weather the equations becomes

$$C_s \frac{\partial T_s'}{\partial t} = B_d T_a' - B_s T_s' \qquad C_a \frac{\partial T_a'}{\partial t} = B_2 T_s' - B_3 T_a' + \sigma \dot{W} \tag{14.57}$$

These equation can be solved with simplifying assumption. As first hypothesis we could keep the sea surface temperature fixed so that $T_s' = 0$. In this case the equation for $T_a$ is simply

$$C_a \frac{\partial T_a'}{\partial t} = -B_3 T_a' + \sigma \dot{W} \tag{14.58}$$

Another way to couple the atmosphere and the ocean is to assume an ocean with zero heat capacity so that the ocean is slaved to the atmosphere. This is equivalent to assume the first of (14.57) to be zero and that $B_d T_a' = B_s T_s'$. This model may represent an atmosphere overlying a surface with small heat capacity so that $T_s^{'}$ can be substituted with $T_s' = B_d T_a'/B_s$ and the second Eq. 14.57 becomes

$$C_a \frac{\partial T_a'}{\partial t} = B_2 \left( \frac{B_d T_a'}{B_s} \right) - B_3 T_a' + \sigma \dot{W} = -B_4 T_a' + \sigma \dot{W} \tag{14.59}$$

with $B_4 = B_3 - B_2 B_d/B_s$. This is the same as Eq. 14.58 but $B_4$ being larger than $B_3$ the damping is reduced. As a matter of fact we have

$$B_4 = B_3 - B_2 B_d/B_s = B_d + B_u - (B_s - C)\, B_d/B_s = B_u + C B_d/B_s$$

which is positive and assures the system is always damped. Values for the parameter are $C = 0.54$ wm$^{-2}$ K$^{-1}$, $B_u = 2.83$ wm$^{-2}$ K$^{-1}$, $B_s = 10.4$ wm$^{-2}$ K$^{-1}$, and $B_d = 11.3$ wm$^{-2}$ K$^{-1}$ so that $B_3/B_4 \approx 4$.

The power spectrum of (14.58) can be calculated as we made before and obtain

$$P_a^U = \frac{\sigma^2}{B_3^2 + C_a^2 \omega^2} \tag{14.60}$$

where $P_a{}^U$ stands for the atmosphere uncoupled . This expression shows for high frequencies the spectrum is white (independent of frequency) and the transition from white to red spectrum happens for $\omega = B_3/C_a = 10^{-6}$ s$^{-1} \approx 0.1$ days$^{-1}$. This means that for time scales longer than a week, the spectrum is expected to be white. The power spectrum for (14.59) has the same form

$$P_a^{C0} = \frac{\sigma^2}{B_4^2 + C_a^2 \omega^2} \tag{14.61}$$

The superscript $C^0$ is indicating zero-capacity ocean. In this case the shoulder spectrum occurs for a time longer than 40 days.

Figure 14.12 illustrates these results.

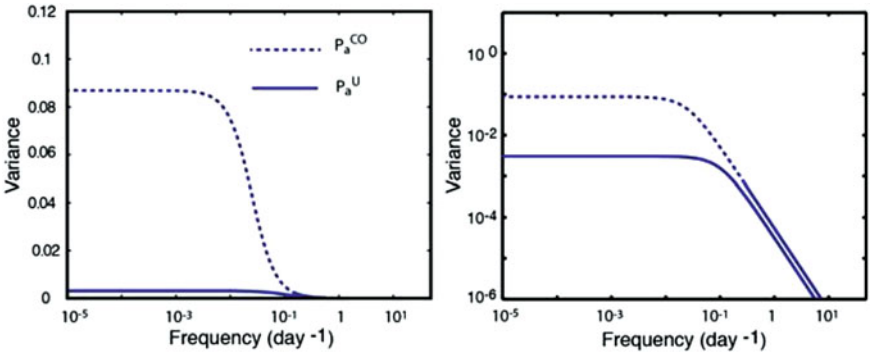



**Fig. 14.12** Power spectra of the model atmosphere obtained for atmosphere only ($P_a^U$) and a zero-capacity ocean ($P_a^{CO}$). The spectra are the same and only the variance scale is changed from linear to logarithm (Vallis 2010)

## 14.4 The Radiative–Convective Models

From the energy balance models, we have learned a few things, especially about the ice–albedo feedback. Actually, the Sellers model introduces a dependence of the water vapor content on the surface temperature that corresponds to another important feedback mechanism known as water vapor feedback. This mechanism can be better explored with a new class of models that include a rather detailed treatment of the infrared radiation.

In this case the passage from zero to one dimension is obtained by considering the vertical dimension. The approach is to divide the atmosphere roughly in two regions: the upper one, in radiative equilibrium, and the troposphere, with an assigned temperature profile consistent with the energy fluxes. As we have seen in Chap. 3 and in the budget sketched in Fig. 14.1, the radiative balance of the stratosphere is not consistent with the condition of nonthermal equilibrium at the ground

In the stratosphere the divergence of the radiative flux is zero so that the radiative flux is constant. Because at the top of the atmosphere there is perfect balance between the net solar flux and the outgoing infrared flux, the net flux (solar and longwave) through the whole stratosphere must be zero. At the surface however the net radiative fluxes are not zero because to keep the surface at thermal equilibrium latent and sensible heat must be considered. The net rate of the change of temperature is then given by (as we have seen in Chap. 3)

$$\frac{\partial T}{\partial t} = -\frac{1}{\rho C_p}\frac{d}{dz}(F_r + F_c) \tag{14.62}$$

where $F_r$ and $F_c$ indicate the radiative and convective flux, respectively.

The way in which the convective term is treated is model dependent and is completely parameterized. For example, we may assume that the flux is different

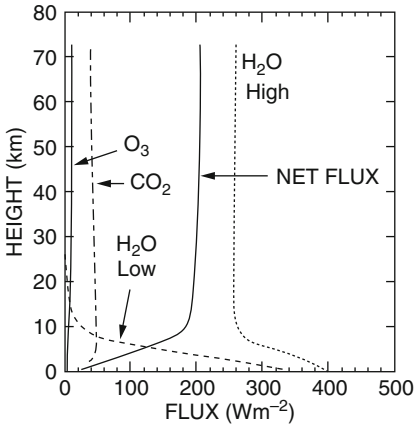



**Fig. 14.13** The infrared fluxes contributed by different gases. For water vapor the *downward* and *upward* fluxes are shown. For ozone and carbon dioxide, only the net fluxes are shown

from zero only when the lapse rate becomes larger than the adiabatic one. In this case the flux is assumed to be diffusive, that is, proportional to the difference between the two lapse rates (real and adiabatic):

$$S = \int_{z_1}^{z_2} \rho s dz = \frac{Cp}{g} \int_{p_2}^{p_1} \ln \theta dp \tag{14.63}$$

Again, the diffusion coefficient $K$ can be taken as a function of the lapse rate. Clearly we cannot expect that Eq. 14.62 could be solved analytically because the radiative flux divergence is a very complex term. The procedure is then to evaluate both the solar and the radiative flux following the method we outlined in the Chap. 13.

The result for the infrared fluxes is given in Fig. 14.13, where both the upward and downward fluxes are shown for water vapor together with the net fluxes for carbon dioxide and ozone. Notice that the net flux has a very rapid variation in the troposphere followed by an almost constant value. Notice also that most of the absorbed flux is due to water vapor. Although the absolute value of the $CO_2$ and $O_3$ fluxes are small with respect to $H_2O$, their contribution to the net flux and the heating rate is relevant. The net flux is evaluated as

$$\mathrm{F}_{\mathrm{net}}^{\mathrm{rad}} = \mathrm{F}_{\mathrm{H_2O}}^{\uparrow} - \mathrm{F}_{\mathrm{H_2O}}^{\downarrow} - \mathrm{F}_{\mathrm{CO_2}} - \mathrm{F}_{\mathrm{O_3}} + F_{\mathrm{net}}^{\mathrm{sw}} = F_{\mathrm{net}}^{\mathrm{lw}} + F_{\mathrm{net}}^{\mathrm{sw}} \tag{14.64}$$

where $F_{\mathrm{net}}^{sw}$ and $F_{\mathrm{net}}^{lw}$ indicate the solar flux and the net IR flux, respectively. From Equation 14.64 it is possible to obtain the radiative heating rate, given by

$$\left(\frac{dT}{dt}\right)_{\mathrm{rad}} = -\frac{1}{\rho C_p} \frac{dF_{\mathrm{net}}^{\mathrm{rad}}}{dz} \tag{14.65}$$

This quantity expressed in degrees per day is shown in Fig. 14.12, where also the heating due to the absorption of solar radiation is shown. The latter has been calculated using the method of Lacis and Hansen illustrated in the Chap. 13. The numerical results of these calculations are illuminating. They show that even if the absolute values of the ozone and carbon dioxide fluxes are small, they actually determine the heating rates (of the order of a few deg/day) in most of the stratosphere. The same figure shows that the heating and cooling rates balance in the stratosphere (radiative equilibrium), while in the troposphere the dominant effect is the cooling induced in the infrared by the water vapor. As we have said earlier, in this region there is no radiative equilibrium and the excess cooling must be balanced by the transport of latent and sensible heat from the surface. Integrating Eq. 14.65, we obtain

$$\frac{C_p}{g}\int_0^{p_s}(\partial T/\partial t)_{\text{rad}}dp = -\int_0^{\infty}\left(\partial F_{\text{net}}^{\text{rad}}/\partial z\right)dz = -\left[F_{\text{net}}^{\text{lw}}(\infty)+F_{\text{net}}^{\text{sw}}(\infty)\right] + \left[F_{\text{net}}^{\text{lw}}(0)+F_{\text{net}}^{\text{sw}}(0)\right]$$

where we have again used the hydrostatic approximation. Because at the top the net flux must be zero, at the ground we have

$$\frac{C_p}{g}\int_0^{p_s}(\partial T/\partial t)_{\text{rad}}dp = \left[F_{\text{net}}^{\text{lw}}(0)+F_{\text{net}}^{\text{sw}}(0)\right] \tag{14.66}$$

This means that at the surface the absorbed solar flux, decreased by the outgoing net IR flux, must be equal to the integral of the net heating of the entire atmosphere. We know however that the integral on the left-hand side is zero in the region in radiative equilibrium, so that the limit of integration may simply include the convective region. In this region we assume that the lapse rate has a fixed value that is denoted as *critical lapse rate*

$$(\partial T/\partial z)^{t+1} = \text{critical lapse rate} \tag{14.67}$$

and for each layer the temperature will be calculated explicitly

$$T^{t+1} = T^t + (\partial T/\partial t)_{net}^t \Delta t \tag{14.68}$$

The net heating rate $(\partial T/\partial t)_{\text{net}}^t$ is determined in such a way that for a layer in contact with the surface and delimited at the bottom by pressure $p_s$ and at the top by pressure $p_t$, we have (Fig. 14.14)

$$\int_{p_t}^{p_s}(\partial T/\partial t)_{\text{net}}dp = \int_{p_t}^{p_s}(\partial T/\partial t)_{\text{rad}}dp + g\left[F_{\text{net}}^{\text{lw}}(0)+F_{\text{net}}^{\text{sw}}(0)\right]/C_p \tag{14.69}$$

This means that, at least in this layer, the radiative flux in excess is compensated by the convective flux

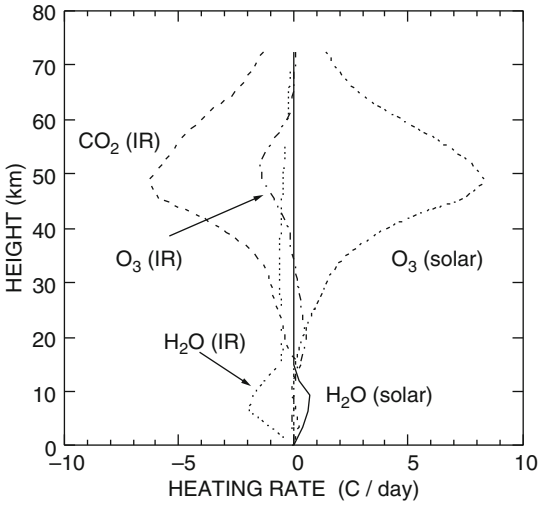



**Fig. 14.14** The contributions to the net radiative heating rate of the atmosphere. Notice that the *IR* contributes mostly to the cooling and the solar radiation mostly to the heating. Ozone in the *IR* warms the lower stratosphere

$$F_c = \left[C_p \left(T - T_{\text{rad}}\right) \Delta p / \left(g \Delta t\right)\right] \tag{14.70}$$

For a layer not in contact with the surface we must have the divergence of the total flux (convective and radiative) be zero. Based on Eq. 14.70 we have

$$\int_{p_b}^{p_t} (\partial T/\partial t)_{\text{net}} dp = \int_{p_b}^{p_t} (\partial T/\partial t)_{\text{rad}} dp$$

Finally, for a layer with a lapse rate lower than the critical one, the net heating rate coincides with the radiative heating rate.

This procedure (also called *convective adjustment*), when implemented on a numerical model, gives results like those shown in Fig. 14.15. This figure also shows the effects of a doubling of a $CO_2$ mixing ratio in the atmosphere. We notice that the calculations reproduce quite well the average temperature profile in the atmosphere. In particular we notice that the temperature maximum in the stratosphere is due to the ozone absorption, because the temperature will increase until the IR emission does not compensate for the solar radiation absorbed in the UV region.

When the carbon dioxide concentration increases, there is a corresponding increase of the infrared flux in the troposphere (and consequently a temperature increase), but also an increase of the infrared flux emitted in the stratosphere with an associated cooling. In practice, in that region the absorbed UV flux is compensated at a lower temperature.

The stratospheric cooling also implies another interesting feedback mechanism. A cooler stratosphere absorbs more infrared radiation coming from the troposphere so that the radiative compensation at the top requires a further increase of the surface temperature.

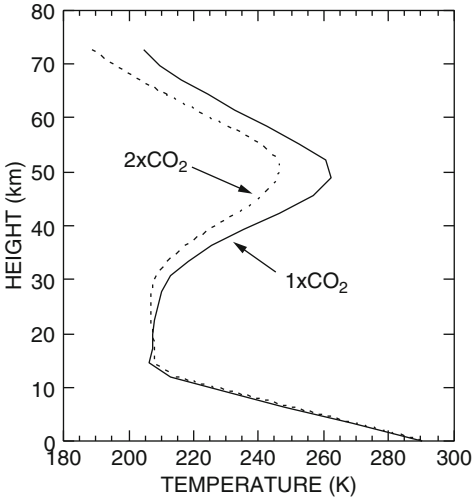



**Fig. 14.15** Temperature as a function of height calculated with a radiative–convective model for the present content of carbon dioxide and for twice the same value. Notice the warming of the troposphere and the cooling of the stratosphere

The origin of the radiative–convective models may be dated to the seminal work by Manabe and Strickler in 1967. For the first time, a quantitative estimation of the different feedback mechanisms was possible.

### *14.4.1 The Radiative–Convective Models and the Greenhouse Effect*

One of the striking results obtained with satellite observations concerns the proof of water vapor feedback and the amplification it has on the greenhouse effect. We should know by now that any effect that increases the temperature (like the increase in the concentration of $CO_2$) increases the saturation pressure for water vapor. Thus more water can evaporate into the atmosphere and increase the absorption of infrared radiation.

This mechanism was conceived or invented through the radiative–convective models but was never proved experimentally. The first measurement of this effect was reported by Raval and Ramanathan 1989 using the data of the ERBE (Earth radiation budget experiment) spaceborne experiment. The greenhouse effect $G$ was defined as the difference between the >> longwave flux at the surface $E$ and the outgoing flux at the top of the atmosphere $F$. Measurements were made over the ocean because the emissivity in that case is within 1 % of that of a blackbody. The surface flux is given by $E = \sigma T^4$, while the flux leaving the atmosphere is given by

$$F = B\,(T_s) - \int_0^1 A(x)\,[dB/dx]\,dx \tag{14.71}$$

In this equation $x$ indicates the ratio between the pressure $p$ and the pressure at the surface $p_s$, and $A(x)$ is the absorptivity between the top of the atmosphere and pressure $p$, while $B$ is the Planck function such that

$$\sigma T^4 = \int B_\lambda(T) d\lambda .$$

Keeping in mind the definition of greenhouse effect, we obtain

$$G = E - F = \sigma {T_s}^4 - B\left(T_s\right) + \int_0^1 A(x)\left[dB/dx\right] dx = 4\sigma \int_0^1 A(x) T(x)^3 \left[dT/dx\right] dx \tag{14.72}$$

The main contribution to this integral is from the troposphere, where the condition is that the temperature decreases with altitude, that is, $dT/dx > 0$. The experimental data show that the global mean annual average greenhouse effect is about 146 W m$^{-2}$ for clear sky conditions, while for a cloudy sky the average is about 179 W m$^{-2}$ of which 33 W m$^{-2}$ are due to the clouds. The outgoing flux for clear sky conditions can be approximated by a linear relationship similar to the one used for the energy balance models:

$$F = 229.36 + 2.31\left(T_s - 273.15\right) \tag{14.73}$$

We can normalize then the greenhouse effect in Eq. 14.72 and obtain a relation linear with the temperature:

$$g = G/\sigma T_s^4 = \frac{4}{T_s^4} \int_0^1 A(x)\left[dT/dx\right] dx = -0.658 + 3.42 \cdot 10^{-3} T_s \tag{14.74}$$

At this point we need to establish that dependence of the greenhouse effect on the temperature must be attributed to an increase in the absorptivity that in turn results from an increase in the water vapor content of the troposphere. Actually the integral Eq. 14.74 does not depend explicitly on the temperature, and assuming an adiabatic gradient we have

$$\left[T(p)/T_s\right] = x^\alpha$$

where $\alpha = \Gamma R/g_a$ so that the integral becomes

$$g = 4\alpha \int_0^1 A(x) x^\beta dx \tag{14.75}$$

where $\beta = (4\alpha - 1)$. Written in this form the temperature dependence can arise either from the lapse rate or the absorptivity. However, using different values of the lapse rate coming from different regions of the Earth gives for $g$ a range of variability

of only 10 %, while for fixed values of the water vapor content, the absorptivity may change only 5 %. The only viable explanation for the increase of the greenhouse effect with temperature is because the water vapor increases.

We can easily see this effect if we consider that the saturation pressure changes with the temperature according to $\exp[-(L/RT)]$, where $L$ is the latent heat. Thus, it is possible to calculate the change in the total water vapor content $W$ (kg m$^{-2}$) as a function of the temperature and the result is a linear relation between the logarithm of $W$ and the temperature

$$d\ln W/dT_s = 6.7 \cdot 10^{-2}\mathrm{K}^{-1}$$

The amount of water vapor can be measured with other satellites (ERBE unfortunately could not gather the data) so that a very similar relation is found: $dlnW/dT_s = 5.5 \cdot 10^{-2}$ K$^{-1}$. This relation enables us to write the dependence of the greenhouse effect on the water vapor, considering that both $A(x)$ and the lapse rate have a weak dependence on the surface temperature:

$$\frac{dg}{dT_s} = \frac{\partial g}{\partial \ln W}\frac{\partial \ln W}{\partial T_s} \tag{14.76}$$

from which we get $\partial g/\partial \ln W = 6.2\ 10^{-2}$. This is actually determined experimentally and may give us an indication of how to evaluate the water vapor feedback in case of an increase of carbon dioxide concentration.

The models and spectroscopic calculations show that for a doubling of $CO_2$ from 345 ppm to 690 ppm the net outgoing IR flux would decrease by 4 W m$^{-2}$ so the greenhouse effect would increase by an equivalent amount. In the absence of the water vapor feedback, the surface temperature should increase as to keep the outgoing flux from the atmosphere at its present mean value 240 W m$^{-2}$. The relation between the fractional change of the flux and the temperature for a blackbody

$$\Delta F/F = 4\Delta T_s/T_s$$

so that for $\Delta F = 4$ w m$^{-2}$ we have $\Delta T_s = 1.2$ K. The experimental data show $\Delta F/\Delta T_s = 2.3$ Wm$^{-2}$ K$^{-1}$ (see Eq. 14.74) so that we have now for the temperature increase $\Delta T_s = 1.7$ K. The inclusion of the water vapor feedback gives a further increase of 0.5 K in the temperature.

As for the greenhouse effect, if its value should remain constant at $g = 0.327$, then the absorbed radiation would increase by

$$\Delta G = 4g\sigma T_s^3 \Delta T_s = 2.1\mathrm{w\,m}^{-2}$$

Keeping in mind the change of $g$ with the surface temperature, the greenhouse effect would increase by 3.1 W m$^{-2}$ which should be added to the 4 W m$^{-2}$ due to the initial $CO_2$ forcing. This comparison is made in Fig. 14.16 both with a spectroscopic

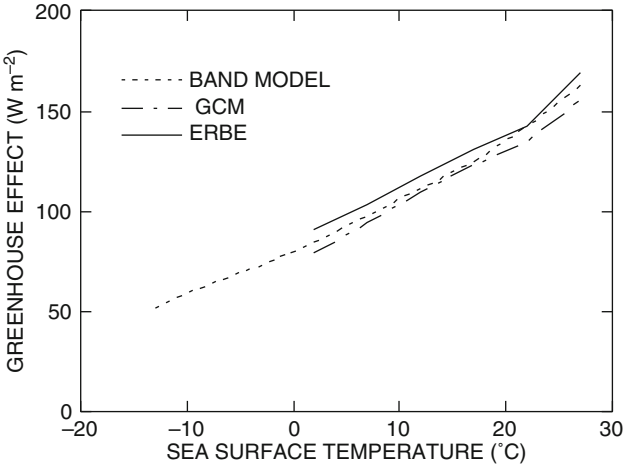



**Fig. 14.16** The comparison between the measured greenhouse effect and the calculations. The band model extends to temperatures that do not make sense for the ocean (Raval and Ramanathan 1989)

model and the flux calculated with a general circulation model. We notice a rather good agreement although the general circulation model shows an underestimation of about 10 W m$^{-2}$ which is mainly due to the fact that in this model the contribution of nitrous oxide is not included. The agreement with the experimental data is confirmed by a comparison with 13 GCM (see also next Chap. 15) which gives $dF/dT_s = 2.3 \pm 0.2 \mathrm{W\,m^{-2}} K^{-1}$.

### *14.4.2 Can We Put Together the Radiative–Convective and Energy Balance Climate Models?*

Sure we can. As we have seen, a radiative–convective model uses global average values for both the albedo and the lapse rate. Such a model can incorporate not only the water vapor feedback but also the ice–albedo feedback. To this end it would be enough to have a relation that connects the global averaged temperature to the albedo value and the latitude of the ice line. In the practical application during the time marching integration we will change the albedo value.

The relation between global temperature and albedo is given by Eq. 14.24, so that the only difficulty is to find a function that gives the dependence of $H_0(x_s)$ on the temperature. This can be done by approximating this function with a second degree polynomial

$$H_0(x_s) = 0.3763 + 0.5021 x_s - 0.1706 x_s^2 \tag{14.77}$$

The procedure is to start with some value of the albedo and then calculate the global temperature through a radiative–convective model. This temperature enables us to evaluate the albedo through Eq. 14.23 and with Eq. 14.77 also the latitude. We need to notice however that the albedo in the two kinds of models is a little different. In the energy balance models, the albedo refers to the total surface–atmosphere value. In the radiative–convective models the surface reflectivity is fixed, and the total results mainly from the clouds used in the model. At this point things get a little confusing. The ice–albedo feedback is based on the relation between the average global temperature and albedo: the lower the temperature, the higher the albedo. However the albedo depends very strongly on the cloudiness and it is intuitive to think that a colder Earth should imply less atmospheric water vapor and then clouds. On the other hand it is also obvious that for the ice to grow it is necessary to have precipitation and then clouds.

The inescapable conclusion is that our simple climate models are much too simple to study such complex problems.

## E.14 Examples

### *E.14.1 Stability of North's Model*

We start from Eq. (14.10) and consider the additional equation

$$B\frac{\partial T_0^0}{\partial q} = H_0 + q\left(\frac{\partial H_0}{\partial T_0^0}\right)\frac{\partial T_0^0}{\partial q} \tag{E.14.1}$$

From this equation, we obtain $B$

$$B = H_0\left(\frac{\partial T_0^0}{\partial q}\right)^{-1} + q\left(\frac{\partial H_0}{\partial T_0^0}\right)$$

which is substituted in the Eq. (14.10) to get

$$\frac{\partial T'}{\partial t} + H_0\left(\frac{\partial T_0^0}{\partial q}\right)^{-1} T' + q\left(\frac{\partial H_0}{\partial T_0^0}\right)T' = q\left(\frac{\partial H_0}{\partial T_0^0}\right)T'$$

and then

$$\frac{\partial T'}{\partial t} + H_0\left(\frac{\partial T_0^0}{\partial q}\right)^{-1} T' = 0 \tag{E.14.2}$$

There is an alternative way to reach the same conclusion and then this to express again the temperature as the solution plus a small deviation:

$$T_0(t) = T_0^0 + T'(t)$$

then from the (14.10) we get

$$\frac{\partial T'}{\partial t} = \left[-BT' + q\left(\frac{dH_0}{dT_0^0}\right)\right] T' \tag{E.14.3}$$

From the equilibrium equation differentiating q with respect to $T_0^0$, we obtain

$$\left[H_0\left(T_0^0\right)\right] q = \left(A + BT_0^0\right) \quad \Rightarrow \quad \left[H_0\left(T_0^0\right)\right] \frac{\partial q}{\partial T_0^0} + q\frac{dH_0}{dT_0^0} = B$$

Substituting in (E.14.3)

$$\frac{\partial T'}{\partial t} = \left[-\left(\left[H_0\left(T_0^0\right)\right] \frac{\partial q}{\partial T_0^0} + q\frac{dH_0}{dT_0^0}\right) T' + q\left(\frac{dH_0}{dT_0^0}\right)\right] T' = -H_0 \frac{\partial q}{\partial T_0^0} T'$$

That is the same as (E.14.2).

## *E.14.2 Time-Dependent Solution of North's Model*

We start from the time-dependent equation for the model

$$C\frac{\partial T}{\partial t} = aQ(x) - (A + BT) + D\frac{\partial}{\partial x}\left[\left(1 - x^2\right)\frac{\partial T}{\partial x}\right] \tag{E.14.4}$$

redefining the quantities

$$q(x) = aQ(x)/C \qquad A = A/C \qquad B = B/C \qquad D = D/C$$

we have

$$\frac{\partial T}{\partial t} = q(x) - (A + BT(x)) + D\left(1 - x^2\right)\frac{\partial^2 T}{\partial x^2} - 2Dx\frac{\partial T}{\partial x}$$

This equation can be transformed in finite difference equation

$$T_j^{k+1} - T_j^k = q_j \Delta t - \left(A + BT_j\right)\Delta t + D\left(1 - x_j^2\right)\frac{T_{j+1}^{k+1} - 2T_j^{k+1} + T_{j-1}^{k+1}}{\Delta x^2}\Delta t - D\frac{T_{j+1}^{k+1} - T_{j-1}^{k+1}}{\Delta x}\Delta t \tag{E.14.5}$$

And grouping the different terms

$$T_{j+1}^{k+1}\left[\frac{\Delta t}{\Delta x^2}D\left(1-x_j^2\right)-Dx_j\frac{\Delta t}{\Delta x}\right]-T_j^{k+1}\left[1+2\frac{\Delta t}{\Delta x^2}D\left(1-x_j^2\right)+B\Delta t\right]$$
$$-T_{j-1}^{k+1}\left[\frac{\Delta t}{\Delta x^2}D\left(1-x_j^2\right)+Dx_j\frac{\Delta t}{\Delta x}\right]=T_j^k+q_j\Delta t-A\Delta t+$$

That can be written in the standard form

$$-A_jT_{j-1}+B_jT_j-C_jT_{j+1}=R_j \tag{E.14.6}$$

where

$$A_j=\frac{\Delta t}{\Delta x^2}\left(1-x_j^2\right)D+Dx_j\frac{\Delta t}{\Delta x}\qquad B_j=1+2\frac{\Delta t}{\Delta x^2}D\left(1-x_j^2\right)+B\Delta t$$
$$C_j=\frac{\Delta t}{\Delta x^2}D\left(1-x_j^2\right)+Dx_j\frac{\Delta t}{\Delta x}\qquad Rj=T_j^k+q_j\Delta t-A\Delta t$$

Difference equations like (E.14.6) can be solved with standard method (see, e.g., Richtmyer and Morton), and the attached program uses a standard routine (tridag). Below we report a possible FORTRAN program with routine inversion taken from Numerica Recipes:

```
dimension b(41),c(41),r(41),xs(41),qs(41),p2(41)
dimension q(41),an(41), bn(41), cc(41)
dimension u(41), ppm(14),ppmac(131)
data ppm/325,337,353,369,391,420,454,491,532,491,454,420,
 391,369/
itime=0.0
 iwrite=0.0
 s=1366.
 aw=206.85
 bw=1.95
 diff=.445
 cp=1.046e9
 dt=3.15e7
 dx=0.05
 do i=1,41
 xs(i)=1.0-(i-1)*dx
 end do
 do i=1,41
 p2(i)=(3*xs(i)*xs(i)-1)/2.
 u(i)=14.5-28.4*p2(i)
 qs(i)=0.25*s*(1.-0.482*p2(i)
 b(i)=diff*(1-xs(i)*xs(i))*dt/(cp*dx*dx)
 c(i)=diff*xs(i)*dt/(cp*dx)
```

```
      end do
      do k=1,13
      ind1=(k-1)*10+1
      ind2=ind1+9
      dppm=(ppm(k+1)-ppm(k))/10.
      do j=ind1, ind2
      ppmac(j)=ppm(k)+dppm*(j-ind1)
      end do
      end do
      ppmac(131)=ppm(14)
20    continue
      if(itime.lt.200)then
      goto 211
      else
      itempo=itime-199.
      ppmco2=ppmac(itempo+1)
      write(6,*) ppmco2,itempo
      phi=alog(ppmco2/300)
      prea=-326.4+9.161*phi-3.164*phi*phi+0.5468*phi*phi*phi
      preb=1.953-0.04866*phi+0.01309*phi*phi-0.002577*phi*phi*phi
      aw =preb*273.15+prea
      bw=preb
      end if
211   continue
      do i=1,41
      an(i)=-b(i)+c(i)
      bn(i)=1+2.*b(i)+bw*dt/cp
      cc(i)=-(b(i)+c(i))
      end do
      cc(1)=-(an(1)/1000+cc(1))
      an(41)=-(an(41)+cc(41)/1000)
      an(1)=0.
      cc(41)=0.
      open(10,file='geoengdata', status='unknown')
101   format(5f12.3)
      do i=1,41
      if(u(i).lt.-10.) then
      q(i)=qs(i)*0.38
      else
      q(i)=qs(i)*(0.697-0.0779*p2(i))
      end if
      r(i)=u(i)+(q(i)-aw)*dt/cp
      end do
1012  format(5e12.4)
1013  format(1e12.3)
```

```
      call tridag(an, bn, cc, r,u, 41)
      itime=itime+1
      iwrite=iwrite+1
      if(itime.lt.200) goto 1001
      if(iwrite.lt.10) goto 1001
      write(10,21)
21    format(10x, 'iwrite, itempo, aw, bw, ppmco2 '/)
       write(10,2100) iwrite, itempo, aw, bw, ppmco2
       2100  format(2i10, 3f10.4)
       write(10,100)(xs(i),u(i),q(i), i=1,41)
       iwrite=0.0
1001   if(itime.lt.331) goto 20
100    format(2f10.5,f12.3)
166    format(2f10.5)
       stop
        subroutine tridag(a,b,c,r,u,n)
        integer n,nmax
          real a(n),b(n),c(n),r(n),u(n)
          parameter (nmax=500)
          integer j
          real bet,gam(nmax)
          if(b(1).eq.0.)pause 'tridag: rewrite equations'
          bet=b(1)
          u(1)=r(1)/bet
          do 11 j=2,n
          gam(j)=c(j-1)/bet
          bet=b(j)-a(j)*gam(j)
          f(bet.eq.0.)pause 'tridag failed'
          u(j)=(r(j)-a(j)*u(j-1))/bet
11        continue
          open(66, file='routine', status='unknown')
          do 12 j=n-1,1,-1
          u(j)=u(j)-gam(j+1)*u(j+1)
12        continue
         return
         end
```

### *E.14.3 Temperature Profile from Maximum Entropy Principle*

We want to mention here a rather unknown approach about the calculation of temperature profile of the atmosphere. It assumes that the atmosphere will settle to the state of maximum entropy production. The approach used by Verkley and Gerkema (2004) starts by considering the entropy calculation for an atmospheric layer between levels $z_1$ and $z_2$ given by

$$S = \int_{z_1}^{z_2} \rho s \, dz = \frac{C_p}{g} \int_{p_2}^{p_1} \ln \theta dp \tag{E.14.7}$$

Then we evaluate the total mass

$$M = \int_{z_1}^{z_2} \rho \, dz = \frac{1}{g} \int_{p_2}^{p_1} dp = \frac{p_1 - p_2}{g}$$

The total energy (internal plus potential energy)

$$E = \int_{z_1}^{z_2} \rho \left(C_v T + gz\right) dz = p_1 z_1 - p_2 z_2 + \frac{C_p}{g} \int_{p_2}^{p_1} T \, dp \tag{E.14.8}$$

If we change the limit of the layer by amounts $\delta z_1$, $\delta z_2$ without changing the pressure levels, the work done by the system will be $\delta W = p_2 \delta z_2 - p_1 \delta z_1$ and so

$$\delta E + \delta W = \frac{C_p}{g} \int_{p_2}^{p_1} T \, dp \tag{E.14.9}$$

Note that $\delta E + \delta W = \delta Q$ is the heat added to the atmospheric column. If in the variational process no heat is added or subtracted to the column, the enthalpy

$$H = \frac{C_p}{g} \int_{p_2}^{p_1} T \, dp \tag{E.14.10}$$

will stay constant. Further we assume that the quantity

$$L = \frac{C_p}{g} \int_{p_2}^{p_1} \theta \, dp \tag{E.14.11}$$

We can now formulate a variation principle such as to maximize S keeping M, L, and H fixed. We write the equation for the Lagrangian multipliers $\lambda$ and $\mu$ so that

$$\delta S + \lambda \delta H + \mu \delta L = 0 \tag{E.14.12}$$

We may express all the quantity appearing terms of T and p. We have

$$S = \frac{C_p}{g} \int_{p_2}^{p_1} \ln T \, dp + \frac{C_p}{g} \int_{p_2}^{p_1} \ln \left(\frac{p_s}{p}\right)^k dp \quad \Rightarrow \delta S = \frac{C_p}{g} \int_{p_2}^{p_1} \frac{\delta T}{T} dp$$

$$\delta H = \frac{C_p}{g} \int_{p_2}^{p_1} \delta T dp$$

$$\delta L = \frac{C_p}{g}\int_{p_2}^{p_1}\delta\theta dp = \frac{C_p}{g}\int_{p_2}^{p_1}\left(\frac{p_s}{p}\right)^k \delta T dp$$

and the variation will be

$$\int_{p_2}^{p_1}\left[\frac{\delta T}{T} + \lambda\delta T + \mu\left(\frac{p_s}{p}\right)^k\right]dp = 0$$

Putting $\alpha = \lambda/\mu$ we have

$$\frac{1}{\mu}\frac{1}{T} + \alpha + \left(\frac{p_s}{p}\right)^k = 0 \;\Rightarrow\; -T = \left[\lambda\left(1 + \alpha\left(\frac{p_s}{p}\right)^k\right)\right]^{-1}$$

To determine $\lambda$ we impose that $T = T_s$ for $p = p_s$ so that

$$T = T_s\frac{1+\alpha}{1+\alpha(p_s/p)^k} \tag{E.14.13}$$

The temperature profile goes from isothermic ($\alpha = 0$) to the isentropic for ($\alpha \to \infty$).

### *E.14.4 Entropy Production and Energy Balance Models*

The atmosphere and the ocean of the Earth can be considered as part of a heat machine that uses a high temperature source (tropics) and release it in region at lower temperature (middle latitudes and polar regions). We can assume a system like the one shown in Fig. E.14.1 where we consider two boxes, one extending to latitude of 30° at temperature $T_0$ and the other of the same horizontal surface at temperature $T_1$. Between the two boxes, a diffusive flux is established proportional to the temperature difference between the two boxes (Lorenz et al. 2001).

$$F = 2D\,(T_0 - T_1) \tag{E.14.14}$$

Where D is the diffusion coefficient measured in w $m^{-2}$ $K^{-1}$. In North's model the diffusion coefficient is of the order of 0.6 w $m^{-2}$ $K^{-1}$. In the same model the flux is written in spherical coordinates:

$$F = -D\left(1 - x^2\right)^{1/2}\frac{dT}{dx}$$

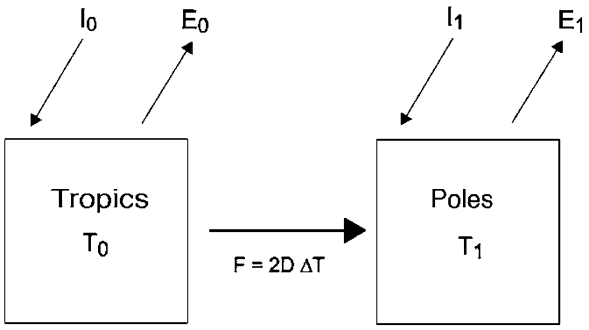



**Fig. E.14.1** The simple box model

Assuming the distance between the two boxes to be $\Delta x = 0.5$, we obtain (E.14.14). If we write the same expression as a function only of latitude we get

$$F = -D\frac{dT}{d\varphi} \tag{E.14.15}$$

The flux of sensible heat is given by

$$F_s = -\frac{\partial}{\partial y}\left(\rho c_p v T\right) \tag{E.14.16}$$

We have the flux in w $m^{-2}$

$$F = -H\frac{\partial}{a\partial\varphi}\left(\rho c_p v T\right) \approx -\frac{\rho c_p v H}{a}\frac{\partial T}{\partial\varphi} \tag{E.14.17}$$

When compared with (E.14.15), it gives us the diffusion coefficient:

$$D \approx -\frac{\rho c_p v H}{a} \tag{E.14.18}$$

Using $v \approx 1\,\mathrm{m\,s^{-1}}$, $\rho = 1.25\,\mathrm{kg\,m^{-3}}$, $c_p = 1000\,\mathrm{J\,kg^{-1}K^{-1}}$, $H = 8\,\mathrm{km}$, $a = 6370\,\mathrm{km}$

we have $D \approx 1.56$ w $m^{-2}$ $K^{-1}$ in the right range.

If the two boxes absorb energy from the sun $I_0$ and $I_1$ and emits infrared radiation $E_0$ and $E_1$, the energy balance requires

$$I_0 - E_0 - F = 0 \qquad I_1 - E_1 + F = 0 \tag{E.14.19}$$

Now we can discuss the entropy production and we can write the change in entropy:

$$\frac{dS}{dt} = F\left(\frac{1}{T_1} - \frac{1}{T_0}\right) \tag{E.14.20}$$

If the flux is zero the entropy production is zero and the planet is in radiative equilibrium. If the heat flux is maximized, then $T_1 = T_0$, and the planet is isothermal, and the entropy production is still zero. Between these two states there is the possibility that the entropy production is not zero and has a single, maximum value.

In order to apply the principle of maximum entropy production (MEP), we need to express the emission fluxes as a function of temperature:

$$E_i = A + BT_i \tag{E.14.21}$$

Where the $B$ coefficient can be calculated with the Eddington approximation. With $\tau$ indicating the optical thickness we have

$$T^4 = \frac{T_e^4}{2}\left(\frac{3}{2}\tau + 1\right) \tag{E.14.22}$$

where $T_e$ is the emission temperature and gives the outgoing flux

$$E(0) = \sigma T_e^4$$

The outgoing flux is related to the upward flux by the following relation

$$F^{\uparrow}(\tau) = \frac{1}{2}\sigma T_e^4\left(\frac{3}{2}\tau + 2\right) = \frac{E(0)}{2}\left(\frac{3}{2}\tau + 2\right) \tag{E.14.23}$$

And from this it is possible to obtain $E(0)$ as

$$E(0) = F^{\uparrow}(\tau) \Big/ \left(\frac{3}{4}\tau + 1\right) = \sigma T^4 / (0.75\tau + 1)$$

And then from the definition of $B$ we have

$$B = 4\sigma T^3 / (0.75\tau + 1) \tag{E.14.24}$$

We can now determine the value for $D$ maximizing the entropy production. Subtracting (E.14.15) we have

$$I_0 - I_1 - B(T_0 - T_1) = 2F = 4D\Delta T \tag{E.14.25}$$

From which we get the temperature difference $\Delta T = T_0 - T_1$

$$\Delta T = \Delta I / (4D + B) \tag{E.14.26}$$

We now combine (E.14.20) with the definition of flux to obtain

$$\frac{dS}{dt} = 2D\Delta T\left(\frac{1}{T_1} - \frac{1}{T_0}\right) = 2D\Delta T^2/\left(T_1 T_0\right)$$

And substituting $\Delta T$ from (E.14.22) obtain

$$\frac{dS}{dt} = 2D\Delta I^2/\left[\left(T_1 T_0\right)(4D + B)^2\right] \tag{E.14.27}$$

Which derived with respect to $D$ gives the maximum of entropy production for $D \approx B/4$

Before this model was formulated, *G .W.* Paldridge demonstrated that based on the minimum entropy production, he could obtain the temperature as a function of latitude. If we indicate with $F_s(x)$ the absorbed solar flux as a function of latitude and with $F_l(x)$ the emitted longwave flux, the total entropy production could be assumed as

$$\frac{dS}{dt} \propto \int_{-1}^{1} \frac{F_s(x) - F_l(x)}{T_l(x)} dx$$

where $T_l(x)$ is the temperature as a function of latitude (with $x$ sin(latitude)). Because the emitted radiation is related to temperature, the minimization process could simply involve the quantity

$$\int_{-1}^{1} \frac{F_s(x) - F_l(x)}{F_l(x)} dx = \int_{-1}^{1} \left(\frac{F_s(x)}{F_l(x)} - 1\right) dx \tag{E.14.28}$$

Using the Lagrange calculator the minimization process is equivalent to

$$\frac{\partial}{\partial F_l}\left[\frac{F_s(x)}{F_l(x)} - 1 - \lambda\left(F_s(x) - F_l(x)\right)\right] = 0 \tag{E.14.29}$$

with a Lagrange multiplier. After performing the derivative we get

$$-\frac{F_s(x)}{F_l(x)^2} + \lambda = 0 \tag{E.14.30}$$

With the solution

$$\frac{F_s(x)}{\lambda} = F_l(x)^2 \tag{E.14.31}$$

the global radiative balance requires that

$$\int_{-1}^{1} [F_s(x) - F_l(x)]\, dx = 0$$

after substituting (E.14.27) we have

$$\int_{-1}^{1} F_s(x) dx = \int_{-1}^{1} \left[\frac{F_s(x)}{\lambda}\right]^{1/2} dx$$

and defining the averages with $\langle\,\rangle$ we have

$$\langle F_s(x)\rangle\, \lambda^{1/2} = \left\langle F_s(x)^{1/2} \right\rangle \Rightarrow \langle F_l(x)\rangle = \left\langle \lambda^{1/2} F_s(x)^{1/2} \right\rangle \tag{E.14.32}$$

From the first equivalence when $F_s$ is known it is possible to obtain $\lambda$ and then $F_1$ from (E.14.31). This conclusion could have been reached using Eqs. E.14.19 and E.14.20 with the assumption that $E \propto T$. The second could be rewritten

$$\frac{dS}{dt} = \left(\frac{E_1 - I_1}{E_1} - \frac{I_0 - E_0}{E_0}\right) = 2 - \frac{I_1}{E_1} - \frac{I_0}{E_0}$$

If this is derived with respect to $E_1$ and maximized we get

$$\frac{dE_1}{dE_0} = -\frac{I_0}{I_1}\left(\frac{E_1}{E_0}\right)^2$$

And the derivative could be obtained from $I_0 - E_0 = E_1 - I_1$ and that $dE_1/dE_0 = -1$ so that

$$\frac{I_1}{I_0} = \left(\frac{E_1}{E_0}\right)^2$$

This result is easily verified. The results of the MEP approach are impressive, but a more accurate discussion shows that the extension to other planets other than Earth poses some problem. For a critical discussion, the work of Goody (2007) is very useful.

# References[1]

## *Books*

Marshall J, Plumb RA (2008) Atmosphere ocean and climate dynamics. Academic, Amsterdam
Palmer T, Williams P (eds) (2010) Stochastic physics and climate modelling. Cambridge University Press, Cambridge

## *Articles*

Budyko MI (1969) The effect of solar radiation variations on the climate of the earth. Tellus 21:611
Crowley TJ, North GR (1991) Paleoclimatology. Oxford University Press, New York
Goody R (2007) Maximum entropy production in climate models. J Atmos Sci 64:2725
Lorenz RD, Lunine JI, Withers PG, McKay CP (2001) Titan, Mars, and Earth: entropy production by latitudinal heat transport. Geophys Res Lett 28:415–418
Manabe S, Wetherald RT (1967) Thermal equilibrium on the atmosphere with a given distribution of relative humidity. J Atmos Sci 24:241
North GR (1975) Theory of energy-balance climate models. J Atmos Sci 32:2033
North GR, Cahalan RF, Coakley JA (1981) Energy balance climate models. Rev Geophys Space Phys 19:91
Paltridge GW (1975) Global dynamics and climate- A system of minimum entropy exchange. Quart J Royal Meterol Soc 101:475
Pujol T, Fort J (2002) States of maximum entropy production in a one dimensional vertical model with convective adjustment. Tellus 54A:363
Ramanathan V, Coakley JA (1978) Climate modeling through radiative convective models. Rev Geophys Space Phys 16:465
Raval A, Ramanathan V (1989) Observational determination of the greenhouse effect. Nature 342:758
Rodgers CD (1976) Comments on Paltridge minimum entropy exchange principle. Quart J Royal Meterol Soc 102:455
Sellers WD (1969) A global climate model based on the energy balance of the earth - atmosphere system. J Appl Meteorol 8:392
Vallis GK (2010) Mechanisms of climate variability from years to decades. In: Palmer T, Williams P (eds) Stochastic physics and climate modelling. Cambridge University Press, Cambridge
Verkley WTM, Gerkema T (2004) On maximum entropy profiles. J Atmos Sci 61:931–936

[1]This chapter is based almost entirely on review or research papers. Each one of them is very useful and instructive. There is a new paragraph on the time dependence of EBM. Some of the examples at the end of the chapter deal with the maximum entropy principle.

# Chapter 15
# The Application of Climate Models

The actual title of this chapter could be the application of simple climate models to the great climatic events. In the previous chapters, we have acquired tools that should enable us to study the evolution of climate on the Earth (and also on the other planets), at least for the most important features. A typical example is the ice age problem, which helps us to understand the complexity of the climate system and the nature of the different forcing and feedback. However, considering the ongoing debate on global warming, some attention is also given to main tool of the practitioners that is General Circulation Models (GCM).

## 15.1 The Climate System

The climate is defined normally as an average of weather in one particular region. The average must be performed for periods long enough with respect to the time of predictability of the weather, for which two weeks may be taken as an upper limit. The characteristics of the climate of a particular region (or at the global level) are determined by a number of factors and mechanisms that constitute what we call the climatic system. In a very simplistic way, we can imagine that the main components of this system are the atmosphere, the ocean, the land, the cryosphere, which includes the polar ice, and the high altitude regions. All these components interact strongly and may determine the weather and its variability.

The climate has temporal scales that, based on its definition, are of the order of several years to thousands of years. A climatic change may be meaningful only if it is statistically consistent with respect to "past normality". In order to define such

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_15) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_15

a reference period, usually a time average of at least 30 years is used. Strangely enough, this interval roughly coincides with a generation and is likely the cause of all the jokes and stories of our parents and grandparents that tell us how things were different during their time. Actually this is such a long period that it is very hard to make an average based only on memory.

To circumvent such semantic difficulties, there is a different way (and probably more rigorous) to define the climate. We may think of a number of "Earth" planets that have all the same physical characteristics and are also subject to the same external influences, like the same solar radiation or the same atmospheric composition. The atmosphere, the ocean, and the other elements of the climate system, although they have the same characteristics, are turbulent media so that we may expect for example that the weather in June on these planets, although not identical, will have the typical summer characteristics. This means that in the Mediterranean basin of these planets the weather will be more or less warm or more or less dry, but in none of them (or on a very few) will there be snow. This corresponds to an average weather that is called "ensemble average" in statistics. This concept of ensemble average implies at least two others, that is *meteorological noise* and *external influences*. The noise is responsible for the uncertainties in the average due to the daily fluctuations of the weather system. The external influence terms include actually both external and internal factors. What separates an internal from an external cause is very often the time scale. For example, a clear external cause is the solar radiation reaching the Earth; on the other hand, the snow cover could be an external cause on a short time scale (on an interannual basis) and an internal factor on a long time scale. If the snow cover lasts for such a long time as to grow into ice sheets then it can change the local orography (as happened in the case of the ice age) and the ice mass may influence the rotational regime of the Earth. In Fig. 15.1, the main components of the climate system are shown and also their main paths of interaction. We will deal here only with those forcings that may change the global climate, and these normally have to do with changes in the energy fluxes within the climatic system. This will be clarified along the way. One of the external causes we have mentioned is the energy coming from the Sun, that is, the solar constant. Time scales for changes in the solar constant are very long, while the shortest (of the order of days and years) refers to change in the flux of UV radiation and microwave. At the moment, there are no definite proofs that the solar constant may change on time scales of the order of one thousand years. Similarly, it is not possible to establish any cause–effect relation between the sunspot frequency and the climate variations. Changes in the solar constant between 0.05 and 0.1 % have been observed during periods of several years but these do not seem to produce any climatic changes. The Sun, on the other hand, has typical periodicities such as the 11–year cycle, which again does not produce any appreciable effects on the climate.

It is quite interesting to compare the effects of increasing greenhouse gases with the changes of the solar constant. The Earth-atmosphere system absorbs about 240 $\mathrm{Wm}^{-2}$ so that a 1 % change in the solar constant corresponds to a change of 2.4 $\mathrm{Wm}^{-2}$. On the other hand, a doubling of the $CO_2$ content implies an extra forcing of about 4 $\mathrm{Wm}^{-2}$ so that a 2 % change in the solar constant is equivalent to the change of all greenhouse gases. Accurate measurements of the solar constant are

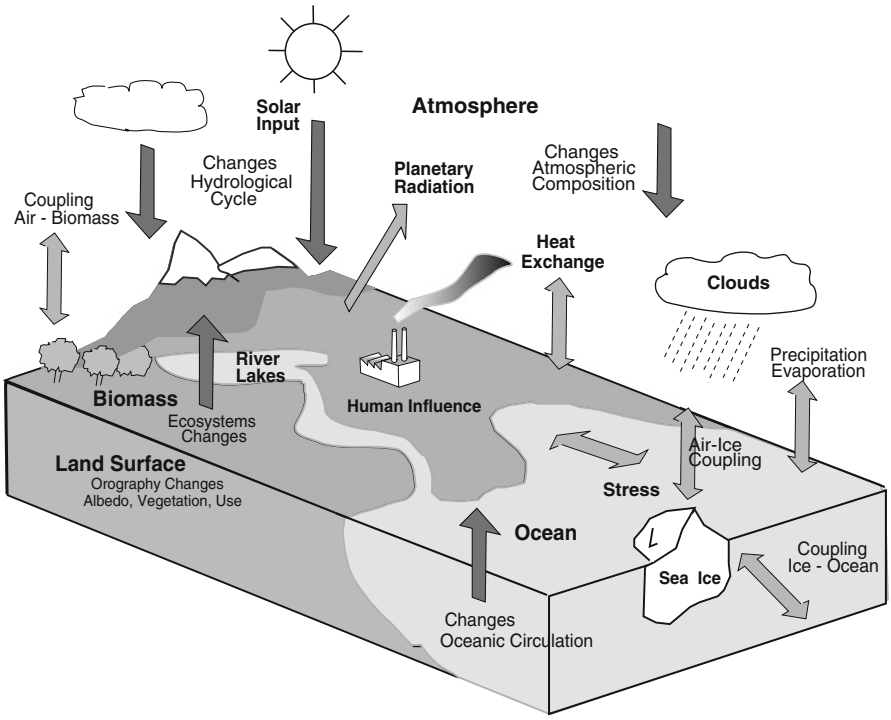



**Fig. 15.1** A simple illustration of the main components of the climatic system. The external influences are indicated with *black arrows*, while *gray arrows* indicate the interactions between internal variables

not available for the last century. However, there are very precise measurements for the period 1978–1997 obtained with two satellites, *Nimbus 7* and *Solar Maximum*. During this period, the solar constant changed only 0.1 % so that the forcing changed by 0.24 $Wm^{-2}$. Also it is possible to reconstruct with different methods the power output from the Sun during the last 300 years, and these results show deviations of only a few percent with respect to the average.

The explosive volcanic eruptions may inject into the stratosphere considerable amounts of dust and sulfate aerosols. These particles are very efficient in reflecting the solar radiation and may cause surface cooling that may be of the order of 0.5–1 °C. It has been argued that major eruptions like Tambora (1815) and Krakatoa (1883) and, more recently, Agung (1963), El Chichon (1982), and Pinatubo (1991) may have produced short-term perturbations to the global climate. Actually for any one of these events but especially for the oldest is very hard to prove anything.

Variations in the orbital parameters of the Earth (eccentricity, obliquity, and the precession of the equinoxes) may influence the amount of solar radiation reaching the high latitude regions. Locally these changes may even reach 10 %. The changes in the eccentricity cause a modulation of the Earth–sun distance producing variations in the radiation flux of less than 1 %.

The obliquity of the rotation axis varies between 22° and 25°, and this produces a modulation in the seasonal differences between summer and winter. In practice, when the obliquity increases, winters are colder and summers are warmer and so the difference is enhanced. The opposite happens when the obliquity decreases.

This last effect due to the equinox precession determines changes in the dates of the equinoxes with a period of about 22,000 years, so that about 11,000 years ago the spring equinox happened on the 21st of October. The most important modulation in the amount of solar radiation received by the Earth is due to the coupling between the obliquity changes and the precession of the equinoxes.

All these things are quite interesting but refer only indirectly to the greenhouse effect, which depends on the atmospheric composition. The concentration of some of the gases contributing to the greenhouse effect has changed drastically. For example, carbon dioxide during the ice age had a concentration of about 190 ppm compared to the present one of 345 ppm. This change may be explained by the complex mechanisms that regulate the concentration of this gas in the atmosphere and which involve also the biosphere. It is rather unlikely however that simply the lower concentration of $CO_2$ may have caused the ice age.

A number of techniques have been used to reconstruct the climatic history of the Earth. It is quite evident that the farther we go in the past, the less precise is the climatic picture. We have direct temperature measurements only for the last 300 years, and in any case, it is only in the last 100 years that these data have been taken by a sufficient number of stations to establish, for example, a temperature trend.

From written documents, it is possible to reconstruct in a qualitative way (sometimes we know that mood can change depending on the weather and not vice versa) the climate in the last millennium. To go farther in time, it is necessary to use climate proxy which is a series of ordinate parameters that somehow may be related to meteorological parameters. One of these is the measurement of the ratio between isotopes 16 and 18 of oxygen in ice cores. A preliminary analysis of the ratio $^{16}O/^{18}O$ can give an estimate of the ice sheets volume and of the surface temperature of the ocean. Different proxies can be used for different time intervals. For example, the tree rings method may be useful to reconstruct climate in the last thousand years. This method (called also dendroclimatology) is based on the fact that each year trees adds a layer of cortex to their trunks. The thickness of this layer may depend on the available humidity in the semi-arid regions and on the temperature in the temperate and alpine regions. Similarly, the flora of some mountain regions is controlled by the climate so that study of the fossil pollen may be useful in analyzing the last 10,000 years. The ice cores, on the other hand, may be useful for the last 100,000 years or more. Cores of the ocean floor may help reconstruct the climate of the last hundred million years. Finally, some sedimentary rocks may give information regarding up to a billion years ago.

Data obtained from climate reconstruction includes the global temperature behavior as a function of time. Some of the data, for example, refer to the last million years, and they show clearly the glacial and interglacial period, that is, the build up of the ice age and its waning. The temperature difference between these periods is of the order of 3–5 °C less than the seasonal variation that we may have

at high latitude. However, such temperature changes imply changes in the physical appearance of extended geographical regions, with the appearance of the continental ice sheets and the lowering by tens of meters of the sea level.

In the most recent times (the post-glacial period), the temperature changes have remained within a 1 °C range which may have caused significant effects on the length of the growing season for plants or in the precipitation regime.

All these nice facts can be found in any popular book on climate (except for those that claim that Christmas may arrive in August because of the equinox precession). However, in the most recent years, a new approach has been used in the study of the climatic system. The suggestion may have arrived from the evident oscillation in the volume of the ice sheets. Actually geochemistry enables the reconstruction of at least four or five complete glacial cycles that are very similar and have a dominant period of 100,000 years. Oscillations usually are a clear sign of at stability of a system, but in the case of the climatic system the extremes of such oscillations correspond either to a comfortable climate (the present one) or to the tough environment of an ice age.

Also, in this case there is a report, written by Edward Lorenz, which is rather unknown because it appeared in 1976 in a journal (*Quaternary Research*) that is not very popular. Here is what Lorenz wrote in the Abstract.

> A basic assumption in some climatic theories is that, given the physical properties of the atmosphere and the underlying ocean and land, specified environmental parameters (amount of solar heating, etc.) would determine a unique climate and that climatic changes therefore result from changes in the environment. The possibility that no such climate exists and that nondeterministic factors are wholly or partly responsible for long period fluctuations of the atmosphere-ocean earth system, is considered. A simple difference equation is used to illustrate the phenomena of transitivity, intransitivity and almost – intransitivity. Numerical models of moderate size suggest that almost – intransitivity might lead to persistence of atmospheric anomalies for a whole season. The effects of this persistence could be to allow substantial anomalies to build up in the underlying ocean or land, perhaps as abnormal temperatures or excessive snow or ice: these anomalies could subsequently influence the atmosphere, leading to long period fluctuations. The implications of this possibility for the numerical modeling of climate, and for the interpretation of the output of numerical models, are discussed.

Now except for some slang terms like intransitivity, the Lorenz thought is quite clear. In practice, he showed in one of the examples that a long series of frigid winters (an unlikely but not impossible event) could push the climatic system toward a cold equilibrium state. This is because such a series would imply an accumulation of snow and so extensive changes in the albedo.

Now we have introduced almost everything we would like to treat in this chapter, including mentioning non-deterministic theories. Lorenz's great achievement is to have shown that it is possible to have aperiodic (that is, without any apparent regularity) behavior of the climate system even if we know the equations of the system.

We will start by applying the simple climate models to the study of the ice ages. However in order to do that we need to go back to study in detail the solar radiation and how this is modulated by the motion of the Earth in its orbit around the sun. We will assume that the Kepler laws are known.

## 15.2 The Solar Radiation and the Orbital Parameters

We can refer to Fig. 3.3, where we assumed that the distance of the Earth from the sun at time $t$ makes an angle $\nu$ with the longitude of the perihelion. The Earth orbit is described by the law of mechanics

$$r^2 \frac{d\nu}{dt} = a^2\left(1-e^2\right)^{1/2} \tag{15.1}$$

where $e$ is the eccentricity of the orbit and $a$ is the semi-major axis, while

$$r(t) = \frac{a\left(1-e^2\right)}{\{1+e\cos[\nu(t)]\}} \tag{15.2}$$

is the instantaneous distance of the Earth from the sun. In this case, the unit for the time is (1 year)/2p. If we define $\omega = \nu + \Pi$, for a fixed value of the longitude of the perihelion $\Pi$ we have

$$dt = \frac{r^2 d\omega}{a^2(1-e^2)^{1/2}} \tag{15.3}$$

The solar flux incident at the top of the atmosphere is then

$$S(t,\phi,i,e,\Pi) = \frac{S_0 a^2}{4r^2}\tilde{S}[\omega(t,e,\Pi),\phi,i] \tag{15.4}$$

where $\phi$ is the latitude, $i$ is the obliquity, that is, the angle between the normal to the ecliptic plane and the rotation axis of the Earth. $S_0$ is the solar constant while the quantity

$$\left(\frac{S_0}{4}\right)\tilde{S}(\omega,\phi,i)$$

can be considered as the flux relative to a circular orbit with $r = a$ and $\omega = t$. We already know that

$$\frac{1}{2}\int_{-\pi/2}^{\pi/2} \tilde{S}(\omega,\phi,i)\cos\phi d\phi = 1 \tag{15.5}$$

The quantity $S(\omega, \varphi, i = 23.3°)$ is shown in Fig. 3.5. The issue now is that the obliquity changes between 22.1° and 24.5° (the present value is 23.5°) with a period of about 40,000 years. The eccentricity changes between 0.00 and 0.04 (present value 0.0167) with a period of about 100,000 years, and finally the longitude of the perihelion returns to the same value roughly every 20,000 years (present value 102°). These quantities are shown schematically in Fig. 15.2 as a function

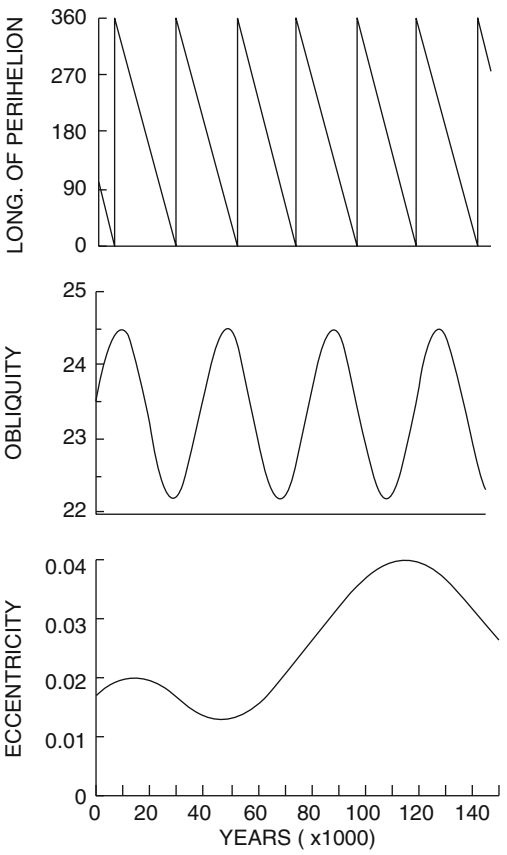



**Fig. 15.2** Plot of the longitude of the perihelion (*top*), the obliquity and the eccentricity for the Earth as a function of the time before the present (in thousands of years)

of time. It is interesting to evaluate the annual average of the solar flux that for an eccentric orbit is given by

$$\frac{S_0}{4}\int_0^{2\pi}\tilde{S}\frac{a^2}{r^2}dt = \frac{S_0}{4(1-e^2)^{1/2}}\int_0^{2\pi}\tilde{S}\,d\omega$$

The annual average is then proportional to the quantity $(1–e^2)^{-1/2}$ and does not depend either on the obliquity or on the longitude of the perihelion. However, the annual average radiation at a given latitude other than the eccentricity also depends weakly on the obliquity (it changes like $\sin(i)$) but again does not depend on the longitude of the perihelion; this, as a matter of fact, only determines the flux distribution during the year at a fixed latitude. If we want to evaluate the absorbed radiation we need to introduce the albedo $\alpha$ so that the radiation absorbed at some latitude is given by

$$\frac{S_0}{4(1-e^2)^{1/2}}\int_0^{2\pi}\tilde{S}\,\alpha\,d\omega \tag{15.6}$$

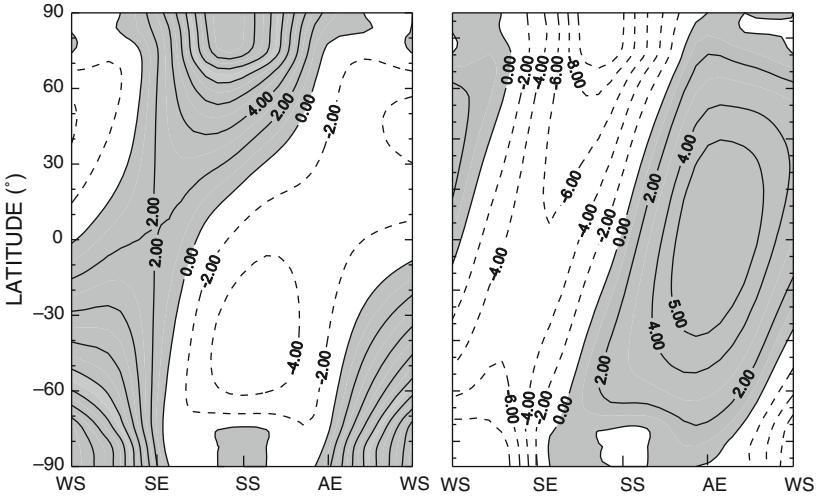



**Fig. 15.3** The difference (in %) with respect to the average value of the solar radiation between two different epochs. At *left* are shown the differences between 10,000 and 30,000 years ago, and on the *right* between 75,000 and 85,000. Negative values (less radiation at the most recent epoch) are indicated with *dashed lines* while the positive differences are *shaded*. WS and SS indicate winter and summer solstice respectively. SE and AE are summer and autumn equinox

This means that for the absorbed flux to change the albedo must depend on the season ($\omega$). This is the reason why it is so important to have models that can calculate the albedo as a function of the season and not just energy balance models because they must include seasonal effects.

The present value of the obliquity is 23.5° so that the axis is roughly in the middle of its oscillation. On the other hand, the eccentricity is around its minimum value. To have a more precise idea of how the solar flux is influenced by the astronomical parameters we can refer to Fig. 15.3, where the flux difference between two distinct epochs is reported as a function of the latitude and the season. The values shown in the figure are normalized to $S_0/4$. The left part of the figure shows the difference between 10,000 and 30,000 years ago. If we consider Fig. 15.2, we see that the most recent epoch corresponds to the maximum value of the obliquity while the most distant epoch corresponds to the minimum. This implies that 10,000 years ago the northern hemisphere in the summer was receiving roughly 12 % more solar flux than 30,000 years ago. The two hemispheres are not completely symmetric because the difference has not been chosen to be exactly the period of the equinox precession (23,000 years).

The figure on the right shows the same difference between 75,000 and 85,000 years ago. In this case, the obliquity has an opposite effect (the most remote epoch receives more radiation) but now the most important effects are the precession of the equinox and the eccentricity of the orbit. As a matter of fact, the eccentricity increases and so the average annual flux increases, while the 10,000 years difference enhances the effects of the equinox precession.

An interesting conclusion of this rather superficial examination is that the seasonal effects are predominant. For thousands of years, some regions of the Earth received a very different quantity of radiation during winter and summer. The simplest mechanism is that the temperature changes that result may change the albedo which amplifies the initial temperature effect (ice-albedo feedback).

## 15.3 Some Experimental Data on the Ice Ages

The detailed history of the ice ages can be obtained from studying the ocean floor sediments. Here the calcareous shells of the organisms that lived in the surface waters may give a clear indication of the sea temperature.

One of the methods is based on measuring the ratio of the oxygen isotopes in the sediments. Oxygen has three stable isotopes $^{16}$O (99.7 %), $^{17}$O (0.04 %), and $^{18}$O (0.2 %). The ratio between $^{18}$O and $^{16}$O is not constant and, for example, the water vapor contains about 7 ‰ more $^{18}$O with respect to the water from which it originated. The reason is that when a water molecule contains the lighter isotope ($H_2{}^{16}O$) because of its higher thermal velocity, it has a greater probability to evaporate. By contrast, in the condensation process, the heavier isotope will be lost preferentially. The net result for a humid air mass that moves from low to high latitudes and that loses some of the water as rain is that it becomes progressively enriched in the $^{16}$O isotope. This implies that Greenland ice contains about 30 ‰ more $^{16}$O than seawater. During the ice age, we may expect the continental ice sheets to have been enriched in the lighter isotope, while the seawater may have been enriched in the heavier isotope (by about 1 ‰). Related to this phenomenon is the effect of the temperature, in the sense that the colder the seawater the higher is the $^{18}$O content in calcareous shells with respect to the seawater. These two effects during an ice age add up in the shells so that they contain about 1 ‰ more of the heavier isotope because of the "heavier water" and about 1 ‰ more because of the colder water.

Differences of about 2 ‰ can be measured with mass spectrometry in ice cores or in the sediments. Today, it is possible to obtain "pistons" 20 m long, and assuming a deposition velocity of the order of 2.5 cm every 1000 years, we can obtain the ice volume until 800,000 years ago. The change in isotopic composition is represented with the symbol $\delta^{18}$O ‰, which gives the deviation with respect to a standard value for seawater. If this deviation is positive, the ice volume increases and vice versa. A typical record of this kind is plotted in Fig. 15.4 up to 0.5 million years ago. It is quite clear from this figure that the rapid warming in the last 20,000 years follows a slow cooling initiated about 100,000 years ago. Also evident from the figure is the prevalence of the 100,000 year period, while the terminal phase of the ice age (deglaciation) is always very fast.

When the signal is Fourier analyzed a spectrum is obtained as is illustrated in Fig. 15.5. In this case, we have reported the power density which is an indication of the energy contained in the different frequencies. Except for the details, which

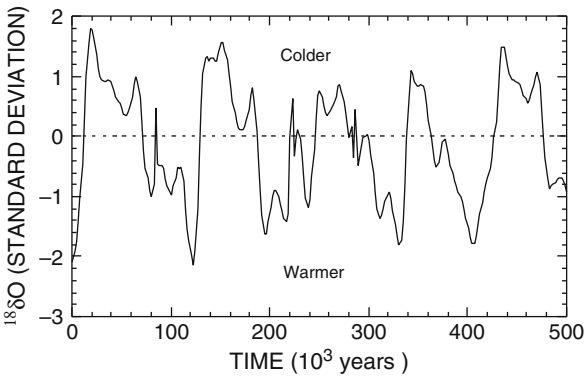



**Fig. 15.4** The change in the isotopic ratio for oxygen in an ice core. In this case, the change is plotted as standard deviation with respect to the mean. This particular sample can go back up to 750 k years, although here only the first half million years are reported. 0 on the time scale is the present epoch

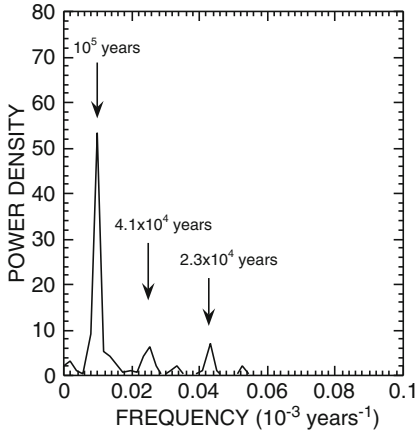



**Fig. 15.5** The power spectrum for the signal of Fig. 15.4. Notice how the peaks coincide with the periods of astronomical modulation of the solar radiation

may depend on the particular sample or ice core, the dominant frequencies are very similar to those mentioned in connection with orbital motion and that is obliquity, eccentricity, and equinox precession. It is quite surprising that most of the energy is contained in the 100,000 year period considering that the modulation of the absorbed radiation introduced by the eccentricity is very small. This is a common characteristic of all the samples. This characteristic may indicate the existence of some mechanisms capable of amplifying the small forcing. In the following pages, we will see a couple of them of a very different nature.

The first of these mechanisms could be an example of how the Earth system may work because it involves a component of the climatic system like the Earth's crust (i.e. the lithosphere) that is very different from the atmosphere. The idea is that once the continental ice sheets reach the "mature stage" their weight could force the lithosphere to subside on the underlying elastic mantle. In this way, the top of the ice would be exposed to a warmer environment so that it could melt. While the initial forcing for the growth of the ice sheets could be the insolation, the dominant period would be determined by the time constant for the deformation of the crust, which can be measured.

The other class of theories has to do with a quite new concept in physics, stochastic resonance. If a dynamic system has a number of equilibrium states corresponding, for example, to the minima of potential curves, then the system may oscillate between these states. The most interesting part, however, is that even statistical physics gives us a finite probability for the system to escape the potential well even if its energy is below the height of the well. This is because a thermodynamic system (like a gas) may have components (like the molecules) distributed along a certain range of energy. We have seen this in the case of Jean's escape. The climatic system can be described by a potential function, and the role of the thermal energy in this case is played by noise. The states of the system could be the glacial (cold solution) and interglacial (warm solution) climate and the noise would be the vagaries of the weather. The passage between the two states (or we may call it the transition) could be facilitated if on the strong noise we superimpose a weak signal at some fixed frequency: then the transition will be mostly at this frequency. This in short is what we call *stochastic resonance*. We will talk at length later on about these theories but we would like to give now a glimpse of other interesting data extracted from the ice cores. When the group of Hans Oeschger, in the early 1980s, developed a technique to extract the gas that was contained in little bubbles in the ice cores, another enigma started. This was actually a way to obtain the atmospheric composition during the ice ages, with the result that is shown in Fig. 15.6, where the content of two greenhouse gases ($CH_4$ and $CO_2$) is shown up to 150,000 years ago together with the temperature difference with respect to the present epoch. The figure is based on much more detailed data and it is smoothed at 5,000 years. The variations of the temperature follow almost perfectly those of the two gases. This is not a proof of the greenhouse effect but something that is much more complex and includes the interactions between biosphere, atmosphere, and the climatic system: a quite good example of the Earth system at work. If a section of the curve is enlarged (like that of the last 20,000 years) during the deglaciation, sudden cooling can be observed that in a few years causes the climate of entire regions to go back to glacial conditions. This is accompanied by covariant changes of the greenhouse gas content. The explanation in this case may involve the deep ocean circulation, and for the time being, we just do not have the ways and means to go ahead on this road. We may try later on.

## 15.4 The 100 Kyear Cycle and the Lithosphere–Atmosphere Coupling

The model (due to J. Weertman) used in this case is sketched in Fig. 15.7. The shape of the ice sheet remains fixed as it grows and adapts as a perfect plastic material. The shape is fixed with a relation between the height of the ice sheet and its length according to

$$h = [\lambda\,(L - y)]^{1/2} \tag{15.7}$$

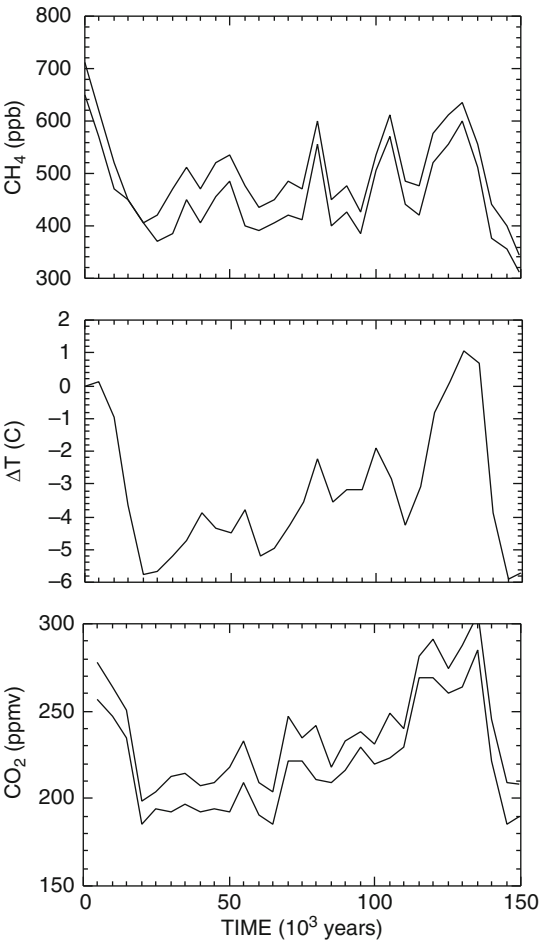



**Fig. 15.6** The content of methane (*top*) and carbon dioxide (*bottom*) as a function of time before the present (in thousands years) compared with the temperature difference with respect to the present epoch

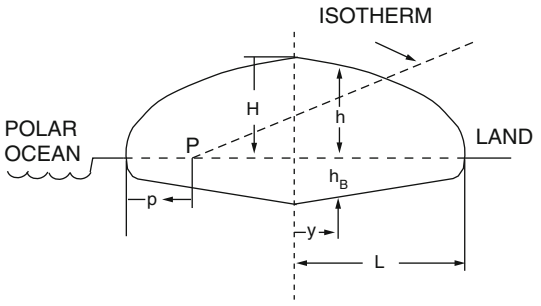



**Fig. 15.7** Parameters for the study of the growth of a continental ice sheet

where, always referring to the figure, the height of the ice sheet is given by

$$H = (\lambda L)^{1/2}$$

The quantity $h_R$ is the buried height, while $2L$ is the latitudinal extension of the ice sheet. The other quantities relevant to the problem must be introduced in connection with the isotherm shown in the figure. This isotherm actually is the snow line: above this line we will assume an accumulation of ice, while below we will have ablation. The north in this figure is on the left so that going to the left we will have lower temperatures. Considering that the temperature decreases also with altitude, it is clear that the snow line will slope upward on the right side. Once the slope is fixed, the position of this line can be obtained by fixing the distance between the intersection ($P$) of the snow line and the surface to the north. We would expect that if the global temperature decreases, the point $P$ will move to the south so that the surface available for the growth of the ice sheet will increase.

The portion of the ice sheet above and to the north of the snow line is denoted as *accumulation zone* and the portion to the south *ablation zone*. Other than the mentioned contribution by J.Weertman, these models are also due to J. Oerlemans and recently to C. J. Van der Veen. In 1983, R. G. Watts and E. Hayder added to the ice model a simple energy balance model. From observations (and also from a bit of intuition), the rate of growth for the ice sheet can be written as

$$\begin{aligned} A &= a\,[H - s\,(2L - p)]\,L - b[H - s\,(2L - p)]^2 L \quad H - s\,(2L - p) < a/2b \\ A &= a^2 L/4b \qquad\qquad\qquad\qquad\qquad\qquad H - s\,(2L - p) > a/2b \end{aligned} \tag{15.8}$$

The growth rate of the ice sheet is actually the rate of increase of the ice sheet (assumed to have a unit width) cross section.

Equation (15.8) needs some comment. The quantity $H - s(2L - p)$ is the difference between the top of the ice sheet and the ice line measured on the southern edge. The net ice accumulation increases with this difference until we have

$$H - s\,(2L - p) = \frac{a}{2b}$$

This maximum accumulation rate corresponds to the southernmost position of the ice line (that starts from the southern edge of the ice sheet) so that beyond this point the growth rate can only be given by the second part of Eq. (15.8).

The Earth's crust bends under the weight of the ice sheet as it grows. We may assume that the crust adjusts to this weight with a time constant $\alpha^{-1}$ according to

$$\frac{dh_B}{dt} = \alpha \left( \frac{h}{3 - h_B} \right) \tag{15.9}$$

so that at steady state $h_R = h/3$. We also assume that $h_R$ has the same form of $h$ so that the ratio

$$\frac{h_B}{h} = \eta(t)$$

is only a function of time. The rate of change of the volume for a unit width of the part of the ice sheet that results toward the equator is

$$\frac{d}{dt}\int_0^L (h + h_B)\, dy = \frac{2\lambda^{1/2}}{3}\frac{d}{dt}\left[(1+\eta)\, L^{3/2}\right] \tag{15.10}$$

To simplify the discussion is convenient to normalize the variables following the rules

$$\begin{array}{lll} R^2 = 4s^2 L/\lambda & p = 2s^2 P/\lambda & \tau = at/2 \\ \kappa = b\lambda/2as & \mu = 2\alpha/a & \Gamma = \eta R \end{array} \tag{15.11}$$

so that Eq. (15.8) becomes

$$\begin{aligned} \frac{1}{3R^2}\frac{d}{d\tau}\left[\left(1+\frac{\Gamma}{R}\right)R^3\right] &= \left(R - R^2 + p\right)\left[1 - \kappa\left(R - R^2 + p\right)\right] \\ \frac{1}{3R^2}\frac{d}{d\tau}\left[\left(1+\frac{\Gamma}{R}\right)R^3\right] &= \frac{1}{4\kappa} \end{aligned} \tag{15.12}$$

The first is valid for $(R{-}R^2 + p) < 1/2$ k and the second for $(R{-}R^2 + p) < 1/2$ k. Equation for $\Gamma$ becomes

$$\frac{d\Gamma}{d\tau} = \mu\left(\frac{R}{3-\Gamma}\right) \tag{15.13}$$

In this case, we need to assume $R > 0$

Now we can verify that the model works by supposing that the position of the climate point $P$ changes periodically with the law

$$P = P_0 + \Delta P \sin\left(\frac{2\pi t}{T}\right)$$

And by putting

$$\beta_0 = 2s^2 P_0/\lambda \quad \beta_1 = 2s^2 \Delta P/\lambda \quad \omega = 4\pi/aT \tag{15.14}$$

we obtain the non-dimensional form of the forcing

$$p = \beta_0 + \beta_1 \sin\omega\tau \tag{15.15}$$

Now we can integrate numerically Eqs. (15.12) and (15.13), once the values of the different parameters are fixed. As usual, these are the result of accurate extrapolations of the growth rate of the ice sheets but also of some degree of fine tuning, in this case

$$s = 10^{-3} \quad \lambda = 7\text{m} \quad P_0 \cong -150\,\text{km} \quad \Delta P \cong 300\,\text{km}$$

$a \cong 7.3 \cdot 10^{-4}$ year$^{-1}$ $b = 2.7 \cdot 10^{-7}\text{m}^{-1}\text{year}^{-1}$ $T = 2 \cdot 10^4$year $\alpha^{-1} = 5 \cdot 10^3$year

When these values are inserted in Eq. (15.14), we have

$$\beta_0 = -0.0429 \quad \beta_1 = 0.0714 \quad \kappa = 1.29 \quad \omega = 0.938 \quad \mu = 0.55$$

Regarding these numbers we can make some considerations. First of all, the slope of the isotherm is simply the ratio between the temperature lapse rate and its horizontal gradient and can be easily determined. In this way, the maximum growth is for $a/2b = 1350$ m.

To obtain the real time from the normalized one, we need to multiply by 2740 years. We may ask with all the approximations we have made if it makes any sense to write down numbers to the fourth decimal digit but we will discover the reason for this in a while. Figure 15.8 shows the results of the integration. In the figure are shown as a function of time, the maximum height of the glacier for two different values of the forcing function that corresponds to a change in the middle position of the climate point of only 8.4 km. We notice clearly a bifurcation because a small change in $P_0$ corresponds to the growth or the decay of the ice sheet. The initial

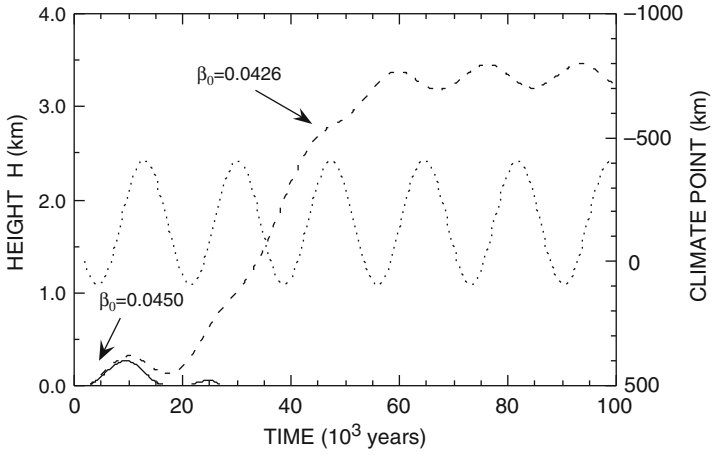



**Fig. 15.8** The growth glacier shown in Fig. 15.7, for the parameters given in the text. The *dotted line* is the position of the climate point. The other curves show the height $H$ in km

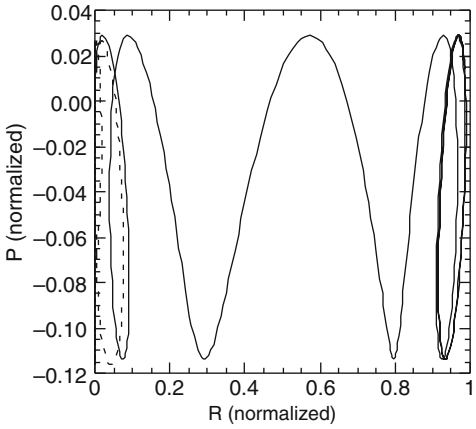



**Fig. 15.9** The phase space diagram for the growth of the ice sheet. On the axis is the position of climate point $p$ and the dimension $R$. Both are normalized

growth is rather slow and the modulation due to the solar radiation (that in this case has a period of 20,000 years) is quite evident.

Further insight in to the mechanism can be obtained by building a diagram in the phase space defined by the dimension of the glacier and the climate point, as shown in Fig. 15.9. Starting from an initial value $p(\tau)$ zero as the climate point moves toward south (that is, the Earth cools), the glacier will grow. Once the maximum value of $p(\tau)$ is reached, the growth slows down even when $p$ becomes negative. If $p$ stays negative long enough, the ice sheet will melt completely and in order to start to grow again $p$ must become positive again. In this way, we enter a limit cycle and the ice sheet grows but not to a large size.

In the case that $p$ becomes positive before the ice sheet is completely melted the size will increase even when $p$ again becomes negative. Then the ice sheet will grow to a maximum size and a new limit cycle will begin. This bifurcation does not depend only on the value of $\beta_0$, which corresponds to the initial position of the climate point, but also on $\beta_1$ which corresponds to the amplitude of the oscillation.

To understand this point we may consider the case in which the crust can adapt instantaneously to the ice sheet weight ($\mu = \infty$). We can then study the system near its equilibrium point for small values of $R$. Equation (15.12) will give

$$\frac{4}{3}\frac{dR}{d\tau} = R + p \tag{15.16}$$

With the initial condition $R = 0$, when $p = 0$ we have

$$\tau_1 = \omega^{-1} \arcsin\left(\frac{-\beta_0}{\beta_1}\right); \quad \tau_2 = 2\pi + \omega^{-1} \arcsin\left(\frac{-\beta_0}{\beta_1}\right)$$

while a particular solution of Eq. (15.16) is of the form

$$R = A \sin\omega\tau + B\cos\omega\tau + C$$

that substituted in Eq. (15.16) gives the equations for the determination of the constants

$$A\omega - 3B/4 = 0; \quad -B\omega - 3A/4 = \beta_1; \quad -3C/4 = \beta_0$$

And considering $R = 0$, we obtain

$$\frac{-\beta_0}{\beta_1} = 1/\left[1 + (4\omega/3)^2\right]^{1/2} \tag{15.17}$$

As a consequence, if $\beta_0/\beta_1$ is larger than the right-hand side, then the ice sheet will remain small. On the other hand, if $\beta_0/\beta_1$ is smaller, then $R(\tau_2) > 0$ and the ice sheet will grow reaching a limit cycle around $R = 1$. To understand how this second limit cycle works we put $R = 1 + \varepsilon$, with $\varepsilon$ a small quantity, so that we can write a linearized form for Eq. (15.12)

$$\frac{d\varepsilon}{d\tau} = \frac{3\,(p - \varepsilon)}{4}$$

And a steady state solution for this equation is given by

$$\varepsilon = \beta_0 + \beta_1 \sin\left[\omega\tau - \arctan\left(\frac{4\omega}{3}\right)\right] / \left[1 + \left(\frac{4\omega}{3}\right)\right]^{1/2} \tag{15.18}$$

From this we see that the center of the limit cycle is given by $R = 1 + \beta_0 \approx 1$ so that the maximum dimension of the ice sheet is given by $L \approx 1/4\, s^2$.

The theory then says that in order for the ice sheet to grow the cooling is not enough (i.e. $\beta_0$). This could, for example, explain why until 2.4 billion years ago there is no trace of glaciation on the Earth, because the climate was too warm. When further cooling produced a colder climate, a threshold was reached for the bifurcation to occur and consequently the ice sheets grew.

This theory does not explain the sudden termination of the ice ages. However, a variation of the same theory, from D. Pollard, suggests a mechanism for the sudden warming. In practice, when the ice sheet starts to melt on its southern edge, it may form a glacial lake that may accelerate its ablation or calving.

Let us face the truth: we are a little bit tired of ad hoc theories and so we would like to turn to a more general approach. We will resume some of the chaotic aspect of climate in the last chapter of this book, when we will learn a little more about limit cycle and all the technical jargon of the dynamic systems.

## 15.5 Stochastic Resonance

The stochastic resonance is a concept introduced in connection with the problem of explaining the timing and periodicity of the ice ages. To learn how it works we can refer to Fig. 15.10, which shows the results of a very simple mechanism.

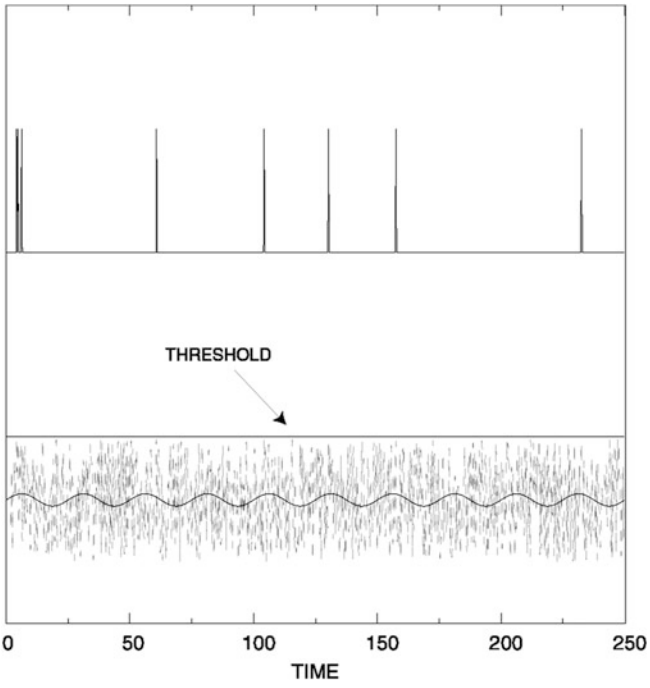



**Fig. 15.10** An illustration of the mechanism for stochastic resonance. The noise (*grey*) added to the small sinusoidal signal produces the spikes at the *top* of the figure each time the signal exceeds the threshold. The spikes have a frequency that is a multiple of that of the periodic signal

We imagine having a periodic signal that is electronically added to a white noise that in this case is actually given by a random number generator. The maximum amplitude (actually the mean square value) is about ten times the amplitude of the periodic signal. Every time the resulting signal is greater then a fixed threshold a spike is produced. Only the noise will be marginal in producing any spike, while the periodic signal, because of its small amplitude, never goes above the threshold. The spike occurrence will happen preferentially when the periodic signal reaches its maximum positive value so that the resulting frequency of the spikes will be tuned to the frequency of the sinusoidal signal. The net result is the possibility to use the noise energy to go over some potential barrier at a frequency that is basically that of a small periodic signal.

One would think that a larger noise would increase the production of spikes but that would be no longer at the frequency of the periodic signal. In a sense there is an optimum ratio between noise and signal that produces a tuned response.

This concept can now be applied to the climatic system. Earlier in Chap. 14, we mentioned that the ice ages might have been produced by oscillations of the climatic system between two equilibrium states that we found studying the simple climate models. In particular if we use Eq. (14.9), we can write a relation like

$$C\frac{dT}{dt} = g\,(t, T) \tag{15.19}$$

where $g\ (t,\ T) = -dF/dt$, with $F$ given, for example, by Eq. (14.12). Actually the right-hand-side term can be written as

$$Q\,(1 - \alpha(T)) - A - BT$$

This function is represented in Fig. 14.4 and is equivalent to a potential having two equilibrium points (warm and cold climate) and one unstable point. This is not a very realistic situation because the temperature interval separating the two states is too large. Only for the sake of an exercise we can think of a very simple form of the potential that again has two minima that are symmetrical. A possible candidate would be of the form

$$c + 0.5\alpha T^2\left(\frac{T^2}{2-1}\right)$$

If the potential is treated as in mechanics we could write out of this potential an "equation of motion" for the temperature given by

$$\frac{d^2T}{dt^2} + k\dot{T} + \alpha\left(T^3 - T\right) = f\cos\omega t \tag{15.20}$$

The term proportional to $dT/dt$ is the damping due to the emission of infrared radiation, while the term proportional to $\alpha$ is in fact the potential gradient (in mechanics term is the acceleration). The potential has two minima that are located at $T = \pm 1$ and a maximum located at $T = 0$, while the depth of the minima is $\alpha/4$. The right-hand-side term in Eq. (15.20) is the forcing and if its amplitude is large enough the system may oscillate between the two potential wells. Following what we have learned before about stochastic resonance, even if the forcing is not large enough we still could expect a transition provided that we add a noise to the forcing term

$$\frac{d^2T}{dt^2} + k\dot{T} + \alpha\left(T^3 - T\right) = f\cos\omega t + q^2A(t) \tag{15.21}$$

where $A(t)$ is the stochastic forcing that has the characteristics of a white noise

$$\langle A(t)\rangle = 0 \quad \text{and} \quad \left\langle A(t)A\left(t'\right)\right\rangle = \varepsilon^2\delta\left(t - t'\right)$$

with $\varepsilon$ being the variance. Usually the system will tend to oscillate in one of the wells unless the sum of the deterministic forcing (i.e. the periodic term) and the noise is greater than the potential barrier. In this case, the system will jump from one well to another.

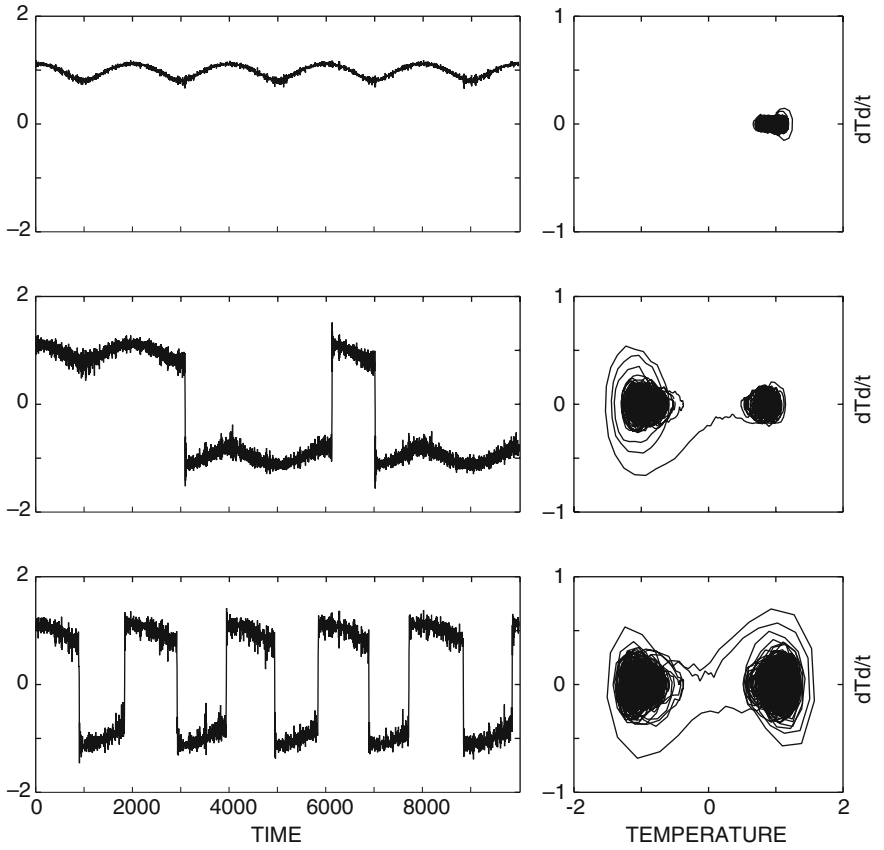



**Fig. 15.11** The amplitude of a non-linear oscillator forced by a deterministic signal and white noise. The three cases (from *top* to *bottom*) are for $q^2 = 0.2$, 0.5, and 0.8. On the *right*, the corresponding phase space diagrams are shown

This is shown in Fig. 15.11, where three case are reported for the parameters in Eq. (15.21), given by $k = 0.15$, $\alpha = 0.5$, $\omega = 10^{-3}\pi$ and $f = 0.14$. We notice that when $q^2$ is small, starting from arbitrary initial conditions, the system will be trapped in one of the two potential wells. The amplitude of the oscillation is always positive (or negative) with a value around 1(or $-1$). Increasing the value of $q^2$ the system will occasionally jump to the other well, but as is clearly shown in the phase space diagram, it will stay mostly in the well where the system is initialized. Increasing further the value of the variance, the transitions become regular and the frequency is just about that of the deterministic signal.

If we look carefully at the phase space diagrams, we see that the transitions happen when the "velocity" (i.e. $dT/dt$) is almost zero because the system is then on the edge of the well. If we start from $T = 1$, the transition can then happen only

if $dT/dt < 0$ so that in the phase space diagram the system goes toward where both $dT/dt$ and $T$ are positive. The opposite happens when we start from $T = -1$.

Transitions between the two states occur only when noise is present and when its variance is large enough. Actually it is possible to notice two frequencies that correspond to the deterministic forcing and to the frequency due to the fact that the system oscillates in one of the potential wells. Around the minimum, the potential can be written as $\approx (T-1)^2$ so that the frequency for small oscillations is of the order of $\alpha^{1/2}$ so that in our case it is about hundred times larger than the frequency of the deterministic forcing

The relevance of stochastic resonance for the ice age problem is rather evident.

In this case, the forcing could be related to the astronomical modulation of the solar radiation. The signal itself is unable to produce any effect but when coupled to a "climatic noise" it could produce transitions between the different climatic states. This could be a manifestation of a quasi-intransitive system and the origin of the noise could be in the seasonal variations. If one of the climatic regimes were to correspond to a frigid winter, then once this state is reached it could be self maintained because at the beginning of summer there could be still an appreciable snow cover.

The scheme we have adopted is extremely simple and its only justification is to illustrate that one of the climatic mechanisms which might have been responsible for the ice ages has to do simply with the non-linearity of the climatic system. The origin of the non-linear characteristics is much harder to locate although it could be traced to the ice-albedo feedback and ice sheet dynamics.

A very simple system of this kind is reported in the Appendix for this chapter and reproduces the original work of Sutera et al. The results shown in Fig. 15.12 are obtained from that model and are compared with the general behavior of the ice volume obtained from ice core data analysis.

## 15.6 The Global Warming: A Simple Exercise

Nothing is more annoying than to see disguised as a very sophisticated problem something that any second year physics student can do as an exercise. In the realm of climate models, there are many things like that. For example, one of the simplest problems is to apply the energy balance concept to the calculation of the average global temperature. In this case, the Earth is reduced to something like a box that receives additional heat from the so-called greenhouse gases that warm the surface of the planet. In this case, the feedback mechanisms are also very important. The deep ocean and the continental crust also absorb heat so that the Earth in a sense behaves like a capacitor. This may explain our mentioning second year in physics, because we will try to imitate the IPCC (Intergovernmental Panel for Climatic Change) through an electrostatic exercise and very elementary electronics.

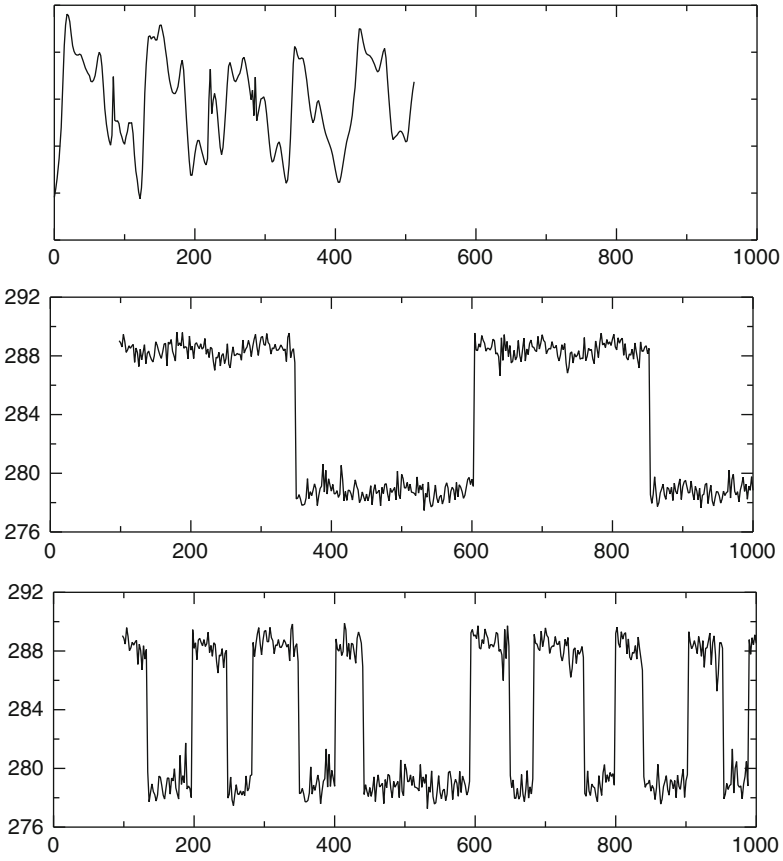



**Fig. 15.12** The results of the simulation of ice age occurrence with a stochastic resonance model. The *top figure* shows the data (same as in Fig. 15.3), while the *panel below* shows the results with an increasing amount of noise

### *15.6.1 The Near Future Climate of the Earth as a Problem of Electrical Engineering*

It is now time to analyze with some detail the feedback concept by refreshing what we learned in some elementary theory of electrical circuits. We refer to Fig. 15.13 which is a typical feedback arrangement. In this case, $V_s$ is the initial input signal. The effective input signal is $V_1$ so that the gain $G$ is given as the ratio between the output signal $V_2$ and the effective input signal

$$G = \frac{V_2}{V_1} \tag{15.22}$$

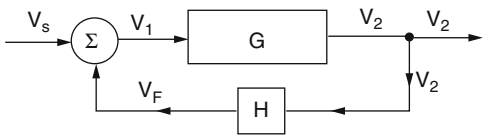



**Fig. 15.13** The feedback circuit. See text for the meaning of the symbols

If a fraction of the output signal ($V_F$) is summed in the input, we can define a feedback factor $H$ as

$$H = \frac{V_F}{V_2} \tag{15.23}$$

The input signal is now $V_1 = V_S + V_F$ and the output signal becomes

$$V_2 = GV_1 = G\,(V_S + V_F) = G\,(V_S + HV_2) \tag{15.24}$$

The gain with the feedback is then

$$G_F = \frac{V_2}{V_S} \tag{15.25}$$

and eliminating $V_2$ between Eqs. (15.23) and (15.24), we get

$$G_F = \frac{G}{1 - GH} = \frac{G}{1 - f} \tag{15.26}$$

where $f = GH$ is the feedback of the system.

The connection with the climatic system is quite simple. We assume a change in the net radiative flux at the top of the atmosphere $\Delta Q$ so that the corresponding temperature change at the surface, taking into account all the feedback mechanisms, will be

$$\Delta T = G_F \Delta Q \tag{15.27}$$

In this case, the role of $\Delta Q$ is the same as $V_s$ while $\Delta T$ substitutes for $V_2$. The corresponding quantity to $V_1$ will be a change in the flux $\Delta Q'$ given by

$$\Delta Q' = \Delta Q + \left(\sum_i H_i\right) \Delta T \tag{15.28}$$

This flux may be written in the form

$$\Delta Q' = \sum_e \frac{\partial Q'}{\partial x_e} \Delta x_e + \left(\sum_i \frac{\partial Q'}{\partial x_i} \frac{\partial x_i}{\partial T} + \frac{\partial Q'}{\partial T}\right) \Delta T \tag{15.29}$$

where it is assumed that the variations of the flux depend on *external variables* ($x_e$) like the solar radiation and *internal variables* ($x_i$) like the albedo and directly on the temperature. Comparing Eqs. (15.28) and (15.29), we obtain

$$\sum_i H_i = \sum_i \frac{\partial Q'}{\partial x_i} \frac{\partial x_i}{\partial T}$$

As done previously Eq. (15.27) can be written as

$$\Delta T = \frac{G}{1 - \Sigma f_i} \Delta Q \tag{15.30}$$

where $G$ is the gain in the absence of feedback. Actually the quantity $G_F$ is what we have previously called climate sensitivity $\lambda$.

A simple example of the feedback factor can be made with the greenhouse effect. We have seen that for an initial forcing of 4 $\mathrm{Wm}^{-2}$ and a climate sensitivity of 0.3 (in absence of any feedback), we have a temperature change of 1.2 °C that may go to 1.7 °C when the water vapor feedback on the IR radiation is considered and to 1.9 °C when the feedback is extended to the solar radiation. If we apply Eq. (15.25), we get $H_{\mathrm{ir}} = 0.98$ and $H_{\mathrm{sol}} = 0.248$, and these values imply a climate sensitivity

$$\lambda = \frac{0.3}{[1 - 0.3\,(0.248 + 0.98)]} \approx 0.48$$

The feedback factor can be converted in the perturbation of infrared or solar flux so that we have

$$\frac{\Delta F_{\mathrm{ir}}}{\Delta T} = 0.98\ \mathrm{Wm}^{-2}\ \mathrm{K}^{-1} \quad \frac{\Delta F_{\mathrm{sol}}}{\Delta T} = 0.248\ \mathrm{Wm}^{-2}\ \mathrm{K}^{-1}$$

Other types of feedback (ice–albedo, clouds, etc.) can be treated in the same way, keeping track of the sign.

A very simplified picture of the climate system is shown in Fig. 15.14, where the changing atmospheric composition produces a change $\Delta Q$ of the radiative forcing. This additional energy implies a temperature change $\Delta T$ of the ocean surface. The surface layer is well mixed and reacts much more rapidly than the underlying ocean through which the heat diffuses with a flux $\Delta F$. The surface of the Earth is heated directly by the atmosphere but also through the energy exchanged with the oceans. We may assume a change in the equilibrium temperature $\Delta T = \lambda\ \Delta Q$, and in this case, the land temperature could be found by assuming a balance between the energy received by the atmosphere and that lost (or gained) to the ocean through conduction. Indicating with $f$ the portion of the earth surface occupied by land (0.29), we obtain

$$f\Delta Q + k\,(\Delta T_O - \Delta T_l) = \frac{f\Delta T_l}{\lambda} \tag{15.31}$$

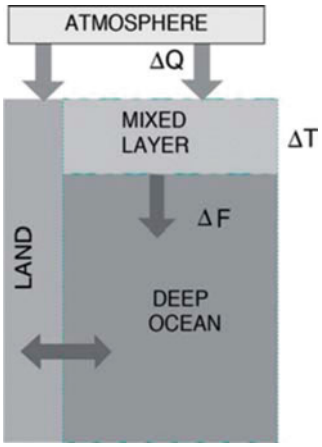



**Fig. 15.14** A very simple sketch of the land–atmosphere–ocean system. The radiative forcing acts both on the land and the ocean. Heat is transported in the ocean and land through diffusive processes (*arrows*)

where $\Delta T_0$ is the temperature change of the atmosphere over the ocean that can be assumed to be the same as the ocean surface layer. $\Delta T_l$ denotes the temperature change over land and $k$ is the exchange coefficient between the solid earth and the ocean. From Eq. (15.31), we obtain easily

$$\Delta T_l = \frac{(f\lambda \Delta T_i + k\Delta T_O)}{(f/\lambda + k)} \tag{15.32}$$

The evaluation of the temperature change of the mixed layer is a little bit more complicated because the heat capacity of the ocean (even of a relatively thin layer) is much larger than a corresponding layer of solid earth. For this, we may assume that the layer of interest has a 6 m thickness with a density of about 1.8 $10^3$ kg m$^{-3}$ and a specific heat of 0.34 $10^3$ J kg K$^{-1}$ so that the heat capacity per unit surface is 3.7 $10^6$ J m$^{-2}$. The corresponding heat capacity for 100 m depth is 4.2 $10^8$ J m$^{-2}$, about two orders of magnitude higher. We can obtain an average heat capacity through a weighted mean that takes into account the relative surface of the ocean and land and we arrive at a value that is about 70 % of the heat capacity of the ocean. The consequence is that the time constant for the temperature change is of the same order of magnitude as the characteristic time for the forcing (that is, the increase of the greenhouse gases). This implies that to calculate the temperature change we must recur to a time-dependent equation. As shown in Fig. 15.14, the input term is again the radiative forcing $\Delta Q$ produces a temperature change $\Delta T$ in the mixed layer. From this layer, heat is lost via diffusion to the deep ocean so that the temperature change in the layer becomes

$$C_m \frac{dT}{dt} = \Delta Q - \frac{\Delta T}{\lambda} - \Delta F \tag{15.33}$$

where $C_m$ is the equivalent heat capacity and $\Delta F$ the diffusive flux. If the temperature gradient at the interface mixed layer-deep ocean is $\partial T_0/\partial z$ and $\kappa$ is the diffusion coefficient in $m^2\ s^{-1}$ the diffusive flux will be

$$\Delta F = \rho c_O \kappa \frac{\partial T_0}{\partial z} \tag{15.34}$$

where $\rho$ and $c_o$ are the density and specific heat of the ocean respectively. This equation shows that it is necessary to obtain not only the surface temperature but also the distribution with depth $T_0(z, t)$. The temperature is determined by the diffusion equation

$$\frac{\partial \Delta T_0}{\partial t} = \kappa \frac{\partial^2 \Delta T_0}{\partial z^2} \tag{15.35}$$

where $z$ is the vertical coordinate with the origin at the bottom of the mixed layer.

Solutions to this equation are obtained with boundary conditions such that at the bottom of the ocean the temperature perturbation must be zero all the time while at the surface the change is instantaneous

$$z \to \infty; \quad \Delta T_0\,(\infty, t) = 0; \quad \Delta T_0\,(0, t) = \Delta T(t).$$

This is a classical problem that has a rather well-known solution

$$\Delta T_0\,(z, t) = 1 - \mathrm{erf}\left[z/2(\kappa t)^{1/2}\right] \tag{15.36}$$

where the error function (erf) is defined as

$$\mathrm{erf}(x) = \left(2/\pi^{1/2}\right) \int_0^x \mathrm{e}^{-\eta^2} d\eta$$

The flux into the deep ocean is then

$$\Delta F = -\rho c_0 \kappa (\partial T_0/\partial z)_{z=0}, \tag{15.37}$$

which may be expressed as a function of the temperature perturbation

$$\Delta F = -\rho c_0 \kappa \frac{\Delta T}{\sqrt{\pi \kappa t}}$$

We introduce a characteristic time for diffusion given by

$$\tau_d = \frac{\pi h^2}{\kappa}$$

where $h$ is the thickness of the mixed layer and Eq. (15.33) becomes

$$\gamma \frac{d\Delta T}{dt} + \Delta T \left(\frac{1}{\tau_f} + \frac{\mu\gamma}{\sqrt{\tau_d t}}\right) = \frac{\Delta Q}{\rho c_0 h} \tag{15.38}$$

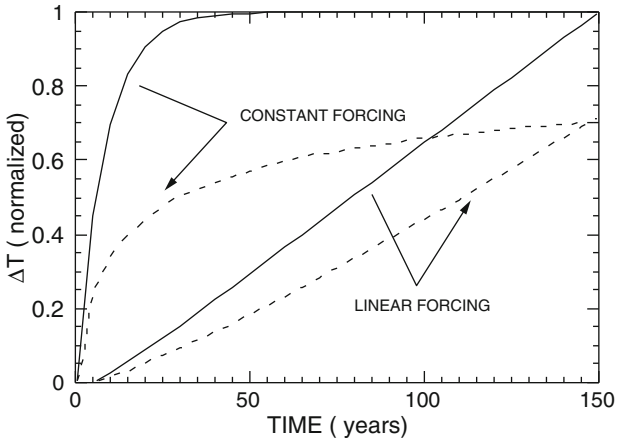



**Fig. 15.15** The response of the climatic system to a constant forcing (step function) and a linear forcing. The temperature in the first case is normalized to its asymptotic value, while the linear forcing amounts to 0.015 tWm$^{-2}$ years$^{-1}$. The *solid line* takes into account only the mixed layer, while the *dashed line* considers the effect of the deep ocean

where

$$\tau_f = \frac{\rho c_0 h}{\lambda}$$

and $\gamma$ is the ratio between the effective thermal capacity and that of the ocean $\rho c_o h$. Equation (15.38) presents a little surprise which is the appearance of the factor $\mu$.

Actually we have used only an approximate method to solve Eq. (15.33) because we have treated the diffusion as a separate process. The factor $\mu$ is then a simple correction to this solution and its value ranges between 1.4 and 2.7 according to the different forcing. The determination of the time constants is one of the many uncertainties in this game. We have $\kappa = (1 \div 3)\ 10^{-4}$ m$^2$s$^{-1}$ so that with $h = 100$ m we have $\tau_\mathrm{d} = 3 \div 10$ years. Even more vague is the value for $\tau_f$ because it depends on the climate sensitivity $\lambda$ and changes linearly with it. For an accepted value of 0.5 w m$^{-2}$ K$^{-1}$, we have $\tau_\mathrm{f} = 6.5$ years. Equation (15.38) is similar to that of a charging capacitor with a time constant that depends on time. To understand better the time dependence we can solve two cases, first using a step function for the forcing that starts at $t = 0$ and then a forcing that increases linearly with time.

Figure 15.15 show the results of this exercise, and it is clear that when the diffusion in the ocean is neglected, the response time of the system is quite short and is equal to the time constant chosen, that is, 7.5 years.

The introduction of the ocean diffusion changes drastically the result because now the time constant increases with time. In the case shown in the same figure, we have an "equivalent" constant that is of the order of 50 years. Always working with the capacitor analogy, we see that the two capacitors (deep ocean and mixed layer) are in parallel so that it is the largest capacity that determines the total value.

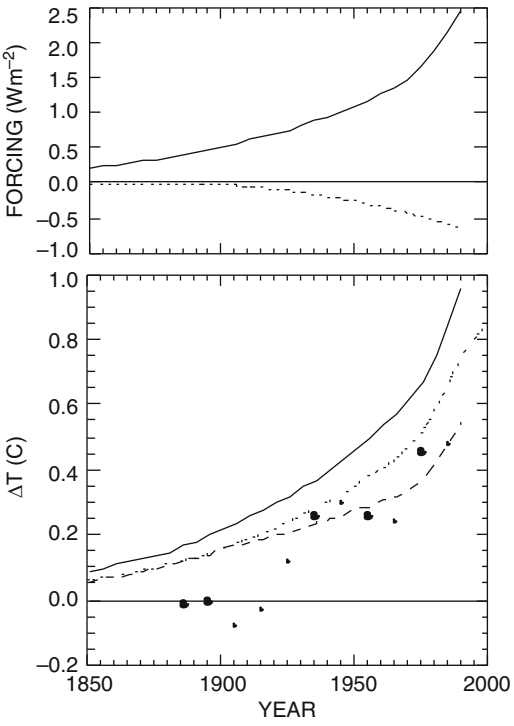



**Fig. 15.16** The total greenhouse gases forcing (*solid line on the top*) and the corresponding temperature change (*solid line below*). The experimental data referred to the 1880–1920 period are shown by the *full circles*. *Dotted line* shows the temperature change when the effect of the ocean is taken into account. *Dashed line* includes the aerosols effect whose radiative forcing is shown on the *top* with the *dashed line*

We can now introduce a more realistic forcing like those used in IPCC scenarios, that take into account, besides the carbon dioxide, also the other greenhouse gases. This is shown in the upper part of Fig. 15.15, while the lower part shows the corresponding temperature change. Comparing this result with the experimental data, we see that our model gives a much larger warming. It should be noted that in this case warming starts in 1765 so that in 1850 it is already much different from zero. Also here we have both cases with and without the ocean (Fig. 15.16).

To reconcile the experimental data with the model results, in the last few years a negative forcing has been added to the scenario that is attributed to the presence of aerosols. When fossil fuels are burned sulfur dioxide is produced and this can form (we will see in the next chapter) sulfuric acid which goes into solution in the raindrops. The resulting aerosol has very nice reflective properties. Solar radiation is then reflected and this process could compensate for the increasing greenhouse forcing. In the upper part of Fig. 15.15, this negative forcing is also shown and the relative temperature simulations look very much like the experimental data.

We can play around with these things, but besides the different constants and the simplicity of the model, a very important point is to decide the period that corresponds to the average. This has different implications. As we have remarked before going too far in the past implies a comparison with data that are quite different in quality. On the other hand, an indication may be good enough, but still it

will depend on the period chosen for the average. In our case, the deviations refer to the average between 1880 and 1920. To change this period and repeat the simulation for it may require some adjustment in the forcing. In any case it may be encouraging to know that such simple calculations are confirmed by more complex simulations, like those performed by the general circulation models (GCM) of the atmosphere, which are the most sophisticated tools for the simulation of climate.

A good way to close this chapter is to talk about them. Of course, we will not be able (due to their complexity) to be very detailed and so we will report only some of their results. These evaluations may be a little outdated but it does not seem that in recent years too much improvement has been made.

## 15.7 The General Circulation Models

The most complete approach to the study of climate dynamics can be identified with the general circulation models (GCM). These can simulate the circulation of the atmosphere and the ocean and in principle should be able to reproduce the climates of the different regions of the Earth and their evolution with time.

We must say that GCMs in a sense are a spin-off from something that is supposed to be more rigorous which is the numerical forecast models. These models are used in completely different way (as we already said) because they solve an initial value problem. Once initialized with roughly observed data, they are able to forecast the weather for several days. When used in the climate mode, the GCMs solve a boundary value problem and may produce reasonable results only when some kind of steady state is reached; this may take forever for the grid resolution of weather forecast models. For this reason in a GCM used as a climate model, the resolution usually is degraded and most of the physical processes are highly parameterized.

To have a rough idea of how many processes must be taken into account we can refer to Fig. 15.17, which is already impressive without going in to any further detail. Usually the atmosphere is treated as an adiabatic gas in hydrostatic equilibrium. We can refer to a single variable like the surface temperature and see how this is determined by the radiative flux (solar and IR). In turn the temperature may influence these same fluxes. Another important component of surface energy balance is the sensible heat which is determined by the winds and how these interact with the small-scale characteristics of the surface.

What is not apparent from this scheme are the different temporal and spatial scales involved in these processes. We go from the planetary scale for the Rossby waves to the microphysical scale for the growth processes of droplets in the clouds. This large span in the scales sometimes requires that these processes be simplified to the point that they can be described by very few variables. For example, the growth of clouds is not treated in detail, but rather each time that the relative humidity reaches a threshold, it is decided that a cloud is formed with an appropriate amount of liquid water and particles of specified size distribution. Another example is the treatment of the soil water content. In this case, a very crude approach is the *bucket*

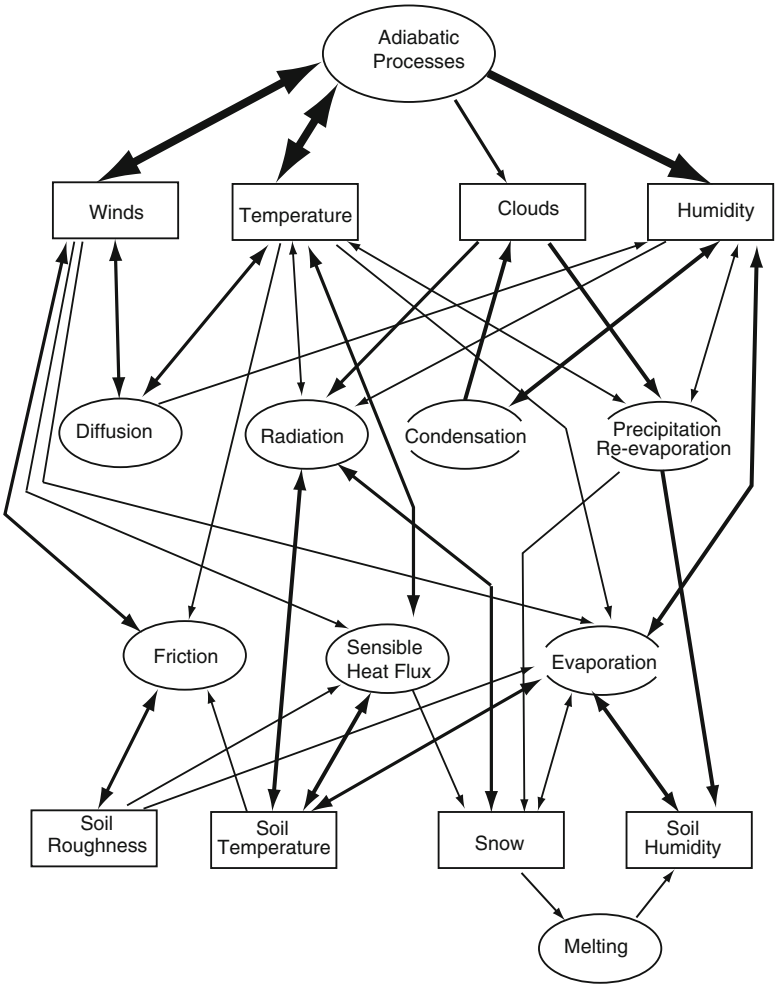



**Fig. 15.17** A schematic illustration of the processes included in a general circulation model. The *arrows* gives the direction of the process and their thickness is proportional to the importance of the different processes (From Simmonds and Bengtsson 1984)

*model*. It is assumed that a certain type of soil can only absorb a finite amount of water with the excess going as run off. In most cases, the function of the vegetation is neglected. This approach is what before we have called parameterization.

The GMC is then more like a numerical machine made of different parts that are called codes or packets. Each of these parts deals with a different process and its interaction with the others. These codes simply translate the equations we have studied (radiative transfer, thermodynamics, etc.) in to numerical equations. The solution of these equations, the flow of input and output data is such a large amount of work that it can be carried out only by the most powerful computers.

### 15.7.1 The Model Equations

The equations on which most of the models are based are called *primitive*, and we have used them in different and simplified forms. The variables that are solved in the models are usually the horizontal velocity $\mathbf{v}(u, v)$, the temperature $T$, the specific humidity $q$, and the surface pressure $p_s$. The vertical coordinate system used in most cases is denoted *sigma*, and in this case, the vertical coordinate for the pressure level $p$ is given by

$$\sigma = \frac{p}{p_s}$$

In this coordinate system, the relevant equations are the momentum

$$\frac{D\mathbf{v}}{Dt} + f\mathbf{k} \times \mathbf{v} + \nabla\varphi + RT\nabla \ln p_s = \mathbf{P_v} + \mathbf{K_v} \tag{15.39}$$

the thermodynamic equation

$$\frac{DT}{Dt} - \frac{RT\omega}{C_p p_s \sigma} = P_T + K_T \tag{15.40}$$

the conservation of water vapor

$$\frac{Dq}{Dt} = P_q + K_q \tag{15.41}$$

the conservation of mass

$$\frac{Dp_s}{Dt} + p_s\left(\nabla \cdot \mathbf{v} + \frac{\partial\dot{\sigma}}{\partial\sigma}\right) = 0 \tag{15.42}$$

and the equations of hydrostatic equilibrium

$$\frac{\partial\varphi}{\partial\sigma} = -\frac{RT}{\sigma} \tag{15.43}$$

The total derivative is also calculated in the sigma system. Besides all the other variables already defined, $P_x$ is the generalization of the force due to all parameterized processes while $K_x$ is a diffusion coefficient. Diffusion is introduced in order to smooth out the small-scale perturbations or those due to numerical integration. The surface pressure is obtained by integrating Eq. (15.42) from $\sigma = 0$ to $\sigma = 1$ so that we get $\dot{\sigma} = 0$

$$\omega = \frac{Dp}{Dt} = \sigma\mathbf{v} \cdot \nabla p_s - \int_0^\sigma \nabla \cdot (p_s\mathbf{v})\, d\sigma \tag{15.44}$$

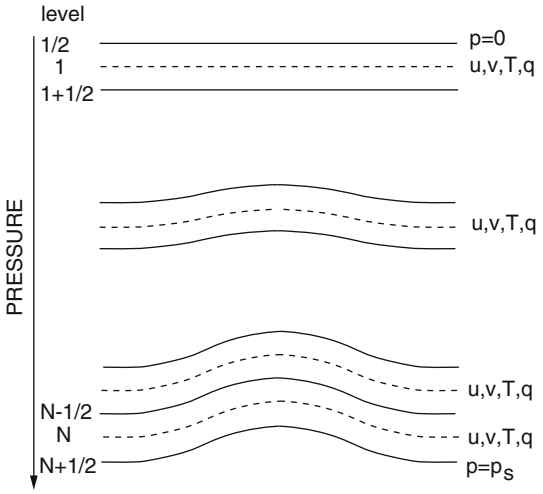



**Fig. 15.18** The vertical coordinates system in a general circulation model

The solution to these equations is obtained by transforming them into finite difference equations in the same way as was done for the two levels model. As shown in that case, the vertical coordinate system refers to layers delimited by pressure levels, as shown in Fig. 15.18. The forecast for all the variables is performed at the center of the layer.

For the horizontal grid, the choice is possible between the rectangular grid and the so-called spectral method. By referring to Fig. 15.18, we see that a rectangular grid just covers the space with grid points (where the variables are predicted) along latitude and longitude.

The spectral method in a sense is more obvious because any variables are developed in spherical harmonics

$$X(\lambda, \phi, \sigma, t) = \sum_{m=-M}^{m=M} \sum_{l=|m|}^{L} X_l^m P_l^m (\sin \phi) \, \mathrm{e}^{im\lambda} \tag{15.45}$$

where $\lambda$ and $\varphi$ are the longitude and the latitude and $P_l^m(\sin \phi)$ are the Legendre associated polynomials. With this method, the prediction is performed on the coefficients $X_l^m$. One of the advantages of this method is that it does not produce singularity at the poles or at the equator. Also the use of the FFT (fast Fourier transform) makes this method particularly efficient from the point of view of computer time. The truncation is one of the critical points of the spectral method because it can be seen from Eq. (15.45) that for each number $m$ two different summations can be performed and consequently two different truncations can be adopted. The first truncation is represented in the upper part of Fig. 15.19 and is called romboidal truncation ($R$). In it, for each zonal number $m$ an equal number of zonal terms are summed, that is $L = |m| + M$. In the triangular truncation, we have

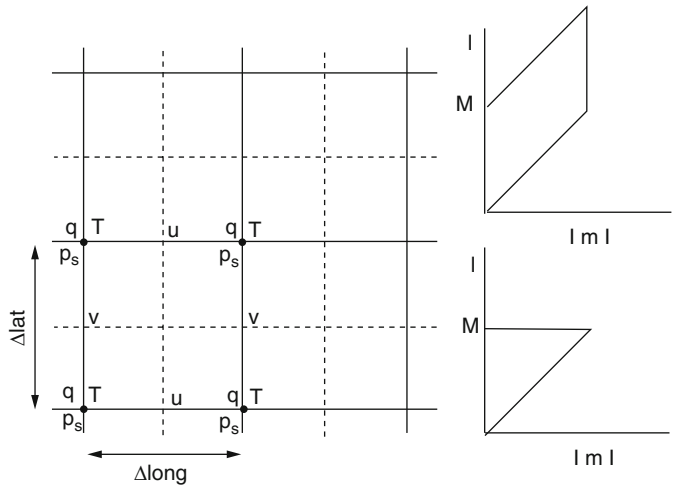



**Fig. 15.19** The grid for solving the equation with a finite difference scheme is shown at the *left*, while at the *right* the region in wave number space is shown for the rhomboidal (*up*) and triangular truncation (*bottom*)

simply $L = M$ and normally the models are indicated for example as $T21$, which means a spectral model with triangular truncation with $M = 21$.

## 15.8 The Performances of GCMs

As is customary in physics, theories are compared with experimental data. It must be stated that GCMs do not represent the expression of any theory on how the climate system works. Models are "engineering" construction used mainly to simulate past present and future climate. In the last application, the appropriate world is not *forecast* but rather but rather *projection*. The simulation of past (recent) climate is called *hindcast* and it not very popular among the practitioners. In any case, there are different problems related to test the GCM projection. One is the comparison with the data and the other is the comparison among models. In both cases because we are talking about the field of variables, methods must be found to assess globally the model results. We report a few methods suggested recently in the literature.

### *15.8.1 The Taylor Diagram*

The Taylor diagram can easily summarize the degree of correspondence between simulated and observed field. In this diagram, the correlation coefficient and the root–mean–square (RMS) difference between the two fields, along with the ratio of

the standard deviations of the two patterns, are all indicated by a single point on a two-dimensional plot. We should give now some statistical definition.

Consider two variables $f_n$ and $r_n$ which are defined at $N$ discrete points (in time and/or in space). The correlation coefficient $R$ is defined as

$$R = \frac{\sum_{n=1}^{N} \left(f_n - \overline{f}\right)\left(r_n - \overline{r}\right)}{N\sigma_r\sigma_f} \tag{15.46}$$

where $\overline{f}$ and $\overline{r}$ are the mean values, and $\sigma_r$ and $\sigma_f$ are the standard deviations for $r$ and $f$. Weight could be used to taken into account the different grid cell area or vertical pressure spacing. Another quantity of importance is the RMS difference $E$, which for the two fields $f$ and $r$ is defined as

$$E = \left[\frac{1}{N}\sum_{n=1}^{N}(f_n - r_n)^2\right]^{1/2} \tag{15.47}$$

Again weights can be used for cells of different areas.

To isolate difference in the patterns from the difference in the mean of the two fields, E can be resolved in two components. The overall "bias" is defined as

$$\overline{E} = \overline{f} - \overline{r} \tag{15.48}$$

and the centered pattern RMS difference is defined as

$$E' = \left\{\frac{1}{N}\sum_{n=1}^{N}\left[\left(f_n - \overline{f}\right) - \left(r_n - \overline{r}\right)\right]^2\right\}^{1/2} \tag{15.49}$$

If we sum the square of (15.49) and (15.48), we obtain the square of (15.46)

$$E^2 = \overline{E}^2 + E'^2$$

The correlation coefficient and the RMS difference provide complementary information about the correspondence between the two patterns. However, a more complete characterization requires the standard deviations of the fields. All these quantities can be displayed in a single diagram, if we recognize from (15.49) that

$$E'^2 = \sigma_f^2 + \sigma_r^2 - 2\sigma_f\sigma_r R$$

This relation is the same used to calculate the sides of the rectangle shown in Fig. 15.20. So that the geometric representation of Fig. 15.20.

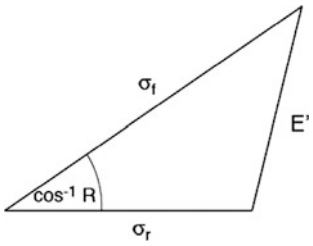



**Fig. 15.20** Geometric relationship between the correlation coefficient *R*, the centered pattern RMS error E', and the standard deviation on the test and reference field

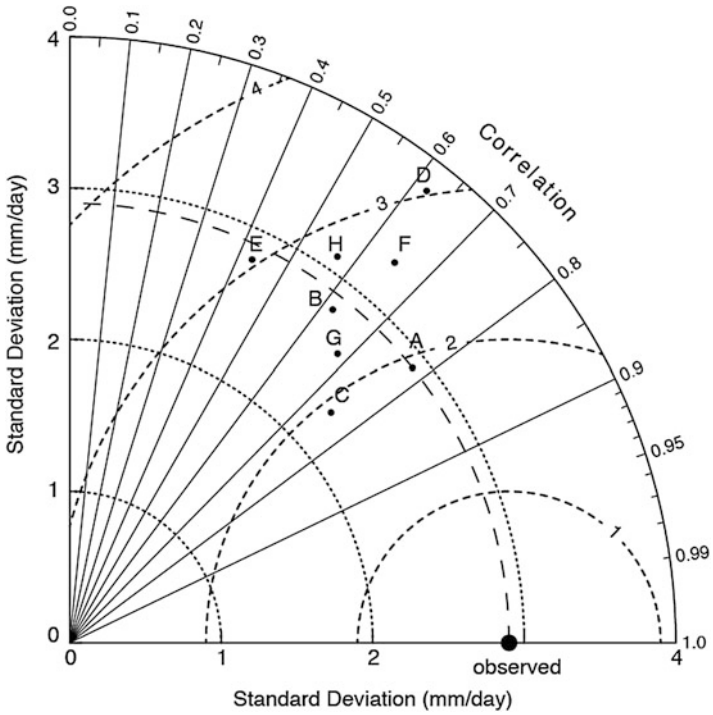



**Fig. 15.21** Diagram for displaying pattern statistics. The radial distance from the origin is proportional to the standard deviation of a pattern. The centered RMS difference between the test and the reference field is proportional to their distance apart (in units of the standard deviation). The correlation between the two fields is given by the azimuthal position of the test field (Adapted from Taylor 2001)

It is now possible to construct a diagram that statistically quantifies the degree of similarity between the filed *f* and *r*. One is defined as the "reference" filed and represents some observed state. The other will be called the "test" field and refer usually to some model simulation. The purpose is to quantify how the test field resembles to the reference field. In Figure 15.21, this kind of diagram is outlined. This diagram refers to the precipitation so that the polar axis report the standard deviation of the precipitation.

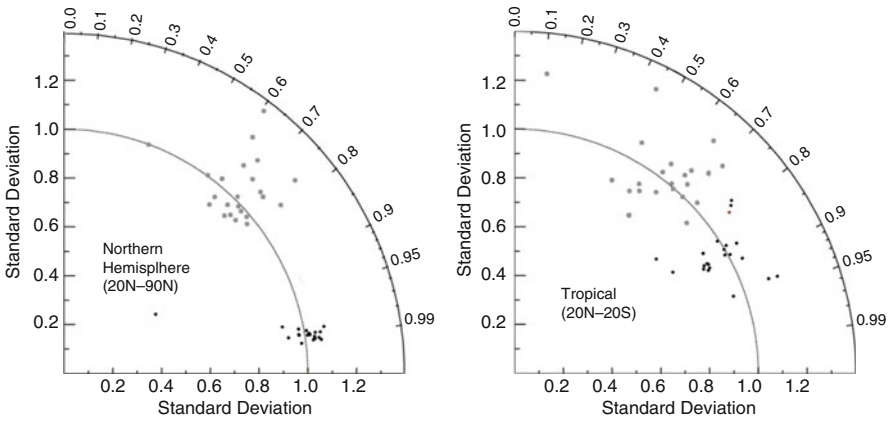



**Fig. 15.22** Taylor diagrams of the twentieth century CMIP3 annual cycle climatology (1980–1999). On the *left*, data are shown for the Northern Hemisphere for the variables, 850 hPa air temperature (*black dots*) and total precipitation. On the right, the same data are shown for the tropical region (Adapted from Taylor 2001)

The observed data is represented by the black circle on the horizontal axis. The radial distance from the origin is proportional to the standard deviation of a pattern. The centered RMS difference between the test and the reference filed is proportional to their distance apart (in units of the standard deviation). The correlation between the two fields is given by the azimuthal position of the test field. The dashed lines circle represent the distance from the reference point and then based on Fig. 15.20 indicates the RMS error. On the same diagram labeled as A, B, C . . . .are represented a number of models. For example, if we take model F we found that the correlation with the reference is 0.65 with a standard deviation of 3.2 mm/day and a RMS of 2.60 mm/day.

A possible application of this method is reported in Fig. 15.22. In this case, the diagram shows two different variables, the air temperature at 850 hPa and the total precipitation rate. These variables are taken from the collective exercise called CMIP3 (Coupled Model Intercomparison Project) and compared with the data set ERA40 (ECMWF Re-Analysis). The latter is a reanalysis of the global atmosphere and surface conditions for 45 years, over the period from September 1957 to August 2002. The only difference with respect to what said before is that the data are normalized around the standard deviation of the reference. It can be seen that for the northern hemisphere that the dispersion of the different models, there are seven of them and the correlation especially for precipitation is quite poor. The temperature shows a much better agreement with data with a lower dispersion. On the other hand, the tropical region shows for both temperature and precipitation a quite poor agreement and again especially for precipitation.

The difference between models is of the same order of magnitude as what is expected from a carbon dioxide doubling and so is not a good characteristic allowing these models to simulate anything. However even the poor agreement may be misleading. In the past, the capability of these models to reproduce the

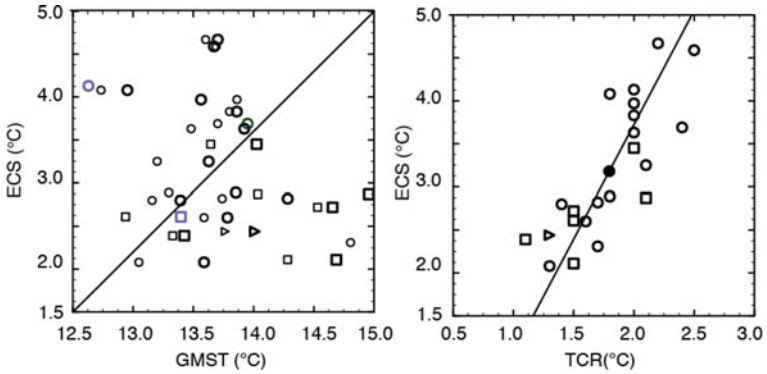



**Fig. 15.23** The Equilibrium Climate Sensitivity (ECS) as a function of Global Mean Surface Temperature (GMST) on the *left* and as a function of the Transient Climate Response (TCR) on the *right*. The small symbols are from the pre-industrial control runs. Data are from 23 GCMs

transport along the latitude has been studied. It was found that, although they show differences of a factor of 2 in the heat fluxes, they still give a rather good agreement in predicting the temperature as a function of latitude. The reason for this apparent paradox is that the models are not equal because they omit any process that may alter the reproduction of the experimental temperature profile. GCMs are then subject by their authors to a certain amount of fine tuning. Following the work of pioneers like Robert Cess, now the model evaluation is carried out by armies of researcher with final products like the IPCC reports. From the latest assessment (2013), we take Fig. 15.23 where the Equilibrium Climate Sensititivity (ECS) is shown as a function of the Global Mean Surface Temperature (GMSAT) and the same ECS as a function of the Transient Climate Response (TCR). The ECS is defined as the equilibrium change in global and annual mean surface air temperature after doubling the atmospheric concentration of carbon dioxide. The TCR, on the other hand, is the change in the same temperature when the concentration of $CO_2$ is increased by 1 % a year and calculated using the difference between the start of the experiment and a 20 years period centered on the time of $CO_2$ doubling. TCR is smaller than ECS because ocean heat uptake delays surface warming. It can be noticed that the sensitivities range between 2 and more than 4.5 °C and the amplitude has not decreased in the last 20 years. The ECS being larger than TCR it is evident from the straight line that fit the data.

### *15.8.2 The Feedback Factor*

Another important parameter that can be obtained from the models is the feedback factor defined earlier. Equation (15.30) can written as

$$\Delta T = \frac{\Delta T_0}{1 - f} \tag{15.50}$$

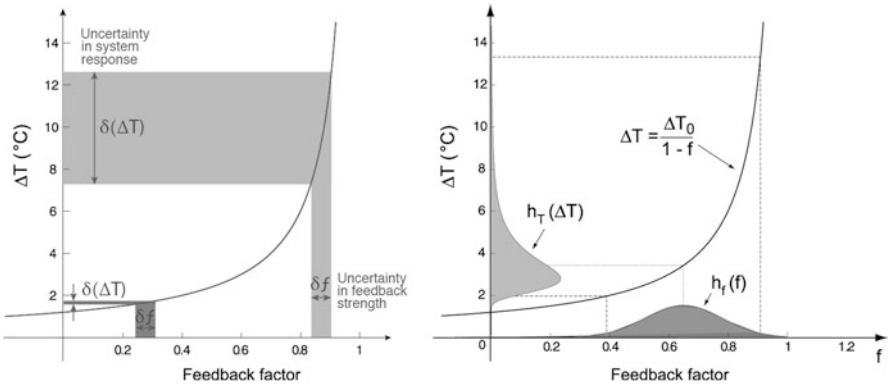



**Fig. 15.24** The relation between the uncertainty in the feedback factor and the resulting uncertainty in the climate sensitivity. To a gaussian distribution in $f$ corresponds a skewed distribution in the sensitivity (Adapted from Roe et al.)

Figure 15.24 shows the behavior of $\Delta T$ as a function of the feedback factor for a $\Delta T_0 = 1.2$ °C. Values of $f$ beyond unity are not meaningful. As shown in the figure, an uncertainty in the feedback factor ($\delta f$) will produce an uncertainty in the temperature ($\delta T$) whose amplitude will depend on the values of $f$. The same uncertainty will produce a larger $\delta T$ for feedback factor near 1. This can be easily shown for a Gaussian distribution $h_f(f)$

$$h_f(f) = \frac{1}{\sigma_f \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{(f-\bar{f})}{\sigma_f}\right)^2\right] \tag{15.51}$$

Then we can obtain the distribution of the sensitivities $h_T(\Delta T)$ using the relation

$$h_T(\Delta T) = h_f(f(\Delta T))\,(df/d\Delta T) = \Delta T_0/(\Delta T)^2 h_f\left(1 - \frac{\Delta T_0}{\Delta T}\right)$$

So that we obtain

$$h_T(\Delta T) = \frac{1}{\sigma_f \sqrt{2\pi}} \frac{\Delta T_0}{\Delta T^2} \exp\left[-\frac{1}{2}\left(\frac{\left(f - \bar{f} - \frac{\Delta T_0}{\Delta T}\right)}{\sigma_f}\right)^2\right] \tag{15.52}$$

As shown in Fig. 15.24, the skewness in the sensitivity distribution is a result of the non-linear dependence of the sensitivity on the feedback factor. The possible

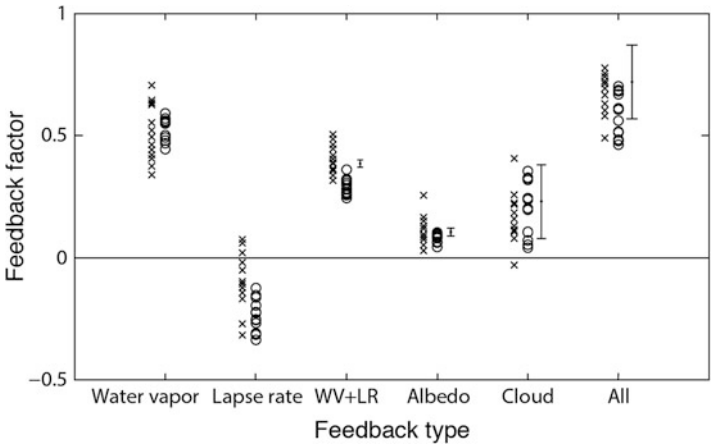



**Fig. 15.25** The feedback factors as determined by a number of models (See text) (Roe 2009)

uncertainty in the feedback factor is shown in Fig. 15.25, where $f$ is reported for the different mechanisms. We remember that the water vapor feedback (WV) is related to the increase of the water vapor as the temperature increase which in turn intensifies the greenhouse effect. The lapse rate feedback can be explained as follows. Suppose there is an enhanced warming in the upper troposphere of tropical regions in response to an increase in the concentration of greenhouse gases. Because of this change in the lapse rate, the outgoing longwave radiation will be more than in a homogenous temperature change over the vertical. The system will then lose more energy, so inducing a negative feedback. On then other hand, a larger low-level warming projected as a response to the positive radiative warming will provide a positive feedback. Figure 15.25 shows that most of the models will give a negative lapse rate feedback. As for the cloud feedback, the mechanism is mostly related to the height of the clouds. Low clouds usually have a greater albedo and higher temperature so they will reflect more solar radiation and increase the emitted long-wave radiation. The net result will be a decrease of the downward radiative flux at the top of the atmosphere. High cloud will have a lower albedo and will emit less long-wave radiation, and the net results will be an increase in downward radiation. The cloud feedback is the major uncertainty in climate models but this is not apparent in Fig. 15.25 that shows a positive feedback factor for this process.

Although the conclusions may look a bit discouraging, the role of the GCMs in predicting global change has been very important because they have given a warning at the political level and have promoted a debate at the global level about ways and means to reduce the impact of the greenhouse gases.

### 15.8.3 The Bayesian Point of View

There is a statistics that is not very well known and goes under the name of Bayes statistics. It is rather different from the classical frequentist approach because it change the probability based on the accumulation of data. The basic rule of the Bayesian statistics can be summarized as follows

$$prob\left(hypothesis \,\middle|\, data, I\right) = prob\left(data \,\middle|\, hypothesis, I\right) \times prob\left(hypothesis \,\middle|\, I\right) \tag{15.53}$$

The various terms in Bayes' theorem have formal names. The quantity on the far right, *prob*(*hypothesis*|*I*), is called the prior probability; it represents our state of knowledge (or ignorance) about the truth of the hypothesis before we have analyzed the current data. This is modified by the experimental measurements through the likelihood function, or *prob(data|hypothesis, I)*, and yields the posterior probability, *prob(hypothesis|data,I)*, representing our state of knowledge about the truth of the hypothesis in the light of the data. In a sense, Bayes' theorem encapsulates the process of learning.

There is a very simple interpretation of (15.53). We can represent the sample space where a certain number of microscopic events (N) appear as a circle (Fig. 15.26). The different microscopic possibilities results in a macrosco events like A and B, and the number A is represented by the fraction of the circle area $N_A$ and the number B by the area $N_B$. The number of possibilities resulting from both the events A and B is $N_{AB}$ and is represented by the common area of the two portions of the circle. The probability of the different events is then

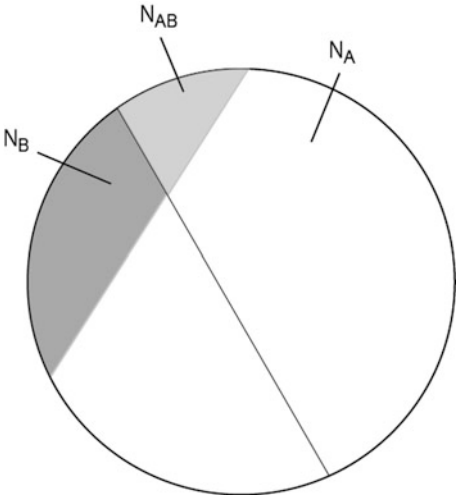



**Fig. 15.26** The sample space of microscopic events that occupies the entire circle as N. Events of type A and B occupy the indicated sector. The superposition of the two types of events is indicated by the common space $N_{AB}$ (Sivia 2006)

$$\Pr(\mathrm{A}) = \mathrm{N_A}/\mathrm{N};\ \ \Pr(\mathrm{B}) = \mathrm{N_B}/\mathrm{N};\ \ \Pr(\mathrm{A}\,|\mathrm{B}) = \mathrm{N_{AB}}/\mathrm{N_B};\ \ \Pr(\mathrm{B}\,|\mathrm{A}) = \mathrm{N_{AB}}/\mathrm{N_A};$$

Then the probability Pr(A, B) that both A and B happen is given by

$$\Pr(\mathrm{A},\ \mathrm{B}) = \mathrm{N_{AB}}/\mathrm{N} = (\mathrm{N_{AB}}/\mathrm{N_B})\,(\mathrm{N_B}/\mathrm{N}) = \Pr(\mathrm{A}\,|\mathrm{B})\Pr(\mathrm{B}) = \Pr(\mathrm{B}\,|\mathrm{A})\Pr(\mathrm{A})$$

From which we obtain the Bayes theorem

$$\Pr(\mathrm{A}\,|\mathrm{B}) = \Pr(\mathrm{B}\,|\mathrm{A}) \times \Pr(\mathrm{A}) \,/\, \Pr(\mathrm{B}) \tag{15.54}$$

There are a number of entertaining applications of this theorem.

We can make a number of examples on the application of this formula. We can evaluate the probability of extracting a figure (*J*, *Q* or *K*) of hearts from a deck of cards. The probability to extract a specific color (heart) is given by

$$\Pr(\text{heart}) = \frac{13}{52} = \frac{1}{4}$$

To evaluate that the card is also a heart, we have

$$\Pr(\text{figure}\,|\text{heart}) = \frac{3}{13}$$

So that the probability that our card be a figure of heart based on (15.54) is given

$$\Pr(\text{figure}\,|\text{heart}) = \Pr(\text{figure}\,|\text{heart})\Pr(\text{heart}) = \left(\frac{3}{13}\right) \times \left(\frac{1}{4}\right) = \frac{3}{52}$$

Actually there are 3 cards over 52 that are hearts. The same result could be obtained determining the probability to extract a figure (12/52) and then that such a figure being a heart (3/12). We have the same result

$$\Pr(\text{figure}\,|\text{heart}) = \Pr(\text{heart}\,|\text{figure})\Pr(\text{figure}) = \left(\frac{3}{12}\right) \times \left(\frac{1}{52}\right) = \frac{3}{52}$$

A more adequate example comes from medicine because based on Bayes it is possible to establish the probability of a patient to have a particular disease based on the results of clinical test. We take as an example one of the most terrible diseases, AIDS applied to Italian citizens. If the test is positive, the probability to be infected is 100 %. However, there is the possibility for the test to be positive even if one is not infected. Let's imagine that such a probability is just 0.2 % than the probability to be infected based on a positive test is now 1–0.002 = 0.98. Suppose now that based on epidemiology data, the incidence of the disease is 1 over 600 (0.1666 %) and imagine to test all the 60 million Italians so that about 6 $10^7/600 = 10^5$ Italian will result positive. The remaining people (59.9 million) will

give $59.9 \times 0.0002 = 119.800$ false positive. The percentage of the solid infected is then $10^5/(10^5 + 1.19\ 10^5) = 45.6\,\%$. This calculation can be put in a Bayesian form if we start by defining the percentage of false positive and true positive

$$\mathrm{P}\,(\mathrm{pos}) = \mathrm{P}\,(\mathrm{pos, AIDS\ yes}) + \mathrm{P}\,(\mathrm{pos, AIDS\ no})$$

And as conditional probability

$$\mathrm{P}\,(\mathrm{pos}) = \mathrm{P}\,(\mathrm{pos}\,|\mathrm{AIDS\ yes})\,\mathrm{P}\,(\mathrm{AIDS\ yes}) + \mathrm{P}\,(\mathrm{pos}\,|\mathrm{AIDS\ no})\,\mathrm{P}\,(\mathrm{AIDS\ no})$$

This result can be used in the Bayes formula

$$\mathrm{P}\,(\mathrm{AIDS\ yes}\,|\mathrm{pos}) = \mathrm{P}\,(\mathrm{pos}\,|\mathrm{AIDS\ yes})\,\mathrm{P}\,(\mathrm{AIDS\ yes})\,/\mathrm{P}\,(\mathrm{pos})$$

Substituting the numerical values P(pos|AIDS yes) = 1 (100 %), P(AIDS yes) = 1/600, P(pos|AIDS no) = 0.002 (0.2 %) e P(AIDS no) = 0.998 (1–0.002) = 99.8 % we obtain

$$\mathrm{P}\,(\mathrm{AIDS\ yes}\,|\mathrm{pos}) = 45\ \%$$

The final result tells us that being positive to a test means to have a 45 % probability to be infected (instead of 100 %) and this important decrease is due to the 0.2 % uncertainty.

Several other applications of the Bayes theorem could be mentioned and some of them is given as exercise but for the time being let us concentrate on how Bayes could impact the evaluation of models.

### *15.8.4 The Bayesian Evaluation of Models: Part 1*

We have already found Bayes even if we did not knew it when discussing climate sensitivity and feedback factor. We could assume that $h_{\mathrm{post}}(\Delta T)$ is the posterior probability for the climate sensitivity and then based on the Bayes theorem

$$h_{\mathrm{post}}\,(\Delta T) = \frac{p\left(f_{\mathrm{obs}}\,\middle|\,\Delta T\right)}{p(f)} h_{\mathrm{prior}}\,(\Delta T) \tag{15.55}$$

where $p\left(f_{\mathrm{obs}}\,\middle|\,\Delta T\right)\,/p(f)$ is the normalized probability that a observation yielding $f_{\mathrm{obs}}$ is consistent with a climate sensitivity ΔT. We can assume a Gaussian distribution for $p\left(f_{\mathrm{obs}}\,\middle|\,\Delta T\right)$ similar to what we assumed for $h(f)$ and for any the prior distribution can be converted as $h_{\mathrm{prior}}(z)dz = h_{\mathrm{prior}}\,(\Delta\ T)\,dT$. A natural choice is that all the feedback are equally likely ($z = f$, $h_{\mathrm{prior}}(f)$ is a constant). Now we

have $f = 1 - \Delta T_0/\Delta T$ then $h_{\text{prior}}(\Delta T) = df/d\Delta T = \Delta T_0/(\Delta T)^2$. If we compare (15.55) with (15.52), we have that

$$p\left(f_{\text{obs}} \left| \Delta T\right.\right) / p(f) = \frac{1}{\sigma_f \sqrt{2\pi}} \exp\left[ -\frac{1}{2}\left( \frac{\left(f - \bar{f} - \frac{\Delta T_0}{\Delta T}\right)}{\sigma_f} \right)^2 \right] \tag{15.56}$$

The most promising application of the Bayes statistics has to do with classification of the models and we will proceed with two examples. The first one is to find evidence for the models when applied to find the change in temperature in a number of region. We can think about the Global Historical Climatology Network (GHCN) of the Goddard Institute for Space Studies (GISS). The historical data cover the period 1951–2010. We have used three geographical domains, Mediterranean, USA, and China and for each model (we consider 5 of them); the time average of the deviation between 1981–2010 and 1951–1980 was evaluated

$$\widehat{x} = \frac{1}{m}\sum_{j=1}^{m} x_j \tag{15.57}$$

where $m$ is the number of run fort he model. Then to describe the width of the distribution, we calculate the covariance matrix

$$\Sigma_x = \frac{1}{m-1}\sum_{j=1}^{m} \left(x_j - \widehat{x}\right)\left(x_j - \widehat{x}\right)^T \tag{15.58}$$

where $()^T$ indicate the transpose matrix. The vector $x$ has $N = 3$ dimension corresponding to the geographical regions. In one of the exercises of this chapter, we show that the prior for this problem could be written as

$$p\left(x \left| M_i\right.\right) = (2\pi)^{-1/2} |\Sigma_x|^{-1/2} \exp\left[ -\frac{1}{2}\left(x_j - \widehat{x}\right)^T \Sigma_x^{-1} \left(x_j - \widehat{x}\right) \right] \tag{15.59}$$

in this case $|\Sigma_x|$ is a matrix determinant. We now need to determine the likelihood function. The best way is to imagine that it is associated with the observing instrumentation alone and not at all with any model. We could indicate it as $p(d|x)$, where both $\boldsymbol{x}$ and $\boldsymbol{d}$ represent a vector of regional temperature change that is the 1981–2010 average to the 1951–1980 average. However, $\boldsymbol{x}$ is the simulated change while $\boldsymbol{d}$ is the measured change. The likelihood function answer the question, if the true regional temperature changes were $\boldsymbol{x}$, what is the probability that we obtained the measurement $\boldsymbol{d}$? The only way the two data could differ is that because the data set could have an uncertainty $s$. The likelihood function would then be

$$p\left(d \left| x\right.\right) = \left(2\pi\sigma^2\right)^{-N/2} |\Sigma_x|^{-1/2} \exp\left[ -\frac{1}{2\sigma^2} |d - x|^2 \right] \tag{15.60}$$

The Bayesian evidence function is then given by

$$p\left(d\left|M_i\right.\right) = \int_0^\infty p\left(x, d\left|M_i\right.\right)dx = \int_0^\infty p\left(d\left|x\right.\right)p\left(x\left|M_i\right.\right)dx \tag{15.61}$$

where we have been using (15.53) or (15.54). Now we assume that all the distributions are Gaussian so that we get a nice analytical solution to the integral

$$p\left(d\left|M_i\right.\right) = (2\pi)^{-N/2}\left|\Sigma_x + \sigma^2\mathbf{I}\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{d}-\widehat{\mathbf{x}}\right)^T\left(\Sigma_x + \sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{d}-\widehat{\mathbf{x}}\right)\right] \tag{15.62}$$

where $\mathbf{I}$ is the unit matrix. Properly speaking, this is the probability of having obtained data $\boldsymbol{d}$ on the condition that model $i$ is the true model. Each model will have its own $\widehat{\mathbf{x}}$ and $\Sigma_x$ corresponding to its mean simulation of historical change and that model uncertainty is that estimate due to natural variability. It is possible to use this evidence to rank the models if you assume that the prior on each model is the same (Fig. 15.27). As a matter of fact, we have

$$\frac{p\left(M_i\left|d\right.\right)}{p\left(M_j\left|d\right.\right)} = \frac{p\left(d\left|M_i\right.\right)}{p\left(d\left|M_j\right.\right)}\frac{p\left(M_i\right)}{p\left(M_j\right)} = \frac{p\left(d\left|M_i\right.\right)}{p\left(d\left|M_j\right.\right)} \tag{15.63}$$

Sometime the ratio is called Bayes Factor, if the denominator is the model that performs better.

### 15.8.5 The Bayesian Evaluation of Models: Part 2

In the exercise section, we have made some simple example about the application of Bayes statistics. Now starting from that we can say something more about climate models evaluation. We need to fix some variables of the problems, which are summarized in Table 15.1 (taken from Leroy 1998). We then define based on (E.15.5) a normal distribution generalized to $n$ dimension

$$\mathcal{N}\left(\mathbf{x}, \mathbf{X}\right) = (2\pi)^{-n/2}|\mathbf{X}|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{X}^{-1}\mathbf{x}\right) \tag{15.64}$$

To obtain the prior $p(E)$, we use the fact that it consists of the predicted signal amplitude with an associated uncertainty determined by the errors in the parameters of the model. The error covariance matrix for the signal amplitude ($\boldsymbol{\alpha}$) is the matrix $\mathbf{A}$. The prior most probable value for the signal amplitude is $\boldsymbol{\alpha}_p$ so that the prior distribution is

$$p\left(E\right) = P\left(\alpha\left|M_i\right.\right) = \mathcal{N}\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}_p, \mathbf{A}\right) \tag{15.65}$$

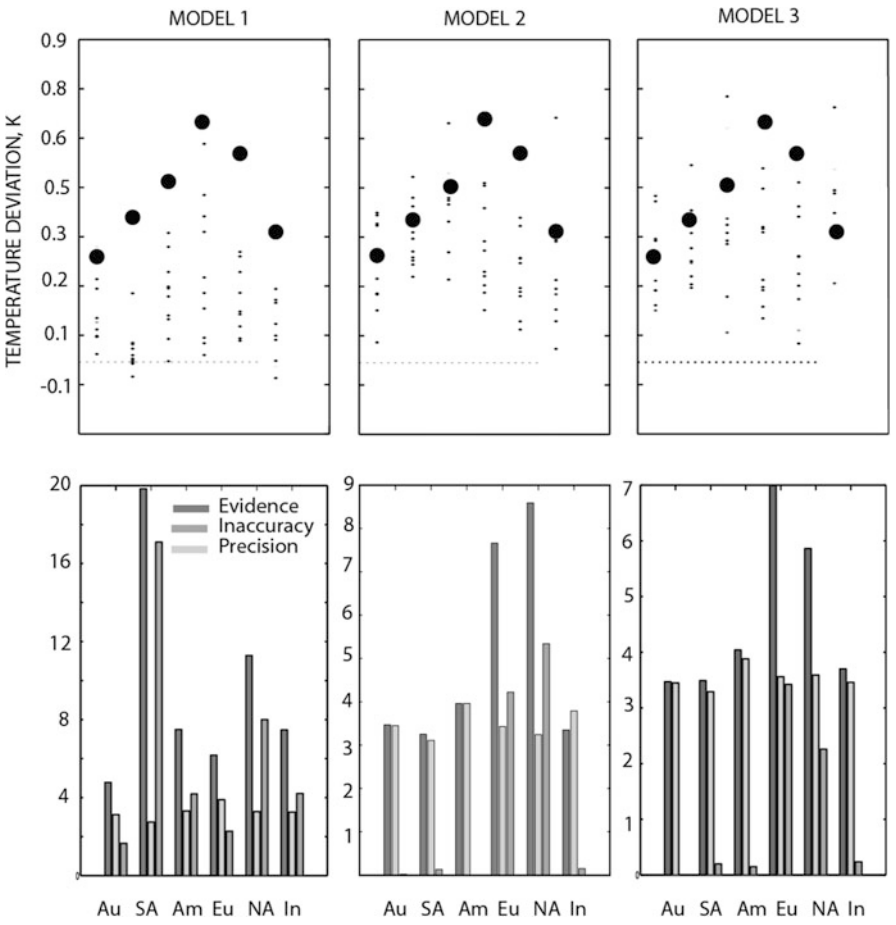



**Fig. 15.27** Figure labeled with models show the average temperature deviation between the 2010–1981 and 1980–1951 for six different regions. The last figure shows the value of the evidence, inaccuracy, and precision computed with the same data. The *black dots* are the average from the real data, and the *black diamond* refers to different run of the model

**Table 15.1** Description of the algebraic notation

| Variable | Dimensions | Description |
|---|---|---|
| $\mathbf{d}$ | $N$ | A set of post-processed observed data |
| $\mathbf{N}$ | $N \times N$ | The covariance matrix of natural climate variability and measurement error |
| $\mathbf{S}$ | $N \times m$ | $S_{ij}$ is the normalized amplitude of signal j at data coordinate i |
| $\boldsymbol{\alpha}$ | $m$ | A variable array of signal amplitudes |
| $\boldsymbol{\alpha}_\mathbf{p}$ | $m$ | The model prior for climate signal amplitudes |
| $\mathbf{A}$ | $m \times m$ | The covariance matrix of the uncertainties in the signal amplitude prediction $\boldsymbol{\alpha}_\mathbf{p}$ |
| $\boldsymbol{\alpha}_\mathbf{m}$ | $m$ | The posterior "most probable" estimate of the signal amplitudes |
| $\mathbf{A}_\mathrm{m}$ | $m \times n$ | The uncertainty covariance of the posterior estimate of the signal amplitude |

The likelihood for the data P(D|E) where the event E represents a possible set of signal amplitudes $\boldsymbol{\alpha}$ and the data D represents a set of measurements **d**. The relation between the signal amplitudes and the data is given by

$$\mathbf{d} = \mathbf{S}\boldsymbol{\alpha} + \mathbf{n}$$

where **n** is the sum of natural variability and measurement errors. This is a random with zero mean and a covariance matrix **N.** The probability of obtaining the measurements **d** given the signal amplitudes $\boldsymbol{\alpha}$ is then

$$p\left(D\,|E\right) = P\left(\mathbf{d}\,|\boldsymbol{\alpha}\right) = \mathcal{N}\left(\mathbf{d} - \mathbf{S}\boldsymbol{\alpha}, \mathbf{N}\right) \tag{15.66}$$

The posterior probability is obtained by multiplying the prior (15.65) for the likelihood (15.66) to obtain

$$p\left(E\,|D\right) = P\left(\boldsymbol{\alpha}\,|\mathbf{d}, M_i\right) \propto \mathcal{N}\left(\mathbf{d} - \mathbf{S}\boldsymbol{\alpha}, \mathbf{N}\right) \mathcal{N}\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}_p, \mathbf{A}\right) \tag{15.67}$$

The most probable posterior estimate for the signal amplitudes is the $\boldsymbol{\alpha}$ which maximizes the distribution. By varying $\boldsymbol{\alpha}$, the most probable signal amplitude is found

$$\boldsymbol{\alpha}_m = \left(\mathbf{S}^T\mathbf{N}^{-1}\mathbf{S} + \mathbf{A}^{-1}\right)^{-1}\left(\mathbf{S}^T\mathbf{N}^{-1}\mathbf{d} + \mathbf{A}^{-1}\boldsymbol{\alpha}_p\right) \tag{15.68a}$$

with the associated uncertainty covariance

$$\mathbf{A}_m = \left(\mathbf{S}^T\mathbf{N}^{-1}\mathbf{S} + \mathbf{A}^{-1}\right)^{-1} \tag{15.68b}$$

The last two equations could be used to find the signal-to-noise ratio of a possible signal detection. For example, the signal-to-noise ratio of the detection of signal $i$ is simply $|\alpha_i|/(A_m)_{i,i}^{1/2}$. Given the uncertainty matrix in the prior **A**, the precision of the prediction is proportional to $|\mathbf{A}|^{1/2}$. Likewise the precision of the detection is proportional to $|\mathbf{A}_m|^{1/2}$.

The evidence for the model given the data, $P(M_i|D)$ is obtained by multiplying the evidence of the data given the model $P(D|M_i)$ by the prior of the model $p(M_i)$. We assume a flat prior $[p(M_i) = 1]$. The evidence for the data is then obtained by integrating the data likelihood and the signal prior over all the possible events. This amounts to integrating the right side of Eq. (15.67) over all signal amplitudes $\boldsymbol{\alpha}$ with the result

$$P\left(D\,|M\right) = P\left(d\,|M_i\right) = (2\pi)^{-N/2}\left(\frac{|\mathbf{A}_m|}{|\mathbf{A}|\,|\mathbf{N}|}\right)^{1/2}\exp\left(-\frac{\mathcal{A}}{2}\right) \tag{15.69a}$$

where the accuracy A of the prediction is given by

$$\mathcal{A} = (\mathbf{d} - \mathbf{S}\boldsymbol{\alpha}_m)^T \mathbf{N}^{-1} (\mathbf{d} - \mathbf{S}\boldsymbol{\alpha}_m) + \left(\boldsymbol{\alpha}_m - \boldsymbol{\alpha}_p\right)^T \mathbf{A}^{-1} \left(\boldsymbol{\alpha}_m - \boldsymbol{\alpha}_p\right) \tag{15.69b}$$

The first term on the right describes the mismatch between the data and the posterior signal, and the second term describes the mismatch between the prediction and the posterior signal. The ensemble average of the first term on the right is the number of data points, or the rank of **N** if the variability accurately describes the properties of the real atmosphere. When the accuracy $A$ is less or equal than the data points, it is said that the model prediction and the data agree with each other.

The relative probability of a model can be evaluated given a dataset and that it depends on the *precision* and the *accuracy* of the model prediction. The prediction of the accuracy is given by $|\mathbf{A}|^{1/2}$. Because this quantity appear in the denominator of (15.69a) that means that a model that cannot give a precise prediction is less probable that the one it does. The accuracy of the model is A and if that is small the model is said to be accurate.

## E.15 Examples

### *E.15.1 100 Kyear Glacial Cycle: Details*

The numerical solution of Eqs. (15.12) and (15.13) is quite simple but requires some further elaboration. In the two Eq. (15.12), we can perform the derivative and obtain

$$\begin{aligned} \frac{dR}{d\tau}\left[\frac{3R+2\Gamma}{3R}\right] &= \left(R - R^2 + p\right)\left[1 - \kappa\left(R - R^2 + p\right)\right] - \frac{1}{3}\mu\left(\frac{R}{3} - \Gamma\right) \\ \frac{dR}{d\tau}\left[\frac{3R+2\Gamma}{3R}\right] &= \frac{1}{4\kappa} - \frac{1}{3}\mu\left(\frac{R}{3} - \Gamma\right) \end{aligned} \tag{E.15.1}$$

where we have substituted for the $\Gamma$ derivative Eq. (15.13).

These equations can be readily integrated with standard routine or (as in the following FORTRAN program) with a simple incremental method.

```
program 100kyr
    real p, r, tau, gamma, aux, k, mu,  cost,   dtau, h
    parameter (beta0=-0.0426)
    parameter (beta1=0.0714)
    parameter (omega=0.938)
    parameter (k=1.29)
    parameter (mu=0.55)
```

```
      parameter (dtau=0.005)

      open (20,status='unknown',file="icesheet.dat ")

c     initial conditions:

      gamma=0.0
      r=0.02
      tau=0.0
      p=beta0
      h=(7*r)/(2*0.001)

c     solving the equations:

      do while (tau.le.75)
      write (20,*) tau*2740, r, h/1000,p

      tau=tau+dtau
      p=beta0+beta1*sin(omega*tau)
      gamma=mu*(r/3-gamma)*dtau+gamma
      aux=r-r**2+p
      cost=(3*r)/(3*r+2*gamma)

      if (aux<1/(2*k)) then
      r=cost*(aux*(1-k*aux)-mu/3*(r/3-gamma))*dtau+r
      else
      r=cost*(1/(4*k)-mu/3*(r/3-gamma))*dtau+r
      end if

      h=(7*r)/(2*0.001)
      end do

      stop
      end
```

## *E.15.2 A Multi-state Climate Model for the Timing of Glaciations*

Didier Paillard in 1998 proposed a simple model to explain the timing of the glaciations. The basic hypothesis is that climate could assume three different states as shown in Fig. E.15.1. The three distinct states are indicated as **i** (interglacial), **g** (mild glacial), and **G** (glacial). Among these three states, transitions are regulated by the insolation and the ice volume according to the following rules

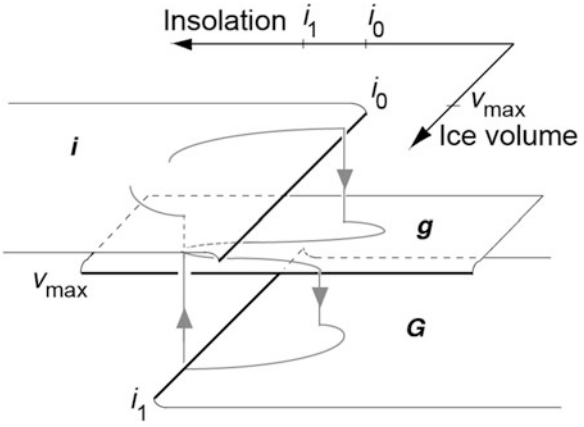



**Fig. E.15.1** Schematic structure of the three state systems in the Paillard model. The climate is assumed to have three distinct regimes **i**, **g**, and **G**. The **i G** transition is triggered when the insolation falls below the threshold $i_0$, the **g G** transition occurs when the ice volume exceeds vmax and **G i** transition occurs when the insolation increases above the level $i_1$ (Paillard 1998)

Transition from **i** to **g** if the insolation goes below $i_0$
Transition from **g** to **G** if the ice volume becomes greater than $v_{\max}$
Transition from **G** to **i** if the insolation becomes greater than $i_1$

The status **g** is localized between **i** and **G** only when the ice volume is low. That volume will be low at the end of regime **i** and high at the end of regime **G**. This mean that the direct transitions **i-G** and **g-G** are not permitted.

Without any explicit control on the ice volume, the only requirement for the volume to become greater that $v_{\max}$ is the growth time $t_g$ (that is the duration of regime **g** is longer than $t_g$) beyond the fact that the maximum value for the insolation must remain below $i_3$ before a possible transition **g-G**. The transition **g-G** is permitted only when the insolation decreases below $i_2$.

This simple model is made a little more sophisticated by allowing a continuous change of ice volume. This is given by the simple relationship

$$\frac{dv}{dt} = \frac{(v_R - v)}{\tau_R} - \frac{F}{\tau_F} \tag{E.15.2}$$

Where $v_R$ is the ice volume in one of the three regimes (**i, g, G**), $v$ is the current volume, $F$ is the forcing, and $\tau_R$ and $\tau_F$ are time constant. The volume is normalized to unity so that

$$v_g = v_G = v_{\max} = 1, \quad v_i = 0$$

The equation for the ice volume is solved numerically in a very simple way. Using the index $i + 1$ for the successive time, the difference equation becomes

$$v\,(i+1) - v\,(i) = \frac{v_R\,(i) - v\,(i)}{\tau_R} - \frac{F}{\tau_F}$$

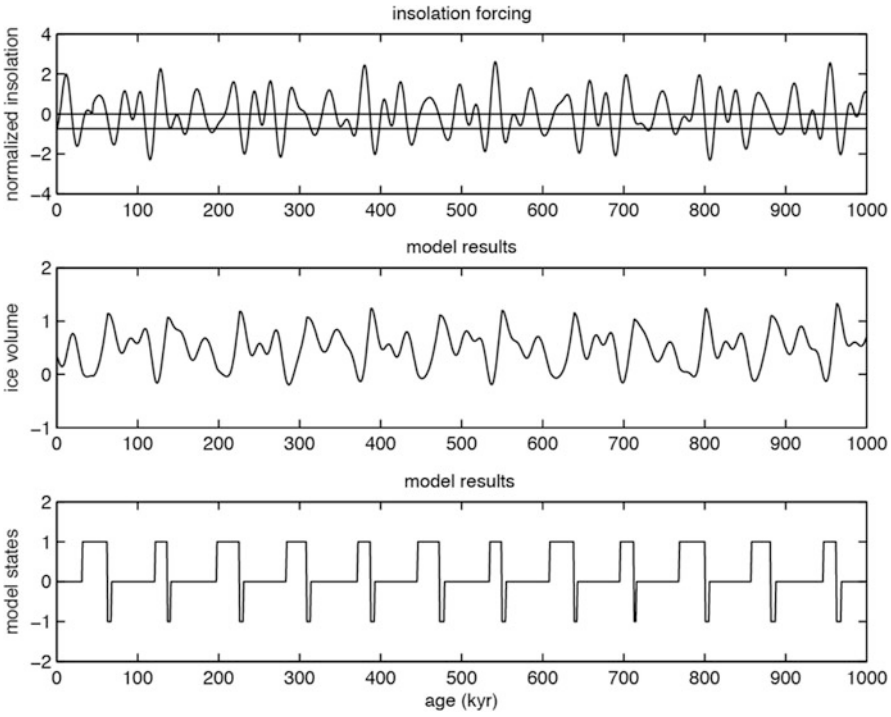



**Fig. E.15.2** A typical result of the program simulating the ice ages. The figure at the *bottom* gives the most simple models where the system is simply triggered among the three states, i, g, and G

From which we obtain the ice volume at the previous time

$$v\,(i) = \left[v\,(i+1) + v_R\,(i)\,/\tau_R - F\,(i)\,/\tau_F\right]\tau_R/\,(1+\tau_R)$$

And this is the expression used in the program. In the original paper by Paillard, there are all the details concerning the choose of the time and the forcing. In the program reported below, the file insolam.dat has been obtained with a separate program also reported below (iceinso). A typical result of this program is reported in Fig. E.15.2 and shows many of the features observed in the ice core data.

```
%program paillard
load insolam.dat
x=insolam(:,1);
inst=insolam(:,3);
i0=-0.75;
i1=0;
vmax=1;
tF=23;
stato(1001)=0;
```

```
v(1001)=0.75;
vR(1001)=1;
tR(1001)=50;
for i=1000:-1:1,
    if(inst(i)<i0 & stato(i+1)==1)
        stato(i)=0;
        vR(i)=1;
        tR(i)=50;
        v(i)=(v(i+1)+vR(i)/tR(i)-inst(i)/tF)*tR(i)/(1+tR(i));
    elseif (inst(i)>i1 & stato(i+1)==-1)
        stato(i)=1;
        vR(i)=0;
        tR(i)=10;
        v(i)=(v(i+1)+vR(i)/tR(i)-inst(i)/tF)*tR(i)/(1+tR(i));
    elseif (stato(i+1)==0 & v(i+1)>vmax)
        stato(i)=-1;
        vR(i)=1;
        tR(i)=50;
        v(i)=(v(i+1)+vR(i)/tR(i)-inst(i)/tF)*tR(i)/(1+tR(i));
    else
        stato(i)=stato(i+1);
        vR(i)=vR(i+1);
        tR(i)=tR(i+1);
        v(i)=(v(i+1)+vR(i)/tR(i)-inst(i)/tF)*tR(i)/(1+tR(i));
    end
end
  subplot(3,1,1)
  plot(x,inst,'r');
  hold on
  a =[0 1000];
  b =[i0 i0];
  plot(a,b,'k')
  hold on
  c=[i1 i1];
  plot(a,c,'k')
  hold on
  axis([0 1000 -4 4])
  hold off
  title('insolation forcing')
  ylabel('normalized insolation')
  subplot(3,1,2)
  plot(v,'c')
  axis([0 1000 -1 2])
  title('model results')
  ylabel('ice volume')
```

```
  subplot(3,1,3)
  plot(stato,'b')
  axis([0 1000 -2 2])
  title('moldel results')
  xlabel('age (kyr)')
  ylabel('model states')

% Program ICEINSO

clear all
close all

s = 1440*1.92;
lat = 65*pi/180;

for l = 0:1000

    t = l*1000;
    dmax = (23.35+1.15*sin(2*pi*t/41000. + 7.5*pi/180.))
         *pi/180.;
if (t < 45000)
        e = 0.0165+0.0035*sin(((360.*t/64000.)+8.21)*pi/180.);
    else
        e = 0.0165+0.0135*sin(((360.*(t-45000)/140000.)-90.)
         *pi/180.);
    end
g=2*t*pi/23000;
    t = 181;
    delta = dmax*sin(360*pi*(t-89)/(365*180));
    theta = (360*pi*(t-13)/(365*180))+g;
    r = (1-e^2)/(1+e*cos(theta));
    f = r^-2;
    c = tan(lat)*tan(delta);
    a = pi/2-(abs(delta));
if (abs(lat) > a)
b=delta*lat;
        if (b > 0)
            qs = s*f*sin(lat)*sin(delta);
        else
            qs=0.;
        end

    else
h = acos(-c);
        d = cos(lat)*cos(delta);
```

```
        qs = (s/pi)*f*(h*sin(lat)*sin(delta)+d*sin(h));
        qst = 0.5*(qs+sqrt(4.+qs*qs));

    end
% Output variables
    qsn(l+1) = qs;
    qstn(l+1) = qst;

end

% Mean and std of output variables
qm  = mean(qsn);
qmt = mean(qstn);
devq = std(qsn);
devqt = std(qstn);

% WRITE OUT
[fid,msg] = fopen('insolam.dat','wt');
if (fid<0), error(msg), end
fprintf(fid,'%d %f %f\n',[(0:length(qsn)-1)',((qsn-qm).
  /devq)',((qstn-qmt)./devqt)']');
fclose(fid);

% FIGURES
figure
plot((qsn-qm)./devq)
title('(qsn-qm)/devq')
```

### *E.15.3 The Wigley – Schlesinger Model (15.6.1)*

We report below also the program for the model by Wigley and Schlesinger described in paragraph 15.6.1. The program is made of two section a *runwigley*, which drives the simple program that integrates the simple differential Eq. (15.38) (Figs. E.15.3 and E.15.4).

```
% runwigley
gamma=0.71;
y0=0.01;
Cp=4.18e3;
k=0.0003;
h=100;
rho=1000;
lamda=0.6;
mu=2.2;
```

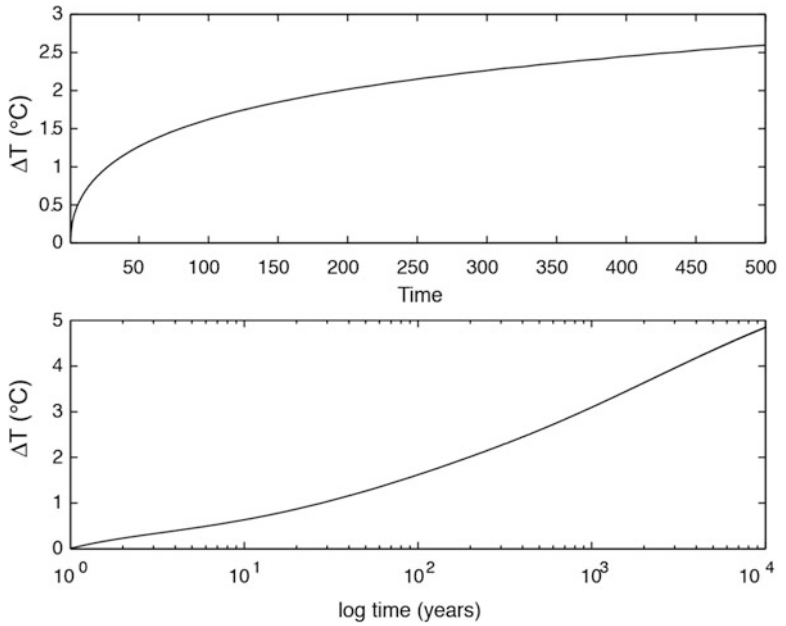



**Fig. E.15.3** Some result from the Wigley–Schlesinger model

```
taud=pi*h*h/(3.15e7*k)
tauf=rho*Cp*h/(3.15e7*lamda)
[t,y]=ode23(@(t,y) wigley(t,y,gamma,tauf,taud,mu),
  [1 20000],y0);
subplot(2,1,1)
        plot(t,y)
        set(gca,'xlim',[1,500])
        title(sprintf('y vs time '))
subplot(2,1,2)
        semilogx(t,y)
        set(gca,'xlim',[1,10000])
        title(sprintf('y vs time log'))
[fid,msg] = fopen('runwig.txt','wt');
fprintf(fid,'%12.3e %12.3\n',[t,y]');
fclose(fid);

%function wigley
function ydot=wigley(t,y,gamma,tauf,taud,mu)
Qf=4;
h=100;
rho=1000;
k=0.0003;
```

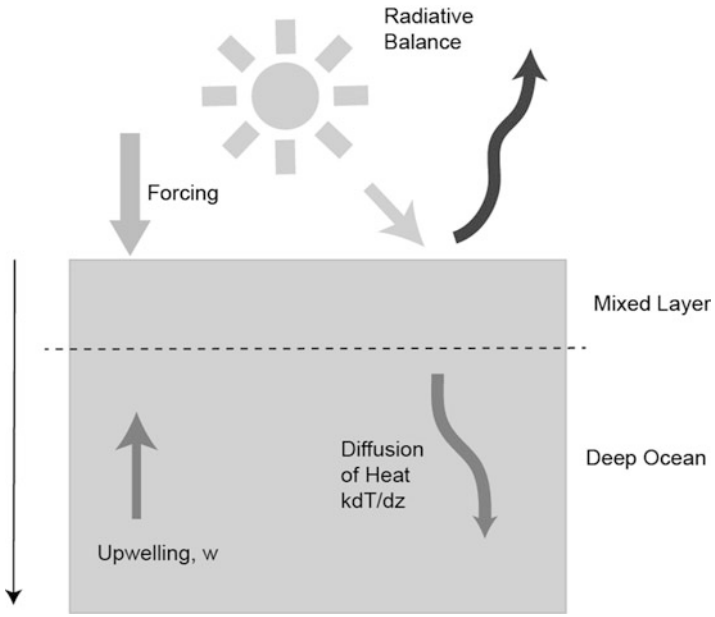



**Fig. E.15.4** Schematic model to evaluate the climate sensitivity

```
w=2.0;%in metri all'anno
Cp=4.13e3;
CC=rho*h*Cp;
CCg=rho*h*Cp*gamma;
Q=Qf*3.15e7/CCg;
ydeep=5000;
eta=0.5*ydeep/(sqrt(3.15e7*k*t));
tdeep=y*erfc(eta);
taudt=taud*t;
    deepf=w*(y-tdeep)/(h*gamma);
%ydot=Q-(y/gamma)*(1/tauf+mu/sqrt(taudt));
ydot=Q-(y/gamma)*(1/tauf+mu/sqrt(taudt))-deepf;
```

## *E.15.4 A Model to Explore Climate Sensitivity*

Baker and Roe (2009) have studied the climate sensitivity with a very simple model illustrated in Fig. E.15.4. It is made up by a oceanic mixed layer above a semi infinite ocean. The mixed layer has depth, density $\rho$, specific heat $C_p$ and thermal conductivity $\kappa$. The transport in the ocean below the mixed layer is dominated by diffusivity $\chi$ (in $m^2\ s^{-1}$) and upwelling velocity $w$. We assume a uniform, time

dependent $T(t)$ temperature in the mixed layer and a temperature at depth $T_{th}(z,t)$. The mixed layer temperature obeys to the following energy balance equation

$$\rho C_p h \frac{dT}{dt} = \kappa \frac{dT_{th}}{dz}\bigg|_{z=0} - \frac{T\,(1-f_a)}{\lambda_0} + \Delta R_F(t) \tag{E.15.3}$$

In this equation $f_a$ is the sum of all feedback factors and $\lambda_0$ is the climate sensitivity, that is $\Delta T_0 = \lambda_0 \Delta R_F$, while the real change in temperature is given by

$$\Delta T = \lambda_0 \left(\Delta R_F + C\Delta T\right) \quad \Rightarrow \quad \frac{\Delta T = \lambda_0 \Delta R_F}{(1 - C\lambda_0)}$$

If we assume $f_a = \lambda_0 C$ we get easily (E.15.3). For the deep ocean we assume that transport is due to advection and diffusion so that the equation is the following

$$\frac{\partial T_{th}}{\partial t} = \chi \frac{\partial^2 T_{th}}{\partial z^2} - w \frac{\partial T_{th}}{\partial z} \tag{E.15.4}$$

where $w$ is the upward advection velocity. The last two equations can be solved numerically by dividing the entire ocean region in layers of assigned thickness plus the mixed layer. As done with problems of the same kind we assume that for each layer the temperature will be a function of time and $z$, $T_{th}(z,t)$. The index $i$ will indicate the layer while the index $j$ will indicate the time. The novelty here is that we will calculate the right-hand side of (E.15.4) as an average between time $j$ and $j+1$. Equation (E.15.3) it is changed a little bit remembering that $G = 1/(1-f_a) = \lambda/\lambda_0$

$$\Delta T = \frac{\kappa \Delta t}{\rho C_p h} \frac{\Delta T_{th}}{\Delta z} - \frac{\Delta t}{\rho C_p h \lambda} T + \frac{\Delta R_F(t) \Delta t}{\rho C_p h}$$

now we call

$$\gamma = \frac{\kappa \Delta t}{\rho C_p h \Delta z}; \quad \delta = \frac{\Delta t}{\lambda \rho C_p h} T$$

An the difference equation between layer 1 (mixed layer) and the first layer of deep ocean (layer 2) becomes

$$T^1_{j+1}\left(1 + \frac{\delta}{2} + \frac{\gamma}{2}\right) - T^2_{j+1}\left(\frac{\gamma}{2}\right) = T^1_j \left(1 - \frac{\delta}{2} - \frac{\gamma}{2}\right) + \frac{\gamma}{2} T^2_j + \frac{\Delta R_F(t) \Delta t}{\rho C_p h} \tag{E.15.5}$$

This is the first line of a tridiagonal matrix. Notice that the left-hand side has been evaluated as an average between time $j$ and $j+1$. In the same way equation (E.15.4) can be transformed in a difference equation

**Table E.15.1** Parameters appearing in the model by Baker and Roe and their standard deviation

| Parameter | Average value | Standard deviation |
|---|---|---|
| $h$ | 75 m | 25 m |
| $w$ | $-1.3\ 10^{-7}\text{ms}^{-1}$ | $0.5\ 10^{-7}\text{ms}^{-1}$ |
| $\chi$ | $1.5\ 10^{-4}\ \text{m}^2\ \text{s}^{-1}$ | $0.5\ 10^{-4}\ \text{m}^2\ \text{s}^{-1}$ |
| $f_a$ | 0.65 | 0.13 |

$$T_{j+1}^i - T_j^i = \frac{\alpha}{2}\left[\left(T_{j+1}^{i+1} - 2T_{j+1}^i + T_{j+1}^{i-1}\right) + \left(T_j^{i+1} - 2T_j^i + T_j^{i-1}\right)\right] - \frac{\beta}{4}\left[\left(T_{j+1}^{i+1} - T_{j+1}^{i-1}\right) + \left(T_j^{i+1} - T_j^{i-1}\right)\right]$$

where $\alpha = \chi\Delta t/\Delta z^2$ and $\beta = w\Delta t/\Delta z$. Then we have

$$-T_{j+1}^{i+1}\left(\frac{\alpha}{2} + \frac{\beta}{4}\right) + T_{j+1}^i(1-\alpha) - T_{j+1}^{i-1}\left(\frac{\alpha}{2} + \frac{\beta}{4}\right) = -T_j^{i+1}\left(\frac{\alpha}{2} - \frac{\beta}{4}\right) + T_j^i(1-\alpha) - T_j^{i-1}\left(\frac{\alpha}{2} - \frac{\beta}{4}\right) \tag{E.15.6}$$

We can see that the left-hand side of the matrix has the value at time $j+1$ while the right-hand side has values at time $j$. This is then a linear system which can be solved at each time step. The last equation for $i = N-1$, where $N$ is the number of layers, will be solved using the boundary condition either zero flux of null value for $T$.

The above scheme can be solved with a Matlab program but more interesting is the characterization of the uncertainty of the results. This depends on the fact that some of the parameters are know with some standard deviation. The Table E.15.1 summarizes the situation.

The problem now is to sample them in a normalized Gaussian way. A possible method is known as Box Muller approach. If $X$ and $Y$ are independent unit normal random variables, their polar coordinates $R = \sqrt{X^2 + Y^2}$ and $\Theta = \tan^{-1}(Y/X)$ are independent with $R^2$ being exponentially distributed with mean 2 and $\Theta$ being uniformly distributed on $(0, 2\pi)$. It can be shown that if $U_1$ and $U_2$ are random numbers then we can set

$$R = (-2\ln U_1)^{1/2}$$
$$\Theta = 2\pi U_2$$

from which it follow that

$$X = R\cos\Theta = (-2\ln U_1)^{1/2}\cos(2\pi U_2)$$
$$\Theta = R\sin\Theta = (-2\ln U_1)^{1/2}\sin(2\pi U_2) \tag{E.15.7}$$

are independent unit normals.

The following program gives the ways to construct such variables. The program generates initially pseudorandom variables that are transformed in gaussian with average 0 and variance 1 with Box Muller method.

```
%% Box-Muller
%-----------------------------------------------------------------
%Generator
%-----------------------------------------------------------------
clear all;
clc;
n = 4;%number of samples
num_iter=5e3; %number of iterations
num1 = zeros(n,num_iter);
num2 = zeros(n,num_iter);
sigma=1; %standard deviation
j = 1;
for i=1:num_iter
u1=rand(n,1); %~U[0,1]
u2=rand(n,1); %~U[0,1]
% Box-Mueller Transformation
w1=sqrt(-2*log(u1)).*cos(2*pi*u2); %~N(0,1)
w2=sqrt(-2*log(u1)).*sin(2*pi*u2); %~N(0,1)
num1(:,j) = w1;%matrix put in column w1
num2(:,j) = w2;
j = j+1;
end
pl = zeros(n,num_iter);
out = zeros(n,num_iter);
check = out;
%-----------------------------------------------------------------
%Plot test
%-----------------------------------------------------------------
%shows n normal distribution
%blue line = estimate
%red line = effective gaussian with the same mean and standard
 deviation
figure(1),clf
for j=1:n
p = num1(j,:);%write a vector for each n with all random w1
m = mean(p);%mean
st = std(p);% standard deviation
[f,x] = ksdensity(p,linspace(m-4*st,m+4*st,num_iter));%ksdensity
 calculates the distribution dist = 1./sqrt(2*pi*st^2)*
exp(-(x-m).^2./(2*st^2));
% dist _e is a gaussian input for the plot
pl(j,:) = f;
check(j,:) = dist;
limit = [m-4.1*st m+4.1*st 0 0.5];
subplot(4,2,j)
```

```
plot(x,pl(j,:),'linewidth',1.5) %plot function of f
 hold on
plot(x,check(j,:),'r') %variable plot as a function of
 distribution
axis(limit)
end
%-----------------------------------------------------------------
%Modifies mean and std in relation to the variables of the problem
%-----------------------------------------------------------------
figure(2),clf
m2 = [75,-1.3e-7,1.5e-4,0.65];
st2 = [25,0.5e-7,0.5e-4,0.13];
%Choice of the distribution
choose = [1,2,3,4];
k = 1;
for j=1:n
p = num1(j,:); %write a vector w1 for each n
m = mean(p);
st = std(p);
if any(choose == j) %if element of choose is j
m = m2(k);
st = st2(k);
p = p*st + m;
k = k +1;
end
[f,x] = ksdensity(p,linspace(m-4*st,m+4*st,num_iter));
dist = 1./sqrt(2*pi*st^2)*exp(-(x-m).^2./(2*st^2));
%output
out(j,:) = p;
%input for the plot
pl(j,:) = f;
check(j,:) = dist;
limit = [m-4.1*st m+4.1*st 0 max(f) + 0.1/st];
subplot(4,2,j)
plot(x,pl(j,:),'linewidth',1.5) %variable plot as a function of f
 hold on
plot(x,check(j,:),'r') % variable plot as a function of
 distribution
axis(limit)
xlabel('Variable','fontsize',14)
ylabel('PDF','fontsize',14)
%%%
% check on the normalization
if abs(sum(f)*(x(2)-x(1)) -1) > 0.01
disp('distribution:')
```

```
disp(j)
disp('non normalised')
end
end
trasp=out';
%Skewness e Kurtosis
s=skewness(trasp);
kr=kurtosis(trasp)-3;
```

An example of the results is shown in Fig. E.15.5 for two different value of the gain G = 2.5 and 10 and a fixed forcing of 4 w $m^{-2}$

## *E.15.5 Properties of Two-Dimensional Gaussian Distribution*

For two independent variables (*x, y*) the joint probability distribution *P*(*x, y*) it is just the product of the two distributions

$$P(x,y) = P(x)P(y) = \frac{1}{(2\pi)^{1/2}\sigma_x}\exp\left[-\frac{(x-\bar{x})^2}{2\sigma_x^2}\right]\frac{1}{(2\pi)^{1/2}\sigma_y}\exp\left[-\frac{(y-\bar{y})^2}{2\sigma_y^2}\right]$$

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{1}{2}\left[\frac{(x-\bar{x})^2}{\sigma_x^2}+\frac{(y-\bar{y})^2}{\sigma_y^2}\right]\right\} \tag{E.15.8}$$

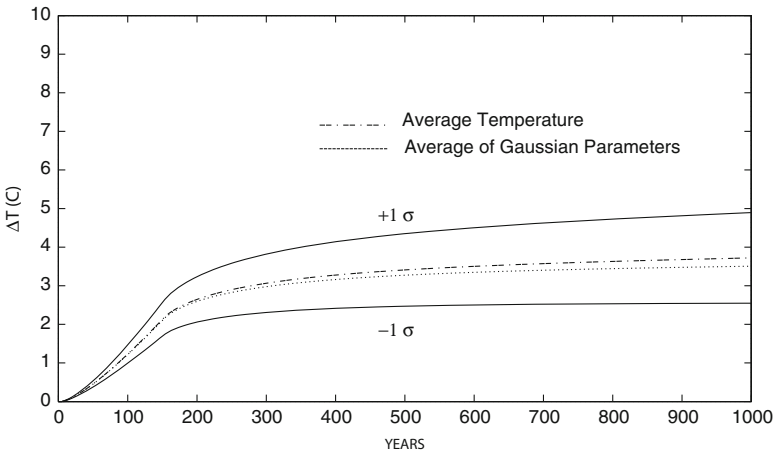



**Fig. E.15.5** The change in temperature for a fixed forcing switched at time 0 of 4 W $m^{-2}$

Now suppose we have two variables $X$ and $Y$ are expressed as rotation with respect the $x$ and $y$ axis

$$X = cx + sy \quad Y = -sx + cy$$

Then the correlation between these quantities

$$\langle XY \rangle = \left\langle scy^2 - scx^2 + \left(c^2 - s^2\right) xy \right\rangle = sc\left(\sigma_y^2 - \sigma_x^2\right).$$

In this case we have assumed averages to be zero and consequently $\sigma_x^2 = \left\langle x^2 \right\rangle$, $\sigma_y^2 = \left\langle y^2 \right\rangle$ and $\langle xy \rangle = 0$. Then we remember the relation between covariance and correlation coefficient $\rho$

$$\text{cov}\,(X, Y) = \rho_{XY}\sigma_X\sigma_Y$$

So we have

$$\rho_{XY}\sigma_X\sigma_Y = sc\left(\sigma_y^2 - \sigma_x^2\right) \tag{E.15.9}$$

We can find the standard deviations using

$$\begin{aligned}\sigma_X^2 = \left\langle X^2 \right\rangle &= \left\langle (cx + sy)^2 \right\rangle = c^2\sigma_x^2 + s^2\sigma_y^2 \\ \sigma_Y^2 = \left\langle Y^2 \right\rangle &= \left\langle (cy - sx)^2 \right\rangle = c^2\sigma_y^2 + s^2\sigma_x^2\end{aligned}$$

So that we have

$$\sigma_X^2\sigma_Y^2 = s^2c^2\left(\sigma_x^4 + \sigma_y^4\right) + \left(c^4 + s^4\right)\sigma_x^2\sigma_y^2 \tag{E.15.10}$$

While from (E.15.2), we have

$$\rho^2\sigma_X^2\sigma_Y^2 = s^2c^2\left(\sigma_y^4 + \sigma_x^4 - \sigma_x^2\sigma_y^2\right)$$

We readily obtain by subtracting

$$\left(1 - \rho^2\right)\sigma_X^2\sigma_Y^2 = \sigma_x^2\sigma_y^2\left(c^4 + 2s^2c^2 + s^4\right)$$

Remembering now the transformation c and s are just the cosine and sine of the rotational angle, and we get

$$\left(1 - \rho^2\right)\sigma_X^2\sigma_Y^2 = \sigma_x^2\sigma_y^2 \tag{E.15.11}$$

Now we rewrite (E.15.1) with zero averages and obtain

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{x^2}{2\sigma_x^2}-\frac{y^2}{2\sigma_y^2}\right\}$$

and substitute for the quantities

$$x = cX - sY; \quad y = sX + cY; \quad P(x,y)\,dxdy = P(X,Y)\,dXdY$$

we get

$$\begin{aligned} P(X,Y) &= \frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{(cX-sY)^2}{2\sigma_x^2}-\frac{(cY+sX)^2}{2\sigma_y^2}\right\} \\ &= \frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{X^2}{2}\left[\frac{c^2\sigma_y^2+s^2\sigma_x^2}{\sigma_x^2\sigma_y^2}\right]-\frac{Y^2}{2}\left[\frac{c^2\sigma_x^2+s^2\sigma_y^2}{\sigma_x^2\sigma_y^2}\right]\right. \\ &\quad \left.+\frac{2XY}{2}sc\left[\frac{\sigma_y^2-\sigma_x^2}{\sigma_x^2\sigma_y^2}\right]\right\} \end{aligned}$$

Substituting the identities

$$\langle X^2\rangle = \sigma_X^2 = c^2\sigma_x^2+s^2\sigma_y^2 \quad \langle Y^2\rangle = \sigma_Y^2 = c^2\sigma_y^2+s^2\sigma_x^2 \quad (1-\rho^2)\,\sigma_X^2\sigma_Y^2 = \sigma_x^2\sigma_y^2$$

We obtain

$$P(X,Y) = \frac{1}{2\pi\sqrt{(1-\rho^2)}\sigma_X\sigma_Y}\exp\left\{-\frac{1}{2}\frac{1}{1-\rho^2}\left[\frac{X^2}{\sigma_X^2}+\frac{Y^2}{\sigma_Y^2}+\frac{2\rho XY}{\sigma_X\sigma_Y}\right]\right\}$$

And for the correlated quantities

$$\begin{aligned} P(x,y) &= \frac{1}{2\pi\sqrt{(1-\rho^2)}\sigma_x\sigma_y} \\ &\quad\times\exp\left\{-\frac{1}{2}\frac{1}{1-\rho^2}\left[\frac{(x-\bar{x})^2}{\sigma_x^2}+\frac{(y-\bar{y})^2}{\sigma_y^2}-2\frac{\rho\,(x-\bar{x})\,(y-\bar{y})}{\sigma_x\sigma_y}\right]\right\} \end{aligned}$$

This joint probability can be put in a more general form defining the error matrix

$$\mathbf{M} = \begin{pmatrix} \langle x^2\rangle & \langle xy\rangle \\ \langle xy\rangle & \langle y^2\rangle \end{pmatrix} \quad \mathbf{M} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

and the discrepancy vector

$$\mathbf{x} = \begin{pmatrix} x - \bar{x} \\ y - \bar{y} \end{pmatrix}$$

and again

$$\mathbf{M}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x \sigma_y} \\ -\frac{\rho}{\sigma_x \sigma_y} & \frac{1}{\sigma_y^2} \end{pmatrix} \quad \det \mathbf{M} = \left(1-\rho^2\right) \sigma_x^2 \sigma_y^2$$

we can write

$$P(x, y) = \frac{1}{2\pi |\mathbf{M}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}^{\mathbf{T}} \mathbf{M}^{-1} \mathbf{x}\right) \tag{E.15.12}$$

## *E.15.6 A Simple Example (Leroy 1998)*

Assume that a model M predicts an ocean temperature to be $y_p \pm \sigma_p$ with the standard deviation obtained by varying the parameters of the model. The result that can be considered a prior can be written

$$P(y\,|M) = \frac{1}{\sigma_p \sqrt{2\pi}} \exp\left(-\frac{\left(y - y_p\right)^2}{2\sigma_p^2}\right) \tag{E.15.13a}$$

This means that the modeler believes the probability that an experiment will find a value between $y$ and $y + dy$ will be $P(y|M)dy$. On the other hand, the measurement will give a different result with a Gaussian statistics

$$P(y_m\,|M) = \frac{1}{\sigma_m \sqrt{2\pi}} \exp\left(-\frac{(y - y_m)^2}{2\sigma_m^2}\right) \tag{E.15.13b}$$

This can be considered the likelihood and the posterior distribution can be considered to be the product of the two

$$P(y\,|y_m, M) \propto \frac{1}{2\pi \sigma_m \sigma_p} \exp\left(-\frac{\left(y - y_p\right)^2}{2\sigma_p^2} - \frac{(y - y_m)^2}{2\sigma_m^2}\right) \tag{E.15.14}$$

The posterior distribution is normal in $y$, and the center of the distribution can be found by maximizing with respect to $y$

$$y_c = \left(\frac{y_p}{\sigma_p^2} + \frac{y_m}{\sigma_m^2}\right)\left(\frac{1}{\sigma_p^2} + \frac{1}{\sigma_m^2}\right)^{-1} \tag{E.15.15}$$

An the variance associated with it

$$\sigma_c = \left(\frac{1}{\sigma_p^2} + \frac{1}{\sigma_m^2}\right)^{-1/2} \tag{E.15.16}$$

The posterior probability distribution is then a temperature of the ocean $y_c \pm \sigma_c$. In Bayesian statistics, it can be determined the probability that this data is obtained given that the model is true. This is obtained by integrating (E.15.14) over $y$ with the result for the evidence function

$$P\left(y_m\,|M\right) = \left[2\pi\left(\sigma_m^2 + \sigma_p^2\right)\right]^{-1/2} \exp\left(-\frac{\mathcal{A}}{2}\right) \tag{E.15.17a}$$

In which the "accuracy" $\mathcal{A}$ is

$$\mathcal{A} = \frac{\left(y_p - y_m\right)^2}{\left(\sigma_p^2 + \sigma_m^2\right)} \tag{E.15.17b}$$

In orthodox statistics, the accuracy is used to make decision on the quality of the data. If it is significantly greater than unity means then either the model or the data are discarded. In Bayesian statistics, the multiplier of the exponential in (E.15.10a) must be considered. Assuming a flat prior probability for all the models $[P(M_i) = \text{constant}]$ and $\sigma_m \ll \sigma_p$, then the posterior evidence for the model $P(M|y_m)$ can be written as

$$-2\log\left(\sigma_{mP}\left(M|\,y_m\right)\right) \simeq \mathcal{A} + \log\left(\frac{\sigma_p}{\sigma_m}\right) \tag{E.15.17c}$$

The smaller this quantity is, the more probable the model is in light of the data. A better accuracy means a smaller $\mathcal{A}$. The second term on the right rewards smaller value of the prediction error $\sigma_p$ meaning a more precise prediction. In fact $\log(sp/sm)$ approximately counts the number of parameters needed to describe the data given the prediction. This number approximates the model of free parameters of model $M$, penalizing models with more free parameters. This is the equivalent of the Occam's razor.

# References[1]

## *Books*

MacKay DJ (2003) Information theory, inference, and learning algorithms. Cambridge University Press, Cambridge
Sivia DS (2006) Data analysis: a Bayesian tutorial. Oxford University Press, Oxford
Van Kampen NG (1992) Stochastic processes in physics and chemistry. North Holland, Amsterdam
Wax N (1954) Selected papers on noise and stochastic processes. Dover, New York

## *Articles*

Baker MB, Roe GH (2009) The shape of things to come: why is climate change so predictable? J Clim 11:640
Cess RD et al (1990) Intercomparison and interpretation of climate feedback processes in 19 general circulation models. J Geophys Res 95:16601
Glecker PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113:D06104
Held I (1982) Climate models and the astronomical theory of the ice ages. Icarus 50:449
Leroy SS (1998) Detecting climate signals: some Bayesian aspects. J Clim 11:640
Nicolis C (1982) Stochastic aspects of climatic transitions – response to a periodic forcings. Tellus 34:1
Paillard D (1998) The timing of Pleistocene glaciations from a simple multiple-state climate model. Nature 391:378
Roe GH (2009) Feedbacks, timescales and seeing red. Ann Rev Earth Planet Sci 37:93
Simmons AJ, Bengtsson L (1984) Atmospheric general circulation models: their design and use for climate studies. In: Houghton JT (ed) The global climate. Cambridge University Press, Cambridge
Sutera A (1981) On stochastic perturbation and long-term climate behaviour. Q J R Meteorol Soc 107:137
Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 106(D7):7183–7192
Washington WM, Parkinson CL (1986) An introduction to three dimensional climate modeling. University Science Books, Oxford University Press, Sausalito
Watts RG, Hayder E (1983) The origin of the 100-Kiloyear ice sheet cycle in the pleistocene. J Geophys Res 88:5163
Wigley TML (1989) Possible climate change due to $SO_2$ derived clouds condensation nuclei. Nature 339:365

[1]This chapter has been changed thoroughly by adding a consistent section on the evaluation of GCMs. These are not "simple" but their evaluation is a topic that gives the occasion to introduce new things like Bayesian statistics. The stochastic approach is treated mainly in the now precious Wax book that has contributions from such people as Chandrasekhar. A more recent treatment can be found in the book by Van Kampen and in the papers by Nicolis and Sutera. One of the best books on Bayesian statistics is Sivia together with the monumental MacKay treatise.

# Chapter 16
# Chemistry of the Troposphere

In the last few chapters, we have seen that an important factor for climatic change is the composition of the atmosphere. This is something we actually learned at the beginning of the first chapter, when it was evident that the differences in the planetary environment were determined both by the distance from the sun and the chemical composition of the atmosphere. It is then written in the stars that an atmospheric physicist must deal with atmospheric chemical processes. In the last 20 years or so, the problems related to the ozone hole and the pollution of our cities have made this topic extremely popular.

## 16.1 Introduction

We need to distinguish immediately what happens in troposphere, which contains most of the atmospheric mass, from what happens in the stratosphere. In the troposphere, chemical processes have a strong interaction with the soil and the oceans, and the main sources (either anthropic or natural) are located in the same region. On the other hand, in the stratosphere, some of these gases, which may be inert in the troposphere, can be activated by the ultraviolet radiation and become the main contributors to the chemistry of this region of the atmosphere.

Figure 16.1 shows the main processes that determine the composition of the troposphere. The gases may be emitted as a natural process from volcanoes (like carbon dioxide), as a result of human activities or the activity of the biosphere (like methane and nitrous oxide). Gases are then subject to quite complex processes in which sometimes the atmosphere has a secondary role. For example, carbon dioxide

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_16) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_16

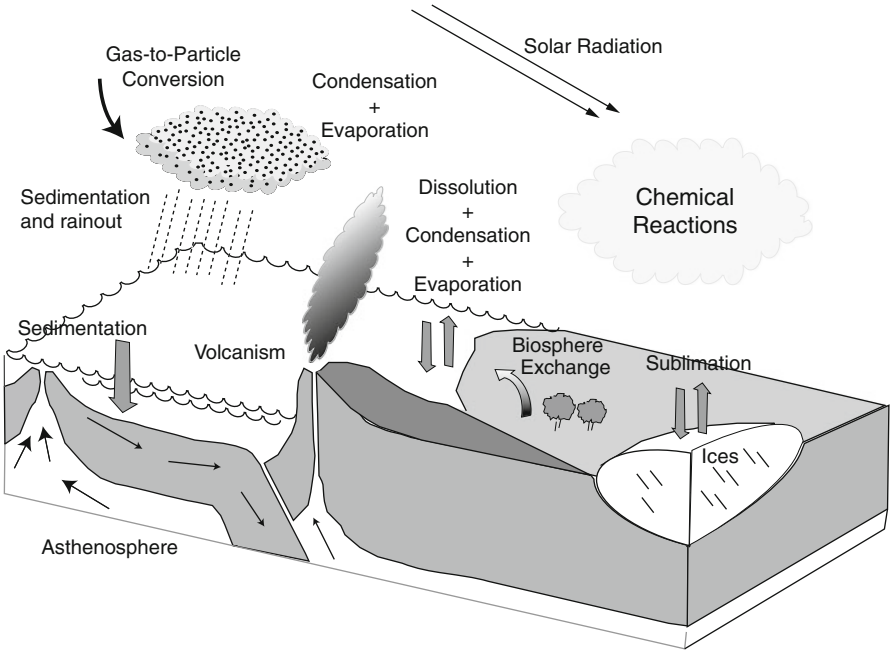



**Fig. 16.1** An illustration of the processes that influence the atmospheric composition. Besides chemical processes, also physical processes like condensation, evaporation, and sublimation must be considered

can be dissolved in the oceans or fixed by photosynthesis, a process that takes place both in the oceans and on land but is almost completely inert in the atmosphere. $CO_2$ in the ocean may be converted in carbonates that constitute the shells of a number of small animals that populate the surface waters. Once these organisms die, they sink to the bottom of the oceans where they accumulate as sediments that are recycled through the very slow processes of the Earth's crust. At the end of these processes, carbon dioxide may be reemitted in the atmosphere through volcanism.

The hydrological cycle is another important and complex process for the troposphere, and this involves almost all the components of the Earth system. To study the complex behavior of the chemical compounds requires a good knowledge of the general chemistry, and our experience is that some physics students do not like chemistry at all. However, we have introduced most of the relevant materials for chemistry in Chap. 5 so that we can move on rapidly.

## 16.2 The Minor Gas Inventory

As in the case of the major gases (oxygen and nitrogen), minor gases also have a roughly constant mixing ratio in the troposphere. Many of the species are destroyed through chemical processes both in the troposphere and in the stratosphere.

A typical example may be chlorofluorocarbons (CFCs), famous for being the main culprit in ozone destruction in the stratosphere. One of the most important is $CFCl_3$ which in the stratosphere is photodissociated at a rate $J_{CFC}$

$$CFCl_3 + hv = CFCl_2 + Cl \quad J_{CFC}$$

The rate of destruction of this CFC is then

$$\int_0^\infty J_{CFC}\,[CFCl_3]\,dz$$

We can then introduce a lifetime for this gas that is simply given by the ratio between the columnar content and the rate of destruction

$$\tau = \int_0^\infty [CFCl_3]\,dz \Big/ \int_0^\infty J_{CFC}\,[CFCl_3]\,dz$$

This is a very important parameter for a gas because it tells us if its influence is global or rather regional. The longer the lifetime, the larger the geographical region affected by the gas. Large-scale dynamics will be more relevant for those gases that have a long atmospheric lifetime.

Not always is the destruction process as simple as in the case of CFC, and in the following table, we have listed not only the lifetime values but also the main processes responsible for the production and the loss of the species. In Table 16.1, first of all we notice those gases whose concentration has been changed by human activity because the change can be found in a very few generations or from the preindustrial epoch.

Of some interest is the fact that all the sources of chlorine, except methyl chloride, have an industrial origin. In this table, the sources for $NO_x$ and the sulfur species are not listed and will be reported later.

The table lists the concentration before the industrial revolution, when this is known. The zero listed for the CFCs actually means that these gases were not present at that time, being of a totally anthropogenic origin. Also reported in the same table is the total burden of the gas expressed in teragram ($1Tg = 10^{12}$ g).

**Table 16.1** The present concentration of some minor gases in the atmosphere

| Gas | ppb (2014) | ppb (preindustrial) | Amount (Tg) | Lifetime (years) |
|---|---|---|---|---|
| $CO_2$ (carbon dioxide) | 395,400 | 280,000 | 3120.000 | See text |
| $CH_4$ (methane) | 1.800 | 720 | 4.850 | 11–17 |
| $N_2O$ (nitrous oxide) | 270 | 325 | 2.325 | 120 |
| CFC-11 | 0.236 | 0 | 6.2 | 50 |
| CFC-12 | 0.527 | 0 | 10.3 | 102 |
| CFC-113 | 0.074 | 0 | 2.6 | 85 |
| $CH_3CCl_3$ (m-chloroform) | 0.160 | 0.007 | 3.5 | 5 |
| $CH_3Cl$ (methyl chloride) | 0.60 | 0.60 | 5 | 1 |

**Table 16.2** Sources and sinks for methane

| Sources | Fluxes (Tg/year) |
|---|---|
| *Natural* | |
| Wetlands | 177–284 |
| Termites | 2–22 |
| Oceans | 5–20 |
| Hydrates | 2–9 |
| *Anthropogenic* | |
| Carbon and gas extraction | 85–105 |
| Rice production | 33–40 |
| Ruminants | 87–94 |
| Waste treatment and landfills | 67–90 |
| Biomass combustion | 32–39 |
| *Sinks* | |
| Reaction with OH | 470–700 |
| Soil removal | 9–47 |
| Atmospheric increase | 8–26 |

We will examine with some detail the different gases except for carbon dioxide. This gas has a number of interactions with the ocean and the biosphere and will be treated separately.

### *16.2.1 Methane*

Table 16.2 gives a rough idea of the sources and sinks for methane. These data are still very uncertain, and it does not make sense to know them with an approximation better than the first decimal digit.

Methane has anthropogenic sources that are roughly one and half times larger than the natural sources. The latter are due to fermentation processes and metabolism of some insects. When animals (that produce methane as a result of their metabolism) are raised in breeding facilities, this source becomes anthropogenic as in the case of cows. This is part of the food industry to which in a sense must be also attributed rice production.

Another important anthropogenic source is the treatment of waste. Dumps or similar facilities are a large producer of methane through fermentation processes.

The destruction processes are dominated by the reaction in the atmosphere with the hydroxyl radical OH

$$CH_4 + OH \rightarrow CH_3 + H_2O$$

This is the first step of the methane oxidation process that produces carbon dioxide, and we will talk about this mechanism in detail. We may conclude that the main methane sources are at the surface, while the main sinks are in the atmosphere.

### 16.2.2 Nitrous Oxide

For nitrous oxide, sources are not well known, while the lifetime has been measured with a good approximation, as is listed in Table 16.1 together with the trend in the atmospheric mixing ratio. It is possible then to determine with some precision the net source (production minus loss). The evaluation of the natural sources can be made on the basis of the change in mixing ratio between the preindustrial epoch and the present. This gives a source of 8.1–8.7 Tg (N) per year. This particular notation indicates that we are referring to the flux of nitrogen atoms. In order to obtain the flux of nitrous oxide, we need to multiply by the stoichiometric ratio (Tg ($N_2O$) = Tg (N)(44/28)). Destruction of $N_2O$ is mainly due to photodissociation in the stratosphere and to the reaction with metastable oxygen $O(^1D)$. This species originates from the ozone photodissociation for wavelengths less than 300 nm

$$O_3 + hv\ (\lambda < 300nm) \rightarrow O_2 + O\left(^1D\right)$$

This form of excited oxygen is extremely reactive and destroys nitrous oxide mainly in the stratosphere where it is more abundant

$$N_2O + O\left(^1D\right) \rightarrow NO + NO$$

In this way, destruction of $N_2O$ becomes a net source for nitrogen oxides in the atmosphere. This loss process can be evaluated with great precision because it depends only on the nitrous oxide mixing ratio at the ground and the ozone concentration. The value obtained compares well with the natural sources.

The principal natural mechanism for the production of $N_2O$ is denitrification, through which the nitrates present in the soil and the oceans are transformed into nitrogen or nitrogen oxides. The results of recent measurements made in the ocean and on land have given the value shown in Table 16.3.

A relatively good estimate of the stratospheric loss process and the present increase of the $N_2O$ concentration in the atmosphere give in turn an estimate of the total sources. The increase may be due to a net source of the order of 3–4.5 Tg/year, while the stratospheric destruction can be estimated at 14.3 Tg/year. The sources to balance these two processes must be in the range 17.3–19.8 Tg/year.

**Table 16.3** Sources and sinks for nitrous oxide

| Process | Fluxes (Tg N/year) |
|---|---|
| Natural sources | 1.2 (0.4–3.1) |
| Oceans | 4.0 (1.9–9.8) |
| Land | 6.6 (3.3–9) |
| Present sources | 5.6–21.9 |
| Stratospheric loss | 14.3 (4.3–27.2) |
| Yearly increase | 3.6 |
| Anthropogenic sources | 5.70 (2.2–7.90) |

The anthropogenic contribution to nitrous oxide is both the biomass combustion and the soil fertilization for agricultural purpose. The latter is important because it accelerates the denitrification process. The final numbers are summarized in Table 16.3. Also for $N_2O$, the sources are essentially at the ground, and the gas is destroyed through atmospheric chemical processes. The uncertainties in its budget make it difficult to indicate measures to control $N_2O$ production.

### *16.2.3 Atmospheric Chlorine*

In Table 16.1, we have listed as the only sources of atmospheric chlorine the CFCs and another natural compound. The sources actually are much larger in number and go from volcanoes to industrial products to the sea aerosols produced by the breaking of waves. The contribution of these sources is even more important, but chlorine produced from them remains in the troposphere and consequently is very rapidly removed. Our interest in the chlorofluorocarbons is that they are able to put chlorine directly in the stratosphere where they can destroy ozone, as we will see in a while. In any case, the study of CFC in the troposphere may be important because their main sink is in the stratosphere, so that just watching how their concentration evolves with time may give indications on the strength of this sink.

This approach can be readily understood with an example based on a box model of the atmosphere that reproduces the one used by D. Cunnold et al. several years ago and which is sketched in Fig. 16.2. In this model, the atmosphere is divided into nine boxes, eight for the troposphere and one for the stratosphere. The tropospheric boxes extend symmetrically from the equator to 30° latitude and from 30 to 90. In altitude, they range from 1000 to 500 hPa and from 500 to 200 hPa. In the stratosphere, there is only one box, between 200 and 0 hPa.

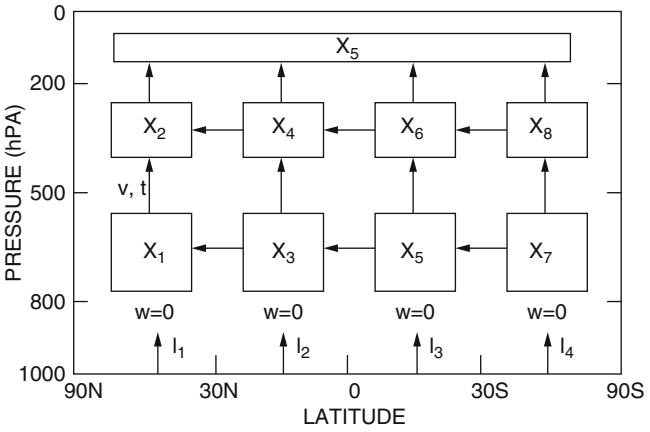



**Fig. 16.2** The box model used for the calculation of the lifetime for chlorofluorocarbons (CFC). *Arrows* indicate the connection between boxes but not necessarily the direction. Emission of CFCs is in boxes 1, 3, 5, 7

The air mass contained in the boxes of the lower troposphere is then 5/3 of the mass in the upper troposphere. It is assumed that two processes regulate the transfer between boxes: diffusion and meridional advection. The diffusive flux is proportional to the difference in concentration between two adjacent boxes, while the advection is proportional to the mean meridional wind. If we indicate with $\chi_1$ the mixing ratio in the box number 1, we can write down a continuity equation

$$\frac{d\chi_1}{dt} = V_{13}\chi_{13} - V_{21}\chi_{21} - \frac{\Delta\chi_{13}}{t_{13}} + \frac{\Delta\chi_{21}}{t_{21}} + \frac{I_1}{M_1}$$
$$\chi_{ij} = 0.5\left(\chi_i + \chi_j\right) \qquad \Delta\chi_{ij} = \left(\chi_i - \chi_j\right) \tag{16.1}$$

where $t_{ij}$ and $V_{ij}$ are the characteristic times for diffusion and the time constant for the meridional transport. Finally, $I_1$ is the rate (in kg year$^{-1}$) at which the component is added to the box $M_1$. A typical equation for the box in the upper troposphere is a little different because it must take into account the difference in mass

$$\frac{d\chi_2}{dt} = \frac{5}{3}V_{24}\chi_{24} + \frac{5}{3}V_{21}\chi_{21} - \frac{\Delta\chi_{24}}{t_{24}} - \frac{5}{3}\frac{\Delta\chi_{21}}{t_{21}} - \frac{2}{3}\frac{\chi_u - \chi_s}{\tau_s} - \frac{\chi_2}{\tau_t} \tag{16.2}$$

where in this case

$$\chi_u = 0.25\left(\chi_2 + \chi_4 + \chi_6 + \chi_8\right)$$

and $\chi_s$ is the average stratospheric mixing ratio. The factor that takes into account the volume is understood because the same flux produces a larger change in mixing ratio in a smaller mass. Finally, for the stratospheric box we have

$$\frac{d\chi_s}{dt} = \frac{\chi_u - \chi_s}{t_s} - \frac{\chi_s}{\tau_s} \tag{16.3}$$

In Eqs. (16.2) and (16.3), $t_s$ and $\tau_s$ are, respectively, the exchange time between troposphere and stratosphere and the residence time due to the photodissociation; $t_t$ is the residence time in the troposphere.

Equations similar to Eqs. (16.1) and (16.2) can be written for all boxes so that the mixing ratio in each box can be obtained by integrating the differential equations. The details of this calculation may be found at https://code.google.com/p/agage-box-model, and here we will give simply some of the results. The objective was to measure the lifetime of chlorofluorocarbons and nitrous oxide. We can find the average mixing ratio in the troposphere using the relation

$$\overline{\chi}_t = \frac{5/3\left(\chi_1 + \chi_3 + \chi_5 + \chi_7\right) + \left(\chi_2 + \chi_4 + \chi_6 + \chi_8\right)}{20/3 + 4}$$
$$= \frac{5}{32}\left(\chi_1 + \chi_3 + \chi_5 + \chi_7\right) + \frac{3}{32}\left(\chi_2 + \chi_4 + \chi_6 + \chi_8\right)$$

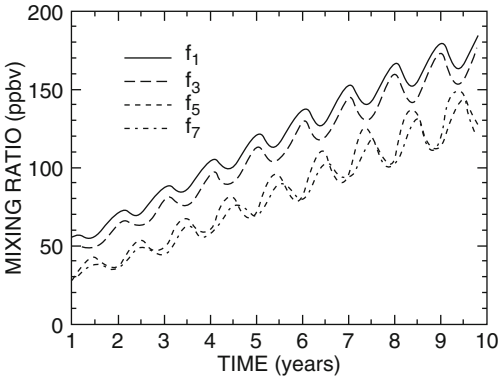



**Fig. 16.3** The increase of the mixing ratio of CFC with time in the four boxes of the lower troposphere. The oscillations are due to the seasonal transport variations

And then the lifetime is given by

$$\tau = \tau_s \left(1 + \chi_s / 4\overline{\chi}_t\right) / \left[(\tau_s/\tau_t) + \chi_s / 4\overline{\chi}_t\right] \tag{16.4}$$

Figure 16.3 shows the behavior of one of the most important CFCs (trichlorofluoromethane $CFCl_3$) starting from 1971 and obtained from the model cited above.

The release rates have varied between 200 and 300 million kg per year with a maximum around 1978.

Notice that the concentration is not the same at all latitudes because most of the compounds are released in the first box of the northern hemisphere (about 78 %). Notice also that the release rate is not the same as the annual production due to the different uses to which CFCs are subjected. When they are used as propellant for spray cans, they are released instantaneously. On the other hand, if they are used as a fluid in a refrigerator, they may be released only when the refrigerator is destroyed, which means a delay of the order of 10 years.

The knowledge of the release rate is quite uncertain, and to this data are related all the errors in the determination of the lifetime for CF-11 which is around 50 years.

## 16.3 The Biogeochemical Cycle for Carbon

On one side, the simplicity in the treatment of the carbon dioxide cycle derives from the fact that it is almost an inert gas in the atmosphere even if it is produced in small quantities by chemical reactions. On the other hand, the same gas has quite complex chemical interactions with both the biosphere and the ocean.

A rather simplified scheme of the $CO_2$ cycle is given in Fig. 16.3. The cycle is made up of a number of reservoirs that exchange carbon through a number of processes. Again, the units are based on the amount of carbon rather than on the amount of carbon dioxide. Knowing the fluxes, it is possible to find the residence time of a particular reservoir which is simply the content divided the flux. Starting

from the atmosphere, carbon dioxide is fixed by photosynthetic processes and released with the opposite processes of respiration and decay. A small part of the carbon fixed by photosynthesis however is subtracted from the respiration process because it may be buried (e.g., forests that are flooded or covered during landslides) so that the fluxes of respiration and photosynthesis are not exactly equal. In this way, a little oxygen is gained, and in this imbalance atmospheric oxygen originates. Things are more complex than this because the oxygen cycle involves other elements like nitrogen, phosphorous, and sulfur.

The buried carbon is released in the atmosphere when fossil fuels are burned or through weathering processes.

The role of the ocean in the carbon cycle is important because $CO_2$ may be dissolved in carbonic acid and bicarbonate ions. The dissolved carbon will form carbonates that precipitate to the bottom of the ocean and contribute to the sediments. The dead organic matter will contribute with a similar process to organic sediments.

A cycle like that shown in Fig. 16.4 can be solved with appropriate software. A typical result is shown in Fig. 16.5 where the model has been run for about 900 years to reach the steady state when a source is activated that amounts to 20 gT year$^{-1}$ of carbon dioxide, equivalent to the burning of the fossil fuels. The sudden increase is unrealistic because actually the source has been introduced gradually. However, the most interesting thing about this result is that a source 1000 times smaller than the smallest natural flux is able to perturb considerably the content of the atmospheric gases, thus affecting climate on a global scale.

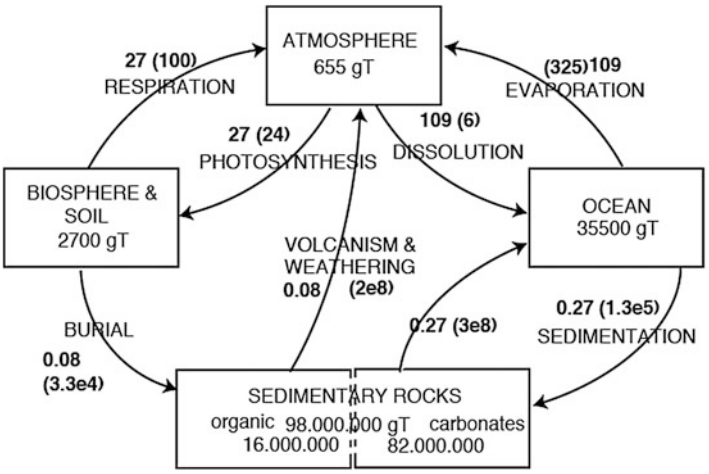



**Fig. 16.4** A simple illustration of the carbon biogeochemical cycle. Reservoirs are represented with *boxes* with their content expressed in $10^9$ tons (gTn) of carbon. The fluxes are indicated with the *arrows*, and the units are gTn year$^{-1}$. Residence times in years are in *parentheses* (from Goody and Walker 1972)

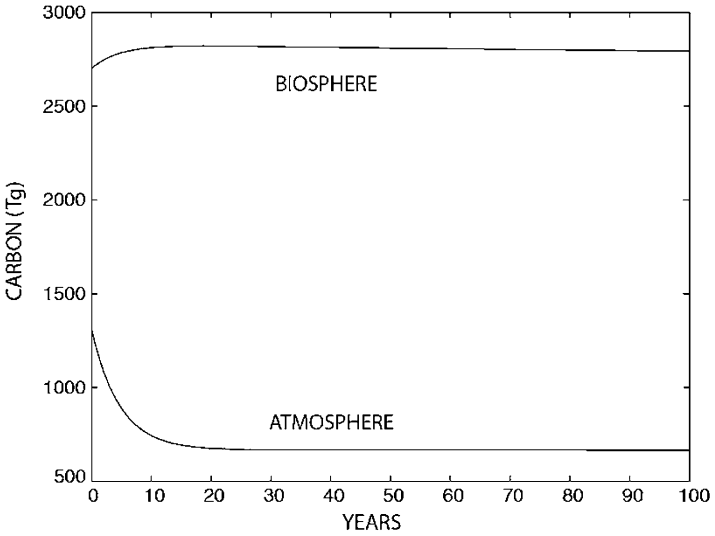



**Fig. 16.5** Carbon dioxide before in the atmosphere (unit of Tg of C) after a source is turned on equivalent to the initial content of the atmosphere (655 Tg). This content is equivalent to about 300 ppm of $CO_2$

The biosphere and the ocean are the components of the climate system that may absorb in the short term most of the carbon dioxide. The ocean in particular has the shortest response time, and it is worth to examine with some detail the interactions between carbon dioxide and the ocean.

### *16.3.1 Carbonate/$CO_2$ System: A Bit of Marine Chemistry*

As rookie scientists of Earth system science, we are not afraid to cope with the chemistry of the ocean also because for the moment it is still quite simple. If you want to learn more, use the Broecker and Peng bible.

We start by looking at Fig. 16.6, where the ocean is simply reduced to the surface water and the deep ocean. In the upper part, carbon dioxide is dissolved. The weak acids that are formed make carbon available for both photosynthesis, responsible for the formation of organic matter, and the carbonates that form the shells of organisms that populate the surface waters.

The reaction scheme starts with the dissolution of carbon dioxide

$$CO_2\,(\text{gas}) + H_2O\,(\text{acq}) \iff H_2CO_3\,(\text{acq}) \tag{16.5}$$

so that at the equilibrium, carbonic acid is formed that in turn is in equilibrium with bicarbonate ion $HCO_3^-$ and hydrogen ion

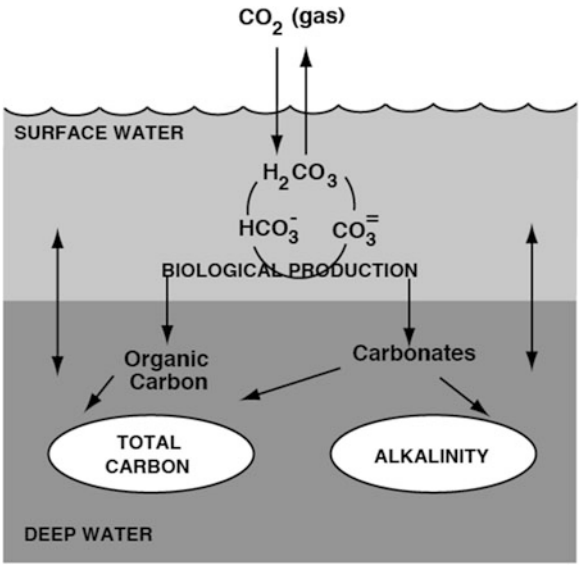



**Fig. 16.6** A rough scheme of a two-reservoir ocean (surface and deep water). The organic particles of biological origin sink to the bottom of the ocean. Carbonates are in equilibrium in the surface water, and they determine the atmospheric pressure of $CO_2$ (Walker 1991)

$$H_2CO_3 \iff HCO_3^- + H^+$$

The main source of bicarbonate ions in the sea is the rocks that, although they have a very limited solubility, produce a certain amount of carbonate ions $CO_3^{2-}$ that are in equilibrium with the bicarbonate ions. If we denote with $k_a$ the equilibrium constant (or more precisely the dissociation constant) for carbonic acid, we have

$$k_a = \left[H^+\right]\left[HCO_3^-\right] / \left[H_2CO_3\right] \tag{16.6}$$

On the other hand, from the reaction

$$HCO_3^- \iff CO_3^{2-} + H^+ \tag{16.7}$$

with equilibrium constant $k_b$, we have

$$k_b = \left[H^+\right]\left[CO_3^{2-}\right] / \left[HCO_3^-\right] \tag{16.8}$$

And eliminating the concentration of the hydrogen ions, we get

$$\left[H_2CO_3\right] = k_b\left[HCO_3^-\right]^2 / k_a\left[CO_3^{2-}\right] \tag{16.9}$$

For dissolved gases, the concentration is given by

$$pCO_2 = \left[CO_2\right] / \alpha \tag{16.10}$$

where $\alpha$ is the solubility constant. Because the carbon dioxide concentration is proportional to that of carbonic acid, we have

$$pCO_2 = k_b\left[HCO_3^-\right]^2/\alpha k_a\left[CO_3^{2-}\right] = k_{CO_2}\left[HCO_3^-\right]^2/\left[CO_3^{2-}\right] \tag{16.11}$$

where

$$k_{CO_2} = k_b/\alpha k_a$$

The determination of the different forms of carbons depends on the total dissolved carbon, indicated by dissolved inorganic carbon (DIC), and this can be approximated as

$$DIC = [H_2CO_3] + \left[HCO_3^-\right] + \left[CO_3^{2-}\right] \tag{16.12}$$

The concentration of the different forms of carbon can be obtained from the condition that the sea is electrically neutral. Considering the variety of the ions present in the seawater, this is not a simple condition. In particular, ions from strong acids and bases give an excess positive charge that is called alkalinity and is indicated with Alk. Actually, it corresponds to the total charge carried by the carbon compounds

$$Alk = \left[HCO_3^-\right] + 2\left[CO_3^{2-}\right] \tag{16.13}$$

Comparing the last two equations and keeping in mind that the contribution of the dissolved carbon dioxide is negligible, we obtain

$$Alk - DIC = \left[CO_3^{2-}\right] \tag{16.14}$$

At this point, it is possible to determine the concentration of the bicarbonate ion as a function of the alkalinity and DIC. From Eqs. (16.13) and (16.14), we have

$$\left[HCO_3^-\right] = 2DIC - Alk$$

Eliminating in Eq. (16.12) the carbonic acid (the carbon dioxide concentration) using Eqs. (16.9) and (16.14), we get

$$\left[HCO_3^-\right] = \left\{DIC - \left[DIC^2 - Alk\,(2DIC - Alk)\,(1 - 4k_{CO_2})\right]^{1/2}\right\} / (1 - 4k_{CO_2}) \tag{16.15}$$

At this point, we should be able to calculate the partial pressure of carbon dioxide in the atmosphere. We assume that the equilibrium between the dissolved carbon dioxide and the atmosphere is established with a characteristic time given by $\tau_{dis}$ so that

$$d\left(p\mathrm{CO}_2\right)/dt = \left[p\mathrm{CO}_2 - \left(p\mathrm{CO}_2\right)_s\right]/\tau_{\mathrm{dis}} \tag{16.16}$$

where $p\mathrm{CO}_2$ and $(p\mathrm{CO}_2)_s$ indicate the atmospheric and surface water partial pressures, respectively. In the surface water, the photosynthesis produces organic matter through the reaction

$$\begin{aligned} &106\mathrm{CO}_2 + 16\mathrm{NO}_3^- + \mathrm{H}_2\mathrm{PO}_4^{2-} + 122\mathrm{H}_2\mathrm{O} \\ &\quad \rightarrow (\mathrm{CH}_2\mathrm{O})_{106}(\mathrm{NH}_3)_{16}\,(\mathrm{H}_3\mathrm{PO}_4) + 138\mathrm{O}_2 \end{aligned} \tag{16.17}$$

In the organic matter, carbon, nitrogen, and phosphorus are in a fixed ratio called the *Redfield ratio* (so that we talk about the Redfield ocean).

Respiration is the opposite of Eq. (16.17). It is quite obvious that in the surface waters, there is a photosynthetic surplus with respect to respiration because a fraction of the material produced sinks to deep waters so that a flux of carbon dioxide is established between the surface and deep ocean. An opposite flux of nutrients is established between the bottom of the ocean and the surface. Surface waters will be impoverished in DIC and enriched in Alk with the opposite happening for deep waters. Nitrates are fixed during photosynthesis and released during respiration.

The result of this process (also denoted *biological pump*) is a depletion of the surface water in carbon dioxide and so of the atmosphere: the higher the biological activity in the ocean, the more severe is the depletion of the surface water. Based on Fig. 16.7, we can establish that the change in surface DIC (that we will call $\mathrm{DIC}_s$) is determined by the sedimentation of organic and inorganic carbon and by the exchange between the surface and deep waters due to the circulation. We have then

$$V_s \frac{d\mathrm{DIC}_s}{dt} = \frac{\left[\left(p\mathrm{CO}_2\right) - \left(p\mathrm{CO}_2\right)_s\right]}{\tau_s} C_m - \left(1 + C_{\mathrm{carb}}\right) P + \left(\mathrm{DIC}_d - \mathrm{DIC}_s\right) F_w \tag{16.18}$$

where with $V_s$ we have indicated the volume of the upper ocean and with $C_m$ a conversion factor between the units of pressure and mass. $C_{\mathrm{carb}}$ is the ratio between organic and inorganic carbon, and $P$ is the sedimentation rate of the inorganic carbon. The units for DIC used here are mole $\mathrm{m}^{-3}$ so that the units for $P$ are mole $\mathrm{year}^{-1}$. In analogy, $\mathrm{DIC}_d$ is the carbon content in deep waters, and $F_w$ is the water flux for the exchange between the surface and deep waters. It is quite easy then to write an equation similar to Eq. (16.18) for deep waters

$$V_d \frac{d\mathrm{DIC}_d}{dt} = \left(1 + C_{\mathrm{carb}}\right) P - \left(\mathrm{DIC}_d - \mathrm{DIC}_s\right) F_w \tag{16.19}$$

Exactly in the same manner, we can write two equations for $\mathrm{Alk}_s$ and $\mathrm{Alk}_d$ considering that, beside the exchange between surface and deep waters, for each mole of carbonate that may end as sediment, there is a change of two equivalent moles of alkalinity due to the term $\mathrm{C_a}^{2+}$ which has two positive charges. Each

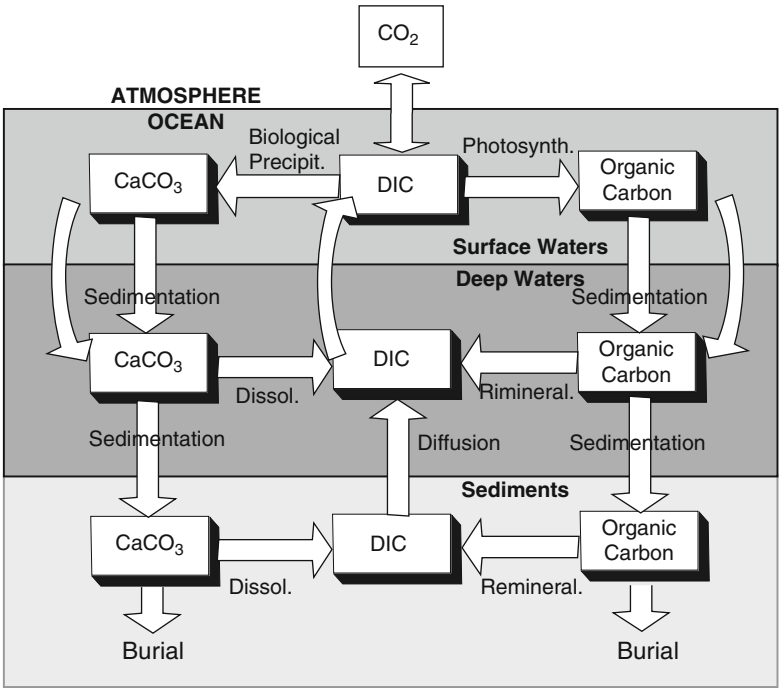



**Fig. 16.7** A schematic diagram of the carbon cycle in the ocean. *Curved arrows* indicate the circulation (From Najjar 1992)

carbon mole is associated, according to Eq. (16.17), with $16/106 \approx 0.15$ mol of nitrogen so that we have

$$V_s \frac{d\text{Alk}_s}{dt} = (\text{Alk}_d - \text{Alk}_s) F_w - (2C_{\text{carb}} - 0.15) P$$
$$V_d \frac{d\text{Alk}_d}{dt} = -(\text{Alk}_d - \text{Alk}_s) F_w + (2C_{\text{carb}} - 0.15) P \qquad (16.20)$$

Equations (16.19, 16.20, and 16.21) are solved in the Examples so that once the alkalinity and the dissolved carbon are calculated, it is possible using Eq. (16.15) to calculate the bicarbonate ion and then the carbon dioxide pressure using Eq. (16.11). The pressure is referred to the present value and is then calculated as preexisting atmospheric level (PAL), while the other data are as follows: $\text{DIC}_s = 2.01$ mol m$^{-3}$, $\text{DIC}_d = 2.23$ mol m$^{-3}$, $\text{Alk}_s = 2.2$ mol m$^{-3}$, $\text{Alk}_d = 2.26$ mol m$^{-3}$ $\tau_s = 8.64$ years, $C_m = 4.95$ 1016 mol,$V_s = 1.2$ $10^{17}$ m$^3$, $V_d = 1.23$ $10^{18}$ m$^3$, $P = 1.75$ $10^{14}$ mol year$^{-1}$, $C_{\text{carb}} = 0.25$, and $F_w = 10^{15}$ m$^3$ years$^{-1}$ = 31.64 Sverdrup. The Sverdrup is a unit used by oceanographers and corresponds to a flux of $10^6$ m$^3$ s$^{-1}$.

An early experiment assumed that initially the mixing ratio of carbon dioxide is about $5 \times 280$ ppm. The results is show that in the first thousand years or so, the concentration of carbon dioxide decreases quite rapidly but does not return

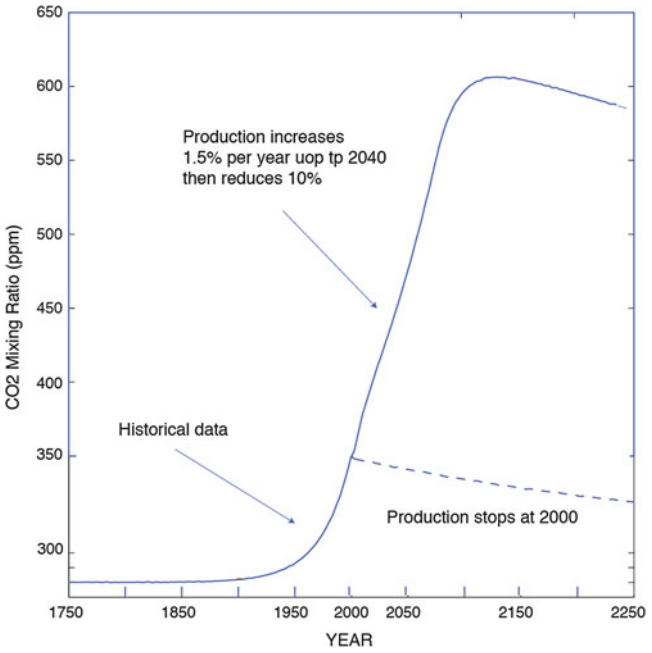



**Fig. 16.8** The change in the carbon dioxide mixing ratio with respect to the preindustrial value due to the burning of fossil fuels according to two scenarios. The preindustrial value is taken as 280 ppm (Broecker and Peng 1984)

to the preindustrial mixing ratio. The important fact is that after a few thousand years (the characteristic time for transport between deep and surface waters), a new equilibrium is reached. This can also be noticed from the fact that the carbon increases in the deep water and decreases in the surface waters, although the changes are not dramatic. In this case, we have considered only the carbon dioxide changes due to the interactions with the ocean. The biosphere should have a very important role, as has been shown previously (see Fig. 16.5).

This is much more evident if we let our model make an incursion into the political world. In Fig. 16.8, the changes in the carbon dioxide mixing ratio due to the burning of fossil fuels are shown. The figure is drawn based on two different scenarios, the first refers to the historical production up to the year 2000 and then by a sudden stops in the production. This goes to show that it will take centuries before the mixing ratio goes back to preindustrial values.

The second hypothesis is based on a constant increase in production by 1.5 % per year up to 2040. After that, there is global and gradual reduction by 10 % per year. No matter how political leaders discuss the moral of the story is the sooner the better. Lifetime of carbon dioxide is so long that needs very urgent actions. Except for these amenities, the important thing is that now the reader can build his own scenario for the buildup of carbon dioxide, and this can be done with a tablet (or laptop).

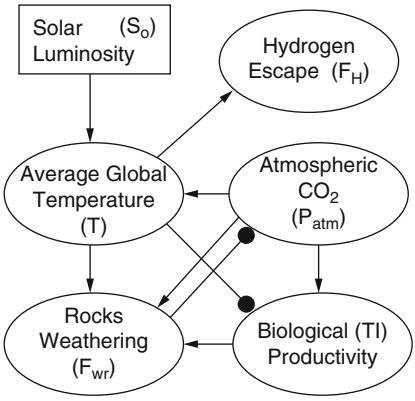



**Fig. 16.9** The links between the variables that control the environment for the biosphere. *Arrows* indicate positive influence and *circles* negative (from Kasting)

What we have seen until now is actually our immediate future or the recent past, but it is interesting to ask what will happen in a more distant time horizon. We are talking here about hundreds of million years, and the role of carbon dioxide will still be fundamental. In the process of answering this question, we can appreciate even more the role of the biosphere.

### *16.3.2 How Long Will the Biosphere Survive?*

James Lovelock (the inventor of Gaia) asked this question for the first time, several years ago, in the following terms. If the sun will evolve, as does any star of its class, in the future, it will emit more and more energy that will heat up the Earth to the point that any kind of life form will be sterilized. Recently, James Kasting has revisited such a question arriving at more optimistic conclusions.

With the help of Fig. 16.9, we can try to set up the terms on the problem. In the figure, arrows indicate a positive influence among variables, while circles indicate a negative influence. For example, if the solar luminosity ($S$) increases, the average global temperature $T$ also increases. To a warmer planet corresponds a larger amount of water vapor in the atmosphere so that precipitation will be more abundant. This may enhance the erosion of silicate rocks, and consequently, there will be a larger consumption of atmospheric carbon dioxide than is necessary to form carbonates. As a matter of fact, this mechanism has been invoked as a way to stabilize the carbon dioxide amount in the atmosphere. A higher temperature will also increase the loss of hydrogen gas from the upper atmosphere.

The decrease in carbon dioxide will affect negatively the biological productivity because the efficiency of the photosynthetic processes will be less. The temperature could become so high that $CO_2$ could be depleted to the point as to inhibit photosynthesis completely. The result would be the complete cessation of biological activity that implies turning off the mechanism that regulates the temperature through erosion processes. Then we may ask based on such premises how long life on Earth could endure.

The block diagram in Fig. 16.9 is used to solve everything through a simple computer program. The first step however is to parameterize some of the processes in very simple terms. We can start from the solar radiation, assuming it increases with time according to the data we have on stellar evolution

$$S(t) = (1 - 0.38t/\tau_0)^{-1} S_0 \tag{16.21}$$

In this relation, we assume that $S_0 = 1368\ \mathrm{Wm}^{-2}$ and the constant is $\tau_0 = 4.55$ Gyr.

The average global temperature is obtained by calculating first the effective temperature (see Chap. 3) and then the increase at the surface due to the greenhouse effect.

$$(1 - \alpha)\, S/4 = \sigma T_e^4$$

where $\alpha$ denotes the albedo. The global average temperature is given by the effective temperature ($T_e$) increase by the amount $\Delta T$ due to the greenhouse effect

$$T = T_e + \Delta T \tag{16.22}$$

The problem now is to establish how the greenhouse warming depends on the carbon dioxide content. Kasting has used a radiative convective model to parameterize the temperature increase $\Delta T$ in terms of the temperature and the carbon dioxide partial pressure, $\psi = \log P_{\mathrm{atm}}$

$$\begin{aligned}\Delta T = {} & 815.17 + \left(4.895 \cdot 10^{-7}\right) T^{-2} - \left(3.9787 \cdot 10^{5}\right) T^{-1} \\ & - 6.7084\psi^{-2} + 73.221\psi^{-1} - 30.882 T^{-1}\psi^{-1}\end{aligned} \tag{16.23}$$

$\Delta T$ also depends on the temperature through the water vapor feedback, and the albedo depends on the temperature through the ice–albedo feedback

$$\alpha = 1.4891 - 6.5979 \cdot 10^{-3} T + \left(8.567 \cdot 10^{-6}\right) T^2 \tag{16.24}$$

The erosion of siliceous rocks is a function of the ground water pH. A typical erosion reaction will include hydrolysis and the combination with carbon dioxide

$$\mathrm{Mg_2SiO_4} + 4\mathrm{CO_2} + 4\mathrm{H_2O} \rightarrow 2\mathrm{Mg^{2+}} + 4\mathrm{HCO_3^-} + \mathrm{H_4SiO_4}$$

that is proportional to the activity of the hydrogen ions ($a_{\mathrm{H^+}}$) and so the carbon dioxide pressure. Experimental data suggest a dependence like

$$\frac{F_{wr}}{F_{wr,0}} = \left(\frac{P_{\mathrm{atm}}}{P_{\mathrm{atm},0}}\right)^{0.3} \exp\left(\frac{T - T_0}{13.7}\right) \tag{16.25}$$

where $T_0 = 15\ ^\circ\mathrm{C}$ is a reference temperature.

The biological activity in the soil has then a very important function in the weathering processes because it maintains in the soil a pressure that can be 30–40 times higher than the atmospheric pressure. For our model, it is very important to establish the dependence of the $CO_2$ pressure in the soil ($P_{soil}$) on the biological activity. The relationship we will use represents a very rough global average and is based on very sparse data. It is then necessary to use a good deal of imagination when using these data. Following Kasting, we can write

$$\frac{P_{soil}}{P_{soil,0}} = \frac{\Pi}{\Pi_0}\left(1 - \frac{P_{atm,0}}{P_{soil,0}}\right) + \frac{P_{atm}}{P_{soil,0}} \tag{16.26}$$

where again with $\Pi$ we have indicated the biological productivity. The final link in the model must be between the productivity and the temperature because we see from Eq. (16.26) that if productivity falls below some threshold than the $CO_2$, pressure in the soil will become equal to the atmospheric pressure, and from that point, the photosynthetic activity will be reduced drastically.

The link between productivity and temperature is given by

$$\frac{\Pi}{\Pi_{max}} = \left[1 - \left(\frac{T - 25}{25}\right)^2\right]\left[\frac{P_{atm} - P_{min}}{P_{1/2} + (P_{atm} - P_{min})}\right] \tag{16.27}$$

This relation points out that the productivity is controlled by pressure and temperature. If the temperature is higher than 50 °C, then the productivity goes to zero. The same thing happens if the atmospheric pressure falls below a threshold ($P_{min}$) that in our case, we have assumed as 10 ppm. Actually, this limit is reasonable for a particular kind of plants, called C4 (e.g., sugar cane, corn), that utilize a photosynthetic process quite different from that of other plants (C3). A few C4 plants may survive even at lower pressure. In Eq. (16.28), $P_{1/2}$ is the pressure at which $\Pi_0 = \Pi$ when $T = T_0$ and $P_{atm} = P_{atm,0}$.

It is a little surprising that the ocean apparently is not present in this model. Actually, things are a little different. Most of the photosynthetic activity takes place in the ocean surface waters. Over the long term, the carbon dioxide added to the atmosphere by the volcanic activity must be balanced by the deposition of carbonates in the sediments. The weathering processes transport in the ocean calcium and magnesium ions that contribute to the alkalinity of the surface waters and consequently to the dissolution of carbon dioxide from the atmosphere.

The model is usually solved for steady-state conditions using an iteration method. If we want to solve it as a function of time, we should take into account that the atmospheric $CO_2$ pressure is regulated on a global scale by volcanic emissions that are long-term source and destroyed by weathering processes. We have

$$\frac{dM_{CO_2}}{dt} = V - F_{wr} \tag{16.28}$$

where $M_{CO_2}$ is the carbon dioxide total mass per unit surface, $V$ is the emission rate by the volcanic activity, and $F_{wr}$ is the weathering rate. In order to use the relations written earlier, we multiply and divide through by the quantities with the subscript 0 to obtain

$$\frac{M_{CO_2,0}}{V_0}\frac{d}{dt}\left(\frac{M_{CO_2}}{M_{CO_2,0}}\right) = \left(\frac{V}{V_0}\right) - \frac{F_{wr,0}}{V_0}\left(\frac{F_{wr}}{F_{wr,0}}\right) \tag{16.29}$$

The ratio between the present $CO_2$ mass ($5.7 \cdot 10^{16}$ mol) and the emission rate from volcanoes ($5 \cdot 10^{12}$ mol year$^{-1}$) gives a time constant of the order of 10 kyr. In the present condition, also the ratio of weathering/volcanoes is unity. We introduce a normalized time as

$$\tilde{t} = tV_0/M_{CO_2,0}$$

so that Eq. (16.30) becomes

$$\frac{d}{d\tilde{t}}\left(\frac{M_{CO_2}}{M_{CO_2,0}}\right) = \left(\frac{V}{V_0}\right) - \left(\frac{F_{wr}}{F_{wr,0}}\right) \tag{16.30}$$

which is the nondimensional equation we may integrate. We make the assumption that the planet temperature will adapt instantaneously to the changing carbon dioxide partial pressure.

In the original work by James Walker, the weathering rate depends on the partial pressure in the atmosphere because the idea is that the hotter the planet, the more abundant the precipitation. In the present work, the weathering rate depends on the partial pressure of carbon dioxide in the soil. This assumption ensures the interaction between the biosphere and the atmospheric pressure. We do not expect very much difference because if the partial pressure in the atmosphere and the soil are in equilibrium, then the ratios will not change too much. The results for this integration are given in Figs. 16.10 and 16.11. The first shows the partial pressure of $CO_2$ in the soil and the atmosphere (in ppm). As the solar constant increases (and then the temperature), both pressures decrease because on one side the biological productivity decreases and on the other the weathering increases.

When the atmospheric pressure falls below 10 ppm, the photosynthetic activity ceases, and then we may assume the death of the biosphere.

To interpret a little better these results, it is useful to examine the behavior of the biological productivity and temperature as shown in Fig. 16.11. The most evident result is a rapid temperature increases when the weathering or the biological activity is absent. One of the cases shown in the figure is when the atmospheric pressure is maintained constant at 320 ppm. The weathering has the effect of stabilizing the temperature because the reduction in the partial pressure of $CO_2$ compensates, with a lower greenhouse warming, for the increase in the solarconstant. It is quite

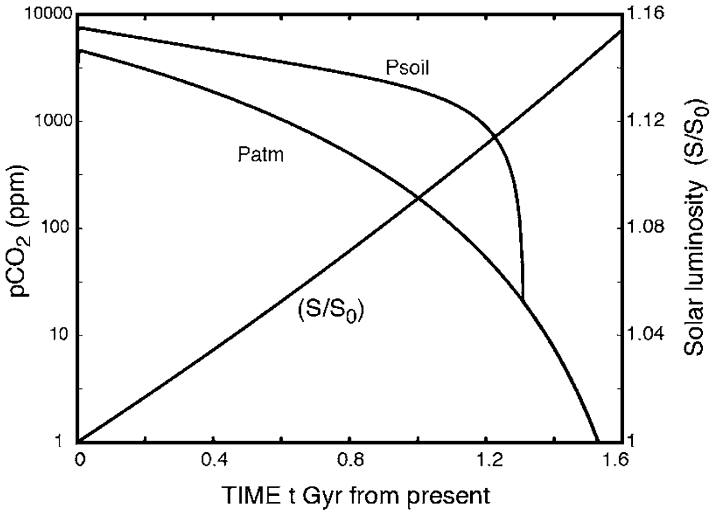



**Fig. 16.10** Carbon dioxide partial pressure in the soil and in the atmosphere as a function of time (or solar constant)

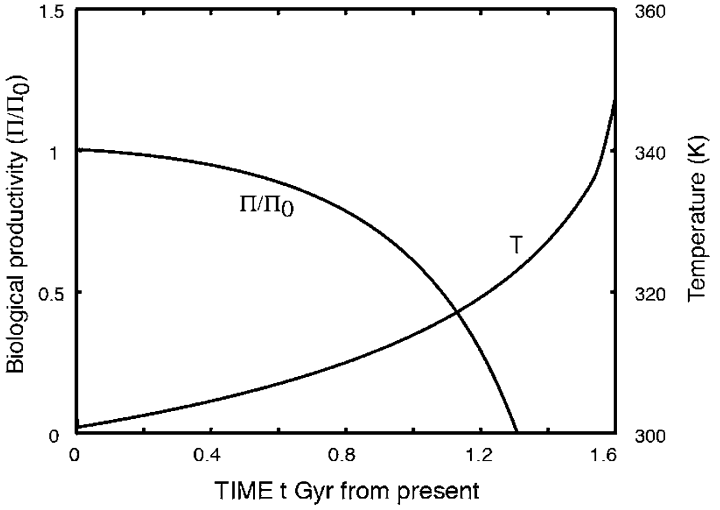



**Fig. 16.11** The behavior of the biological productivity and temperature for the model of the previous figure

interesting in any case how the biological activity can stabilize more efficiently the temperature, at least in the short term, because it may accelerate the weathering rate through a higher pressure in the soil.

Our exercise on the Earth system may be assumed completed so that we can proceed to the chemistry of the troposphere.

## 16.4 Chemistry of the Troposphere

The gases that are emitted in the atmosphere as a result of the different biogeochemical cycles have a certain degree of reactivity so that the most important passage in the cycle may be what happens in the atmosphere. Very important reactions in the troposphere are those that increase the oxidation state of different compounds. The more interesting are then carbon monoxide (CO), hydrocarbons, sulfur dioxide ($SO_2$), nitrogen oxides ($NO_x$), hydrogen sulfide ($H_2S$), ammonia ($NH_3$), and so on.

Once the oxidation is completed, the species formed are removed more rapidly than the precursor gases. In the gas phase, this process of oxidation and removal is dominated by the reactions with the hydroxyl radical OH (also called the "sweeper" of the atmosphere). The average hydroxyl concentration during the daytime is of the order of $1.5 \cdot 10^6$ cm$^{-3}$. The reactivity of OH is such that it may form or can be destroyed in time frames of the order of seconds.

The primary sources for this radical is a dissociative reaction of water with metastable oxygen that in turn is produced in very small quantities in the troposphere by the ozone photodissociation

$$\begin{aligned} & O_3 + h\nu \rightarrow O_2 + O\left(^1D\right) \\ & O\left(^1D\right) + H_2O \rightarrow OH + OH \\ \text{net} \quad & O_3 + h\nu + H_2O \rightarrow O_2 + 2OH \end{aligned} \tag{16.31}$$

The hydroxyl radicals, once formed, are converted into members of the same *odd hydrogen family* ($HO_x$) which includes H, OH, $HO_2$, and $H_2O_2$. As we have mentioned before, the family approach is quite convenient because enables us to calculate very rapidly the mixing ratio of the different components.

The main processes that determine the tropospheric chemistry are shown in Fig. 16.12. From it, we see that, once formed, the OH radical is transformed into hydroperoxy ($HO_2$) through a reaction that involves the conversion of carbon monoxide in carbon dioxide

$$\begin{aligned} & OH + CO \rightarrow CO_2 + H \\ & H + O_2 + M \rightarrow HO_2 + M \\ \text{net} \quad & OH + CO + O_2 \rightarrow CO_2 + HO_2 \end{aligned} \tag{16.32}$$

$HO_2$ is converted back into OH through a reaction that involves ozone and nitrogen oxide

$$\begin{aligned} & HO_2 + NO \rightarrow OH + NO_2 \\ & HO_2 + O_3 \rightarrow OH + 2O_2 \end{aligned} \tag{16.33}$$

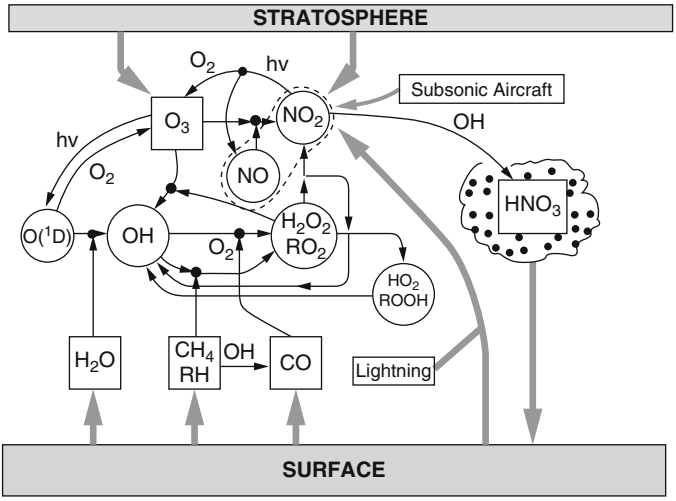



**Fig. 16.12** The main reactive processes in the gas phase in the troposphere. The *gray arrows* indicate the sources, while the *thin arrows* indicate the reactions between the single components (Adapted from Prinn 1992)

The first of these reactions (but also the other) is very important in the continental regions where the abundance of nitrogen oxide is considerable. In the remote ocean areas, the conversion utilizes radicals indicated with *R* represents either hydrogen or alkyl radicals (e.g., deriving from methane)

$$\begin{aligned} &RO_2 + HO_2 \rightarrow ROOH + O_2 \\ &ROOH + hv \rightarrow RO + OH \end{aligned} \tag{16.34}$$

The primary sources for OH dominate in the areas with low $NO_x$ abundance.

Hydroxyl radicals have a very important role in the oxidation of methane and other hydrocarbons. The process has such implications that are worth to study them in detail. It implies the transformation of methane into carbon monoxide and the further oxidation into carbon dioxide as shown by Eq. (16.33). The role of nitrogen oxide is very important in this process so that we need to know something more about it. The sources in the lower troposphere are the combustion processes and lightning, while in the upper troposphere, an important source is direct injection from commercial and military air traffic. Also important is the intrusion from the stratosphere. The main sink is the formation of nitric acid that can be dissolved in the raindrops.

### *16.4.1 Methane Oxidation*

The main sources of the atmospheric methane can be attributed to both natural and anthropogenic processes, as shown in Table 16.2, which also lists as the main sink

the reaction with hydroxyl radicals, although nitrogen oxides play an important role. The chemical chain starts with the reaction that forms the methyl radical $CH_3$

$$CH_4 + OH \rightarrow CH_3 + H_2O \tag{16.35}$$

The methyl radical forms a peroxy radical $CH_3OO$

$$CH_3 + O_2 \rightarrow CH_3OO$$

that may react with a nitrogen oxide radical, starting a process that has as end products of ozone and hydroxyl radicals following the system of reactions listed

$$\begin{aligned}
&CH_3OO + NO \rightarrow CH_3O + NO_2\\
&CH_3O + O_2 \rightarrow H_2CO + HO_2\\
&HCHO + hv\ (\lambda \leq 330\,nm) \rightarrow HCO + H\\
&H + O_2 + M \rightarrow HO_2 + M\\
&HCO + O_2 \rightarrow CO + HO_2\\
&3\ [HO_2 + NO \rightarrow HO + NO_2]\\
&4\ [NO_2 + hv\ (\lambda \leq 410\,nm) \rightarrow NO + O]\\
&4\ [O + O_2 + M \rightarrow O_3 + M]\\
\text{net}\quad &CH_4 + 8O_2 + 5hv \rightarrow CO + 4O_3 + 2OH + H_2O
\end{aligned} \tag{16.36}$$

Through the formation of formaldehyde (HCHO) and its subsequent photodissociation, for each molecule of methane, four molecules of ozone and two hydroxyl radicals are formed, while carbon is oxidized to carbon monoxide. CO may react, as we have seen already, with $HO_2$, but the result of this new oxidation cycle depends critically on the content of nitrogen oxide

$$\begin{aligned}
&CO + HO \rightarrow H + CO_2\\
&H + O_2 + M \rightarrow HO_2 + M\\
&HO_2 + NO \rightarrow HO + NO_2\\
&NO_2 + hv\ (\lambda \leq 410\,nm) \rightarrow NO + O\\
&O + O_2 + M \rightarrow O_3 + M\\
\text{net}\quad &CO + 2O_2 + hv \rightarrow CO_2 + O_3
\end{aligned} \tag{16.37}$$

If the concentration of $NO_x$ is not high enough, ozone is destroyed by the odd hydrogen cycle. The first two steps of the sequences in Eqs. (16.37 and 16.38) are

identical, and they are faster by a factor of 4000 than the third part of Eq. (16.38) which destroys ozone. Ozone will be produced when the $NO_x$ concentration

$$\begin{aligned} &CO + HO \rightarrow H + CO_2 \\ &H + O_2 + M \rightarrow HO_2 + M \\ &HO_2 + O_3 \rightarrow HO + 2O_2 \\ \text{net} \quad &CO + O_3 \rightarrow CO_2 + O_3 \end{aligned} \tag{16.38}$$

is higher by the same factor with respect to the ozone concentration in the troposphere, 5–10 pptv ($10^{-12}$).

To understand a little better the importance of the nitrogen oxides, we may return to the reaction of Eq. (16.37). In the case of a low $NO_x$ concentration, the first reaction becomes so slow that the following process is favored

$$\begin{aligned} &CH_3OO + HO_2 \rightarrow CH_3O_2H + O_2 \\ &CH_3O_2H + h\nu \rightarrow CH_3O + OH \\ &CH_3O + O_2 \rightarrow HCHO + HO_2 \end{aligned} \tag{16.39}$$

Also, in this case, in principle, formaldehyde is formed; however, the dissociation of $CH_3O_2H$ is so slow that it is more rapidly dissolved in water and then rained out. In this process, two radicals are lost (OH and $HO_2$), and carbon monoxide is not formed. More accurate calculations show that there is the loss of 3–4 radicals for poor $NO_x$ air and a gain of 1–2 in situations of high mixing ratio.

The reactions we have examined are in practice the same that dominate the urban air, where methane is substituted by other hydrocarbons. In this case, the processes are much more complex, but the interaction between hydrocarbons, nitrogen oxides, and light is still responsible for the ozone production.

In the reaction scheme, we have seen that nitrogen oxides would never be destroyed if not for the reaction with hydroxyl radical, through which nitric acid is formed and rained out with a time constant of a few days. This time is enough for the mixing ratio of $NO_x$ to decrease by 10 % every 1000 km away from urban areas.

Another important conclusion we can draw from the study of methane oxidation is that of the central role played by hydroxyls. These are able to eliminate most of the trace gases ($NO_x$ in our case), and without them, the residence time of the same gases would be so long that the result would be a chemical composition of the troposphere quite different from the present one. The problem is that the concentration of methane together with $NO_x$ and carbon monoxide is increasing in the atmosphere. This implies in the long term a reduction of the global hydroxyl concentration and consequently a reduction of the oxidizing capacity of the atmosphere.

### 16.4.2 The Chemistry of Urban Air

One of the most serious problems in the chemistry of urban areas is still ozone formation. The process starts with the production of atomic oxygen from the photodissociation of nitrogen dioxide. This atomic oxygen is then used for ozone $O_3$ production but can be destroyed by the reaction

$$O_3 + NO \rightarrow NO_2 + O_2 \tag{16.40}$$

The role of $HO_2$ in converting nitrogen oxide into dioxide in the urban atmosphere is taken by the peroxy radical $RO_2$ which forms from the reaction between hydrocarbon and hydroxyl. These convert the nitrogen oxide into $NO_2$

$$RO_2 + NO \rightarrow NO_2 + RO \tag{16.41}$$

where the peroxy radical $RO_2$ is produced by

$$\begin{aligned} RH + OH &\rightarrow R + H_2O \\ R + O_2 + M &\rightarrow RO_2 + M \end{aligned} \tag{16.42}$$

The result of these processes is that in urban, air rich in hydrocarbons and $NO_x$ ozone is formed in considerable quantities. This chemical mechanism was known a few years ago as "Los Angeles atmosphere," but today it could very well be known as the atmosphere of every city, even small ones.

To be more precise regarding the mechanism for the formation of radicals from hydrocarbons, we can refer to Fig. 16.13, where the ethylene ($C_2H_4$) structure is shown. In some cases, one or more hydrogen atoms may be substituted by a radical $CH_3$ – that is denoted with *R*. In general, one hydrocarbon may assume the form RHC = CHR so that the reaction proceeds like in the upper part of Fig. 16.13. The radical formed on the right reacts then with oxygen to give a peroxy radical that in turn will oxidize NO into $NO_2$. This may clarify the different significances of *R* that appear in the simplified chemical formulae.

We are now ready to examine in detail the ozone production so that we can go on to solve a simple model for its evolution. We will start from the reaction

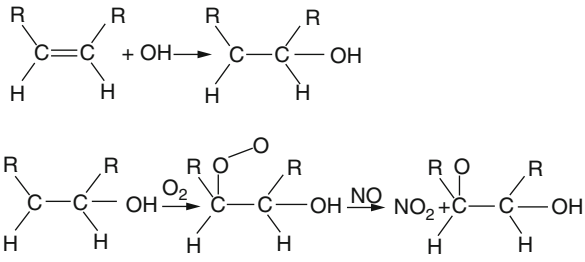



**Fig. 16.13** A scheme for the formation of peroxy radicals from hydrocarbons (From Baird 1995)

scheme (16.43). It is very crude and neglects some reactions but teaches how these models are used. From what we have seen earlier, the most important reaction for the formation of ozone is the production of atomic oxygen from nitrogen dioxide. This in turn is produced in the presence of hydrocarbons that are first converted in $RO_2$

$$\begin{aligned}
&NO_2 + hv \rightarrow NO + O && J_2 \\
&O + O_2 + M \rightarrow O_3 + M && k_1 \\
&NO + O_3 \rightarrow NO_2 + O_2 && k_2 \\
&RH + O \rightarrow R + \text{products} && k_3 \\
&RH + O_3 \rightarrow \text{products} && k_4 \\
&NO + R \rightarrow NO_2 + R && k_5 \\
&NO_2 + R \rightarrow \text{products (incl.PAN)} && k_6
\end{aligned} \tag{16.43}$$

This repeats a very crude scheme proposed by Friedlander in 1969. We can assume photochemical equilibrium and write for the odd oxygen

$$[O] = \frac{J_2\,[NO_2]}{k_1\,[O_2]\,[M] + k_3\,[RH]} \tag{16.44}$$

$$[O_3] = \frac{k_1\,[M]\,[O_2]\,[O]}{k_2\,[NO] + k_4\,[RH]} \tag{16.45}$$

Then we may use the fact that $k_1\,[O_2]\,[M] >> k_3\,[RH]$ and $k_2\,[NO] >> k_4\,[RH]$ so that the above equations simplify to

$$[O] = \frac{J_2\,[NO_2]}{k_1\,[O_2]\,[M]} = \gamma\,[NO_2] \tag{16.46}$$

$$[O_3] = \frac{J_2\,[NO_2]}{k_2\,[NO]} = \beta\frac{[NO_2]}{[NO]} \tag{16.47}$$

We also assume photo equilibrium between all radicals so we get

$$k_3\,[RH]\,[O] + \varepsilon k_5\,[R]\,[NO] = k_5\,[R]\,[NO] + k_6\,[R]\,[NO_2] \tag{16.48}$$

This is because in reaction (5) part of the radicals $R$ are used to destroy NO, and others are regenerated. $\varepsilon$ is then the number of free radicals generated in reaction $k_5$. We get the concentration of radicals

$$[R] = \frac{k_3\,[RH]\,[O]}{k_6\,[NO_2] - (\varepsilon - 1)\,k_5\,[NO]} \tag{16.49}$$

We can assume the denominator to be roughly constant so that

$$[\mathrm{R}] = k_3' [\mathrm{RH}] [\mathrm{O}] \tag{16.50}$$

where

$$k_3' = k_3 / \{k_6 [\mathrm{NO_2}] - (\varepsilon - 1) k_5 [\mathrm{NO}]\}$$

the rate equation for $NO_2$, NO, and RH can be then written

$$\frac{d[\mathrm{NO_2}]}{dt} = [\mathrm{NO_2}] [\mathrm{RH}] \{\alpha [\mathrm{NO}] - \lambda [\mathrm{NO_2}]\} \tag{16.51}$$

$$\frac{d[\mathrm{NO}]}{dt} = -\alpha [\mathrm{NO_2}] [\mathrm{NO}] [\mathrm{RH}] \tag{16.52}$$

$$\frac{d[\mathrm{RH}]}{dt} = [\mathrm{NO_2}] [\mathrm{RH}] \{\theta + \mu / [\mathrm{NO}]\} \tag{16.53}$$

where

$$\alpha = \gamma k_5 k_3'; \quad \lambda = \gamma k_6 k_3'; \quad \theta = \gamma k_3; \quad \mu = \beta k_4$$

Equations (16.51), (16.52), and (16.53) can be easily integrated, and the results are shown in Fig. 16.14. In some of the exercises at the back, we report the simple program and the justification for the numbers. The integration was carried out for two different initial conditions. In the case shown on the left, [RH] = 1.15 ppm, $[NO_2]$ = 0.2 ppm, and [NO] = 0.68 ppm. For the case on the right, [RH] = 2 ppm, $[NO_2]$ = 0.2 ppm, and [NO] = 1.0 ppm. Although the photodissociation coefficient was held constant, the results clearly show the delay in the ozone formation with respect to the hydrocarbons and nitric oxide. This then could be a classical situation for the morning traffic with concentration of hydrocarbons and $NO_x$ rising from the traffic and ozone reaching the maximum a few hours later.

### *16.4.3 Can We Control Air Quality?*

From the previous chapter, we have learned that the amount of ozone produced is strictly related to the NOx and RH amount. There is a classical way to assess this dependence which is no longer very popular but still quite instructive for a student. We start by considering the following reactions (Jacob 1999)

$$\mathrm{O_3} + h\nu \rightarrow \mathrm{O_2} + \mathrm{O}\left(^1\mathrm{D}\right) \qquad J_D$$

$$\mathrm{O}\left(^1\mathrm{D}\right) + \mathrm{M} \rightarrow \mathrm{O} + \mathrm{M} \qquad k_M$$

$$\mathrm{O}\left(^1\mathrm{D}\right) + \mathrm{H_2O} \rightarrow 2\mathrm{OH} \qquad k_3$$

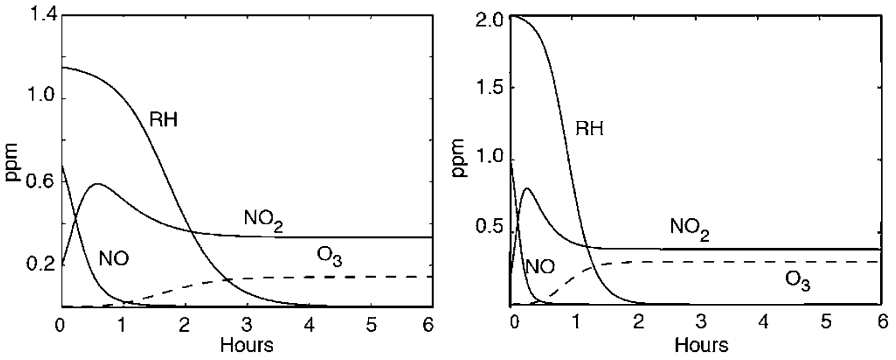



**Fig. 16.14** The results of integration of Eqs. (16.51), (16.62), and (16.53). Mixing ratios in ppm are shown for RH, $NO_2$, NO, and $O_3$. The photodissociation coefficient for $NO_2$ was held constant (Sillman et al. 1990; Sillman 1999)

This basically is the same as (16.31) and this results in the net production of OH

$$P_{OH} = \frac{2J_D k_3}{k_M\,[\mathrm{M}]}\,[\mathrm{O_3}]\,[\mathrm{H_2O}] \tag{16.54}$$

The loss of odd hydrogen is due to the following reactions

$$\begin{aligned} \mathrm{HO_2 + HO_2 \rightarrow H_2O_2 + O_2} \quad &\rightarrow k_8 \\ \mathrm{NO_2 + OH + M \rightarrow HNO_3 + M} \quad &\rightarrow k_9 \end{aligned}$$

So the loss is given by

$$L_{\mathrm{HO}x} = k_8[\mathrm{HO_2}]^2 + k_9\,[\mathrm{NO_2}]\,[\mathrm{OH}]\,[\mathrm{M}] \tag{16.55}$$

Equating (16.54) and (16.55), we have

$$k_8[\mathrm{HO_2}]^2 + k_9\,[\mathrm{NO_2}]\,[\mathrm{OH}]\,[\mathrm{M}] = \frac{2J_D k_3}{k_M\,[\mathrm{M}]}\,[\mathrm{O_3}]\,[\mathrm{H_2O}] \tag{16.56}$$

At this point, we need to express everything as a function of total odd hydrogen and total odd nitrogen. The ratio between OH and $HO_2$ is easily found based on

$$\begin{aligned} \mathrm{HO_2 + NO \rightarrow OH + NO_2} \quad &\rightarrow k_7 \\ \mathrm{RH + OH + O_2 \rightarrow RO_2 + H_2O} \quad &\rightarrow k_4 \end{aligned}$$

so that

$$[\mathrm{OH}] = \frac{k_7\,[\mathrm{O_3}]\,[\mathrm{H_2O}]}{k_4\,[\mathrm{RH}]} \tag{16.57}$$

In the same way, the set of reactions

$$\begin{aligned} &NO_2 + h\nu \rightarrow NO + O && \rightarrow J_N \\ &NO + HO_2 \rightarrow NO_2 + OH && \rightarrow k_7 \\ &NO + O_3 \rightarrow NO_2 + O_2 && \rightarrow k_{10} \\ &NO + RO_2 \rightarrow NO_2 + RO && \rightarrow k_{11} \end{aligned}$$

gives the ratio of $NO_2$ to $NO_x$

$$\frac{[NO]}{[NO_2]} = \frac{J_N}{k_{10}\,[O_3] + k_7\,[HO_2]} \approx \frac{J_N}{k_{10}\,[O_3]} \tag{16.58}$$

Now (16.57) and (16.58) can be substituted in (16.56); in terms of total odd hydrogen $[HO_x] = [HO_2] + [OH]$ and odd nitrogen $[NO_x] = [NO_2] + [NO]$, we get a second degree equation for $[HO_2]$

$$k_8[HO_2]^2 + \frac{k_7 k_9 k_N [NO_x]^2\,[M]}{k_4\,[RH]}\,[HO_2] - \frac{2J_D k_3}{k_M\,[M]}\,[O_3]\,[H_2O] = 0 \tag{16.59}$$

Once $HO_2$ is calculated, the ozone production is obtained by

$$P_{O_3} = [NO]\,\{k_5\,[RO_2] + k_7\,[HO_2]\} \tag{16.60}$$

where

$$[RO_2] = k_4\,[OH]\,[RH]\,[O_2]\,/k_5\,[NO] \tag{16.61}$$

The details are given in one of the exercises. Figure 16.15 shows the results for the ozone isopleths as a function of the RH and $NO_x$ abundance. The gray thick curve separates two regions, the upper left where ozone production is controlled by the $NO_x$ concentration and the lower right where ozone production is controlled by the RH concentration.

There are a number of approximations in tracing a figure like Fig. 16.15. The main one is the lack of consistency between the ozone production rate and the ozone concentration. In order to calculate the chemical compounds, we have adopted an ozone mixing ratio of 0.05 ppm. This may not be consistent with the production reported in the figure. The reader may refer to the simple MATLAB program attached and evaluate the figure for other mixing ratio or even write a better self consistent program.

### *16.4.4 The Atmospheric Sulfur Cycle*

Sulfur, as shown in Fig. 16.16, is emitted in the Earth's atmosphere mainly as sulfur dioxide $SO_2$ from combustion processes and volcanoes (in much less amounts).

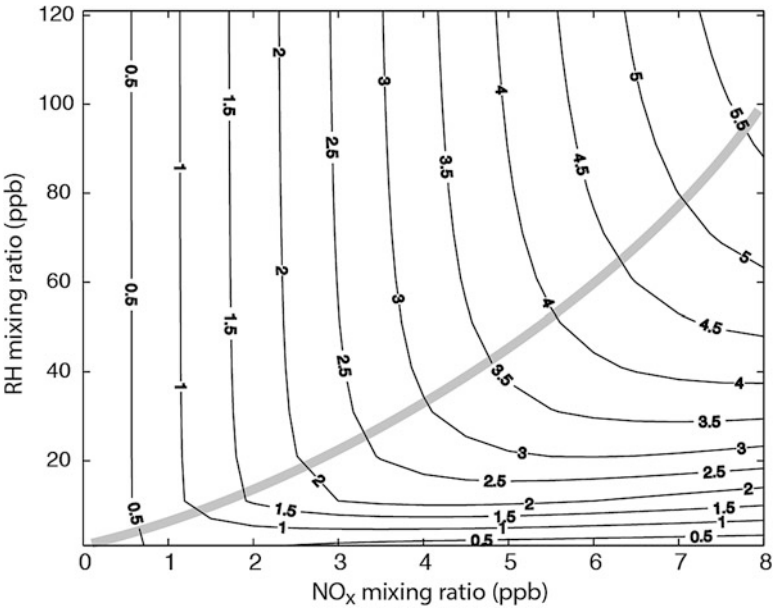



**Fig. 16.15** Qualitative ozone production rate (in units of $10^8$ molecules/cm$^3$ s) as a function of the NOx and RH mixing ratio. The *gray lines* separate the region where production is controlled by NOx (*upper left corner*) from the region where production is controlled by the RH

Also, reduced forms of sulfur are important, like dimethyl sulfide $(CH_3)_2S$, hydrogen sulfide $H_2S$, carbonyl sulfide COS, and carbon disulfide $CS_2$. These species in most cases are of natural origin and are produced during decomposition processes, from algae and in general through microbiological processes in the sea and swamps. Once in the atmosphere, these species are oxidized either through reaction with hydroxyl OH and the radical nitrogen trioxide $NO_3$ or through photodissociation. The conversion happens on time scales that range from a few hours to a few days.

In particular, the conversion of carbonyl sulfide is a very slow process because photodissociation rate is low and the reaction with atomic oxygen is quite slow. The lifetime in this way is long enough for COS to reach the stratosphere, where the photodissociation is much faster so that it can be readily oxidized to sulfuric acid which may dissolve in the very dry stratospheric environment and form a highly concentrated solution.

The same fate follows for the sulfur dioxide in the troposphere, and it is quite interesting to look in detail at such processes

$$\begin{aligned}
&SO_2 + OH \rightarrow HSO_3^- \\
&HSO_3^- + O_2 \rightarrow SO_3 + HO_2 \\
&SO_3 + H_2O \rightarrow H_2SO_4\,(g) \\
&H_2SO_4(g) \overset{H_2O}{\rightarrow} H_2SO_4\,(aq)
\end{aligned} \tag{16.62}$$

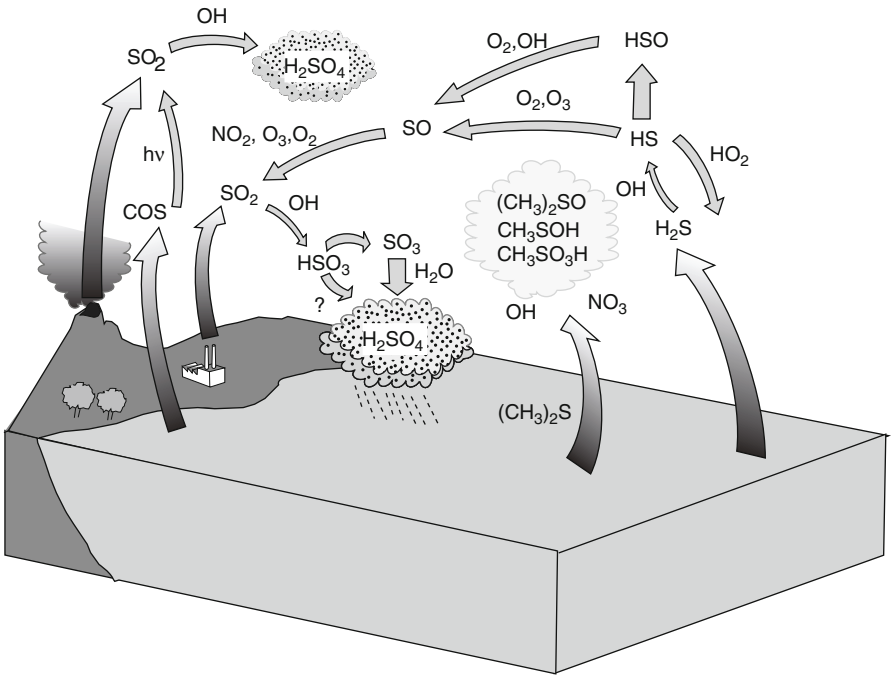



**Fig. 16.16** The scheme for the atmospheric sulfur cycle. The *large arrows* indicate the main fluxes (sinks and sources). The clouds indicate those processes that happen in aqueous phase. The *gray arrows* indicate the atmospheric chemical processes

where (*g*) and (*aq*) mean that the reaction is in the gas phase or aqueous phase. The sulfuric acid formed in this way is, together with nitric acid, the main component of acid rain. Sulfur dioxide can be dissolved in water giving directly sulfurous acid

$$SO_2 + H_2O \rightarrow H_2SO_3\,(aq)$$

This acid may form a bisulfite ion and give a very weak acidity to the rain so that the main contribution to the acidification of rain remains the formation of sulfuric acid. It is interesting to note that even in this case, there is a relationship between the presence of nitrogen oxides and the production of sulfuric acid because the radical formed in the second part of Eq. (16.48) can then form a hydroxyl in the presence of the nitrogen oxide

$$HO_2 + NO \rightarrow NO_2 + OH$$

so that the net effect is a catalytic reaction that forms sulfuric acid

$$SO_2 + NO + O_2 \overset{H_2O}{\rightarrow} H_2SO_4\,(acq) + NO_2 \tag{16.63}$$

In a couple of chapters, we will talk about the formation of stratospheric aerosols from the sulfur species. For the troposphere, a very important product is acid rain, whose effect on the ecosystem will depend very much on the composition of soil or rocks. It may seem a paradox but the oldest and, in a sense, noblest rocks like granite or quartz are those which suffer the greater damage from acid rain. These minerals have no way to neutralize the acid. On the other hand, rocks containing chalk or clay can neutralize the acid because they contain calcium carbonate

$$\begin{aligned} &\mathrm{CaCO_3}(s) + \mathrm{H^+}\,(\mathrm{acq}) \rightarrow \mathrm{Ca^{2+}}\,(\mathrm{acq}) \rightarrow \mathrm{HCO_3^-}\,(\mathrm{acq}) \\ &\mathrm{HCO_3^-}\,(\mathrm{acq}) + \mathrm{H^+}\,(\mathrm{acq}) \rightarrow \mathrm{H_2CO_3}\,(\mathrm{acq}) \rightarrow \mathrm{CO_2}(g) + \mathrm{H_2O}\,(\mathrm{acq}) \end{aligned} \tag{16.64}$$

This implies that lakes with a granite bed are strongly acidified.

The very same reactions are responsible for the weathering of monuments or marble statues.

A very important effect of acid rain is the mobilization of metals in the soil. The combined effect of the acidity of water and the presence of metals like aluminum is the main cause for the damage to the fish population in lakes. It is likely that the same agents are also the cause for the extended damage to the North European forests (although other complex mechanisms may be at work).

## 16.5 Modes of a Chemical System

We have previously defined the lifetime for a gas. There is actually a way to generalize this concept that may be useful in a number of applications.

We will consider the same methane oxidation mechanism which may be reduced to only three reactions

$$\begin{aligned} d\,[\mathrm{CH_4}]\,/dt &= \mathrm{S_{CH_4}} - k_1\,[\mathrm{CH_4}]\,[\mathrm{OH}] \\ d\,[\mathrm{CO}]\,/dt &= \mathrm{S_{CO}} + k_1\,[\mathrm{CH_4}]\,[\mathrm{OH}] - k_2\,[\mathrm{CO}]\,[\mathrm{OH}] \\ d\,[\mathrm{CO}]\,/dt &= \mathrm{S_{OH}} - k_1\,[\mathrm{CH_4}]\,[\mathrm{OH}] - k_2\,[\mathrm{CO}]\,[\mathrm{OH}] - k_3\,[\mathrm{X}]\,[\mathrm{OH}] \end{aligned} \tag{16.65}$$

where the three sources are indicated with $S_i$, while $X$ is any sink much faster than any reaction with OH. Equation (16.65) shows that, while the production terms can be assigned, the loss terms depend on all three species involved in the system. If the net production is indicated with $P_i(X_i)$ where $i = 1{,}3$ we can write

$$dX_i/dt = P_i\,(X_i)$$

We now assume that we have found a stationary solution to this system with values $X_i{}^0$; then for a perturbation $D_i$, we will have

$$d\left[X_i^0 + D_i\right]/dt = P_i\left(X_i^0 + D_i\right)$$

that can be written as

$$d\left[X_i^0\right]/dt + d\left[D_i\right]/dt = P_i\left(X_i^0\right) + \left[\partial P_i\left(X_i\right)/\partial X_k\right] D_k + \mathrm{O}\left(D^2\right)$$

from which we obtain

$$d\left[D_i\right]/dt = J_{ik} D_k + \mathrm{O}\left(D^2\right) \tag{16.66}$$

where

$$J_{ik} = \left[\partial P_i\left(X_i\right)/\partial X_k\right]$$

The eigenstates of the matrix **J** describe the behavior of the system for small perturbations around the equilibrium state. If $A$ is an eigenvector of **J** with eigenvalue $-c$, then we have

$$dA/dt = \mathbf{J}A = -cA$$

so that the perturbation will relax as $A \propto \exp(-ct)$ with a characteristic time $1/c$. If the chemistry is linear and the system completely decoupled, then the matrix would be diagonal with the diagonal elements as eigenvectors. When for the system of Eq. (16.52) we adopt the right values for the constants, we get

$$k_1 = 5.0 \times 10^{-15}\mathrm{cm}^3\mathrm{s}^{-1} \qquad \mathrm{S}_{\mathrm{CH_4}} = 1.6 \times 10^5\mathrm{cm}^{-3}\mathrm{s}^{-1}$$
$$k_2 = 2.0 \times 10^{-13}\mathrm{cm}^3\mathrm{s}^{-1} \qquad \mathrm{S}_{\mathrm{CO}} = 2.4 \times 10^5\mathrm{cm}^{-3}\mathrm{s}^{-1}$$
$$k_3\left[X\right] = 1.0\mathrm{s}^{-1} \qquad \mathrm{S}_{\mathrm{OH}} = 1.12 \times 10^6\mathrm{cm}^{-3}\mathrm{s}^{-1}$$

while the steady-state solutions are

$$[\mathrm{CH_4}] = 5.714 \times 10^{13}\mathrm{cm}^{-3} \quad [\mathrm{CO}] = 3.571 \times 10^{12}\mathrm{cm}^{-3} \quad [\mathrm{OH}] = 5.6 \times 10^5\mathrm{cm}^{-3}$$

The matrix elements can be easily calculated as

$$J_{11} = -k_1[\mathrm{OH}] \qquad J_{12} = 0.0 \qquad J_{13} = -k_1[\mathrm{CH_4}]$$
$$J_{21} = k_1[\mathrm{OH}] \qquad J_{22} = -k_2[\mathrm{CO}] \qquad J_{32} = -k_1[\mathrm{CH_4}] + k_2[\mathrm{CO}]$$
$$J_{31} = -k_1[\mathrm{OH}] \qquad J_{32} = -k_2[\mathrm{CO}] \qquad J_{32} = -k_1[\mathrm{CH_4}] - k_2[\mathrm{CO}] - k_3[X]$$

It is to notice that the diagonal elements of this matrix represent the lifetime of the single species and correspond to the times

$$\tau_{CH_4} = 1/J_{11} = 11.32\,\mathrm{year} \quad \tau_{CO} = 1/J_{22} = 0.283\,\mathrm{year} \quad \tau_{OH} = 1/J_{33} = 0.5\,\mathrm{s}$$

Using the rules for the calculation of the eigenvectors, we find (in $s^{-1}$)

$$c_1 = -1.7691 \times 10^{-9} \qquad c_1 = -8.8630 \times 10^{-8} \qquad c_3 = -2.0$$

which correspond to 17.9 years, 0.358 years, and a half second, respectively. The first result to notice is that the methane lifetime is much longer when we consider the coupled system compared to just the single species. This is interesting because we must consider terms like $J_{31}$ and $J_{32}$ that if neglected would make the eigenvalues to coincide with the diagonal elements and the lifetimes to that of the single species. Once the eigenvalues are known, we can calculate the eigenvectors that are (in units of molecules $cm^{-3}$)

$$\begin{array}{lll} +0.999 & -0.182 & -0.138 \\ +0.039 & +0.983 & -0.208 \\ -3.6 \times 10^{-9} & -5.5 \times 10^{-8} & -0.968 \end{array}$$

To understand better the significance of the eigenvectors, it is more convenient to normalize them to a 100 % perturbation. This does not seem logical because for such a perturbation, the system would not be linear. For example, for methane (the first column), we obtain the perturbation normalized to the carbon monoxide as

$$\left(5.714 \times 10^{13}/0.999\right) \times \left(0.039/3.571 \times 10^{12}\right) = 0.624$$

and in the same way for the others. We obtain in this way a very interesting table

$$\begin{array}{rrr} 100.0 & -1.2 & 0.00000 \\ +62.4 & 100.0 & 0.000003 \\ -36.8 & -35.6 & 100.0 \end{array}$$

From these results, we see that a 100 % perturbation in methane implies a 63 % change in carbon monoxide and a 37 % change in the hydroxyl in the opposite direction.

Another consideration should be made about the sensitivity of the lifetime to the strength of the sources. This is quite important especially for methane which has a source that changes with time. A changing source implies a change in the equilibrium values and then the lifetime. The simplest case is that of an excess source for the hydroxyl OH. At equilibrium, we have

$$\begin{aligned} S_{CH_4} &= k_1 [OH] [CH_4] \\ S_{CO} + S_{CH_4} &= k_2 [OH] [CO] \end{aligned} \tag{16.67}$$

If we consider the third equation for the equilibrium (relative to OH), we have

$$k_3 [X] [OH] = S_{OH} - (2S_{CH_4} + S_{CO})$$

In order to have a solution for OH, we have a condition on the sources

$$S_{OH} > (2S_{CH_4} + S_{CO})$$

that is, the OH source must be strong enough to oxidize the combined sources of methane and carbon monoxide. If the lifetime is defined simply as the inverse of $J_{11}$, we have

$$J_{11} = k_1 [OH] = k_1 \frac{S_{OH} - (2S_{CH_4} + S_{CO})}{k_3 [X]} = \frac{k_1 (2S_{CH_4} + S_{CO})}{k_3 [X]} E$$

where with

$$E = S_{OH} / (2S_{CH_4} + S_{CO}) - 1$$

we have defined the excess source. It is clear at this point that the lifetime changes as the inverse of $E$. If we want the lifetime for the coupled system, we need to calculate the eigenvalues for each value of the excess source. For example, if the concentration of OH increases, then those of methane and carbon monoxide decrease. The result of this calculation is shown in Fig. 16.17, where the lifetime obtained with the two methods is reported. We notice that the lifetime taken as the inverse of the diagonal term scales as the inverse of the hydroxyl source. The lifetime obtained by considering the coupled system tends to the previous one for $E > 0$, while it is very long for small values of the source, that is, for very low OH concentrations. The ratio between the time constants can be interpreted in a different way. From the definition of eigenvalue, we get

$$c_1 \Delta CH_4 = J_{11} \Delta CH_4 + J_{13} \Delta OH$$

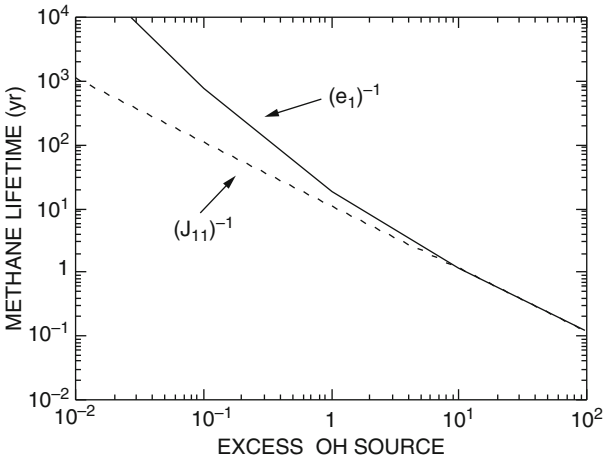



**Fig. 16.17** The methane lifetime compared with that calculated with the eigenvalue

from which we obtain easily

$$c_1 \Delta (\ln CH_4) = J_{11} \Delta (\ln CH_4) + J_{13} \{[OH] / [CH_4]\} \Delta (\ln OH)$$

Considering the ratio between concentrations at the steady state, we have

$$c_1 \Delta (\ln CH_4) \approx J_{11} \{\Delta (\ln CH_4) + \Delta (\ln OH)\}$$

This means that the ratio between the eigenvalue and the diagonal element gives the sensitivity of the methane mixing ratio to a change in the hydroxyl concentration. This approach is based on the modes of a chemical system and can be extended to models that are function of height and latitude. It has the advantage that isolates the mechanism that can potentially perturb the chemical system.

We have neglected in this case the dynamic transport of the chemical species because this complicates considerably the treatment of the modes. Those interested may read the bibliography (look for Prather) after they have read the next section.

$$[NO_2] [OH] [M]$$

For sake of simplicity, we will assume that the family is composed only of NO and $NO_2$ so that if $R$ is the ratio

$$R = [NO] / [NO_2]$$

we have for the residence time

$$\tau_{NO_x} = (1 + R) / k [M] [OH]$$

which gives the result of a few days compared to the few minutes for a single component. In the model, the transport equation will be integrated for the families, while the single components will be calculated at photochemical equilibrium.

All this will be examined in greater detail in the next two chapters which deal with the stratosphere.

## E.16 Examples

### *E.16.1 The Simple Carbon Cycle*

The simple carbon cycle given in Fig. 16.6 can be solved with the following MATLAB program

```
[t,x]=ode45('carbonsim',[0 100],
       [2*655 3.55e4 8.19e7 1.6e7 2.7e3]);
```

```
plot(t,x(:,1),t,x(:,5),'--')
function xdot=carbonsim(t,x)
xdot=zeros(5,1);
xdot(1)=x(5)/100+x(2)/325-x(1)*(1/24+1/6);
xdot(2)=x(1)/6+x(3)/3e8-x(2)*(1/325+1/1.3e5);
xdot(3)=x(2)/1.3e5-x(3)/3.e8;
xdot(4)=x(5)/3.3e4-x(4)/2e8;
xdot(5)=x(1)/24-x(5)*(1/3.3e4+1/100);
```

The program is made up of two parts, the run (run_carbonsim) and the function that specifies the derivatives. The five reservoirs are in the order: x(1), atmosphere; x(2), ocean; x(3), sedimentary rocks; x(4), sedimentary rocks and organic; and x(5), biosphere and soil. For each reservoir, the input and output are given by the appropriate fluxes. These are obtained dividing the reservoir by the relative residence time. Notice that for the atmosphere, the quantity of carbon here specified corresponds to the situation in which the mixing ratio of $CO_2$ was 300 ppm, that is, the situation of the mid-1950s. It is actually a good occasion to establish the relation between the mixing ration of $CO_2$ and its mass in the atmosphere. 1 Tg of carbon corresponds to $1 \cdot 10^{15}$ (44/29) $= 3.66 \cdot 10^{15}$ g of $CO_2$. These with respect to the total mass of the atmosphere ($5.2 \cdot 10^{21}$ g) correspond to a mass mixing ratio of $7 \cdot 10^{-7}$, and to a volume mixing ratio of $4.6 \cdot 10^{-7}$ .1 ppm of $CO_2$ corresponds to 7.90 Tg of the gas.

With this simple model, we can do the most elementary simulation like going to add to the volcanic source the "human" source and reproduce the historical behavior of the industrial contribution to the increase in the $CO_2$ mixing ratio.

## *E.16.2 The Carbon Cycle with the Ocean*

This refers now to the other more sophisticated cycle, given in Fig. 16.7. Also in this, we use a MATLAB program composed of run and the function.

```
[t,x]=ode45('oceancarb',[0 500],[2.26 2.2 1. 2.247 2.03]);
plot(t,x(:,3)*280,'--')
function xdot=oceancarb(t,x)
%alkd=x(1)
%alks=x(2)
%pco2=x(3)
%sigd=x(4)
%sigs=x(5)
xdot=zeros(5,1);
watem=288;
vols=1.2e17;
vold=1.23e18;
corat=0.25;
```

```
kcarb=5.75e-4+6.e-6*(watem-278);
kco2=0.035+0.0019*(watem-278);
matmco2=5.02 e16;
prod=1.75e14
proda=(2*corat-0.15)*prod;
prods=(1+corat)*prod;
dstime=8.64;
if t<250
antr1=3e12*exp(t/26.9);
antr=antr1/(44*matmco2);
else
    antr=0
end
dalk=(x(1)-x(2))*wflux;
dsig=(x(4)-x(5))*wflux;
hco3=(x(5)-sqrt(x(5)^2-x(2)*(2*x(5)-x(2))*(1-4*kcarb)))/
  (1-4*kcarb);
co3=(x(2)-hco3)/2.;
pco2s=kco2*hco3*hco3/co3;
xdot(1)=(proda-dalk)/vold;
xdot(2)=(dalk-proda)/vols;
xdot(3)=(pco2s-x(3))/dstime+antr;
xdot(4)=(prods-dsig)/vold;
xdot(5)=((-(pco2s-x(3))*matmco2/dstime+dsig)-prods)/vols;
```

In this case, the variables are just those listed to solve Eqs. (16.19, 16.20, and 16.21), and we only need to clarify how the carbon dioxide mixing ratio is evaluated. The present atmospheric level (PAL) is assumed to be 280 ppm, and so according to what we had calculated before, this corresponds to $280 \times 7.90 \cdot 10^{15}$ g $= 5.02 \cdot 10^{16}$ mol. This represents the matmco2 parameter in the program. The anthropogenic source has been fitted to the experimental data of emission since the year 1750 (t = 0). The term antr1 for the year 2000 corresponds to $3.26 \cdot 10^{16}$ g of $CO_2$/year and then $7.40 \cdot 10^{14}$/year mole of $CO_2$. To obtain "antr," we need to divide again by matmco2.

### *E.16.3 The Oxygen Cycle Is Connected with the Carbon Cycle*

In the book by Jacob, the connection is quite clear. The total oxygen contained in the atmosphere amounts to $0.21 \cdot (32/28.96)\ 5.2 \cdot 10^{21} = 1.2 \cdot 10^{6}$ Pg. (1 Pg = petagram $10^{15}$ g). If all the carbon contained in the biosphere would be oxidized, it would consume $7.2 \cdot 10^{18}$ g of oxygen, that is, only 0.06 % of the atmospheric oxygen. The conclusion is that the biosphere hardly controls the atmosphericoxygen, and

this control must be found somewhere else. Actually, as we learned long ago in the school, the main source of oxygen should be photosynthesis

$$CO_2 + H_2O \rightarrow CH_2O + O_2 \tag{E.16.1}$$

The reverse of this reaction is respiration or decay. However, the two reactions are not perfectly in balance because some of the organic material produced by photosynthesis is subtracted by burial to oxidation. Respiration and decay are on a time scale of roughly 100 years, and the fraction subtracted to oxidation is around 0.1–0.2 %. However, there are other sinks for atmospheric oxygen that works on time scale shorter than burial. Also there are other redox-sensitive elements buried with carbon the most important being sulfur and iron. During weathering, $O_2$ dissolved in water oxidizes sulfur within continental pyrite ($FeS_2$), making soluble sulfate ($SO_4{}^{2-}$) which is carried out by rivers to the ocean. In the ocean, bacteria reduce sulfate and ferric ions ($Fe^{3+}$) to pyrite. The reducing power of photosynthesized organic carbon is transferred to pyrite so that pyrite buried in sediments is balanced by $O_2$ production.

$$\begin{aligned}
&15CO_2 + 15H_2O \rightarrow 15CH_2O + 15O_2\\
&15CH_2O + 2Fe_2O_3 + 16H^+ + 8SO_4^{2-} \rightarrow 4FeS_2 + 23H_2O + 15CO_2\\
&\text{net} \quad 2Fe_2O_3 + 16H^+ + 8SO_4^{2-} \rightarrow 4FeS_2 + 8H_2O + 15CO_2
\end{aligned} \tag{E.16.2}$$

We see then that for each carbon atom, a molecule of $O_2$ is consumed, while each sulfur atom as $FeS_2$ consumes 19/8 $O_2$ molecules (30 atoms from $CO_2$ and 8 from $H_2O$ are consumed by 8 S atoms as $FeS_2$). The organic carbon in sedimentary rocks and pyrite is estimated to be $1.2 \cdot 10^7$ Pg and $5 \cdot 10^6$ Pg, respectively, and the time it takes for plate tectonics to recycle is about 100 million years. The corresponding weathering rates are 0.12 Pg C/year and 0.05 Pg S/year. The resulting loss for oxygen is then $2(0.12 + 0.05(19/8)) \approx 0.4$ Pg O/year. This would give a lifetime for oxygen of $1.2 \cdot 10^6/0.4 \approx 3$ million year.

## *E.16.4 The Simple Polluted Atmosphere*

From Eq. (16.43), we can write the rate equations

$$\frac{d\,[NO_2]}{dt} = k_5\,[NO]\,[R] - k_6\,[NO_2]\,[R] \tag{E.16.3}$$

$$\frac{d\,[NO]}{dt} = -k_5\,[NO]\,[R] \tag{E.16.4}$$

$$\frac{d\,[RH]}{dt} = -\,[RH]\,\{k_3\,[O] + k_4\,[O_3]\} \tag{E.16.5}$$

[R], [O], and [$O_3$], in favor of [RH], [NO], and [$NO_2$], to obtain

$$\frac{d\,[NO_2]}{dt} = [RH]\,[NO_2]\left\{\gamma k_3' k_5\,[NO] - \gamma k_3' k_6\,[NO_2]\right\}$$

$$\frac{d\,[NO]}{dt} = -\gamma k_5 k_3'\,[NO_2]\,[NO]\,[RH]$$

$$\frac{d\,[RH]}{dt} = -\,[NO_2]\,[RH]\left\{k_4\,(J_2/k_3)\,[NO_2]\,/[NO]^2 + \gamma k_3\right\}$$

With the notation

$$\alpha = \gamma k_5 k_3'; \quad \lambda = \gamma k_6 k_3'; \quad \theta = \gamma k_3; \quad \mu = \beta k_4$$

we obtain (16.61, 16.62, and 16.63)

These equations again can be solved with a simple matlab program as follows

```
[t,x]=ode45('chem',[0 60*6],[0.2 0.68 1.15]);
beta=0.001;
o3=beta*x(:,1)./x(:,2);
plot(t/60,x(:,1),t/60,x(:,2),t/60,x(:,3),t/60,o3,'--')
function xdot = chem(t,x)
% x(1) NO2
% x(2) NO
% x(3) RH
alfa=0.1;
theta=1.83e-3;
mu=2.45e-4;
lamda=0.02;
xdot=zeros(3,1);
xdot(1)=x(1)*x(3)*(alfa*x(2)-lamda*x(1));
xdot(2)=-alfa*x(1)*x(2)*x(3);
xdot(3)=-x(1)*x(3)*(theta+mu/x(2));
```

The program is self-explanatory, and the ozone is calculated in the run program. With a slight change, this program could be used to evaluate the pollution during rush hour just assigning a source function for RH and NO.

## *E.16.5 The Isopleth Diagram for Ozone*

The program we wrote simply follows paragraph 16.5.3, and we have indicated the relevant reaction rates and the constant like density (rho) and so on. As we have explained in the text, there could be some inconsistency between the calculated isopleth and the assumed ozone mixing ratio. But again this is something that can be tested.

```
%program o3isopleth
clear all
close all
r1=1.86e-5; % o3 + uv
r2=3.e-11; % o1d+m
r3=2.2e-10; % o1d+h2o
r4=2.3e-29; % rh+oh+o2
r5=2.5e-12; % no+ro2
r7=8.1e-12;% ho2+no
r8=1.7e-12; % ho2+ho2
r9=2.5e-11; % no2+oh+m
r10=1.9e-14; % no+o3
jno2=0.0081;
rho=2.5e19;
h2o=1.e-2*rho;
o3=0.5e-7*rho/4;
niox=[0:0.5:8];
rhx=[1:10:125];
for i=1:length(niox)
for j=1:length(rhx)
nox(i)=niox(i)*rho*1.e-9;
no2(i)=r10*o3*nox(i)/(r10*o3+jno2);
no(i)=jno2*no2(i)/(r10*o3);
rh(j)=1*rhx(j)*rho*1.e-9;
rn=jno2*r10*o3/((r10*o3+jno2)*(r10*o3+jno2));
a=r8;
b=r9*r7*rn*(nox(i).^2)./(r4*rh(j)*rho);
c=2*r1*r3*h2o*o3/(r2*rho);
werke=b.*b+4*a*c;
ho2(i,j)=(-b+sqrt(werke))/(2*a);
oh(i,j)=r7*ho2(i,j)*no(i)/(r4*rh(j)*rho);
ro2(i,j)=r4*oh(i,j)*rh(j)*0.21*rho/(r5*no(i));
pro3(i,j)=no(i)*(r5*ro2(i,j)+r7*ho2(i,j));
end
end
[x,y]=meshgrid(nox/2.5e10,rh/2.5e10);
[C,h]=contour(x,y,pro3'/1.e8);
set(h,'ShowText','on','TextStep',get(h,'LevelStep')*0.5)
```

### *E.16.6 The Lifespan of the Biosphere*

The program we have used in this case is a slight variation with respect to the one illustrated in the previous paragraphs. The main difference is that we have used the

approximation that the right-hand side of Eq. (16.25) is unity, and this will give immediately the dependence of pressure from temperature

$$P_{\text{atm}} = P_{\text{atm},0} \exp\left(\frac{T - T_0}{4.11}\right)$$

The rest of the program remains the same (calculation of temperature and so on). The MATLAB program is quite straightforward.

```
%program lifespan1
clear all
close all
temp(1)=288.15;
time(1)=0.;
p0=320;
patm(1)=p0;
for i=1:1599
    time(i)=(i-1)*10;
    solar(i)=1368/(1-0.38*time(i)/4.55e4);
    albedo=1.4981-0.0065979*temp(i)+8.567e-6*temp(i)*temp(i);
    teff1=((1-albedo)*solar(i)/(4*5.669e-8));
    teff2=sqrt(teff1);
    teff=sqrt(teff2);
    psi=log10(1.e-6*patm(i));
    dtemp1=4.895e7/(temp(i)*temp(i))-3.9787e5/temp(i)-6.7084/
        (psi*psi);
    dtemp=815.17+dtemp1+73.221/psi-30882/(temp(i)*psi);
    temp(i+1)=teff+dtemp;
    pco2(i)=p0*exp(-(temp(i+1)-288.15)/4.11);
    if pco2(i)>0.003125
    patm(i+1)=p0*pco2(i);
    else
    patm(i+1)=1;
    end
    tcels=1-(((temp(i+1)-273.15)-25)/25)^2;
    if tcels<0 | patm(i+1)<10
    productivity(i)=0;
    else
    productivity(i)=2*tcels*(p0-10)/(300+p0-10);
    end
    if patm(i+1)<10 | (temp(i+1)-273.15)>50
    product=0;
    else
    product=productivity(i);
    end
    psoil0=3200;
```

```
    psoil(i)=product*(psoil0-320)+patm(i+1);
end
patm(end)=[];
temp(end)=[];
plotyy(time(5:end),[log10(patm(5:end))',log10(psoil(5:end))'],
    time(5:end),solar(5:end)/1368)
figure
plotyy(time(5:end),productivity(5:end),time(5:end),temp(5:end))
```

### *E.16.7 An Example on Chemical Modes*

We will follow Prather (2007) by studying the Chapman reaction for an oxygen atmosphere.

| | | |
|---|---|---|
| $O_2 + \text{photon} \rightarrow O + O$ | $R_1 = J_1 [O_2]$ | $J_1 = 1.8 \times 10^{-10} s^{-1}$ |
| $O + O_2 + M \rightarrow O_3 + M$ | $R_2 = k_2 [O] [O_2]$ | $k_2 = 5.26 \times 10^{-17} cm^3 s^{-1}$ |
| $O_3 + \text{photon} \rightarrow O_2 + O$ | $R_3 = J_3 [O_3]$ | $J_3 = 1.4 \times 10^{-3} s^{-1}$ |
| $O + O_3 \rightarrow O_2 + O_2$ | $R_4 = k_4 [O] [O_3]$ | $k_2 = 3.0 \times 10^{-15} cm^3 s^{-1}$ |

The set of reactions can be written as

$$\frac{d[O3]}{dt} = R_2 - R_3 - R_4$$
$$\frac{d[O]}{dt} = 2R_1 + R_3 - R_2 - R_4$$
$$\frac{d[O2]}{dt} = 2R_4 + R_3 - R_1 - R_2$$

The elements of the Jacobian can be easily calculated as

$$J_{11} = \frac{\partial (R_2 - R_3 - R_4)}{\partial [O_3]} = -J_3 - k_4 [O]$$
$$J_{12} = \frac{\partial (R_2 - R_3 - R_4)}{\partial [O]} = k_2 [O_2] - k_4 [O_3]$$
$$J_{13} = \frac{\partial (R_2 - R_3 - R_4)}{\partial [O_2]} = k_2 [O]$$

With the same method, the other Jacobian elements are

| | | |
|---|---|---|
| $-J_3 - k_4 [O]$ | $k_2 [O_2] - k_4 [O_3]$ | $k_2 [O]$ |
| $J_3 - k_4 [O]$ | $-k_2 [O_2] - k_4 [O_3]$ | $2J_1 - k_2 [O]$ |
| $J_3 + 2k_4 [O]$ | $-k_2 [O_2] + 2k_4 [O_3]$ | $-J_1 - k_2 [O]$ |

We can easily determine from the inverse of diagonal elements the relevant lifetimes

$$T(\mathrm{O_3}) = (J_3 + k_4[\mathrm{O}])^{-1} \quad = 12 \ \text{min}$$
$$T(\mathrm{O}) = (k_2[\mathrm{O_2}] + k_4[\mathrm{O_3}])^{-1} = 1.4 \ \text{s}$$
$$T(\mathrm{O_2}) = (J_1 + k_2[\mathrm{O}])^{-1} \quad = 176 \ \text{yr}$$

Beside these time constants, we can derive one for odd oxygen ($\mathrm{O} + \mathrm{O_3}$) by noticing that

$$\frac{d[\mathrm{O}]}{dt} + \frac{d[\mathrm{O3}]}{dt} = -2R_4 + 2R_1$$

and time constant is just

$$T([\mathrm{O}] + [\mathrm{O3}]) = \frac{[\mathrm{O}] + [\mathrm{O3}]}{2R_4} = 1.5 \ \text{day}$$

With the same method, it can be shown that the total oxygen ($\mathrm{O} + 2\mathrm{O_2} + 3\mathrm{O_3}$) is constant, and its lifetime is infinity. At this point, we ask which of these time scales characterizes the system. It is easy to find the eigenvalues of the Jacobian matrix that consist of one zero eigenvalues and two nonzero corresponding to $-1/1.4$ s and $-1/0.74$ day. The zero eigenvalue corresponds to the infinite time scale for oxygen; the second largest coincides with the time scale for odd oxygen, and the smallest is the one of atomic oxygen. The eigenvector for the smallest time is ($[\mathrm{O_3}]$, $[\mathrm{O}]$, $[\mathrm{O_2}]$) = ($-0.995$,$+1.000$, $+0.992$), and the one for 0.74 day is ($+1.000$, $+0.0019$, $-1.5010$). In the odd oxygen family, the ratio $[\mathrm{O}]$:$[\mathrm{O_3}] = 0.0019$ corresponds to the steady-state solution. It is to notice that the characteristic time for the system does not coincide with the photodissociation of ozone.

## References[1]

### *Books*

Baird C (1995) Environmental chemistry. Freeman, New York
Broecker WS, Peng T (1977) Tracers in the sea. Eldigo Press, Palisades

[1]This chapter in large part has been rewritten based mainly on the Jacob book. One of the novelties of this book is that it tries to introduce marine chemistry and its relation to atmospheric chemistry. Good introductory material could also be found either in Baird or Graedel and Crutzen. For marine chemistry, the classical book by Broecker and Peng is the main reference, with the Walker book a very useful companion. Most of the other material is in papers, and a few may look outdated, but they are very useful.

Broecker WS, Peng T (1984) The climate chemistry connection, in climate processes and climate sensitivity. In: Hansen JE, Takahashi eT (eds) AGU monograph 29:327
Graedel TE, Crutzen PJ (1993) Atmospheric change: an earth system perspective. Freeman, New York
Hannon B, Ruth M (1994) Dynamic modeling. Springer, New York
Jacob D (1999) An introduction to atmospheric chemistry. Princeton University Press, Princeton, NJ
Jacobson MZ (1999) Fundamentals of atmospheric modeling. Cambridge University Press, Cambridge/New York
Walker JCG (1991) Numerical adventures with geochemical cycles. Oxford University Press, New York

## *Articles*

Friedlander SK, Seinfled JH (1969) A dynamic model of photochemical smog. Environ Sci Technol 3:1175
Goody RM, Walker JCG (1972) Atmospheres. Prentice Hall, New York
Kasting JF, Caldeira K (1992) The life span of the biosphere revisited. Nature 360:721
Najjar RG (1992) Marine biogeochemistry. In: Trenberth KE (ed) Climate system modeling. Cambridge University Press, Cambridge/New York
Prather MJ (1994) Lifetimes and eigenstates in atmospheric chemistry. Geophys Res Lett 21:801
Prather MJ (2007) Lifetimes and time scales in atmospheric chemistry. Phil Trans R Soc A 365:1705
Prinn RG (1992) Tropospheric chemical models. In: Gille JC, Visconti G (eds) The use of EOS for studies of atmospheric physics. Elsevier, Amsterdam
Sarmiento JL (1992) Biogeochemical ocean models. In: Trenberth KE (ed) Climate system modeling. Cambridge University Press, Cambridge/New York
Seinfeld JH (1989) Urban air pollution: state of science. Science 243:745
Sillman S (1999) The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments. Atmos Environ 33:1821
Sillman S, Logan JA, Wofsy SC (1990) The sensitivity of ozone to nitrogen oxides and hydrocarbons in regional ozone episodes. J Geophys Res 95:1837
Walker JCG, Hays PB, Kasting JF (1981) A negative feedback mechanism for the long term stabilization of the earth's surface temperature. J Geophys Res 86:9776

# Chapter 17
# Dynamics of the Middle Atmosphere

We have now all the necessary tools to study the region that has generated in a sense, most of the popularity to the atmospheric sciences. In this region, we find the ozone which has given a lot of trouble at the international political level.

The middle atmosphere, until a few years ago of mere academic interest, refers to that portion of the atmosphere in the altitude range between the tropopause (10–15 km) up to the homopause (about 100 km). Actually the most interesting region of the middle atmosphere remains the stratosphere which goes up to about 50 km. The stratosphere interacts very strongly with the troposphere because from it is source of most of the energy that determines the circulation and also most of the chemical species that determine the ozone concentration.

In principle, the study of the circulation of the middle atmosphere should be simpler than for the troposphere because, as we often said, this region is roughly in radiative equilibrium so that the temperature should be very easily determined and so the circulation. Actually we will see in a while that the circulation of the middle atmosphere is in part due to the departure from radiative equilibrium.

On the other hand, it is almost impossible to study the dynamics of this region without recurring to the chemistry. This is the real difference with the troposphere, where thermodynamic exchanges of energy are as much important as absorption of solar and planetary radiation by minor gases. In any case in the troposphere, the mixing ratio of the main absorbers (carbon dioxide and water vapor) does not depend very much on the atmospheric chemical processes. The temperature of the stratosphere by contrast results from emission and absorption processes due to gases whose mixing ratio is determined by chemical processes.

Another important difference with the troposphere is that we do not have markers of the circulation like the clouds and we do not have real statistics for the circulation,

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_17) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_17

like those of Oort and Rasmussen of the old days. This is particularly true for the Eulerian circulation. The Lagrangian circulation (this is a new concept), we will show, can be obtained with enough accuracy from radiative data.

We will start then from what is known better about the middle atmosphere, that is, the temperature.

## 17.1 Thermal Structure of the Stratosphere

In Fig. 17.1, the observed temperature of the stratosphere is compared with the one resulting from the radiative equilibrium. We notice immediately that the summer hemisphere is much colder than what we would expect from radiative equilibrium and the winter hemisphere is warmer. In practice, if the region around 50 km should be in radiative equilibrium the temperature should be as low as 160 K. In the same figure, the radiative equilibrium temperature for the troposphere does not make any sense. It is clear, however, that the temperature around 50 km has a considerable seasonal variation, as it is most of the middle atmosphere. The reason for the temperature to be higher in summer is that it receives more solar radiation, which increases with latitude so that the warmest region is at the summer pole. A more precise idea of the difference between radiative equilibrium and the real case is given in Fig. 17.2. These differences have an immediate consequence if we remember the Newtonian cooling approximation. If we denote with $T_{\text{rad}}$ the radiative equilibrium temperature and with $T$ the observed temperature we have for the heating rate $Q$ (measured in degrees per unit time)

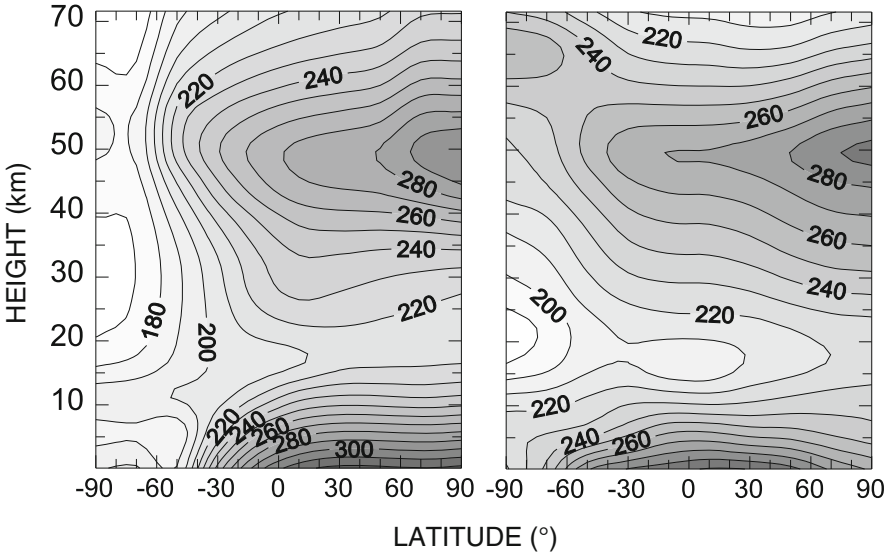



**Fig. 17.1** The radiative equilibrium temperature (*left*) and the observed one (*right*) as a function of latitude and height. The isotherms are at 10 °C intervals

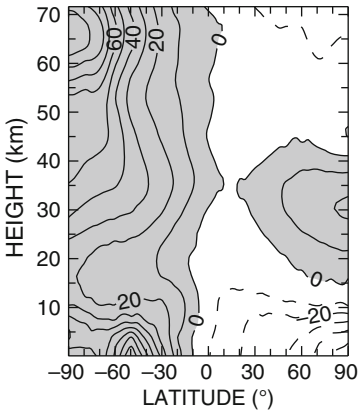



**Fig. 17.2** The difference in degrees between the observed and the calculated radiative equilibrium temperature. The data are those in Fig. 17.1

$$Q = \frac{-(T - T_{\text{rad}})}{\tau_r} \tag{17.1}$$

where $\tau_r$ is a time constant for the radiative relaxation. This equation is written in such a way that the atmosphere relaxes toward the radiative equilibrium, so that if $T < T_{\text{rad}}$ the atmosphere will heat up and vice versa. This implies that if the winter hemisphere has a tendency to radiative equilibrium, actually it will have a positive heating rate while the summer hemisphere will have a negative heating rate (cooling). This behavior has a very simple physical explanation. If the winter hemisphere is warmer than the radiative equilibrium, it will emit more infrared radiation than what it absorbs (solar plus infrared) and thus it will tend to cool. The opposite will happen for the summer hemisphere.

At this point, it is not obvious that cooling air will subside and heating air will rise as often we read in books. The vertical motion is due only to buoyancy effects that depend rather on the temperature difference between air parcels and surrounding air. The explanation must be more complex and is related the momentum budget of the atmosphere. For the time being then we would rather explore why the observed temperature is so different from the radiative equilibrium values.

Before doing that it is worth mentioning another consequence of the radiative equilibrium. Once the $T_{\text{rad}}$ values are known, the zonal wind can be calculated using the thermal wind equation, and the results are shown in Fig. 17.3. We can see that these winds are completely unreasonable. These high winds are not reduced even when we use the gradient wind equation (that is, we take into account the centrifugal force). The reason why, in both figures, the values below 20 km are not shown is because there the radiative equilibrium, as we said before, does not make any sense. The interesting thing is that in the middle atmosphere the winds have a pronounced seasonal variation, with those in the summer hemisphere blowing in the opposite direction to the winds in the winter hemisphere. This is a consequence of the temperature gradient that in summer is directed mainly from the equator to the pole.

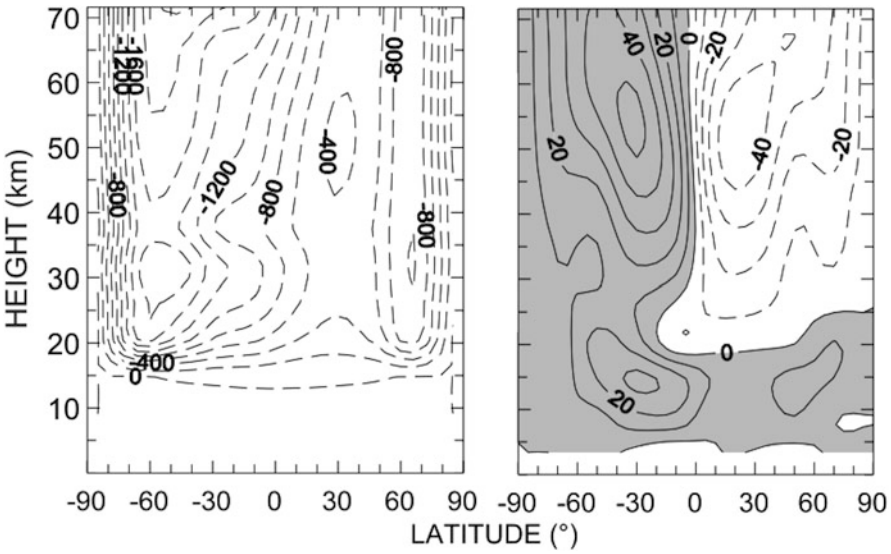



**Fig. 17.3** Zonal wind velocities for an atmosphere in radiative equilibrium (*left*) and the real atmosphere (*right*). The westerly winds are *shaded* and the seasons are the same as in Fig. 17.1

Now we must recur to the dynamics to understand why the summer hemisphere is colder than the radiative equilibrium and the winter hemisphere is warmer. This may imply some new concepts about the dynamics and the abstract Eulerian and Lagrangian definitions will be much more clear (we hope).

## 17.2 The Eulerian Mean Circulation

This is a good occasion to go over what we have learned about dynamics in the previous chapters. We have introduced, among other things, a coordinate system called logpressure that in the vertical had the coordinate

$$Z = -H \ln\left(\frac{p}{p_0}\right)$$

where $H$ is the scale height. We can rewrite the equations of motion, hydrostatic equilibrium, continuity, and thermodynamics

$$\frac{Du}{Dt} + fv + \frac{\partial \Phi}{\partial x} = X \tag{17.2a}$$

$$\frac{Dv}{Dt} + fu + \frac{\partial \Phi}{\partial y} = Y \tag{17.2b}$$

$$\frac{\partial \Phi}{\partial Z} = \frac{RT}{H} \tag{17.2c}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{\rho_0}\frac{\partial (\rho_0 w)}{\partial Z} = 0 \tag{17.2d}$$

$$\frac{DT}{Dt} + \frac{N^2 H}{R} w = \frac{J}{C_p} \tag{17.2e}$$

In particular for the thermodynamic equation, we can refer to Chap. 11. In this case, $J$ is the power absorbed per unit mass and $Y$ forces of every nature other than pressure. We remember that $N^2$ is the Brunt–Vaisala frequency and the total derivative is given by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial Z}$$

while the density is given by

$$\rho_0 = \rho_s \exp\left(\frac{-Z}{H}\right)$$

At this point for each variable, we introduce the zonal mean so that we can write $u = \overline{u} + u'$ is zero. Substituting in Eq. (17.2), we obtain

$$\frac{\partial \overline{u}}{\partial t} - f\overline{v} - \overline{X} = -\frac{\partial \left(\overline{u'v'}\right)}{\partial y} \tag{17.3a}$$

$$\frac{\partial \overline{v}}{\partial y} + \frac{1}{\rho_0}\frac{\partial (\rho_0 \overline{w})}{\partial Z} = 0 \tag{17.3b}$$

$$\frac{\partial \overline{T}}{\partial t} + \frac{N^2 H}{R}\overline{w} = \frac{J}{C_p} - \frac{\partial \left(\overline{v'T'}\right)}{\partial y} \tag{17.3c}$$

In Chap. 11, we have obtained similar equations and the approximation is that the vertical motions can be neglected with respect to the horizontal, and among these the zonal motion is prevalent. The first part of Eq. (17.3) in particular has a very interesting interpretation. First of all, there is no longer the forcing term due to the geopotential because the zonal mean of the gradient is zero. This implies that the zonal mean motion can only be accelerated by the Eliassen and Palm flux divergence. The third equation shows that the temperature change is due not only to the local heating but also to the advection. A very simple consequence of these equations is that, if the flux divergence is negative (positive), there is a positive acceleration of the zonal flux. For continuity, the meridional velocity must increase (decrease) and consequently there will be a change in the vertical velocity.

In the same manner we notice that from the last equation, in the presence of a heating flux (but in absence of an eddy flux), the vertical velocity must increase and for continuity the meridional velocity will change. Similar equations have been developed in Chap. 10 to explain qualitatively the atmospheric circulation. As for the zonal mean value for $u$, it can be shown that it satisfies the geostrophic equilibrium

$$f\overline{u} = -\frac{\partial\overline{\Phi}}{\partial y} \tag{17.4}$$

If this is used in Eq. (17.2c), we have the thermal wind equation

$$f\frac{\partial\overline{u}}{\partial Z} + RH^{-1}\frac{\partial\overline{T}}{\partial y} = 0 \tag{17.5}$$

The thermal wind relation is actually a constraint on the meridional circulation that tends to change independently because of the divergence of the eddy, heat, or momentum flux. Actually, if for any reason the geostrophic equilibrium of the zonal wind is changed, the pressure gradient forces produce a meridional circulation that maintains the thermal wind equilibrium. This mechanism is so efficient that the zonal wind changes very little even in the presence of large changes of the heat, momentum, and eddy flux. Further insight into the role of the meridional circulation can be found by introducing a coordinate transformation.

### *17.2.1 The Transformed Eulerian Mean*

In Eq. (17.3) but also in the experimental data we notice that the change in mean quantities $(\overline{u}, \overline{T})$ results from small differences between mean quantities and eddies. This suggests that an adequate transformation could isolate the smallest terms. Andrews and McIntyre suggested such a transformation in the form

$$v^* = \overline{v} - \frac{R}{H\rho_0}\frac{\partial}{\partial Z}\left(\frac{\rho_0\overline{v'T'}}{N^2}\right) \tag{17.6a}$$

$$w^* = \overline{w} + \frac{R}{H}\frac{\partial}{\partial y}\left(\frac{\overline{v'T'}}{N^2}\right) \tag{17.6b}$$

These are called residual velocities and the transformation corresponds to introducing a streamfunction for the residual velocities given by

$$\overline{X}^* = \overline{X} + \rho_0\frac{R}{H}\frac{\overline{v'T'}}{N^2} \tag{17.7}$$

where $X$ is such that

$$\rho_0 \overline{v}^* = \frac{-\partial \overline{X}}{\partial z} \quad \rho_0 \overline{w}^* = \frac{\partial \overline{X}}{\partial y} \tag{17.8}$$

If the transformation of Eq. (17.6) is substituted in Eq. (17.3) to eliminate $\overline{v}, \overline{w}$, we have

$$\frac{\partial \overline{u}}{\partial t} - f\overline{v}^* = \frac{1}{\rho_0} \nabla \cdot \mathbf{E} + \overline{X} = \overline{G} \tag{17.9a}$$

$$\frac{\partial \overline{T}}{\partial t} + \frac{N^2 H}{R} \overline{w}^* = \frac{J}{C_p} \tag{17.9b}$$

$$\frac{\partial \overline{v}^*}{\partial y} + \frac{1}{\rho_0} \frac{\partial \left(\rho_0 \overline{w}^*\right)}{\partial Z} = 0 \tag{17.9c}$$

where

$$\mathbf{E} = \mathbf{j} E_y + \mathbf{k} E_z$$

is the Eliassen and Palm flux that we have found in Chap. 10 and that we define here in a slightly different way

$$E_y = -\rho_0 \overline{u'v'} \quad E_z = \rho_0 f R \overline{v'T'} / \left(N^2 H\right) \tag{17.10}$$

The residual circulation in a sense unifies the role of the eddy momentum and heat fluxes because the zonal mean circulation is forced by the divergence of the Eliassen and Palm flux.

Of particular interest are Eqs. (17.9a, 17.9b) because at the steady state Eq. (17.9a) tells us that the circulation is determined by the convergence of eddy momentum and heat fluxes. Equation (17.9b) shows that in stationary conditions the vertical velocity is determined totally from the diabatic heating. In principle knowing the net heating of the atmosphere, it is possible to obtain the value of the vertical velocity and through the continuity equation to obtain also the meridional velocity. This circulation is known as *diabatic* and is actually a first approximation to the residual circulation.

In Fig. 17.4, the diabatic circulations for the equinox and solstice are shown. We see that for the equinox we have roughly a single cell that rises from the summer hemisphere and subsides at the winter hemisphere. At the equinox, the warmest region is at the equator so that we have two cells that originate there and subside at the poles. This residual circulation disappears when the divergence of the Eliassen and Palm flux is zero, which means that the waves are linear, steady, and conservative. In this case, the net heating is zero so that the atmosphere is in radiative equilibrium. The essential condition for the existence of the residual circulation is that the atmosphere must not be in radiative equilibrium.

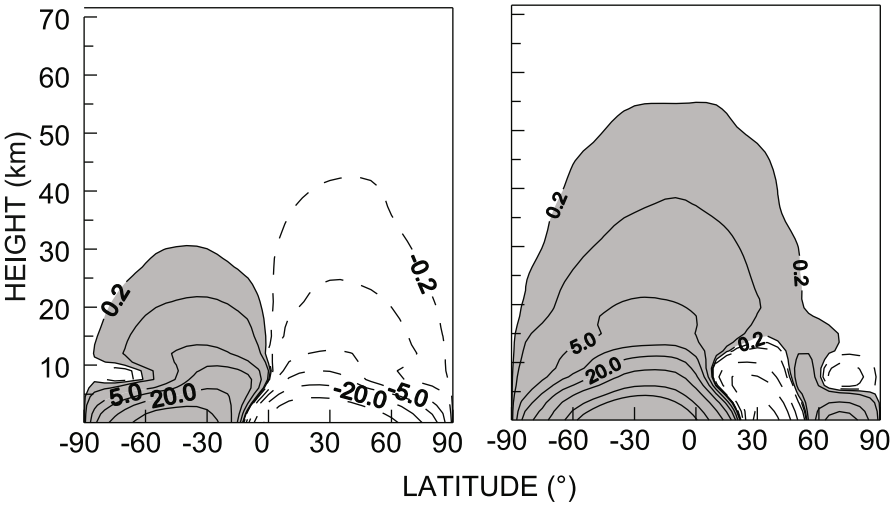



**Fig. 17.4** The residual circulation for the equinox (*left*) and solstice (*right*). Shaded areas indicate circulation toward the southern hemisphere. The mass streamlines refer to a mass flux in Mt $s^{-1}$ of −0, −20, −10, −5, −2, −0,2 e 0,2, 2, 5, 10, 20, 40, 80, 120

### *17.2.2 An Attempt to Understand the Origin of the Residual Circulation*

We must ask now what keeps the atmosphere away from radiative equilibrium. We can start from the steady state solutions of Eq. (17.9)

$$-f\overline{v}^* = \overline{G} \tag{17.11a}$$

$$\frac{N^2 H\overline{w}^*}{R} = \frac{J}{C_p} \tag{17.11b}$$

Substituting the residual velocities obtained in this way we obtain

$$-\frac{\partial \overline{G}}{\partial y} + \frac{fR}{\rho_0 C_p H}\frac{\partial}{\partial Z}\left(\rho_0 \frac{J}{N^2}\right) \tag{17.12}$$

This equation again shows the relation between the forcing term due to the eddies and the diabatic heating. In case the net heating can be approximated by a Newtonian cooling, we have

$$\frac{J}{C_p} = -\alpha\left(\overline{T} - T_{\text{rad}}\right) \tag{17.13}$$

A first interesting conclusion can be obtained substituting this equation in Eq. (17.9), which can be used to eliminate $\overline{u}, \overline{v}^*, \overline{w}^*$ using the thermal wind equation

$$\left[\frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial y}\left(\frac{f^2}{\rho_0}\frac{\partial}{\partial Z}\frac{\rho_0}{N^2}\right)\right]\frac{\partial \overline{T}}{\partial t} + \frac{\partial}{\partial Z}\left\{\frac{f^2}{\rho_0}\frac{\partial}{\partial Z}\left[\alpha\rho_0\frac{\left(\overline{T} - T_{\text{rad}}\right)}{N^2}\right] - \frac{fH}{R}\frac{\partial G}{\partial y}\right\} = 0 \tag{17.14}$$

We can see from Eq. (17.14) that the rate of change of the zonal temperature depends on the relaxation of the temperature deviation from the radiative equilibrium value. This deviation is "diffused" because of the elliptic term that appears in the left-hand side of the same equation.

If we consider time scales of the order of the season, the temperature difference with respect to radiative equilibrium is small and then Eq. (17.12) may be considered valid so that

$$\frac{fR}{\rho_0 H}\frac{\partial}{\partial Z}\left(\rho_0\frac{\alpha\delta\overline{T}}{N^2}\right) \cong \frac{\partial \overline{G}}{\partial y} \tag{17.15}$$

where $\delta\overline{T} = \left(\overline{T} - T_{\text{rad}}\right)$ is the difference between the zonal mean and the radiative equilibrium temperature. The almost equal sign is due to the fact that the continuity equation is satisfied only for seasonal averages. The temperature difference $\delta\overline{T}$ can be directly related to the horizontal term of the eddy divergence. With $\tau_r = 1/\alpha$, we have

$$\delta\overline{T} \cong \frac{\tau_r H^2}{fR}\frac{\partial}{\partial Z}\left(N^2\right)\frac{\partial \overline{G}}{\partial y} \tag{17.16}$$

This shows that the atmosphere is kept away from radiative equilibrium by the convergence of the eddy term. In practice, this means that the energy absorption that follows the breakdown of planetary waves is actually the term that heats up or changes the temperature of the atmosphere. This argument may be convincing for the winter pole but is reasonable to ask how the eddies can cool the summer stratosphere. In this case the explanation must be a little more complex and this is something we will see in the next section studying a very old friend of the stratospheric dynamicist: the sudden stratospheric warming.

### *17.2.3 The Sudden Stratospheric Warming*

A very interesting application of what we have learned so far are the sudden stratospheric warmings (SSW or stratwarms). From Fig. 17.3b, we can see that the winter stratosphere around 20 km altitude and 60° north latitude shows a westerly wind that increases rapidly with height. On the other hand, the temperature decreases toward

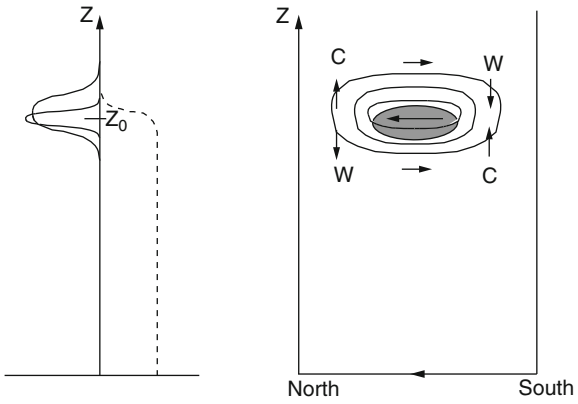



**Fig. 17.5** The interaction between wave and zonal flow in the case of a stratwarm. At *left*, the *dashed line* shows the EP flux that propagates upward and is absorbed at $Z_0$. The flux divergence at *left* is represented by the narrow Gaussian and at *right* by the *shaded area*. The zonal flow under this forcing is slowed down, and the deceleration isopleths are represented at right like ellipses and at *left* by the wider gaussian. The *arrows* indicate the induced circulation with the warm (W) and cold (C) regions (Adapted from Andrews et al. 1987)

the pole after reaching a maximum around 45° latitude. Sporadically in the winter it happens that such a situation changes completely in 1–2 weeks so that the wind changes direction and the temperature increases up to 40 C. Other observations confirm that responsible for this heating is the upward propagation of planetary waves, especially the long ones (wave number 1 and 2). These waves produce an EP flux that drives the mean circulation in the stratosphere.

To understand a little better, we refer to Fig. 17.5, where we assume an EP flux that propagates upward until it is absorbed at level $Z_0$. Around this altitude, the divergence of the EP flux is negative so that based on Eq. (17.9a) the zonal flow is slowed down. In the same way as we have seen for the temperature, the atmosphere response is not local so that the negative acceleration produces effects on a region that is wider than the narrowest Gaussian shown in the figure. Because of this slow down of the zonal flux, the Coriolis acceleration also decreases so that a residual circulation is established made up of a double cell in the meridional plane. In the region where the motion is downward, there is compression and the atmosphere heats up, while in the region where the motion is upward there is expansion and then cooling. Exactly as in the case of the temperature perturbation, the flux divergence acts as a heat source for the region. In quantitative terms, the diffuse response can be understood by looking at Eq. (17.9) together with the thermal wind equation. Following exactly the same procedure, we have used to obtain Eq. (17.14) for the temperature, we may obtain an equation for the deceleration of the zonal wind

$$\rho_0 \left[ \frac{\partial^2}{\partial y^2} + \frac{1}{\rho_0} \frac{\partial}{\partial z} \left( \rho_0 \frac{f^2}{N^2} \frac{\partial}{\partial z} \right) \right] \overline{u}_t = (\nabla \cdot \mathbf{E})_{yy} - \left( \rho_0 \frac{Rf}{N^2 H} \frac{J}{C_p} \right)_{yz} \tag{17.17}$$

where the subscripts indicate the partial derivatives. We see that in this case the deceleration does not correspond exactly to the forcing but again the presence of an elliptic operator produces a diffuse response. We must also notice the presence of the radiative term. In a similar way, we obtain the meridional velocity

$$\rho_0\left[\frac{\partial^2}{\partial y^2}+\frac{1}{\rho_0}\frac{\partial}{\partial z}\left(\rho_0\varepsilon\frac{\partial}{\partial z}\right)\right]f\overline{v}^*=-\left[\rho_0\varepsilon\left(\rho_0{}^{-1}\nabla\cdot\mathbf{E}\right)_z\right]_z-\left(\rho_0\frac{R}{fH}\varepsilon\frac{J}{C_p}\right)_{yz} \tag{17.18}$$

where $\varepsilon=(f/N)^2$. In both cases, the forcing deriving from the divergence of the EP flux or the radiation corresponds to an operator for the zonal acceleration or for the residual meridional velocity that, because of the presence of the height derivative of the density, is not symmetrical with respect to height and latitude. This may explain why in Fig. 17.5, the isopleths for the acceleration are moved upward.

We have examined another way to deal with the interaction between zonal flux and eddies in Chap. 10. In that case we found the way to obtain the meridional flux of the eddy potential vorticity using Eq. (10.67)

$$\overline{q'v'}=-\frac{1}{2}\frac{\partial\left(\overline{q'^2}\right)/\partial t}{\partial\overline{q}/\partial y}$$

If the gradient of mean zonal vorticity is positive then for a growing wave, the eddy vorticity will increase with time so that the meridional flux will be negative, that is, directed toward low latitudes. On the other hand, the rate of change of the mean zonal vorticity is related to the flux by (see Eq. (10.66))

$$\frac{\partial\overline{q}}{\partial t}=\frac{-\partial\left(\overline{q'v'}\right)}{\partial y}$$

so that because the flux is zero at the pole, the term on the right is positive (negative flux gradient) and the mean zonal vorticity decreases with time and so the zonal velocity.

Through the eddy terms, the dynamics can change the mean circulation of the atmosphere and keep it away from radiative equilibrium. In particular in the case of sudden warming, the circulation is disrupted by the propagation of energy from the troposphere. The break down occurs mainly for the zonal current which in the winter is found in the low polar stratosphere between 15 and 20 km. Wave activity is much more pronounced in the northern hemisphere with respect to the southern hemisphere because of the more complex orography. This means that the polar vortex will break more often around the north than the south pole. This implies that the Antarctic polar vortex is more stable and will break only at the onset of spring.

This pronounced asymmetry between the two poles has a very important role in explaining the ozone depletion in Antarctica, as we will see in the next chapter.

## 17.3 Tracers Transport in the Stratosphere

The introduction of the residual circulation, in principle, has greatly simplified the treatment of how a tracer (that is an inert chemical compound) is transported in the stratosphere. In the years following this minor revolution, there was the illusion that finally we could get rid of the eddy terms. Actually the residual circulation would have absorbed such terms. However, the first attempts to directly use the residual circulation made clear that there was still a long way to go before the closure of the problem could be achieved.

To better understand this point we can refer to Fig. 17.6, where the main processes that determine the transport in the stratosphere are shown. The residual circulation is represented here by the thick lines and may be responsible for the rising of air at the tropics and its subsidence at the poles. This implies that the lines of constant mixing ratio are forced to slope toward the pole causing a horizontal gradient in the mixing ratio. This gradient is smoothed out by the large-scale eddy mixing processes.

We consider the continuity equation for a species with a mass mixing ratio $\chi$

$$\frac{D\chi}{Dt} = S; \quad \frac{\partial \chi}{\partial t} + u\frac{\partial \chi}{\partial x} + v\frac{\partial \chi}{\partial y} + w\frac{\partial \chi}{\partial z} = S \tag{17.19}$$

where $S$ is the net chemical source. Again the quantities appearing in this equation can be written as the sum of the zonal mean and a small deviation from it

$$\chi = \overline{\chi} + \chi', \quad u = \overline{u} + u', \quad v = \overline{v} + v', \quad w = \overline{w} + w', \quad S = \overline{S} + S'$$

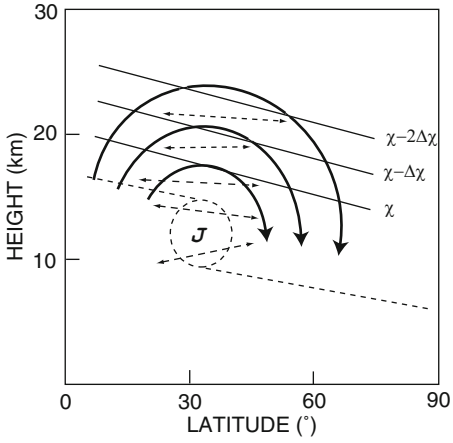



**Fig. 17.6** A simple scheme of the processes responsible for the transport in stratosphere. The *thin solid lines* are isopleths for the mixing ratio $\chi$. The *heavy arrows* indicate the residual circulation around the jetstream core J. The large scale mixing processes are represented by the *dashed lines* (Adapted from Holton 1992b)

If we use log pressure as the vertical coordinate system, we have $W = dZ/dt$ so that on the sphere

$$\frac{\partial \overline{\chi}}{\partial t} + \frac{\partial \left(\overline{\chi}\,\overline{v}\cos\phi\right)}{\partial y} + \frac{1}{\rho_0}\frac{\partial \left(\rho_0 \overline{\chi}\,\overline{W}\right)}{\partial Z} + \frac{\partial \left(\overline{v'\chi'}\cos\phi\right)}{\partial y} + \frac{1}{\rho_0}\frac{\partial \left(\rho_0 \overline{\chi' W'}\right)}{\partial Z} = S \tag{17.20}$$

where $\phi$ is the latitude and $y = a\,\sin\phi$. At this point, we can introduce the residual circulation defined by Eq. (17.6) keeping in mind the spherical geometry

$$\overline{v}^* = \overline{v} - \frac{1}{\rho_0}\frac{\partial}{\partial Z}\left(\rho_0 \frac{\overline{v'T'}}{\Gamma}\right), \quad \overline{W}^* = \overline{W} + \frac{1}{a\cos\phi}\frac{\partial}{\partial \phi}\left(\cos\phi \frac{\overline{v'T'}}{\Gamma}\right)$$

where $a$ is the radius of the Earth and $\Gamma = N^2 H/R$ is the static stability. With these substitutions Eq. (17.19) becomes

$$\frac{\partial \overline{\chi}}{\partial t} + \frac{1}{a\cos\phi}\frac{\partial}{\partial \phi}\left(\overline{v}^* \overline{\chi}\cos\phi\right) + \frac{1}{\rho_0}\frac{\partial}{\partial Z}\left(\rho_0 \overline{W}^* \overline{\chi}\right) = -\nabla \cdot \mathbf{F}^* + \overline{S} \tag{17.21}$$

where the components of the flux $\mathbf{F}^*$ are

$$F_\phi^* = \frac{\overline{v'\chi'}}{a} - \frac{\overline{\psi}}{a}\frac{\partial \overline{\chi}}{\partial Z}, \quad F_Z^* = \overline{W'\chi'} + \frac{\overline{\psi}}{a}\frac{\partial \overline{\chi}}{\partial \phi}, \quad \overline{\psi} = \frac{\overline{v'T'}}{\Gamma} \tag{17.22}$$

We can see very easily that we are again confronting a closure problem with the changes of the average mixing ratio related to eddy quantities that in this case are deviations from the zonal mean. This problem arises because we are interested in average quantities considering that in a general circulation model, the model itself produces the eddy terms.

Equation (17.20) is a different way of writing the continuity equation, and it is called *flux form* because just those terms are put in evidence. Another interesting characteristic is that the eddy flux does contain not only the correlation between velocity and mixing ratio but also the one between velocity and temperature.

Expressing the eddy terms as a function of the average quantities again solves this closure problem. This opens for the stratosphere one of those constructs of which nobody knows the origin but which everybody uses: the diffusion coefficients.

### *17.3.1 The Two-Dimensional Diffusion Coefficients*

In Chap. 10, we mentioned that the gravity wave breakdown may be accompanied by the generation of turbulence whose effects can be parameterized in terms of a diffusion coefficient (see Eq. (10.103)). For simplicity, we will write the continuity

equation in a Cartesian coordinate system that uses $Z$ as the vertical coordinate. In the flux form, our equation becomes

$$\frac{\partial \chi}{\partial t} + \frac{\partial}{\partial x}(\chi u) + \frac{\partial}{\partial y}(\chi v) + \frac{1}{\rho_0}\frac{\partial}{\partial Z}(\rho_0 \chi W) = S \tag{17.23}$$

At this point, we introduce a new average that we will call isobaric, defined for a generic quantity $\Psi$ as

$$\langle \Psi(Z,t) \rangle = \int_{\text{globo}} \Psi(x,y,Z,t)\, dx\, dy \tag{17.24}$$

that when applied to Eq. (17.23) gives

$$\frac{\partial \langle \chi \rangle}{\partial t} + \frac{1}{\rho_0}\frac{\partial}{\partial Z}(\rho_0 \langle \chi W \rangle) = \langle S \rangle \tag{17.25}$$

The quantity $\langle \chi W \rangle$ is the globally averaged vertical flux, and Eq. (17.25) has a closure problem itself because it relates an average mixing ratio to an average flux. Following a linear theory, the flux can be written in the form

$$\langle \chi W \rangle = -K(z)\frac{\partial \langle \chi \rangle}{\partial Z} \tag{17.26}$$

In this case, the average mixing ratio is determined by a simple diffusion equation

$$\frac{\partial \langle \chi \rangle}{\partial t} = \frac{1}{\rho_0}\frac{\partial}{\partial Z}\left[\rho_0 K(z)\frac{\partial \langle \chi \rangle}{\partial Z}\right] + \langle S \rangle \tag{17.27}$$

Assuming that $K(z) = K_{zz}$ is determined only by the gravity waves breakdown, an expression similar to Eq. (10.103) can be adopted for it. This very simple approach was used in the old days simply to evaluate the effects of the very complex system of reactions on the source term and to calculate the vertical distribution. Most of the time, the real exercise was to find an appropriate form for $K(z)$ in order to explain the very sparse data on the distribution.

The same problem can be treated in a two-dimensional space (latitude–altitude) using Eq. (17.21), when again the relation between eddy flux and mean concentration is known. When the eddies are attributed to steady or dissipative waves, the first step is to consider the passage of a stationary wave on a parcel located at the initial position $x_0$, $y_0$, as shown in Fig. 17.7. If the wave is steady then there is no net transport and the parcel undergoes an oscillatory motion in the two directions so that its trajectory is elliptical. Amplitude and tilting of the ellipse depend on the amplitude and phase of the motion in height and latitude. If we are dealing with a tracer, we will not observe changes in the chemical composition. On the other hand, if we consider a highly reacting species its concentration will adapt very fast to that

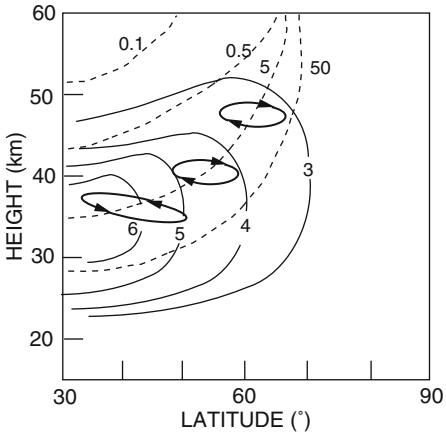



**Fig. 17.7** The trajectories in the $z$, $y$ plane (*ellipses*) of an air parcel at the passage of a steady wave. The *thin solid lines* represent qualitatively the isopleths of ozone mixing ratio (ppm) and *dashed lines* the chemical lifetime in days (From Garcia and Solomon 1983)

of the other species along the trajectory. The most interesting case is that of a species with a lifetime comparable with the dynamic, time scale, that is the time it takes for the particle to go around the elliptical trajectory.

This time is of the order of a few days so that at the end, the concentration will be an average along the trajectory. This kind of average is known as Lagrangian and will be extremely important for what we are going to say in a while. The displacements along the $Z(\zeta)$ and $y(\eta)$ coordinates can be linearly related to the velocity by

$$v' = \left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\eta' \quad W' = \left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\zeta' \tag{17.28}$$

The displacements are linearly related to the deviations of the mixing ratio $\chi'$

$$\chi' = -\frac{\eta'}{a}\frac{\partial\overline{\chi}}{\partial\phi} - \zeta'\frac{\partial\overline{\chi}}{\partial Z} \tag{17.29}$$

so that the flux components of Eq. (17.22) can be written as

$$\begin{aligned} F^*_\phi &= -\frac{\overline{\eta' v'}}{a^2}\frac{\partial\overline{\chi}}{\partial\phi} - \left(\frac{\overline{v'\zeta'}}{a} + \frac{\overline{\psi}}{a}\right)\frac{\partial\overline{\chi}}{\partial Z} \\ F^*_Z &= -\left(\frac{\overline{\eta' W'}}{a} - \frac{\overline{\psi}}{a}\right)\frac{\partial\overline{\chi}}{\partial\phi} - \overline{W'\xi'}\frac{\partial\overline{\chi}}{\partial Z} \end{aligned} \tag{17.30}$$

and in more compact form can be written as

$$\mathbf{F}^* = -\left(\mathbf{K_s} + \mathbf{K_a}\right)\nabla\overline{\chi} \tag{17.31}$$

where $\mathbf{K_s}$ and $\mathbf{K_a}$ indicate a symmetric and anti-symmetric matrices whose elements are given by

$$\mathbf{K_s} = \begin{pmatrix} K_{yy}/a^2 & K_{yz}/a \\ K_{yz}/a & K_{zz} \end{pmatrix} \quad \mathbf{K_a} = \begin{pmatrix} 0 & -B/a \\ B/a & 0 \end{pmatrix} \tag{17.32}$$

Comparing Eqs. (17.31) and (17.30), we obtain

$$K_{yy} = \overline{\eta' v'} \quad K_{yz} = \frac{\left(\overline{v'\zeta'} + \overline{\eta' W'}\right)}{2} \quad K_{zz} = \overline{W'\xi'} \quad B = \overline{\psi} + \frac{\left(\overline{v'\zeta'} - \overline{\eta' W'}\right)}{2} \tag{17.33}$$

The tensor $\mathbf{K_s}$ represents the effects of dispersion due to the correlation between velocities and displacements. The interpretation of the anti-symmetric part is more complex but it can be shown that it is related to the meridional advection.

This quite elegant theory has taken us back to the starting point. In the old days (in the 1970s), people simply used the symmetric tensor, tuning its value to fit the observed species distributions, although the data were really limited. At this point, it should be necessary to know the displacements and the eddy velocities to calculate the coefficients. Actually some attempt has been made in this direction using either observations or data produced by models, but the scarcity of the former and the imperfections of the latter have frustrated all the attempts.

In any case, a basic difficulty with this approach is really the residual circulation. If we should use Eq. (17.6) again, we are going nowhere because we still need to know the eddy terms. This difficulty can be apparently resolved because it is possible to show that the residual circulation can be approximated by the *diabatic* circulation obtained from Eq. (17.9b) by putting $\partial T/\partial t = 0$ so that we have for the vertical velocity

$$\overline{W}^* = \frac{J}{\Gamma C_p}$$

The meridional component is obtained through the continuity equation. Again we are fooling around because the net heating depends on the temperature which cannot be determined without taking into account the dynamics, as clearly shown by Eq. (17.14). Thus if we want to insist on the two-dimensional models we have to make a drastic decision and follow the hard road. We will follow the approach *R*. Garcia has invented that brings us back to gravity and planetary waves.

### *17.3.2 Self Consistent Transport in Two Dimensions*

As we have seen before, dynamics affects directly the temperature which feeds back into the dynamics determining the diabatic circulation. To fix the problem,

we start from the Eq. (17.7) referring to the streamfunction for the Transformed Eulerian Mean (TEM) circulation that can be used to reduce Eq. (17.9) to one elliptic equation

$$\frac{\partial^2 \overline{X}^*}{\partial y^2} + \rho_0 \left(\frac{f}{N}\right)^2 \frac{\partial}{\partial Z}\left(\frac{1}{\rho_0}\frac{\partial \overline{X}^*}{\partial Z}\right) = \frac{\rho_0}{N^2}\left[\frac{\partial}{\partial y}\frac{R}{C_p}\left(\frac{J}{H}\right) + f\frac{\partial \overline{G}}{\partial Z}\right] \tag{17.34}$$

In this equation, the radiative term ($J$) and the dynamic forcing term ($\overline{G}$) are independent so that only explicating those terms we may determine the streamfunction.

The method used by Garcia starts from the equation for the zonal wind

$$\overline{u}_t - \overline{v}^*\left(f - \overline{u}_y\right) + \overline{w}^*\overline{u}_z = -K_R\overline{u} + \left(\frac{1}{\rho_0}\right)\nabla\cdot\mathbf{E} \tag{17.35}$$

and for the average temperature

$$\overline{T}_t + \overline{v}^*\overline{T}_y + \overline{w}^*\left(\overline{T}_z + \frac{HN^2}{R}\right) = Q_s - B - \alpha\overline{T} \tag{17.36}$$

with the subscripts indicating the partial derivatives. There are important differences with respect to what we have discussed before. In Eq. (17.34), there is a term for the momentum advection $-\overline{v}^*\overline{u}_y$ and a term for the viscous stress, known as Rayleigh friction, $-K_R\overline{u}$ due mainly to the turbulence introduced by the gravity wave breakdown. Actually the Eliassen and Palm flux already contains terms due to the gravity waves that are of the form $\rho_0\overline{u'w'}$. In (17.36) the main novelty is the presence of the temperature $T(z, \varphi)$ that actually is the deviation with respect to a globally averaged distribution $T_0(z)$. The net radiative heating is now decomposed in a solar heating term $Q_s$, and a Newtonian cooling term linear with the temperature. We introduce now a streamfunction $\overline{X}^*$ defined as

$$\overline{v}^* = -\frac{1}{\cos\phi}\left(\frac{\partial \overline{X}^*}{\partial Z} - \frac{\overline{X}^*}{H}\right), \quad \overline{w}^* = \frac{1}{\cos\phi}\frac{\partial \overline{X}^*}{\partial y} \tag{17.37}$$

so that we arrive at an elliptic equation like Eq. (17.34)

$$C_{zz}\overline{X}^*_{zz} + C_z\overline{X}^*_z + C_{zy}\overline{X}^*_{zy} + C_y\overline{X}^*_y + C_{yy}\overline{X}^*_{yy} = C_F\cos\phi \tag{17.38}$$

where the coefficients at left are a function only of mean quantities while the forcing term at the right-hand side is given by

$$C_F = f\frac{\partial}{\partial z}\left(-K_R\overline{u} + \frac{1}{\rho_0}\nabla\cdot\mathbf{E}\right) + \frac{R}{H}\frac{\partial}{\partial y}\left(-B - \alpha\overline{T}\right) \tag{17.39}$$

In the divergence of the EP flux, there are terms like $\overline{u'v'}$ or $\overline{u'w'}$ due to the planetary and gravity waves respectively. In Chapter 10, we have found for the gravity waves

$$\overline{u'w'} = \frac{-k(\overline{u}-c)^3}{2N}$$

that is equivalent to a diffusion coefficient

$$K_{zz} = \frac{k(\overline{u}-c)^4}{2HN^3}$$

For planetary waves, things are a little more complicated. First of all, it is simpler to express directly the divergence of the EP flux like in (10.60)

$$\nabla \cdot \mathbf{E_p} = \overline{q'v'}$$

instead of correlating the zonal and meridional velocity. As a matter of fact, the potential vorticity can be obtained as a solution of Eq. (10.55)

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right) q' + v'\overline{q}_y = D$$

where $D$ is some dissipative process. Actually to determine the potential vorticity we must determine the streamfunction and then the velocity. An example may clarify everything. If the dissipative term is taken as a time constant $1/\delta$ then the equation of the potential vorticity becomes

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right) q' + v'\overline{q}_y = -\delta q'$$

and assuming for $q'$ a behavior like $\exp[ik(x-ct)]$ we obtain

$$ik(\overline{u}-c)\,q' + v'\overline{q}_y = -\delta q' \tag{17.40}$$

from which it is possible to obtain $q'$ so that

$$q'v' = -\frac{v'v'}{[\delta + ik(\overline{u}-c)]}\overline{q}_y$$

and averaging and taking the real part

$$\overline{q'v'} = -\frac{\overline{v'v'}\delta}{\left[\delta^2 + k^2(\overline{u}-c)^2\right]}\overline{q}_y \tag{17.41}$$

If we assume a diffusive flux of the form $\overline{q'v'} = -K_{yy}\overline{q}_y$, we can write the horizontal diffusion coefficient as

$$K_{yy} = \frac{\overline{v'v'}\delta}{\left[\delta^2 + k^2(\overline{u} - c)^2\right]} \tag{17.42}$$

With this approach we need to solve both the elliptic equation and the potential vorticity equation specifying the horizontal wave number. The solution for the potential vorticity gives also the phase and thus the phase velocity and the other two wave numbers $l$ and $m$. The meridional eddy velocity can also be found and through the equation of continuity $K_{zz}$ can also be determined.

The coefficients determined in this way can then be used in the continuity equation for a chemical species with mixing ratio $\chi$

$$\frac{\partial\overline{\chi}}{\partial t} + \overline{v}^*\frac{\partial\overline{\chi}}{\partial y} + \overline{w}^*\frac{\partial\overline{\chi}}{\partial Z} = \overline{S} + D_\chi \tag{17.43}$$

In this $D_\chi$ is the divergence of the eddy flux with components

$$\overline{\chi'w'} = -K_{zz}\overline{\chi}_z \quad \overline{\chi'v'} = -K_{yy}\overline{\chi}_y.$$

We can also write an equation for the deviation from the zonal mean

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\chi' + v'\overline{\chi}_y + w'\overline{\chi}_z = -\frac{\chi'}{\tau} \tag{17.44}$$

where $\tau$ is the chemical lifetime. Solving Eq. (17.44) as we did for Eq. (17.40), we can obtain the horizontal diffusion coefficient for each species

$$K^c_{yy} = \frac{\overline{v'v'}/\tau}{\left[\tau^{-2} + k^2(\overline{u} - c)^2\right]} \tag{17.45}$$

This confirms that for those species that have a very short lifetime (and can be considered at photochemical equilibrium), the diffusion coefficient is small and the diffusion is negligible.

### *17.3.3 Eddies and the Troposphere–Stratosphere Flux*

If we know the residual circulation $\left(\overline{v}^*, \overline{w}^*\right)$ then the vertical mass flux from some latitude $\phi_0$ to the pole is given by

$$F_m = 2\pi a^2 \int_{\phi_0}^{\pi/2} \rho_0 \overline{w}^* \cos\phi d\phi \tag{17.46}$$

Expressing the velocities through the mass streamfunction, we get

$$a^{-1}\frac{\partial X}{\partial \phi} = \rho_0 \overline{w}^* \cos\phi \quad \frac{\partial X}{\partial Z} = \rho_0 \overline{v}^* \cos\phi$$

So that Eq. (17.46) becomes, with the condition that $X(\pi/2) = 0$

$$F_m = \pm 2\pi a X(\phi_0) \tag{17.47}$$

where the positive sign is for the northern hemisphere and vice versa. We need to remember that Eq. (17.47) expresses the total flux poleward of some latitude, so that the total extratropical flux will be obtained by finding the latitude for which Eq. (17.47) is maximum. Indicating with $\phi_N$ and $\phi_S$ these two latitudes

$$F_m(N) = +2\pi a X(\phi_N) \quad F_m(S) = -2\pi a X(\phi_S)$$

At the steady state, the net extratropical flux of the two hemispheres must be equal to the tropical flux $F_m(T)$ so that

$$F_m(T) = 2\pi a \left[X(\phi_N) - X(\phi_S)\right] \tag{17.48}$$

To obtain the streamfunction in terms of the eddies, we write Eq. (17.9) as

$$\overline{u}_t - \left[f - \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}(\overline{u}\cos\phi)\right]\overline{v}^* = \frac{1}{a\cos\phi}\nabla\cdot\mathbf{E}$$

where, as we have seen before, the Coriolis term is substituted by the zonal mean absolute vorticity to obtain a better approximation at middle latitudes. At the steady state $\overline{u}_t = 0$, so that from the definition of streamfunction we get

$$\left[f - (a\cos\phi)^{-1}(\overline{u}\cos\phi)_\phi\right]\frac{\partial X}{\partial Z} = \nabla\cdot\mathbf{E}$$

Substituting the definition of the EP flux and integrating from the altitude $Z_T$ to the top of the atmosphere, we get

$$X(\phi, Z_T) = \left(\rho_0 \cos\phi \overline{v'T'}/\Gamma\right)_{Z_T}(\overline{\eta} a\cos\phi)^{-1}\int_{Z_T}^{\infty}\frac{\partial}{\partial\phi}\left(-\rho_0\cos^2\phi\overline{u'v'}\right)dZ \tag{17.49}$$

where

$$\eta = \left[f - (a\cos\phi)^{-1}(\overline{u}\cos\phi)_\phi\right]$$

According to this relation, the mass flux at the height of the tropopause, for example is proportional to the vertical component of the EP flux plus the integral of the horizontal divergence of the same flux. An intriguing aspect of this equation is that

apparently the flux at some altitude depends only on the eddy dissipation above that level. This conclusion in the literature is called "downward control" and was found by Michael McIntyre. The intriguing part is that the integral could be made just as well from the altitude $Z_T$ to the bottom, so that we could talk about "upward control". A very simple explanation for the previous choice is that it would be very much complicated to calculate the integral of Eq. (17.55) from the ground up because in that case we should consider the eddies in the boundary layer: a very complicated matter indeed. That this is the case can be appreciated by writing Eq. (17.35) as a function of the mass streamfunction and the angular momentum per unit mass

$$\overline{m} = a\cos\phi\,(\overline{u} + a\Omega\cos\phi)$$

At the steady state, it can be shown that

$$\frac{\partial X}{\partial \phi}\frac{\partial \overline{m}}{\partial Z} - \frac{\partial X}{\partial Z}\frac{\partial \overline{m}}{\partial \phi} = a^2\nabla\cdot\mathbf{E}\cos^2\phi \tag{17.50}$$

We can see that the left-hand side is the Jacobian

$$\frac{\partial\,(X,\overline{m})}{\partial\,(\phi,Z)}$$

It can also be shown that

$$\frac{\partial \overline{m}}{\partial \phi} = \frac{\partial\,(\overline{m},Z)}{\partial\,(\phi,Z)} = \frac{\partial \overline{m}}{\partial \phi} - \frac{\partial \overline{m}}{\partial Z}\frac{\partial Z}{\partial \phi}$$

Dividing (17.56) by $\overline{m}_\phi = \partial\overline{m}/\partial\phi$, we get

$$\frac{\partial\,(X,\overline{m})}{\partial\,(\phi,Z)}\frac{\partial\,(\phi,Z)}{\partial\,(\overline{m},Z)} = \frac{\partial\,(X,\overline{m})}{\partial\,(\overline{m},Z)} = -\left(\frac{\partial X}{\partial Z}\right)_{\overline{m}} = \frac{a^2\nabla\cdot\mathbf{E}\cos^2\phi}{\overline{m}_\phi} \tag{17.51}$$

where the derivative of the streamfunction with respect to $Z$ is calculated keeping the angular momentum constant. Equation (17.51) can be easily integrated to give

$$X(\phi,Z) = \int_Z^\infty \left(a^2\nabla\cdot\mathbf{E}\cos^2\phi/\overline{m}_\phi\right)_{\phi=\phi(Z')} dZ'$$

Again the integral must be made along a contour at $\overline{m}$ constant. We can the obtain easily the vertical velocity

$$\overline{w}^* = \frac{1}{\rho_0\cos\phi}\frac{\partial}{\partial\phi}\left[\int_{Z'}^\infty \left(a^2\nabla\cdot\mathbf{E}\cos^2\phi/\overline{m}_\phi\right)_{\phi=\phi(Z')} dZ'\right]$$

If we keep in mind that in the quasi-geostrophic approximation, we have

$$\overline{u} << a\Omega \cos\phi$$

so that

$$\overline{m}_{\phi} \approx -2a^2\Omega \sin\phi \cos\phi$$

is only a function of latitude we get

$$\overline{w}^* = \frac{1}{\rho_0 \cos\phi} \frac{\partial}{\partial\phi} \left[ \int_{Z'}^{\infty} (\nabla \cdot \mathbf{E} \cos\phi / 2\Omega \sin\phi)_{\phi=\cos t} dZ' \right] \tag{17.52}$$

From this equation, we see how the integral over $Z$ is the same if made at constant latitude or at constant angular momentum. This confirms that the angular momentum isopleths, at least at extratropical latitudes, are vertical lines (Fig. 17.8).

In case only gravity waves contribute to the EP flux in Eq. (17.52). We have

$$\nabla \cdot \mathbf{E} = \partial \left( \rho_0 \overline{u'w'} \right) / \partial Z$$

and the residual vertical velocity becomes

$$\overline{w}^* = -\frac{1}{a \cos\phi} \frac{\partial}{\partial\phi} \left( \frac{\overline{u'w'} \cos\phi}{2\Omega \sin\phi} \right) \tag{17.53}$$

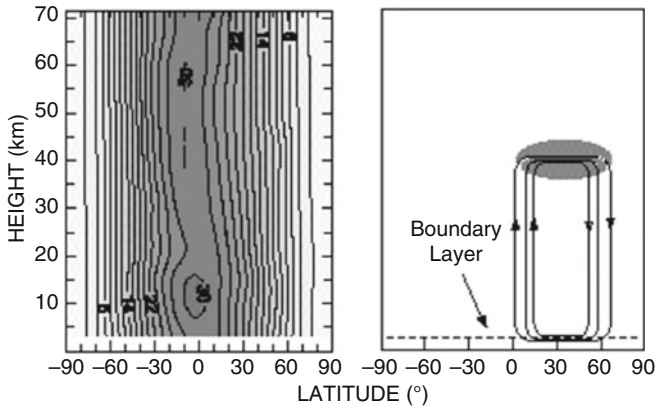



**Fig. 17.8** The isopleths for the angular momentum (*left*) for the winter season in the northern hemisphere. Units are $10^8$ m$^2$ s$^{-1}$. Notice the almost vertical lines in the extratropical regions. At *right*, a schematic circulation cell is shown produced by a zonal force corresponding to the *shaded region*

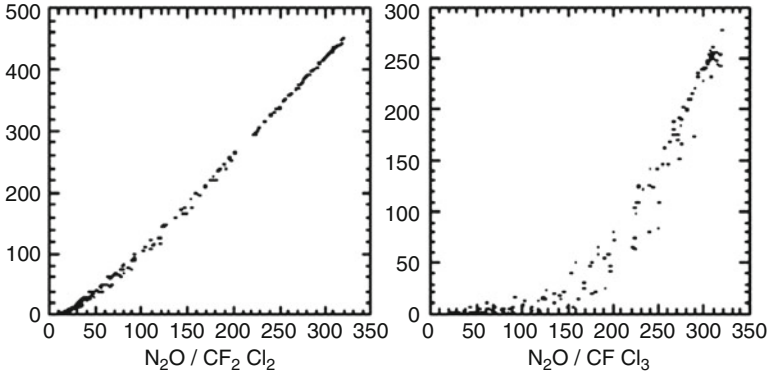



**Fig. 17.9** The correlation between mixing ratios of CF12 (*left*) and CF11 (*right*) with nitrous oxide. Notice that the units are different, ppt for CFC and ppb for $N_2O$

This means that at some height the value of the residual velocity is only a function of the latitudinal gradient of the vertical flux of momentum. The most interesting thing comes from the solution of Eq. (17.50) when the EP divergence is zero. The solution for the streamfunction is given by $\partial X/\partial Z = 0$, and in case that $\overline{m}$ is only function of the latitude we have $\overline{v}^* = 0$, so that the streamlines are parallel to the isopleths of the angular momentum. As shown in Fig. 17.9, this implies that in a steady state situation the momentum gained at some altitude must be dissipated in the boundary layer at the surface. The resulting circulation cell will be below the accelerated region.

This result can be seen as the asymptotic situation of Fig. 17.5 in the sense that, just at the beginning of the forcing, the circulation assumes the shape of a double cell. As the time progress, the upper branch of the cell disappears while the lower branch is extended until it hits the surface, where the momentum gained above is dissipated.

We see actually that downward control is more a mathematical convenience rather than a physical concept. As a matter of fact, the calculation of the circulation with this method may be more difficult than the one based on the heating.

Actually the initial idea was to calculate with the downward control the residual circulation in the region that separates troposphere and stratosphere, where the diabatic terms are small (because the atmosphere is almost in radiative equilibrium) and difficult to calculate. On the other hand, the eddy statistics in the same region are also very sparse. The other disadvantage is that the evolution of the circulation forced by a change in angular momentum is not very intuitive and the moment can only be explained by the fact that Fig. 17.5 is described by an elliptic equation and Fig. 17.9 by an hyperbolic equation: however, we usually try to understand things first and then write down the mathematics.

## 17.4 Transport in Isentropic Coordinates

In what follows we want to refer to some alternative way to formulate the transport problem. This analysis should have some advantages especially when in presence of a strong zonal current like the polar vortex. The starting point is to write the continuity equation in the isentropic coordinate system as we did in Chap. 7 and in particular Eq. (7.55) which we write in the extended form

$$\sigma_t + \nabla_\theta \cdot (\sigma \mathbf{V}) + \left(\sigma \dot{\theta}\right)_\theta = 0$$

We notice that the equivalent density is

$$\sigma = g^{-1}\left(\frac{\partial p}{\partial \theta}\right)$$

Also we have

$$\dot{\theta} = \frac{D\theta}{Dt} = \frac{J\theta}{C_p T} = Q$$

so that the continuity equation becomes

$$\sigma_t + \nabla_\theta \cdot (\sigma \mathbf{V}) + (\sigma Q)_\theta = 0 \tag{17.54}$$

and for a generic tracer with a mass mixing ratio $\chi$

$$\chi_t + u\chi_x + v\chi_y + Q\chi_\theta = S \tag{17.55}$$

where $S$ is the net chemical source. If we multiply Eq. (17.54) for $\chi$ and Eq. (17.55) by $\sigma$ and add, we get the flux form of the continuity equation

$$(\sigma\chi)_t + (\sigma u\chi)_x + (\sigma v\chi)_y + (\sigma Q\chi)_\theta = \sigma S \tag{17.56}$$

At this point, each variable is written as the sum of the zonal mean and the deviation from it so that Eq. (17.56) becomes

$$\overline{\sigma}\,\overline{\chi}_t + (\overline{\sigma v})\,\overline{\chi}_y + \left(\overline{\sigma Q}\right)\overline{\chi}_\theta = \overline{\sigma S} - \left(\overline{\sigma'\chi'}\right)_t - \left[\overline{(\sigma v)'\chi'}\right]_y - \left[\overline{(\sigma Q)'\chi'}\right]_\theta \tag{17.57}$$

Also in this case we have a closure problem to express the eddy terms as a function of average quantities. The first step is to write Eq. (17.57) as a function of weighted means like

$$\overline{A}^* = \frac{\left(\overline{\sigma A}\right)}{\overline{\sigma}}$$

where A is a generic quantity. Equation (17.57) becomes in this case

$$\overline{\chi}_t + \overline{v}^* \overline{\chi}_y + \overline{Q}^* \overline{\chi}_\theta = \overline{S}^* - \sigma^{-1} \left\{ \left(\overline{\sigma' \chi'}\right)_t - \left[\overline{(\sigma v)' \chi'}\right]_y - \left[\overline{(\sigma Q)' \chi'}\right]_\theta \right\} \quad (17.58)$$

To close the problem at this point we write the deviation of the mixing ratio as

$$(\partial/\partial t + \overline{u}\partial/\partial x)\, \chi' + v' \overline{\chi}_y + Q' \overline{\chi}_\theta = S' \quad (17.59)$$

so that putting

$$(\partial/\partial t + \overline{u}\partial/\partial x)\, \eta' = v', \quad (\partial/\partial t + \overline{u}\partial/\partial x)\, q' = Q', \quad (\partial/\partial t + \overline{u}\partial/\partial x)\, \gamma' = S'$$

We obtain for the deviation of the mixing ratio

$$\chi' = -\eta' \overline{\chi}_y - q' \overline{\chi}_\theta + \gamma' \quad (17.60)$$

A drastic simplification at this point can be obtained if we consider adiabatic motions so that $Q' = q' = 0$, and if we neglect the terms containing $\sigma'$ so that Eq. (17.58) becomes

$$\overline{\chi}_t + \overline{v}^* \overline{\chi}_y + \overline{Q}^* \overline{\chi}_\theta = \overline{S}^* + \sigma^{-1} \left[\overline{(\sigma v)' \chi'}\right]_y$$

substituting Eq. (17.66) and performing the zonal means we get

$$\overline{\chi}_t + \overline{v}^* \overline{\chi}_y + \overline{Q}^* \overline{\chi}_\theta = \overline{S}^* - \overline{\sigma}^{-1} \left(\overline{\sigma} \overline{v' \gamma'}\right)_y + \overline{\sigma}^{-1} \left[\overline{\sigma} \frac{1}{2} \overline{\eta'^2} \overline{\chi}_y\right]_y \quad (17.61)$$

with the notation

$$\overline{S}^\dagger = \overline{S}^* - \overline{\sigma}^{-1} \left(\overline{\sigma} \overline{v' \gamma'}\right)_y \quad D = \left(\frac{1}{2}\right) \overline{\eta'^2};$$

Using for the horizontal, the spherical coordinates we get

$$\overline{\chi}_t + \overline{v}^* \overline{\chi}_y + \overline{Q}^* \overline{\chi}_\theta = \overline{S}^\dagger + (\overline{\sigma} \cos\phi)^{-1} \left(\overline{\sigma} D \overline{\chi}_y \cos\phi\right)_y \quad (17.62)$$

This is the same as in Eq. (17.42) but for the transport on isentropic coordinates. If the motion is adiabatic, we may expect the tracer to move along isentropic surfaces with diffusion that is mainly in the same direction. The big news for the moment is the absence of a vertical diffusion coefficient, while the horizontal component has been defined in the same way as before, that is, based on the deviations of the air parcels with respect to an average position.

### 17.4.1 Stratospheric Dynamics and Ertel Potential Vorticity

The most interesting thing now is to relate the diffusion coefficient to other concepts we have found like downward control, but in an isentropic coordinate system the Ertel vorticity should play an important role. Until now, the application of this concept has been rather limited and before we show its importance we need to introduce some further consideration. We may start simply by the equation of motion in $x$ direction

$$\frac{Du}{Dt} = fv - \rho^{-1}\frac{\partial p}{\partial x}$$

We simply expand the Lagrangian derivative, and keeping in mind that

$$\frac{\partial u}{\partial y} = -\zeta + \frac{\partial v}{\partial x}$$

we obtain

$$u_t - v\zeta + \left(u^2 + v^2\right)_x/2 - fv = -\rho^{-1}p_x$$

With zonal mean we obtain the result

$$\overline{u}_t - \overline{v\zeta} - f\overline{v} = 0$$

that relates the acceleration of the zonal wind to the meridional vorticity advection. The quantity $\overline{v\zeta}$ can be expressed as a function of the eddy terms to obtain

$$\overline{u}_t - \overline{v}\,\overline{\zeta} - \overline{v'\zeta'} - f\overline{v} = 0$$

so that substituting for the zonal mean value of the vorticity $\overline{\zeta} = -\partial\overline{u}/\partial y$, we get

$$\overline{u}_t - \overline{v}\,(f - \partial\overline{u}/\partial y) - \overline{v'\zeta'} = 0 \tag{17.63}$$

The same procedure can be applied to the equations of motion written in the isentropic system and in this case the starting point will be Eq. (7.50), written in the form

$$\frac{\tilde{D}\mathbf{V}}{Dt} = -f\mathbf{k}\times\mathbf{V} - \nabla_\theta\Psi \tag{17.64}$$

where

$$\frac{\tilde{D}}{Dt} = \frac{\partial}{\partial t} + \frac{u\partial}{\partial x} + \frac{v\partial}{\partial y} + \frac{Q\partial}{\partial\theta}$$

while $\Psi$ is the Montgomery streamfunction. The analog of Eq. (17.63) will then be

$$\overline{u}_t - \overline{v}^* f^\dagger + \overline{Qu_\theta} = \sigma \overline{(v'P')}^* + M \tag{17.65}$$

where

$$f^\dagger = f - (\overline{u}\cos\phi)_\phi / a\cos\phi$$

$P$ is the Ertel potential vorticity with the index that gives the deviation from the weighted zonal mean previously defined, and $M$ is any other kind of zonal acceleration At this point, the eddy term can be written in the usual form $P' = -\eta P_v$ so that the flux of the Ertel potential vorticity can be written as $\overline{(v'P')}^* = -D\overline{P}_y$ that can be substituted in Eq. (17.65) to obtain

$$\overline{u}_t - \overline{v}^* f^\dagger + \overline{Qu_\theta} = -\overline{\sigma} D\overline{P}_y + M \tag{17.66}$$

The relation between the vorticity eddy flux and the mean gradient has been used sometime to determine the diffusion coefficient. This is in any case an approximation that can be justified only for time intervals shorter than the average lifetime of a planetary wave. The heating term that appears in Eq. (17.66) can be related to the potential vorticity flux through the downward control. The steady state form of Eq. (17.66) in the absence of external acceleration and neglecting the vertical advection gives

$$\overline{v}^* = \overline{\sigma} D\overline{P}_y / f^\dagger \tag{17.67}$$

From this expression, we can still find the downward control principle if we define a streamfunction such that

$$\overline{Q}^* = -\psi_y / \overline{\sigma}\cos\phi \quad \overline{v}^* = \psi_\theta / \overline{\sigma}\cos\phi$$

Integrating Eq. (17.66) with respect to $\theta$ and putting $\psi(\infty) = 0$, we have

$$\psi(\theta_0) = -\int_{\theta_0}^{\infty} \left(\overline{\sigma}^2 D\overline{P}_y / f^\dagger\right) \cos\phi d\theta$$

that can be differentiated with respect to the latitude to obtain $\overline{Q}^*$

$$\overline{Q}^*(\theta_0) = (\overline{\sigma}\cos\phi)^{-1} \int_{\theta_0}^{\infty} \left(\overline{\sigma} D\overline{P}_y / f^\dagger\right)_y d\theta \tag{17.68}$$

This is similar to Eq. (17.58) and says that the diabatic heating is controlled by the meridional flux of the Ertel potential vorticity. This is a further confirmation of the "radiative spring" and that the eddies are responsible for keeping the atmosphere away from radiative equilibrium particularly at high latitudes. At middle latitudes, the atmosphere is near radiative equilibrium and the eddy activity is negligible.

### *17.4.2 The Slope of the Tracers*

In Fig. 17.6, we have seen how the circulation may determine the slope of the isopleths of the mixing ratio. This is not an academic discussion because we will see that it will have important implications.

We may start to consider a tracer that has no source in the sense that it may be released at some instant and allowed to spread around. In this case, the change in the mixing ratio will be

$$\frac{\partial \overline{\chi}}{\partial t} = -\overline{v}\frac{\partial \overline{\chi}}{\partial y} - \overline{w}\frac{\partial \overline{\chi}}{\partial z} + \frac{\partial}{\partial y}\left(K_{yy}\frac{\partial \overline{\chi}}{\partial y} + K_{yz}\frac{\partial \overline{\chi}}{\partial z}\right) + \frac{\partial}{\partial z}\left(K_{zz}\frac{\partial \overline{\chi}}{\partial z} + K_{yz}\frac{\partial \overline{\chi}}{\partial y}\right) = 0$$

In case the mean gradients of the mixing ratio are assumed to be constant, we obtain easily, keeping in mind the definition of streamfunction

$$\frac{\partial}{\partial y}\left[K_{yy}\frac{\partial \overline{\chi}}{\partial y} + \left(K_{yz} - \psi\right)\frac{\partial \overline{\chi}}{\partial z}\right] = -\frac{\partial}{\partial z}\left[K_{zz}\frac{\partial \overline{\chi}}{\partial z} + \left(K_{yz} + \psi\right)\frac{\partial \overline{\chi}}{\partial y}\right]$$

Usually the vertical diffusion is negligible and we have also

$$\frac{\partial}{\partial y}\left[\left(K_{yz} - \psi\right)\frac{\partial \overline{\chi}}{\partial z}\right] >> \frac{\partial}{\partial z}\left[\left(K_{yz} + \psi\right)\frac{\partial \overline{\chi}}{\partial y}\right]$$

so that at the equilibrium we have

$$\frac{\partial}{\partial y}\left[K_{yy}\frac{\partial \overline{\chi}}{\partial y} + \left(K_{yz} - \psi\right)\frac{\partial \overline{\chi}}{\partial z}\right] = 0$$

And the ratio of the gradients gives the slope

$$\Delta z/\Delta y \cong -\frac{\partial \overline{\chi}/\partial y}{\partial \overline{\chi}/\partial z} = \frac{K_{yz} - \psi}{K_{yy}} \tag{17.69}$$

In a situation like that of Fig. 17.6, starting from the equator toward the northern hemisphere, we will have a negative vertical gradient for the mixing ratio and since it is the meridional gradient the slope will be positive. This argument is quite similar to that of Reed and German, to which the previously discussed invention

of the diffusion coefficients must be attributed. The only change is that now also the streamfunction play a role. If we go to an isentropic coordinate system, the situation is quite different. Here we assume that the mixing ratio is given by a global average value $\chi^0$ that is only a function of the potential temperature plus a small quantity $\chi^\dagger$so that $\overline{\chi} = \chi^0 + \chi^\dagger$. In this case, we have

$$\chi_t^\dagger + \overline{Q}^* \chi_\theta^0 = \overline{S}^* + (\overline{\sigma}\cos\phi)^{-1}\left(\overline{\sigma} D \chi_y^\dagger \cos\phi\right)_y \tag{17.70}$$

From this equation it is possible to obtain a slope with respect to an isentrope

$$\overline{S}^* = \tan^{-1}\left(\overline{\chi}_y / \overline{\chi}_\theta\right) \approx \tan^{-1}\left(\overline{\chi}_y^\dagger / \chi_\theta^0\right)$$

And the way to proceed is that to integrate the steady state solution for Eq. (17.70) with respect to $y$ again recurring to the definition of streamfunction

$$-\psi_y \chi_\theta^0 = \overline{S}^* \overline{\sigma}\cos\phi + \left(\overline{\sigma} D \chi_y^\dagger \cos\phi\right)_y$$

Thus because $\chi_\theta^0$ is only a function of $\theta$, we get

$$S = \tan^{-1}\left[-\psi(\overline{\sigma} D \cos\phi)^{-1}\right] + S_0 \tag{17.71}$$

From this equation as result also from Eq. (17.70), we notice how all the tracers that lay on isentropic surfaces have the same slope. It is very interesting to substitute Eq. (17.68) for the streamfunction because we get

$$S = \tan^{-1}\left[\overline{\sigma}^{-1}\int_{\theta_0}^{\infty}\left(\overline{P}_y \overline{\sigma}^2 / f^\dagger\right) d\theta\right] + S_0$$

that shows how the slope does not depend on the diffusion coefficient at the least when this is constant. This is not surprising because $D$ controls both the net heating and the mixing. If the diffusion coefficient increases then the slope decreases while the net heating increases and so this will increase again the slope.

A rather interesting case is to calculate the slope for two tracers with mixing ratios $\chi_1$ and $\chi_2$ that we define as $S_\mathrm{r} = \partial\chi_1/\partial\chi_2$.

The fact that the two tracers may have different slopes is not ruled out by Eq. (17.71) because the integration constant depends on the net chemical source. We can start from a global mean of Eq. (17.56) defined as

$$\langle \ldots \rangle = \frac{1}{4\pi}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} (\ldots)\cos\phi \, d\phi \, d\lambda$$

where $(\ldots)$ indicates the quantity on which the average is performed

$$\frac{\partial}{\partial t}\langle\sigma\chi\rangle + \frac{\partial}{\partial\theta}\langle\sigma Q\chi\rangle = \langle\sigma S\rangle \tag{17.72}$$

We may now approximate the heating rate with its zonal mean $\overline{Q}^*$ and also substitute $\sigma = \sigma_0(\theta)$ so that Eq. (17.72) becomes

$$\frac{\partial}{\partial t}\langle\chi\rangle + \frac{1}{\sigma_0}\frac{\partial}{\partial\theta}\left\langle\sigma_0\overline{Q}^*\overline{\chi}\right\rangle = \langle S\rangle \tag{17.73}$$

The diabatic term can be evaluated from Eq. (17.62) assuming that the last three terms are at the equilibrium and that the diffusive and chemical terms may be parameterized in terms of time constants

$$\overline{Q}^*\frac{\partial\chi}{\partial\theta} = -\overline{\chi}\left(\frac{1}{\tau_c} + \frac{1}{\tau_d}\right) \tag{17.74}$$

From this equation $\overline{\chi}$ may be obtained which can be substituted in Eq. (17.72) to give

$$\frac{\partial}{\partial t}\langle\chi\rangle + \frac{1}{\sigma_0}\frac{\partial}{\partial\theta}\left\langle\sigma_0\left(\overline{Q}^*\right)^2\left(\tau_c^{-1} + \tau_d^{-1}\right)^{-1}\frac{\partial\chi}{\partial\theta}\right\rangle = \langle S\rangle$$

This notation is equivalent to the introduction of a diffusion coefficient given by

$$K_{zz} = \left\langle\left(\overline{Q}^*\right)^2\left(\tau_c^{-1} + \tau_d^{-1}\right)^{-1}\right\rangle \tag{17.75}$$

In case of long-lived tracers, we have $\tau_d^{-1} >> \tau_c^{-1}$ so that the vertical diffusion coefficient becomes

$$K_{zz} \approx \left(\overline{Q}^*\right)^2\tau_d.$$

and at the steady state we could write

$$-\frac{\partial}{\partial\theta}\left\langle\sigma_0 K_{zz}\frac{\partial\langle\chi\rangle}{\partial\theta}\right\rangle = \sigma_0\langle S\rangle = -\sigma_0\frac{\overline{\chi}}{\tau_c}$$

We see how the gradient of the mixing ratio is a function of the chemical lifetime of the species. For two species having mixing rations $\chi_1, \chi_2$ and relative mixing ratios, we have

$$\frac{d\chi_1}{d\chi_2} = \frac{\overline{\chi}_1}{\overline{\chi}_2}\frac{\tau_{2,c}}{\tau_{1,c}} \tag{17.76}$$

This relation shows that if we know the correlation between two long-lived species it is possible to determine from one species the lifetime of the other. An example is shown in Fig. 17.9 for the most important chlorofluorocarbons. In this case, the tracer with the longest lifetime is the nitrous oxide ($N_2O$), and we notice that the correlation is almost linear for CF12 but not for CF11. This may derive from the origin of the different points. In the second case, it appears that we should have two different lifetimes, and this could happen if the points taken are from the high troposphere or the stratosphere, where the lifetimes can be different and give raise to a different slope of the correlation curve. Notice that the mixing ratios of the species shown are very much different, being ppb for $N_2O$ and ppt in the case of CFCs. In the case of CF12, the correlation may be expressed as (Fig. 17.10)

$$\chi_{\mathrm{CF12}}\,(\mathrm{ppt}) = -59.4 + 1.6\chi_{\mathrm{N_2O}}\ \ (\mathrm{ppb})$$

Assuming a lifetime for nitrous oxide of 130 years, we get for Eq. (17.76) a lifetime

$$\tau_{\mathrm{CF12}} = 420.6 \cdot 130/\,(1.6 \cdot 300) = 113\ \mathrm{year}$$

where as reference concentration we have taken 300 ppt for $N_2O$. We will see that $N_2O$ could be used as a "coordinate", being an almost conservative quantity.

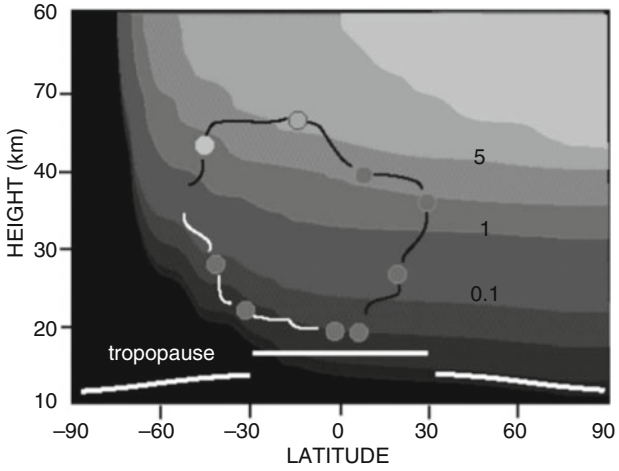



**Fig. 17.10** The path of air parcel to a sample point. Parcels leaving from the same point arrive at the same time at the sample point with a different content of the tracer. This situation could represent $N_2O$ with the *shaded zone* representing regions with different loss rates (in years$^{-1}$) (From Schoeberl et al. 2000)

### 17.4.3 The Tracer Correlation: Age of Air and Transport

A paper by M. Schoeberl gives a very nice picture of the tracer correlation method in relation to the age of the air. This is defined as the time elapsed for an air parcel since it left the troposphere. In practice, if $\mathbf{x}_0$ indicates the position that initially is occupied by the air parcel and $\boldsymbol{x}_s$ the location where the same parcel is sampled, the age of the air parcel is just the time it takes to go between the two points. Figure 17.11 shows in a very qualitative way how age can be related to photochemical process. Air parcels leaving the tropopause at the same time and from the same point may reach the sample point at the same time (having the same age), although they go through a quite different path. One of the parcels penetrates a region where the loss rate for a generic species is very high, while the other stays in a region where the loss is limited. As a result parcel having the same age have a different content of the species. This case could refer for example to nitrous oxide. The sample will be made up of a number of irreducible (that is they do not mix along the path) parcels that have taken many different paths to reach the sample point. The amount of constituent measured in the sample will be the average over the different parcels. The parcels will have a distribution of age so we talk about *age spectrum*, and this implies that within the sample there will be a distribution in composition that we will call *tracer spectrum*.

For a parcel that is released at point $\mathbf{x}_0$ and it is sampled at $\mathbf{x}_s$, we can define the quantity $G_0$ ($T$, $\mathbf{x}_s$. $\mathbf{x}_0$) as the probability that the transit time is in the range $T$, $T + dT$ and we will have

$$\int_0^\infty G\Big(T, \mathbf{x}_s, \mathbf{x}_0\Big) dT = 1 \tag{17.77}$$

$G$ can also be identified as the Green function for an inert tracer with a source $B_0(t, \mathbf{x}_0)$ such that the concentration of the tracer at the sample point $B_s$ will be

$$B_s\left(\mathbf{x}_s, \mathbf{x}_0, t\right) = \int_0^t B_0\left(T, \mathbf{x}_0\right) G\Big(t - T, \mathbf{x}_s, \mathbf{x}_0\Big) dT$$

We can then define the mean age for the sample

$$\Gamma = \int_0^\infty G\Big(T, \mathbf{x}_s, \mathbf{x}_0\Big) T dT = 1$$

For inert tracers, the mean age and the constituent mixing ratio are directly related so that because $B_0(T, \mathbf{x}_0) \approx T$ then the mixing ratio will be proportional to the mean age. However, if we consider a constituent that is subject to a loss process, things may be a little different. We consider that all the air parcels are released at the tropical tropopause with a mixing ratio $B_0$ that we may normalize to unity. The

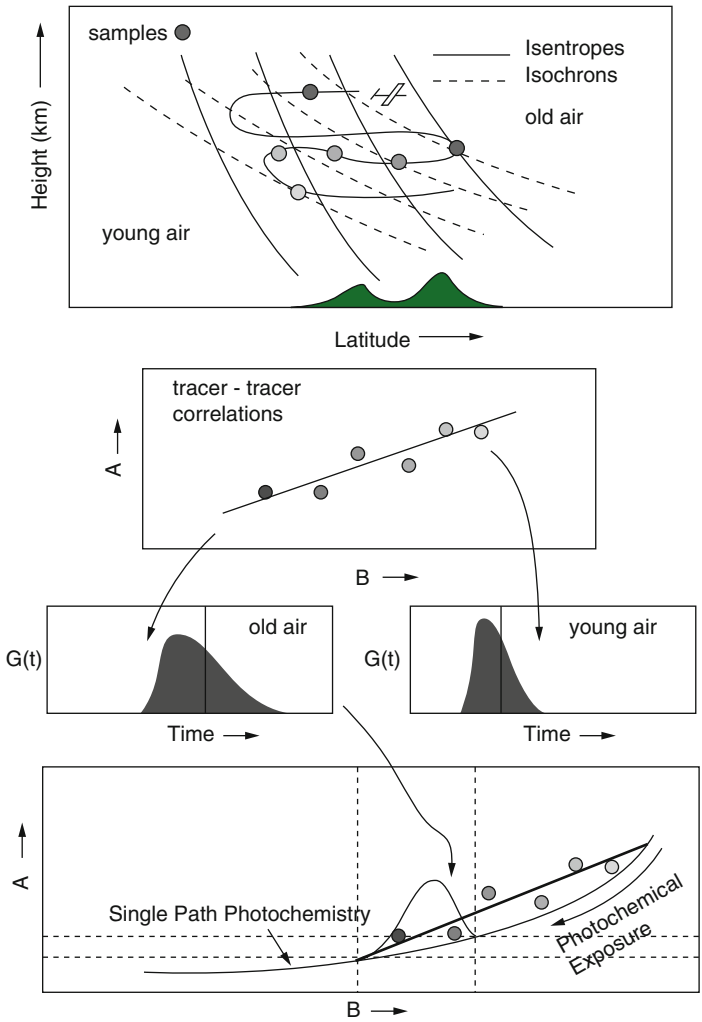



**Fig. 17.11** This figure, taken from Schoeberl et al. (2000), illustrates the atmospheric sampling. The aircraft (*top*) measure air sample with different ages that give rise to the correlation plot just below the *top panel.* The age spectrum for two samples is also shown. The *panel at the bottom* shows qualitatively the single path photochemistry curve (SPPC) and the case of two constituents with ages $T_1$ and $T_2$. The *dotted lines* represent the relative values $A_1$, $A_2$ and $B_1$, $B_2$ for which the chord gives the air sample concentration

parcel $i$ follows a path $\mathbf{x}_i(t)$ with a chemical loss rate $\beta[\mathbf{x}_i(t), t]$ so that the parcel amount of the constituent will be given by

$$DB_i/Dt = -\beta\,[\mathbf{x}_i(t), t]\,B_i \tag{17.78}$$

Integrating this equation over the time $T_i$ (the age of the parcel) it takes for the parcel to arrive at the sample point we have

$$\log(B_i) = -\int_0^{T_i} \beta\Big[x_i(t), t\Big]dt = -T_i\overline{\beta}_i(T_i) \tag{17.79}$$

where

$$\overline{\beta}_i(T_i) = \frac{1}{T_i}\int_0^{T_i} \beta\Big[x_i(t), t\Big]dt$$

The overbar indicates the average on a single path. As shown in Fig. 17.11, parcels starting from the same place will arrive at the sample point with different amounts of tracer so that a tracer spectrum can be introduced like counting the parcels that have mixing ratios between $B$ and $B + dB$. We assume that $P(B)$ represents such a distribution so that the sample concentration will be

$$B_s = \int_0^1 P(B)BdB \tag{17.80}$$

We considernow Fig. 17.12. An aircraft samples air with different ages and different tracer amount as indicated by the dots. Isochrons, that is isopleths of age, are the solid lines. Based on a tracer–tracer correlation, the different samples are plotted

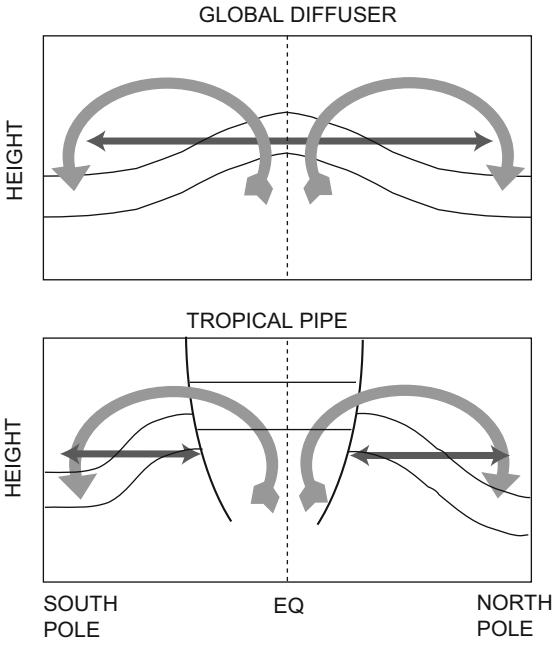



**Fig. 17.12** Two models for diffusion and transport of chemical constituents in the stratosphere. The global diffuser model is at the *top* and the tropical pipe model at the *bottom*

and each of them is characterized by an age spectrum $G(t)$. The amount of tracer contained in each of the irreducible parcels depends on their exposure to the photochemical loss process. If this process is simple photodissociation, the loss rate will be simply proportional to the photodissociation coefficient

$$J_B = \beta_B \int \sigma F d\lambda$$

where the meaning of the symbols should be well known. The photochemical exposure $n$ is defined in such a way that

$$dn = \left( \int \sigma F d\lambda \right) dt$$

which is simply the energy absorbed, so that we have for parcel (i) an expression similar to Eq. (17.83)

$$\log (B_i) = -\beta_B n \tag{17.81}$$

If we now consider another trace gas $A$, we will have $n = -\log(A_i)/\beta_A$ so that eliminating $n$ we get

$$A_i = B_i^{\beta_A/\beta_B} \tag{17.82}$$

This equation is valid only when the functional form of the photodissociation coefficient is the same. In the more general case, the sample concentration will depend of the spectrum $G_0(T)$ and it can be shown that

$$A_s = \int_0^\infty A_i(T) G_0(T) dT = \int_0^\infty B_i^{\left\langle \overline{\beta}_A(T) \right\rangle / \left\langle \overline{\beta}_B(T) \right\rangle} G_0(T) dT \tag{17.83}$$

Equation (17.88) or (17.89) represents the amount of tracer $A$ as a function of $B$ and the photochemical exposure. On a $A$-$B$ correlation plane, they represent a *single path photochemistry curve* (SPPC) on which the constituents moves as they wander through the stratosphere. To illustrate the point, we may consider two equal number of irreducible parcels that have ages $T_1$ and $T_2$. Then we will have

$$G_0(T) = [\delta (T - T_1) + \delta (T - T_2)] /2; \quad \Gamma = (T_1 + T_2) /2; \quad B_s = (B_1 + B_2) /2;$$

$$A_s = \left( B_1^{\overline{\beta}_A(T_1)/\overline{\beta}_B(T_1)} + B_2^{\overline{\beta}_A(T_2)/\overline{\beta}_B(T_2)} \right) \Big/ 2$$

The sample point $A_s$, $B_s$ will be at the midpoint of a chord that intersects the SPPC curve at point $A_1$, $B_1$ and $A_2$, $B_2$. The effect of mixing will be to spread the sample points below the limit chord with intercept (1,1) and (0,0).

The data plotted in Fig. 17.9 have been produced with a model and are reported only as an example. In practice, when real data are plotted with the same system, we notice that the slope of the correlation curve is different depending on the region where data are taken. For example, for $CFCl_3$, the slope tends to increase drastically for data taken in the intertropical region. If the mechanism for transport were the same for all latitudes, we should get the same correlation. We know that the circulation maintains a mixing ratio gradient against the effect of isentropic mixing.

This consideration, together with other experimental data, suggested to Alan Plumb to elaborate two transport models, sketched in Fig. 17.12. The "global diffuser" model considers that transport is due to a single cell that covers from equator to the poles and is responsible for the tracer advection. In this case, the mixing due to diffusion processes is present from the tropics to high latitudes. In the other hypothesis, the tropical region corresponds to some kind of containment vessel from which tracers are advected to the high latitude and the mixing is limited to the extra-tropical regions. In the figure, the advection is indicated with the gray arrows while the diffusion is represented with the black arrows.

These two models give a correlation between long-lived tracers that are very much different. In the first model, the isopleths of the mixing ratio are much higher at the equator and slope gently toward the poles. In the second model (called also tropical pipe), the vertical mixing is inhibited within the tropical pipe so that the mixing ratio is independent on latitude. At middle latitudes, there is not too much difference between the models. The second scheme may depend very much from the season because the position of the pipe is actually determined by the subsolar point.

We make Eq. (17.75) very intuitive, if we neglect the mixing due to the $K_{yx}$ component and assume that the advective flux at some latitude is balanced by the vertical diffusive flux. Then we may write

$$K\frac{\partial \chi}{\partial y} \approx -\frac{\partial \chi}{\partial z}\int w dy \tag{17.84}$$

If we consider that the slope is very small (that is quasi-horizontal), we may assume it to be independent of latitude. Equation (17.84) can be integrated to give

$$K\frac{\partial \chi}{\partial y} = \frac{\partial \chi}{\partial y}\int_{y}^{y_p} (w - \langle w \rangle) dy + K\gamma\,(y_t)\,\frac{y_p - y}{y_p - y_t} \tag{17.85}$$

where $\gamma$ now indicates the slope and the condition is that at the pole $\gamma = 0$. We have denoted with $y_p$, $y_t$ the latitude of the pole and the tropics. With $\langle w \rangle$ we have indicated the velocity averaged between the pole and the equator

$$\langle w \rangle = \int_{y_t}^{y_p} w dy \Big/ \left(y_p - y_t\right)$$

It is possible to show that the second term on the right-hand side of Eq. (17.84) is negligible so that the equation becomes

$$K\frac{\partial \chi}{\partial y} = \frac{\partial \chi}{\partial z}\int_{y}^{y_p}\Big(w - \langle w\rangle\Big)dy \tag{17.86}$$

relating the slope to transport.

Within the tropical pipe we may assume that the vertical diffusion is negligible while the mixing ratio is constant with latitude so that the source is balanced only by the vertical advection

$$\rho W\frac{\partial \chi_T}{\partial z} \approx S$$

This is equivalent to saying that the correlation between mixing ratio is the ratio between the sources

$$\left(\frac{d\chi_1}{d\chi_2}\right) = \left(\frac{S_1}{S_2}\right) \tag{17.87}$$

In the tropical pipe, species have a constant mixing ratio with latitude and the correlation depends on the intensity of the sources.

### *17.4.4 The Conservative Coordinates*

The same equation that describes the mixing ratio for a long-lived tracer (like nitrous oxide) could be used for any other tracer, that is, for a conservative quantity like the Ertel potential vorticity $P$. In this case, the isopleths for the vorticity will follow a similar equation with zero source

$$\Pi_t^{\dagger} + \overline{Q}^{*}\Pi_\theta^0 = (\overline{\sigma}\cos\phi)^{-1}\big(\overline{\sigma}D\overline{\Pi}_y\cos\phi\big)_y \tag{17.88}$$

We should expect the isopleths of the vorticity to follow the constant mixing ratio lines for $N_2O$ so that for all practical purposes we could substitute one conservative quantity with the other. The advantage is that potential vorticity is defined by dynamic data so that when these are not available the mixing ratio data will do the job. Unfortunately, as we shall see in the next chapter, the chemical data are very sparse so that it is very difficult to distinguish between dynamic and chemical contributions in the determination of a tracer distribution. In other occasions, measurements taken at the same time in two atmospheric regions cannot be compared because the dynamical connection between the two is not known. There is also a more general question related to the definition of zonal mean, which should have a very precise physical meaning rather than a simple reduction in the number of dimensions.

Some of these problems can be solved with an appropriate coordinate transformation. As an example we may consider the continuity equation for a tracer

$$\frac{\partial \chi}{\partial t} + V \cdot \nabla \chi = S$$

With the appropriate choice of three independent quantities (i.e. $\alpha \neq f(\beta, \gamma)$ etc.) in the region of interest, we can write

$$\frac{\partial \chi}{\partial t} + \dot{\alpha}\frac{\partial \chi}{\partial \alpha} + \dot{\beta}\frac{\partial \chi}{\partial \beta} + \dot{\gamma}\frac{\partial \chi}{\partial \gamma} = S \tag{17.89}$$

where

$$S(\alpha, \beta, \gamma) \equiv S[\phi(\alpha, \beta, \gamma), \lambda(\alpha, \beta, \gamma), Z(\alpha, \beta, \gamma)]$$
$$\chi(\alpha, \beta, \gamma) \equiv \chi[\phi(\alpha, \beta, \gamma), \lambda(\alpha, \beta, \gamma), Z(\alpha, \beta, \gamma)]$$

and $\phi, \lambda, Z$ are latitude, longitude, and height. We can choose the new coordinates in such a way that their derivative with respect to time is zero, that is, their value is conserved within the air parcel

$$\dot{\alpha} = \dot{\beta} = \dot{\gamma}$$

In this new reference system, the changes in the mixing ratio are only a function of the source and the Lagrangian and Eulerian derivative coincide

$$\partial \chi / \partial t = S$$

A first candidate for the coordinate transformation is the potential temperature because for the adiabatic motion we have $d\theta/dt = 0$ so that $\theta$ could substitute for the height. For latitude, the candidate could be the Ertel vorticity, as is quite evident in Fig. 17.14. Here are reported the zonal mean values for the potential vorticity and the potential temperature. We notice that especially at middle latitude, the surfaces with constant vorticity are almost parallel to the potential temperature surfaces so that both quantities change very little with latitude. On the other hand, if we use the modified potential vorticity (like the one shown at the bottom of the same figure) then we notice a nice improvement because now the vorticity changes with latitude once the potential temperature is fixed. Another important point is that the polar vortex is much more evident in this case. So we may assume the modified potential vorticity (MPV) to be our candidate to substitute for the latitude.

To understand a little better the origin of the MPV, we start from the vorticity equation in isentropic coordinates in the case in which the potential temperature surface are perfectly horizontal. Here the equation can be written as (see Chap. 7)

$$(\zeta + f)_t + \nabla_\theta \cdot (\zeta + f)\mathbf{V} = -Q\zeta_\theta + u_\theta Q_y - v_\theta Q_x \tag{17.90}$$

Using the definition of potential vorticity, we get

$$\frac{D\Pi}{Dt} = \frac{1}{\sigma}\frac{D(\zeta + f)}{Dt} - \frac{(\zeta + f)}{\sigma^2}\frac{D\sigma}{Dt}$$

For an isentropic system, we have

$$D/Dt = \frac{\partial}{\partial t} + \frac{u\partial}{\partial x} + \frac{v\partial}{\partial y} + \frac{Q\partial}{\partial \theta}$$

so that the vorticity equation becomes

$$\frac{D\Pi}{Dt} = \Pi Q_\theta + \sigma^{-1}\left[Q_y u_\theta - Q_x v_\theta\right] \tag{17.91}$$

which corresponds to the conservation of Ertel potential vorticity in the adiabatic case ($Q = 0$). An intuitive interpretation of this equation may be obtained when the second term on the right is negligible. In this case, the conservation can be written in the flux form

$$\frac{\partial}{\partial t}(\sigma\Pi) + \nabla \cdot \mathbf{J} = 0$$

where

$$\mathbf{J} = \mathbf{V}\sigma\Pi - Q(\zeta + f)$$

and considering the mass conservation we obtain again Eq. (17.97). The conservative coordinate are then Ertel vorticity and potential temperature because we have for these variables

$$\dot{\theta} = Q \qquad \dot{\Pi} = \Pi Q_\theta + M \tag{17.92}$$

where $M$ represents all the extra terms including the friction. The conservative coordinates correspond then to a region of the atmosphere where the derivatives appearing in Eq. (17.92) are negligible and that means a negligible heating as well as its change with potential temperature.

To make Eq. (17.91) more useful, we specialize the Ertel vorticity by writing it in the form

$$\Pi = (\zeta + f)/\sigma \approx -g(\zeta + f)\,\partial\theta/\partial p$$

From the definition of potential temperature, we get

$$\Pi = -g(\zeta + f)\left[\left(\frac{p_0}{p}\right)^\kappa \frac{\partial T}{\partial p} - \frac{\kappa}{p_0 T^{1/\kappa}}\theta^{(\kappa+1)/\kappa}\right] \tag{17.93}$$

where $\kappa = R/C_p \approx 2/7$ and $(\kappa+1)/\kappa \approx 9/2$. In case of an isothermal atmosphere, we have

$$\Pi = \frac{g\kappa}{p_0 T_0^{1/\kappa}} (\zeta + f)\, \theta^{(\kappa+1)/\kappa} \tag{17.94}$$

so that in the vertical $\Pi$ changes as $\theta^{9/2}$. At this point, we remember that the Ertel vorticity is defined for any scalar $\lambda$

$$\Pi = \rho^{-1} \zeta_a \cdot \nabla\lambda$$

We could then take any monotonic function of the potential temperature $\lambda = S(\theta)$. This property can be used to make the vorticity more appropriate to represent the meteorological fields. From Eq. (17.94), the vorticity changes too rapidly with altitude because the potential temperature varies as $\theta^{-7/2}$. A more appropriate choice would be

$$h = S(\theta) = -(2/11)\, \theta^{-11/2}$$

so that

$$\frac{\partial h}{\partial p} = \left(\frac{\partial S}{\partial \theta}\right)\left(\frac{\partial \theta}{\partial p}\right)$$

and the Ertel vorticity is given by

$$\Pi_m = \Pi(\theta/\theta_0)^{-9/2} \tag{17.95}$$

For this new variable, we can write the equation similar to Eq. (17.91)

$$\frac{D\Pi_m}{Dt} = \frac{1}{\sigma \cos\phi}\left(\frac{\theta}{\theta_0}\right)^{-9/2} \left(Q_\phi u_\theta \cos\phi - Q_\lambda \upsilon_\theta\right) + \Pi Q_\theta - \frac{9}{2}\frac{Q}{\theta} \tag{17.96}$$

A scale analysis shows that the dominant term is $\Pi Q_\theta$ together with the last term on the right. At the bottom of Fig. 17.13, the isopleths for $\Pi_m$ are shown and we see that isopleths for $\Pi_m$ change very nicely with latitude. With these approximations the points, defined now on the $\theta$, $\Pi_m$ grid, satisfy the condition that a tracer amount depends only on the chemical source along the longitude. Fixing the value of the vorticity and the temperature correspond to following a tube along a latitudinal "circle" with a section defined by $\Delta\theta \times \Delta\Pi_m$.

Figure 17.14 shows a possible application of this coordinate system with the isopleths of the ozone mixing ratio in the two reference system. The new system completely destroys the intuition that was in the old system but gives something more useful, because we may assume now that along the tubes defined above the ozone mixing ratio remains the same. A very nice implication is the possibility now to compare measurements at distant points with the help of meteorological data (vorticity and temperature). This may be very important in treating the experimental data.

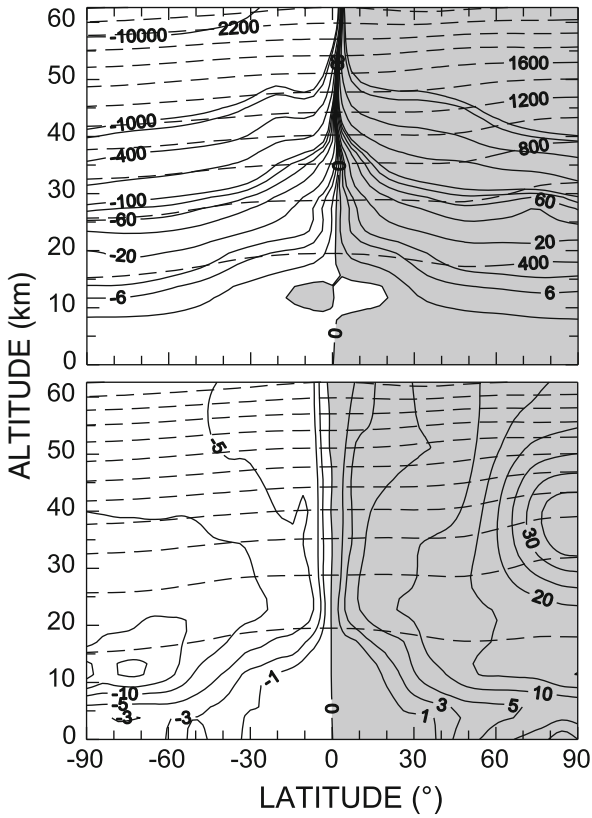



**Fig. 17.13** At the *top*, the isopleths (s*olid lines*) for the Ertel potential vorticity (units $10^{-6}$ $m^2$kg K $s^{-1}$) and the potential temperature (*dashed*) are shown. At the *bottom*, the isopleths are for the modified potential vorticity. Negative values for the latitude refer to the southern hemisphere

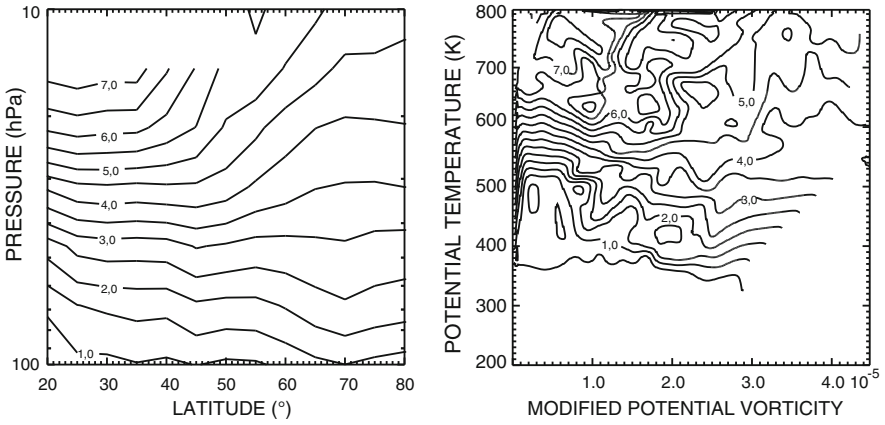



**Fig. 17.14** The ozone mixing ratio in the latitude pressure reference system (*left*) and in the potential temperature, modified potential vorticity system (*right*)

As a conclusion of this excursion into the new reference system, we come back to the possibility of using nitrous oxide as the vertical coordinate. The thermodynamic equation can be written

$$\frac{\partial \theta}{\partial t} + u^n \frac{\partial \theta}{\partial x} + v^n \frac{\partial \theta}{\partial y} + \dot{N} \frac{\partial \theta}{\partial N} = Q \tag{17.97}$$

where $u^{\mathrm{n}}$ is now the zonal velocity calculated along a line of constant $N_2O$ mixing ratio, while $v^{\mathrm{n}}$ is normal to that direction. The quantity $\dot{N}$ plays the same role as the vertical velocity, and because the mixing ratio is conserved along the line, we have $\dot{N} = 0$. Using the usual perturbation method, we have

$$\theta_t + \overline{u'\theta'_x} + \overline{v'\theta'_y} + \overline{v}\overline{\theta}_y = \overline{Q} \tag{17.98}$$

And with good approximation, equilibrium is established between the last term of the left-hand side and the right-hand side

$$\overline{v} = \overline{Q}/\overline{\theta}_y \tag{17.99}$$

These relations allow us to obtain a very simple picture, at least in the polar region, based on the mixing ratio of nitrous oxide, as shown in Fig. 17.15. The polar vortex is indicated by the isopleths of the zonal velocity, while the mixing ratio isopleths are the thick lines. These lines are bent downward by the meridional circulation. The meridional gradient inside the vortex increases because in that region the mixing is negligible, or at least weaker than, outside the vortex. The subsidence inside the vortex means the existence of a strong cooling in that region.

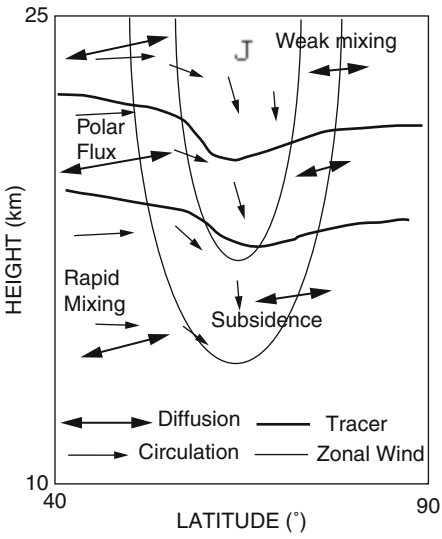



**Fig. 17.15** The dynamical processes that are present in the polar vortex (Adapted from Schoeberl et al. 1992)

In practice, we can establish that velocities are of the order of 0.05 cm $s^{-1}$, corresponding to a cooling rate of a few degrees per day.

The polar region is then a dynamically isolated region and this will be a reason for the existence of the ozone hole as we shall see in the next chapter.

## E.17 Examples

### *E.17.1 Troposphere–Stratosphere Exchange*

A rather interesting application of what we have said so far is the evaluation of the atmospheric residence time of some species. This approach is taken from Holton in the reference list of this chapter. In Fig. 17.8, we show two types of box models similar to the one we have studied in the previous chapter to evaluate the lifetime of chlorofluorocarbons. These models are specific for studying the exchange between troposphere and stratosphere. We assume then $M_N$ to be the mass of a particular species in the northern hemisphere and $M_S$ the corresponding mass in the southern hemisphere. We also indicate with $F_N$ and $F_S$ the flux exchanged between north and south and vice versa. The species is also destroyed at a rate $\alpha$ in both hemispheres. Similarly, $E_N$ and $E_S$ are the rate of emission in the two hemispheres. We have

$$\frac{dM_N}{dt} = E_N + F_S - \alpha M_N \quad \frac{dM_S}{dt} = E_S + F_N - \alpha M_S \tag{E.17.1}$$

The fluxes $F_{N,S}$ can be expressed as a function of the masses as

$$F_{N,S} = \beta M_{N,S}$$

where $b$ is the atmospheric flux through the equator divided by the mass of air in one hemisphere. In steady state conditions, we have

$$\begin{aligned} E_N - \beta\left(M_N - M_S\right) - \alpha M_N = 0 \\ E_S + \beta\left(M_N - M_S\right) - \alpha M_S = 0 \end{aligned} \tag{E.17.2}$$

For the sake of illustration, we can take a specie that is emitted only in the northern hemisphere so that $E_S = 0$. Eliminating $a$ between the (E.17.2), we have

$$\beta^{-1} = \left(M_N^2 - M_S^2\right)/E_N M_S \tag{E.17.3}$$

We can refer now to ethane for which

$$M_N = 2M_S = 9 \cdot 10^{10} \quad E_N = 1.2 \cdot 10^{11} \text{mole year}^{-1}$$

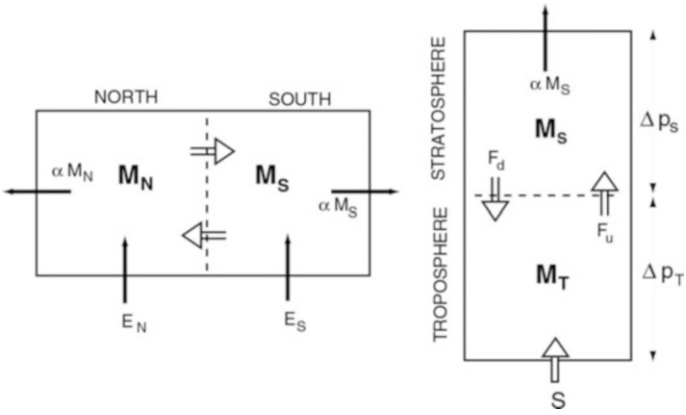



**Fig. E.17.1** Box models for the tropospheric exchange (*left*) and between troposphere and stratosphere (*right*). For the meaning of the symbols see text (From Holton 1992a, b)

In this case, we can calculate that the time for exchange is $\beta^{-1} = 1.125$ years. This data could be evaluated also from Oort and Rasmussen taking the average mass flux through the equator in a year.

For the box model relative to the troposphere stratosphere exchange we could write (Fig. E.17.1)

$$\frac{dM_S}{dt} = F_u - F_d - \alpha M_S \quad \frac{dM_T}{dt} = S - F_u + F_d \tag{E.17.4}$$

In this case, the fluxes are given by

$$F_u = \chi_T W \quad F_d = \chi_S W;$$

where $W$ is the mass flux through the tropopause expressed in kg m$^{-2}$ s$^{-1}$, and $\chi_T$ and $\chi_S$ are the mixing ratios in the troposphere and stratosphere, respectively. The mass per unit surface can be easily calculated

$$M_T = \Delta p_T \chi_T / g \quad M_S = \Delta p_S \chi_S / g$$

so that (E.17.4) can be written as a function of the mixing ratio

$$\begin{aligned} \frac{d\chi_S}{dt} &= -\beta_S (\chi_S - \chi_T) - \alpha \chi_S \\ \frac{d\chi_T}{dt} &= \frac{Sg}{\Delta p_T} - \beta_T (\chi_S - \chi_T) \end{aligned} \tag{E.17.5}$$

where the exchange times in stratosphere and troposphere are defined by

$$\beta_S^{-1} = \Delta p_S / gW \quad \beta_T^{-1} = \Delta p_T / gW;$$

At the steady state, the system (E.17.5) gives

$$\chi_T = S\left(g\Delta p_S^{-1}\right)\left(\alpha^{-1} + \beta_S^{-1}\right) \qquad \chi_S = S\left(g\Delta p_S^{-1}\right)\alpha^{-1} \tag{E.17.6}$$

The most important thing in these relations is that the tropospheric mixing ratio depends on the exchange time in stratosphere. As a matter of fact, these exchange times will depend on latitude, height, and season and their accurate evaluation bring us directly to an important form of control that eddies have on the circulation or in general on the exchange between troposphere and stratosphere.

## *E.17.2 Equatorial Waves*

We have treated extensively the waves in Chap. 10 and then we have all the tools to develop some new application in particular on the equatorial waves. We would like that because we want to talk about the quasi biennal oscillation (QBO) where these kinds of waves seem to have a very important influence. In the equatorial region, the Coriolis parameter will be approximated as $f = \beta y$ so that the momentum equations will be

$$\begin{aligned} \frac{\partial u'}{\partial t} - \beta y v' &= -\partial \Phi' / \partial x \\ \frac{\partial v'}{\partial t} + \beta y u' &= -\partial \Phi' / \partial y \end{aligned} \tag{E.17.7}$$

And the continuity equation

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \rho_0^{-1}\partial\left(\rho_0 w'\right)/\partial z = 0 \tag{E.17.8}$$

The buoyancy equation can be obtained from the (E.10.24) or simply imposing equilibrium between buoyancy and pressure force

$$\rho_0^{-1}\partial p'/\partial z = -g\Delta\theta/\theta_0$$

Substituting for the geopotential $d\Phi' = dp'/\rho_0$ and $\Delta\theta = \Delta z\,(\partial\theta/\partial z)$

$$\partial\Phi'/\partial z + N^2\Delta z = 0$$

where $N^2 = (g/\theta_0)\,\partial\theta/\partial z$. Deriving with respect to time, we get

$$\partial^2\Phi'/\partial t\partial z + w'N^2 = 0 \tag{E.17.9}$$

We assume at this point that the perturbation propagates zonally but also vertically with wave number m. The amplitude growth in height proportionally to $\rho_0^{-1/2}$ and the typical form for the perturbation will be

$$u'(x,y,z,t) = e^{z/2H}\widehat{u}(y)\exp\left[i\left(kx+mz-\nu t\right)\right] \tag{E.17.10}$$

Substituting these forms into (E.17.7), (E.17.8), and (E.17.9), we get a set of ordinary differential equations

$$-i\nu\widehat{u}-\beta y\widehat{v}=-ik\widehat{\Phi} \tag{E.17.11}$$

$$-i\nu\widehat{v}+\beta y\widehat{u}=-\partial\widehat{\Phi}/\partial y \tag{E.17.12}$$

$$\left(ik\hat{u}+\partial\hat{v}/\partial y\right)+i\left(m+i/2H\right)\hat{w}=0 \tag{E.17.13}$$

$$\nu\left(m-i/2H\right)\widehat{\Phi}+\widehat{w}N^2=0 \tag{E.17.14}$$

### E.17.2.1 Kelvin Waves

The first case we can work out refers to Kelvin waves. At the equator, these can propagate only zonally and vertically. We put then $\hat{v}=0$ and eliminate $\hat{w}$ in the last two equations

$$i\nu\widehat{u}=ik\widehat{\Phi} \tag{E.17.15}$$

$$\beta y\widehat{u}=-\partial\widehat{\Phi}/\partial y \tag{E.17.16}$$

$$-\nu\left(m^2+1/4H^2\right)\widehat{\Phi}+\widehat{u}kN^2=0 \tag{E.17.17}$$

These are equivalent to two equations

$$\nu\partial\widehat{u}/\partial y=-k\beta y\widehat{u} \tag{E.17.18}$$

$$-\left(m^2+1/4H^2\right)(\nu/k)^2+N^2=0 \tag{E.17.19}$$

In these equations, the phase velocity is $c=\nu/k$ so that the first gives the simple solution

$$\widehat{u}=u_0\exp\left(-\beta y^2/2c\right) \tag{E.17.20}$$

In order for the solution to be limited, the phase velocity must be positive that is the wave propagates eastward. As for (E.17.19), this represents the dispersion relation

$$c^2 = N^2/\left(m^2 + 1/4H^2\right) \tag{E.17.21}$$

for $m^2 \gg 1/4H^2$ this reduces to $c = \pm N/m$ that is the same as internal gravity waves. This can be generalized for a gravity wave having a wavenumber **k** with component $k$ and $m$ that is

$$c = \pm N/|\mathbf{k}| \Rightarrow \nu = \pm Nk/|\mathbf{k}|$$

The plus sign is taken for eastward propagation and $|\mathbf{k}| = \left(k^2 + m^2\right)$. If we assume $k > 0$ and $m < 0$, then the lines of constant phase must tilt eastward with increasing height. This because for the phase $\phi = kx + mz$ to remain constant as x increase also z must increase.

### E.17.2.2 Rossby Gravity Waves

With no restrictions Eqs. (E.17.11), (E.17.12), (E.17.13), and (E.17.14) can be solved by eliminating $\hat{u}$ and $\hat{w}$ with the approximation that $m^2 >> 1/4H^2$. In this case, we get easily

$$\widehat{v}\left[\beta^2 y^2 - \nu^2\right] = ik\beta y\widehat{\Phi} + i\nu\partial\widehat{\Phi}/\partial y \tag{E.17.22}$$

$$\widehat{\Phi}\left[\nu^2 - gh_e k^2\right] + igh_e\nu\left(\partial\widehat{v}/\partial y - \beta ky\widehat{v}/\nu\right) \tag{E.17.23}$$

where $gh_e = N^2/m^2$. Now (E.17.23) can be substituted in (E.17.22) to eliminate $\widehat{\Phi}$ and obtain a second order differential equation in $\hat{v}$

$$\frac{d^2\hat{v}}{dy^2} + \left[\left(\frac{\nu^2}{gh_e} - k^2 - \frac{k}{\nu}\beta\right) - \frac{\beta^2 y^2}{gh_e}\right]\hat{v} = 0 \tag{E.17.24}$$

Before discussing this equation, we can examine some special case for example when $h_e \to \infty$ that is $m \to 0$ that is the motion in non divergent. In this case, (E.17.24) reduces to

$$\frac{d^2\hat{v}}{dy^2} + \left[\left(-k^2 - \frac{k}{\nu}\beta\right)\right]\hat{v} = 0$$

solutions of the form $\hat{v} \sim \exp(ily)$ do exist if the following dispersion relation holds

$$-l^2 - k^2 - \frac{k}{\nu}\beta = 0 \quad \Rightarrow \quad \nu = -\beta k\Big/\left(l^2 + k^2\right)$$

On the other hand, if rotation is eliminated ($\beta = 0$) equation becomes

$$\frac{d^2\hat{v}}{dy^2} + \left[\left(\frac{\nu^2}{gh_e} - k^2\right)\right]\hat{v} = 0$$

and with a solution of the form $\hat{v} \sim \exp(ily)$, we get a dispersion relation

$$\nu = \pm\left[gh_e\left(k^2 + l^2\right)\right]^{1/2}$$

which corresponds to the shallow water gravity model. In the most general case with the condition $\hat{v} = 0$ per $y \to \pm\infty$, the problem reduces to an eigenvalue case and the solution to (E.17.24) exists only when the coefficient in square brackets satisfies the relationship

$$\frac{\sqrt{gh_e}}{\beta}\left(\frac{\nu^2}{gh_e} - k^2 - \frac{k}{\nu}\beta\right) = 2n + 1; \qquad n = 0, 1, 2, .... \tag{E.17.25}$$

For the $n = 0$ mode, the dispersion relation gives

$$\left(\frac{\nu^2}{gh_e} - k^2 - \frac{k}{\nu}\beta\right) = \frac{\beta}{\sqrt{gh_e}}$$

That can be written as

$$\left(\frac{\nu}{\sqrt{gh_e}} - \frac{\beta}{\nu} - k\right)\left(\frac{\nu}{\sqrt{gh_e}} + k\right) = 0$$

this implies that $\left(\nu/\sqrt{gh_e} - \beta/\nu - k\right) = 0$ and then

$$|m| = N\nu^{-2}(\beta + \nu k) \tag{E.17.26}$$

For no rotation ($\beta = 0$) again we find the hydrostatic gravity waves. However, we see that $\beta \neq 0$ introduces an asymmetry between westward propagating modes ($\nu < 0$) and eastward propagating modes ($\nu > 0$) because eastward propagating waves have shorter vertical wavelengths.

From (E.17.26) we see that in order to have vertical propagation ($m > 0$) must be

$$(\beta + \nu k) > 0$$

so that the phase velocity

$$c = \nu/k > -\beta/k^2$$

The westward $n = 0$ propagating mode is referred as Rossby-gravity wave mode. For upward energy propagation, this mode must have downward phase propagation.

We have then found two possible tropical wave: Kelvin eastward waves and westward Rossby gravity wave. We will find an application for then in studying the quasi biennial oscillation (QBO).

## *E.17.3 The Simplest Theory on Quasi-Biennial Oscillation*

The quasi-biennial oscillation is a quasi-periodic phenomenon that interests the stratosphere. We have seen in this chapter that the zonal winds in the stratosphere have a periodic behavior being easterly in the summer hemisphere and westerly in the winter. However, when the long terms means are subtracted there remains an oscillatory regimes for zonal winds. In particular, zonally symmetric easterly and westerly winds regimes alternate regularly with a period varying from about 24–30 months. Successive regimes first appear above 30 km, but propagate downward at a rate of 1 km month$^{-1}$. These oscillations are symmetric about the equator with a half width of about 12° and an amplitude of roughly 20 ms$^{-1}$ at the equator. Figure E.17.2 shows the behavior of the QBO for the last 30 years.

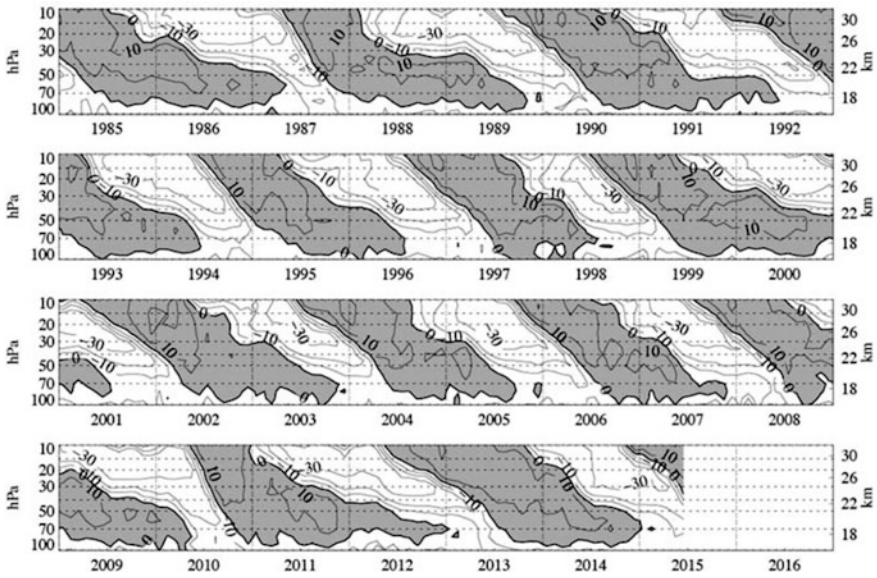



**Fig. E.17.2** Time-height section of monthly mean zonal winds (m/s) at equatorial stations: Canton Island, 3°S/172°W (Jan 1953 – Aug 1967), Gan/Maledive Islands, 1°S/73°E (Sep 1967 – Dec 1975) and Singapore, 1°N/104°E (since Jan 1976). Isopleths are at 10 m/s intervals; westerlies are shaded (Updated from Naujokat 1986)

The theory has received contribution from Plumb (1977) and Lindzen and Holton (1978), and we will refer to a simplified two-level model by Yoden and Holton (1988). All the theories refer to the interaction between waves and the mean flow.

A good starting point for the theory is to consider the acceleration of the mean flow due to the convergence of the eddy wave flux and the upward diffusion.

$$\frac{\partial \overline{u}}{\partial t} = -\frac{\partial F}{\partial z} + K\frac{\partial^2 \overline{u}}{\partial z^2} \tag{E.17.27}$$

where K is equivalent to a diffusion coefficient.

The EP flux for each wave evolves according to the relation

$$F_n(z) = F_{0n} \exp\left(-\int_0^z \Lambda_n\left(z'\right) dz'\right) \tag{E.17.28}$$

where $\Lambda_n$ is the attenuation rate for the $n$th component

$$\Lambda_n = \frac{\alpha}{(\partial\omega/\partial m)_n} = \frac{\alpha N^2}{m_n k(\overline{u} - c_n)^3} \cong \frac{\alpha N}{k(\overline{u} - c_n)^2} \tag{E.17.29}$$

where $\alpha$ is the thermal damping rate. This means that in westerly flow ($\overline{u} > 0$) a wave of westerly phase speed is selectively absorbed and gives enhanced westerly forcing of the mean flow. Actually (E.17.29) refers to Kelvin wave which do have westerly phase speed. For Rossby gravity waves, the coefficient is very much different

$$\Lambda_n = \frac{\alpha N}{k(\overline{u} - c)^2}\left\{\frac{\beta}{k\,(\overline{u} - c)} - 1\right\} \tag{E.17.30}$$

Using (E.17.29) with the same k, the same zonal velocity and just the opposite phase speed, we deduce that the wave with easterly phase speed is not very much attenuated and may give rise to a easterly acceleration of the mean flow. This represents the basic mechanism of the QBO explanation as shown in Fig. E.17.3.

The figure shows a zonal flow slightly perturbed in presence of two waves with equal and opposite phase speed +c and –c. According to what just we said the +c wave will be attenuated and give rise to an acceleration of the zonal flow, while the –c wave will penetrate to a higher altitude and generate an easterly flow there. This process will continue until all the westerly flow will be destroyed and the process could start again to create and initial westerly flow high up.

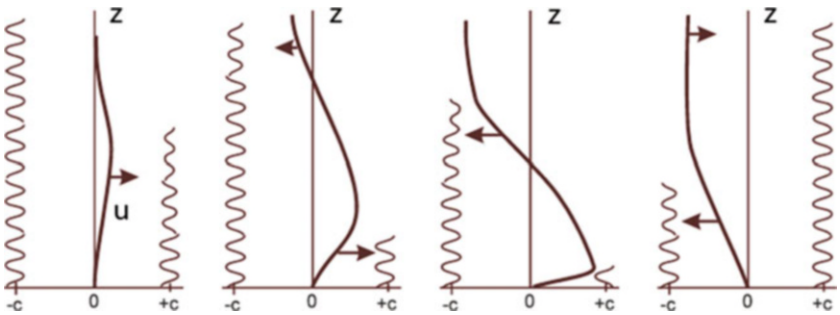
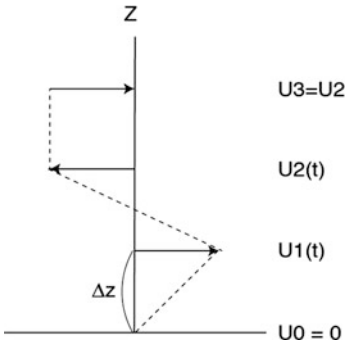



**Fig. E.17.3** Wave propagation as an explanation for QBO

**Fig. E.17.4** The two-level model. *Bottom* ($U_0$) and *top* ($U_3$) are governed by boundary conditions




### E.17.3.1 The Detailed Two Level Model

We will follow the method outlined by Yoden and Holton (1988) where the atmosphere is reduced to four levels where the wind is defined. According to Eqs. (E.17.27), (E.17.28), and (E.17.29), we have

$$\frac{\partial u\left(z,t\right)}{\partial t}=\sum_{n=1}^{N}F_n(0)\Lambda_n\left(z,t\right)\exp\left[-\int_0^z\Lambda_n\left(z',t\right)dz'\right]+K\frac{\partial^2u\left(z,t\right)}{\partial z^2}\qquad\text{(E.17.31)}$$

with the top and bottom boundary conditions given by

$$\frac{\partial u\left(z,t\right)}{\partial t}=0\quad\text{at}\quad z=z_T;\quad u=0\quad\text{at}\quad z=0$$

In general, the evaluation of the integral is made with a simple trapezoidal rule while the second order derivative is reduced to a difference equation. So, in general, we have for the j level $U_j(t)=u\left(j\Delta z,t\right)$ and then Fig. E.17.4

$$\frac{\partial U_j}{\partial t} = \sum_{n=1}^{N} F_n(0) G_{nj} \times \exp\left[ -\left\{ \frac{1}{2}\left(G_{n0} + G_{nj}\right) + \sum_{k=1}^{j-1} G_{nk} \right\} \Delta z \right] + K \frac{U_{j+1} - 2U_j + U_{j-1}}{\Delta z^2} \tag{E.17.32}$$

where

$$G_{nj} = \frac{\alpha}{k_n \left(U_j - c_n\right)^2}$$

and the boundary conditions

$$U_J = U_{J-1} \qquad U_0 = 0$$

We can now apply Eq. (E.17.32) and get

$$\frac{\partial U_1}{\partial t} = \sum_{n=1}^{N} F_n(0) \frac{\alpha}{k_n (U_1 - c_n)^2} \times \exp\left[ -\frac{\alpha \Delta z}{2k_n} \left\{ \frac{1}{(U_1 - c_n)^2} + \frac{1}{c_n^2} \right\} \right] + K \frac{U_2 - 2U_1}{\Delta z^2} \tag{E.17.33}$$

$$\frac{\partial U_2}{\partial t} = \sum_{n=1}^{N} F_n(0) \frac{\alpha}{k_n (U_2 - c_n)^2} \times \exp\left[ -\frac{\alpha \Delta z}{2k_n} \left\{ \frac{1}{(U_2 - c_n)^2} + \frac{2}{(U_1 - c_n)^2} + \frac{1}{c_n^2} \right\} \right] + K \frac{U_1 - U_2}{\Delta z^2} \tag{E.17.34}$$

These are the equation that will be used in the model implemented with a Matlab code. Figure E.17.5 shows the results obtained with the code build around Robinson (2001) Stella program.

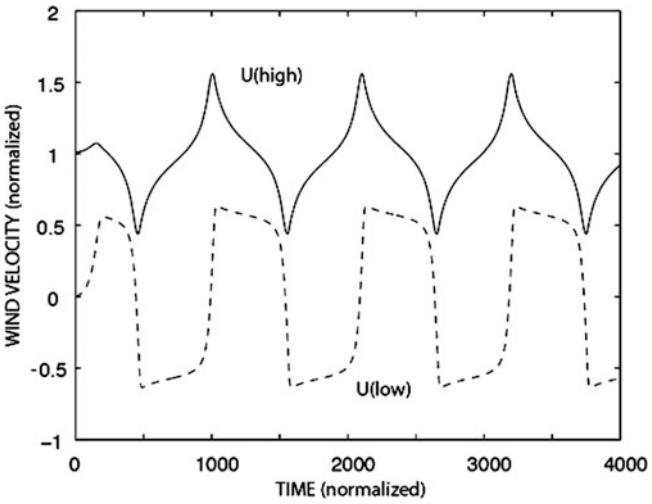



**Fig. E.17.5** The behavior of *upper* (*U*(high)) and *lower level* (*U*(low)) wind obtained with the Matlab code for the Quasi Biennial Oscillation

### E.17.3.2 The Matlab Code for the QBO

```
run.qbo
[t,x]=ode45('qbo',[0 4000],[0.01 0.0]);
plot(t,x(:,1)+1,t,x(:,2),'--')

%program qbo
function xdot=qbo(t,x)
xdot=zeros(2,1);
alfa=0.5;
deltaz=1;
k=1;
taudiff=33;
f=0.08;
alfadz2k=alfa*deltaz/(2*k);
ceast=1;
cwest=-ceast;
falfadk=f*alfa/k;
uhighminceast=(x(1)-ceast)^2;
uhighmincwest=(x(1)-cwest)^2;
ulowminceast=(x(2)-ceast)^2;
ulowmincwest=(x(2)-cwest)^2;
work1=falfadk/uhighminceast;
work2=exp(-alfadz2k*(1/uhighminceast+2/ulowminceast+1/ceast^2));
inputh=work1*work2;
```

```
help1=(x(1)-x(2))/taudiff;
help2=falfadk/uhighmincwest;
help3=exp(-alfadz2k*(1/uhighmincwest+2/ulowmincwest+1/cwest^2));
outh=help1+help2*help3;
werk1=falfadk/ulowminceast;
werk2=exp(-alfadz2k*(1/ulowminceast+1/ceast^2));
inputl=werk1*werk2+(x(1)-x(2))/taudiff;
holp1=x(2)/taudiff;
holp2=falfadk/ulowmincwest;
holp3=exp(-alfadz2k*(1/ulowmincwest+1/cwest^2));
outl=holp1+holp2*holp3;
xdot(1)=inputh-outh;
xdot(2)=inputl-outl;
```

## References[1]

### *Books*

Andrews DG, Holton JR, Leovy CB (1987) Middle atmosphere dynamics. Academic, New York
Holton JR (1975) The dynamic meteorology of the stratosphere and mesosphere, Meteorol. Monogr. No. 37. American Meteorological Society
Holton JR (1992a) An introduction to dynamic meteorology. Academic, New York

### *Articles*

Garcia RR (1991) Parameterization of planetary wave breaking in the middle atmosphere. J Atmos Sci 48:1406
Garcia RR, Solomon S (1983) A numerical model of the zonally averaged dynamical and chemical structure of the middle atmosphere. J Geophys Res 88:1379
Haynes PH et al (1991) On the downward control of extratropical diabatic circulations by eddy induced mean zonal forces. J Atmos Sci 47:651
Holton JR (1986) A dynamically based transport parameterization for the one dimensional photochemical models of the stratosphere. J Geophys Res 91:2681
Holton JR (1992b) Dynamics of the middle atmosphere: its role in transport and troposphere-stratosphere coupling. In: Gille JC, Visconti G (eds) The use of EOS for studies of atmospheric physics. Elsevier, Amsterdam/New York

[1]Except for the books of Andrews et al. and Holton, this chapter is mainly based on research and review papers. The field is evolving too fast for the arguments to be settled in books. I would recommend the papers by Garcia, Holton, Plumb, McIntyre, and Schoeberl. This chapter has been mainly updated through the example.

McIntyre ME (1992) Atmospheric dynamics: some fundamentals with observational implications. In: Gile JC, Visconti G (eds) The use of EOS for studies of atmospheric physics. Elsevier, Amsterdam/New York

Naujokat B (1986) An update of the observed quasi-biennial oscillation of the stratospheric winds over the tropics. J Atmos Sci 43:1873–1877

Plumb RA (1996) A "tropical pipe" model of stratospheric transport. J Geophys Res 101, 3957ù

Plumb RA, Ko MKW (1992) Interrelationship between mixing ratio of long lived stratospheric constituents. J Geophys Res 97:10145

Schoeberl MR et al (1992) The structure of the polar vortex. J Geophys Res 97:7859

Schoeberl MR, Lait LR (1992) Conservative-coordinate transformations for atmospheric measurements. In: Gile JC, Visconti G (eds) The use of EOS for studies of atmospheric physics. Elsevier, Amsterdam/New York

Schoeberl MR, Sparling LC, Jackman CH, Fleming EL (2000) A lagrangian view of stratospheric trace gas distributions, 105, 1537

Yoden S, Holton JR (1988) A new look at Equatorial Quasi Biennal Oscillation Models. J Atmos Sci 45:2703

# Chapter 18
# Stratospheric Chemistry

The basis for studying the stratospheric chemistry has been laid out in the chapter on tropospheric chemistry (Chap. 16). In that case, we dealt mainly with the gas sources, but we have introduced the main tools that are used in the photochemical studies of the atmosphere. In the troposphere, some of the gases have a quite short lifetime because, for example, they are dissolved in water or are rapidly oxidized so that they hardly can penetrate the stratosphere.

As shown in Fig. 18.1, these gases are dissociated to produce radicals of chlorine, nitrogen, and hydrogen (not shown in the figure). The chlorine radicals are produced from the photodissociation of chlorofluorocarbons (CFC) but also from natural compounds. Nitrogen oxides are emitted directly in the atmosphere from combustion processes (in the upper troposphere by the aircraft) or may be produce by the nitrous oxide decomposition. The production of hydrogen radical is due mainly to the decomposition of water vapor through dissociative reaction with metastable oxygen (1O).

The great interest for ozone is because it absorbs most of the ultraviolet solar radiation and so may protect living things from damage to their DNA. In the last 10 years or so, this has been a very hot topic because we now understand that the anthropically produced CFCs are responsible for the ozone depletion.

Ozone is produced in the tropical stratosphere and from there transported to high latitude in the lower stratosphere. Radicals have a very important role in the ozone chemistry because they accelerate its recombination with atomic oxygen and so facilitating ozone destruction. In the last few years and in the context of the ozone problem, a great interest has been growing about heterogeneous chemistry when the chemical reactions may happen between different phases, like gas and liquids or solids. Some of the reactions develop on the surface of sulfate particles that are part of the Junge layer, a thin aerosol layer present in the lower stratosphere. This layer is refurbished in a sporadic manner by the catastrophic volcanic eruptions that produce large amount of sulfur dioxide. The dynamical isolation of the polar region on the other hand determines such low temperatures there in the winter stratosphere



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_18

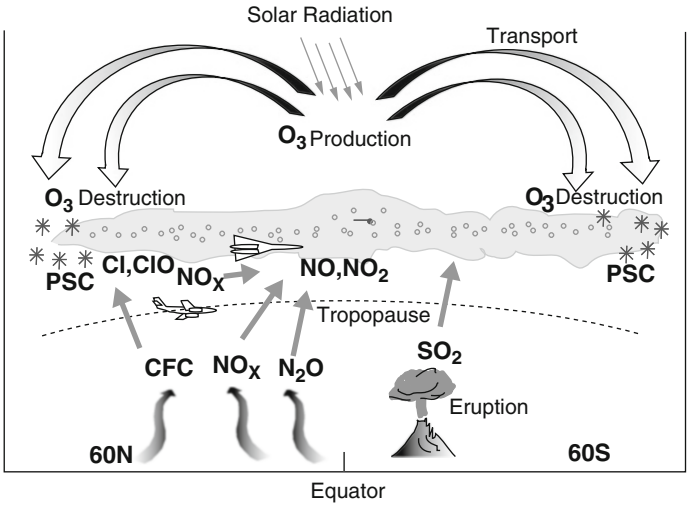



**Fig. 18.1** A cartoon showing the main chemical processes in the stratosphere. In this case the main attention is to the ozone problem. Figure is not in scale

that cause the condensation of nitric acid or even water vapor, that is, the formation of polar stratospheric clouds (PSC). These clouds have a fundamental role in the ozone destruction in the polar regions.

The scientific research on the ozone problem has never been academic, and even in the 40s it had a great relevance in meteorology. Looking at Eq. (16.8) in the simple Chapman model, the ozone lifetime is given by $(2k_1[O])^{-1}$ that because of the very low atomic oxygen concentration in the 20–30 km region is of the order of several weeks. Ozone could then be used as a tracer to study the dynamics of the lower stratosphere. The interest of the meteorologist of the time was then anticipating the important role that would later be recognized to this topic as we have shown in the previous chapter.

## 18.1 The Ozone Distribution

Our starting point could be the study of the ozone distribution as shown in Fig. 17.2. In this figure, we report the ozone columnar density as a function of latitude and season. The columnar density is defined as

$$N(z) = \int_z^\infty \left[O_3\right] dz \tag{18.1}$$

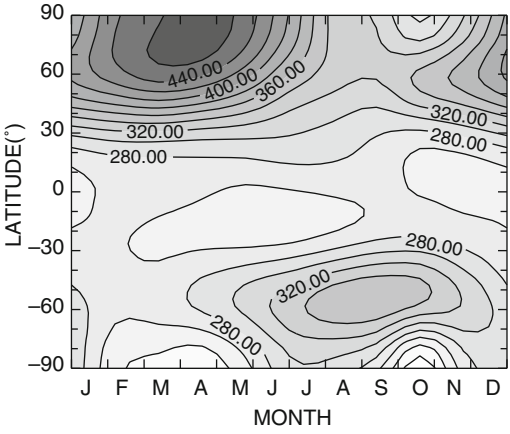



**Fig. 18.2** The columnar ozone distribution as a function of season and latitude. The units are Dobson and the negative latitudes refer to the southern hemisphere

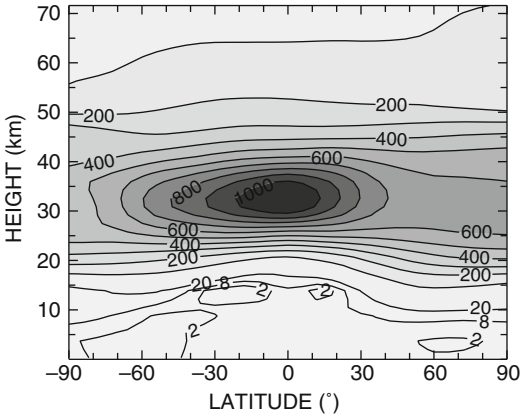



**Fig. 18.3** The zonal mean ozone mixing ratio at the equinox. Spring on the right in the northern hemisphere. Units are 0.01 ppm

so that $N(z)$ is the number of molecules for unit surface above a certain altitude. In the case of ozone, a unit is used to measure the columnar density called *Dobson* (in name of the English physicist George M. Dobson, a pioneer of the ozone studies) that corresponds to $2.68 \cdot 10^{16}$ molecules cm$^{-2}$. Because at STP we have $2.68 \cdot 10^{19}$ molecules cm$^{-3}$, this means that 1 cm – atm of ozone corresponds to 1000 Dobson. As shown in Fig. 18.2, in the atmosphere there is only 0.3–0.4 cm of ozone. The first impression we have from Fig. 18.2 is that the columnar density of ozone has a quite marked variation during the season that is not symmetric between the two hemispheres. The maximum amount of ozone occurs in winter at high latitudes. This is followed by a minimum in spring that is much more evident in the southern hemisphere. The fact that the maximum amount of ozone is found in winter is not surprising because in this season there is a very small amount of atomic oxygen so that the ozone lifetime is quite long.

To have an idea of the role of the dynamics in determining the ozone amount, we can refer to Figs. 18.3 and 18.4. The first of these figures shows that at high latitude

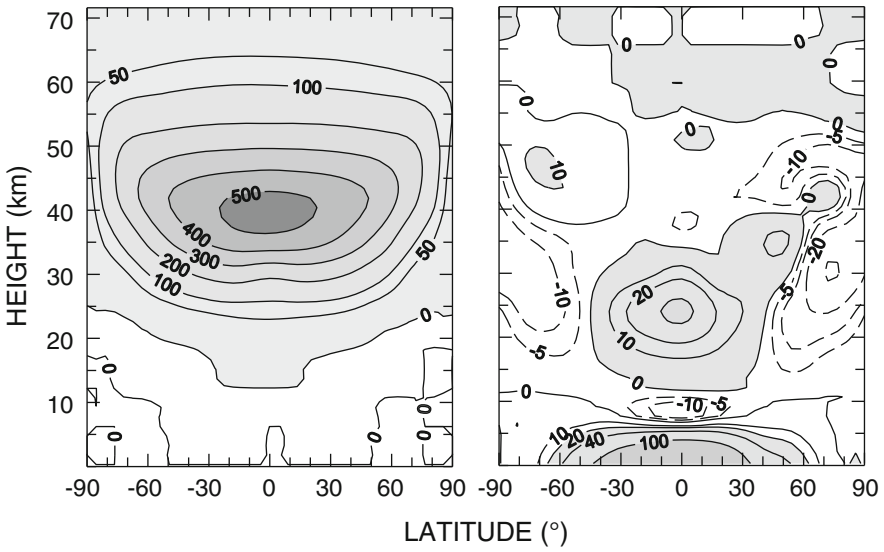



**Fig. 18.4** The production (*left*) and the net source for ozone at the equinox. Units are $10^5$ molecules $cm^{-3}$ $s^{-1}$. *Shaded areas* indicated positive values, while the *dashed lines* show net loss

the isopleths of ozone tend to be lower in altitude as we go to higher latitudes. This means that the ozone maximum at high latitudes is due in large part to the density increase in the lower stratosphere. This data is quite surprising especially if we compare it with Fig. 18.4 that shows the ozone production at the equinox. The production rate of ozone is easy to calculate because as we have seen in Chap. 16, the main mechanism is oxygen photodissociation so that the production depends on the amount of oxygen and solar radiation. From Fig. 18.4 we see that the production has a maximum in the tropical stratosphere while decreases considerably at high latitude and in the lower stratosphere. This suggests that ozone is transported from the "production" region at the tropics to the "reservoir" regions at high latitudes. The dynamics of the stratosphere is responsible for such transport. Actually dynamics redistributes ozone also in regions where it can be more easily destroyed.

## 18.2 The Ozone Homogeneous Chemistry

Time has come to see what are the chemical processes that determine the production and destruction of ozone in the stratosphere. As illustrated in Fig. 18.1, the species that contribute to stratospheric chemistry have their source at the surface like the CFCs.

Once in the stratosphere, these species release the reacting gases through a number of photochemical processes. The solar radiation remains the most important

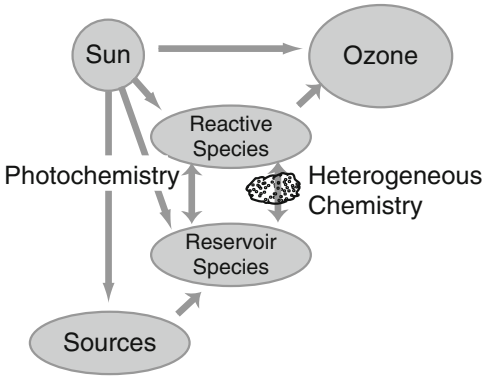



**Fig. 18.5** A cartoon showing the ozone chemistry in the stratosphere. *Arrows* indicated the conversion of sources in reactive or reservoir species (Adapted from Brune 1992)

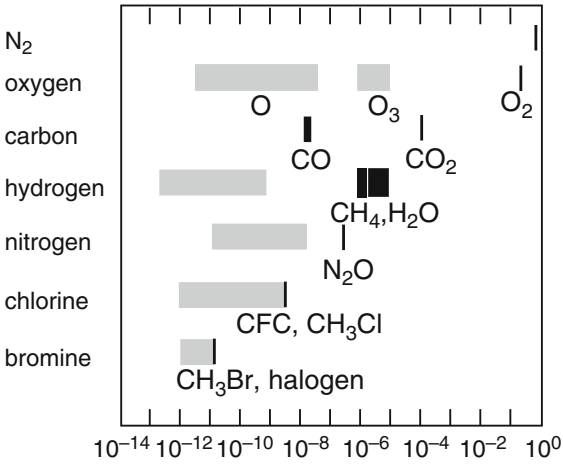



**Fig. 18.6** The volume mixing ratio of the species that determine the stratospheric chemistry. The *bold lines* give the mixing ratio in the troposphere, while the *shaded regions* give the range for the reactive and reservoir species

factor in the activation of the stratospheric chemistry. The main species contributing to the ozone chemistry are either produced directly by the source gases or through intermediate compounds called *reservoirs*. The release of reactive species may be the results of either photodissociation or some heterogeneous process. A schematic view of these processes is given in Fig. 18.5.

The most surprising thing about stratospheric chemistry is that the concentrations of the reacting gases are extremely small and in any case much smaller than the source gases that we find in the troposphere. Figure 18.6 shows with more detail this aspect. We notice that with respect to the most abundant atmospheric components (oxygen and nitrogen), the source gases have already very small mixing ratios that are even smaller in the stratosphere. In particular, the bromine and chlorine sources produce species that have a great efficiency in destroying ozone present with mixing ratios of the order of part per billion (ppb) or even part per trillion (ppt).

Another observation on this figure is about the grouping of the species in families, that is, odd oxygen, odd hydrogen, nitrogen, chlorine, and bromine. The

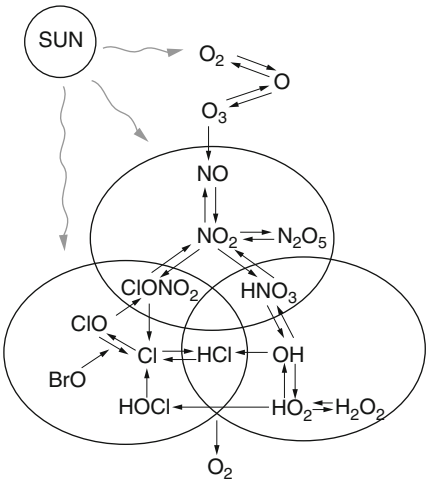



**Fig. 18.7** A simple scheme of the most important gas phase reactions for the stratospheric chemistry. The ellipses include the nitrogen, chlorine, and hydrogen families (From Brune 1992)

most important components of these families are shown in Fig. 18.7 where also the most important gas phase reactions are shown. The photodissociation is the main mechanism through which the radicals are released from the reservoir species ($HCl$, $N_2O_5$, $ClONO_2$, $HNO_3$, $HOCl$). The different families have some "common member" like chlorine nitrate ($ClONO_2$) for the chlorine and nitrogen families, hydrogen chloride ($HCl$), and hypochlorous acid ($HOCl$) that relate hydrogen and chlorine families. The nitric acid ($HNO_3$) is common between the nitrogen and hydrogen families. In the chlorine families, we also include bromine that interacts strongly with the chlorine radicals.

The figure does not show that all the reactions are mediated somehow by other species, and this is quite important because one of the species is just ozone. The final result is that when ozone enters the cycles shown in the figure, it is released as molecular oxygen. We will see this in detail in the next paragraph where we will also understand the meaning of the double arrows shown in the Fig. 18.7.

### *18.2.1 The Catalytic Cycles in the Gaseous Phase*

We have already seen in Chap. 16 that ozone is formed in the reaction between molecular and atomic oxygen that may form itself from the oxygen photodissociation. The complete path is given by

$$\begin{aligned} O_2 + h\nu &\rightarrow O + O \\ O_2 + O + M &\rightarrow O_3 + M \end{aligned} \tag{18.2}$$

$O_3$ formed can be destroyed through photolysis or recombination with $O$

$$\begin{aligned} O_3 + h\nu &\rightarrow O_2 + O \\ O_3 + O &\rightarrow 2O_2 \end{aligned} \tag{18.3}$$

As we mentioned before the first of these reactions does not corresponds to the ozone destruction because atomic oxygen is actually a different form of odd oxygen. The second reaction converts actually an ozone molecule and an oxygen atom in molecular oxygen, and this actually corresponds to a net loss of odd oxygen. The interesting thing is that such a process can be accelerated by some species that act as catalysts. The general scheme is given by

$$\begin{aligned} & X + O_3 \rightarrow XO + O_2 \\ & XO + O \rightarrow X + O_2 \\ \text{net}: \quad & O + O_3 \rightarrow 2O_2 \end{aligned} \tag{18.4}$$

The first and second reactions are always very fast so that if $k$ indicates the reaction rate, the odd oxygen destruction rate can be taken as $2k[O][X]$ so that the model for the pure oxygen atmosphere studied before gives

$$d\left[O_x\right]/dt = 2J_2\left[O_2\right] - 2\left[O\right]\left\{k\left[O\right] + k_3\left[O_3\right]\right\} \tag{18.5}$$

The ratio between oxygen and ozone does not change in presence of the catalyst

$$\left[O\right]/\left[O_3\right] = J_3/k_2\left[O_2\right]\left[M\right]$$

that substituted in (18.5) gives a relation similar to the one found before

$$\left[O_3\right] = \left[O_2\right]\left\{J_2k_2\left[M\right]/k_3J_3\left(1 + A\right)\right\} \tag{18.6}$$

where $A = rk/k_3$ with $r$ ratio between atomic oxygen and ozone. The net effect of the catalyst is to accelerate the recombination by a factor given by $k_3(1 + A)$. It is to notice that normally $k/k_3 \approx 10^4$ so that even in regions where the oxygen abundance is low, the effect on ozone may be important. This simple discussion emphasizes that the cycle (18.4) in order to close needs atomic oxygen that is not very abundant in the lower stratosphere. Actually this has been one of the main difficulties that have been encountered in the explanation of the ozone hole.

### *18.2.2 The Odd Hydrogen Catalytic Cycle*

Historically, the first cycle to be proposed was based on the hydrogen radicals $OH$ an $HO_2$ according to the reactions

$$\begin{aligned} & OH + O_3 \rightarrow HO_2 + O_2 \\ & HO_2 + O \rightarrow OH + O_2 \\ \text{net} \quad & O + O_3 \rightarrow 2O_2 \end{aligned} \tag{18.7}$$

The second reaction may also acts directly on ozone

$$\begin{aligned} & OH + O_3 \rightarrow HO_2 + O_2 \\ & HO_2 + O_3 \rightarrow OH + 2O_2 \\ \text{net} \quad & O_3 + O_3 \rightarrow 3O_2 \end{aligned} \tag{18.8}$$

The second cycle does not depend on the atomic oxygen so that it will be particularly efficient below 25 km. At higher altitudes also the cycle based on atomic hydrogen could be important

$$\begin{aligned} & H + O_3 \rightarrow OH + O_2 \\ & OH + O \rightarrow H + O_2 \\ \text{net} \quad & O + O_3 \rightarrow 2O_2 \end{aligned} \tag{18.9}$$

Sources for the hydrogen radicals are water vapor, methane, and molecular hydrogen that are subject to the dissociative reaction with metastable oxygen $O^1(D)$:

$$\begin{aligned} & O\left(^1D\right) + H_2O \rightarrow OH + OH \\ & O\left(^1D\right) + CH_4 \rightarrow CH_3 + OH \\ & O\left(^1D\right) + H_2 \rightarrow OH + H \end{aligned} \tag{18.10}$$

Photodissociation on the other hand may be important for hydrogen peroxide and hydroxyl radical production from nitric acid:

$$\begin{aligned} & HNO_3 + h\nu \rightarrow NO_2 + OH \\ & H_2O_2 + h\nu \rightarrow OH + OH \end{aligned} \tag{18.11}$$

The peroxyl radicals form from the reaction

$$H + O_2 + M \rightarrow HO_2 + M$$

A first consideration to make about catalytic cycles is that in principle a single catalyst molecule could destroy all the ozone. This does not happen because the radicals may react with species that may sequester them and consequently slow down their effect. In the case of hydrogen radicals, the main reaction is the formation of nitric acid according to the scheme

$$OH + NO_2 + M \rightarrow HNO_3 + M \tag{18.12}$$

This reaction has the effect of destroying both the nitrogen oxides and the hydroxyl although both species reform through the first of (18.9). Actually the hydrogen radicals are removed because the nitric acid transported down in the troposphere is dissolved in the raindrops and is rained out. A similar process

happens also in the polar atmosphere where nitric acid may form iced crystals or hydrates. In this case the process is called *denitrification*.

Radicals may also react between themselves and are destroyed. The most important reactions are besides (18.12)

$$\begin{aligned} &OH + OH \rightarrow H_2O + O \\ &OH + HO_2 \rightarrow H_2O + O_2 \\ &OH + HNO_3 \rightarrow H_2O + NO_3 \\ &OH + HO_2NO_2 \rightarrow H_2O + NO_2 + O_2 \\ &HO_2 + HO_2 \rightarrow H_2O_2 + O_2 \\ &H_2O_2 + OH \rightarrow HO_2 + H_2O \end{aligned} \tag{18.13}$$

The odd hydrogen concentration is given by $[HO_x] = [H] + [OH] + [HO_2]$. As we said earlier, $[HO_x]$ can be treated as a specie, while its components can be calculated at photochemical equilibrium.

### 18.2.3 The Odd Nitrogen Catalytic Cycle

The effect of nitrogen radicals on ozone is essentially that seen in Chap. 16 for the troposphere:

$$\begin{aligned} &NO + O_3 \rightarrow NO_2 + O_2 \\ &NO_2 + O \rightarrow NO + O_2 \\ \text{net} \quad &O + O_3 \rightarrow 2O_2 \end{aligned} \tag{18.14}$$

The second reaction competes with the photodissociation of nitrogen dioxide:

$$\begin{aligned} &NO_2 + h\nu \rightarrow NO + O \\ &O + O_2 + M \rightarrow O_3 + M \end{aligned} \tag{18.15}$$

Without atomic oxygen, the ozone loss can be attributed to the cycle

$$\begin{aligned} &NO + O_3 \rightarrow NO_2 + O_2 \\ &NO_2 + O_3 \rightarrow NO_3 + O_2 \\ &NO_3 + h\nu \rightarrow NO + O_2 \\ \text{net} \quad &O_3 + O_3 \rightarrow 3O_2 \end{aligned} \tag{18.16}$$

Sources of nitrogen oxides are at the surface (combustion processes) or in the atmosphere (lightning and aircraft) and in the dissociation of nitrous oxide

$$N_2O + O\left(^1D\right) \rightarrow NO + NO \tag{18.17}$$

The main sink is again reaction (18.12) with the formation of nitric acid. The odd nitrogen concentration however includes also species we have not mentioned so far so that the total concentration is given by

$$NO_x = NO_2 + NO + NO_3 + N_2O_5 + HNO_4 + ClONO_2 + BrONO_2$$

Nitrogen pentoxide ($N_2O_5$), nitrous acid ($HNO_4$), chlorine nitrate ($ClONO_2$), and bromine nitrate form from the reactions

$$\begin{aligned} NO_2 + NO_3 + M &\rightarrow N_2O_5 + M \\ HO_2 + NO_2 + M &\rightarrow HO_2NO_2 + M \\ NO_2 + ClO + M &\rightarrow ClONO_2 + M \\ NO_2 + BrO + M &\rightarrow BrONO_2 + M \end{aligned} \tag{18.18}$$

Some of these species are rapidly dissociated so that they are present only at night. A few important reactions must be also noted between nitrogen oxides, chlorine, and peroxyl radicals:

$$\begin{aligned} NO + HO_2 &\rightarrow NO_2 + OH \\ NO + ClO &\rightarrow NO_2 + Cl \end{aligned} \tag{18.19}$$

Actually in the lower stratosphere, the ratio between $NO_2$ and $NO$ is given by

$$[NO_2] / [NO] = \{k_{NO+O_3} [O_3] + k_{NO+ClO} [ClO] + k_{NO+HO_2} [HO_2]\} / J_{NO_2} \tag{18.20}$$

The nitrogen radicals are also responsible for the partition in the odd hydrogen family, that is, reactions (18.8) and (18.19). We have

$$[HO_2] / [OH] \approx k_{OH+O_3} [O_3] / k_{NO+HO_2} [NO] + k_{HO_2+O_3} [O_3] \tag{18.21}$$

The coupling between nitrogen and chlorine cycle is through the third of reactions (18.18) and the photodissociation of chlorine nitrate:

$$[ClO] = J_{ClONO_2} [ClONO_2] / k_{NO+HO_2} [NO_2] \tag{18.22}$$

Together with the odd hydrogen, it is convenient to introduce a large family known as $NO_y = NO_x + HNO_3$. The convenience of this notation is that the new family has essentially only one production term given by the dissociation of nitrous oxide and a loss term that is the nitric oxide rain out. The net production of $NO_y$ is then

$$d\left[NO_y\right] / dt = k_{N_2O+O(^1D)} [N_2O] \left[O\left(^1D\right)\right] - (\text{rainout}) \tag{18.23}$$

In this way is possible to solve two continuity equations (for $NO_y$ and $HNO_3$) and obtain $NO_x$ as a difference. As we have seen, the nitrogen oxides have a primary importance in the heterogeneous chemistry because some species like nitrogen pentoxide can be hydrolyzed so they can be dissolved in the water contained, for example, in volcanic aerosols.

### 18.2.4 The Bromine and Chlorine Catalytic Cycles

Chlorine is produced from the photodissociation of industrial gases like chlorofluorocarbons (CFCs) or also from natural gases like methyl chloride ($CH_3Cl$). Processes are the following:

$$\begin{aligned} &CH_3Cl + h\nu \rightarrow CH_3 + Cl \\ &CH_3Cl + OH \rightarrow Cl + \text{products} \\ &CFCl_3 + h\nu \rightarrow CFCl_2 + Cl \end{aligned} \tag{18.24}$$

Other industrial compounds like carbon tetrachloride or methyl chloroform may produce free chlorine with similar mechanisms. Other compounds may produce bromine. Chlorine atoms may be very efficient catalysts for the ozone destruction according to the process

$$\begin{aligned} &Cl + O_3 \rightarrow ClO + O_2 \\ &ClO + O \rightarrow Cl + O_2 \\ \text{net}\quad &O + O_3 \rightarrow 2O_2 \end{aligned} \tag{18.25}$$

Once again this cycle needs atomic oxygen to close so that in the lower stratosphere the prevailing cycles are

$$\begin{aligned} &Cl + O_3 \rightarrow ClO + O_2 \\ &OH + O_3 \rightarrow HO_2 + O_2 \\ &ClO + HO_2 \rightarrow HOCl + O_2 \\ &HOCl + h\nu \rightarrow OH + Cl \\ \text{net}\quad &2O_3 \rightarrow 3O_2 \end{aligned} \tag{18.26}$$

or based on the $NO_x$ radicals

$$\begin{aligned} &Cl + O_3 \rightarrow ClO + O_2 \\ &NO + O_3 \rightarrow NO_2 + O_2 \\ &ClO + NO_2 + M \rightarrow ClONO_2 + M \\ &ClONO_2 + h\nu \rightarrow Cl + NO_3 \\ &NO_3 + h\nu \rightarrow NO + O_2 \end{aligned} \tag{18.27}$$

that has the same result of the previous cycle.

Bromine may have a similar scheme as (18.25), but actually the main loss for ozone is determined by an interaction between bromine and chlorine radicals according to the following scheme:

$$\begin{aligned} & BrO + ClO \rightarrow Br + Cl + O_2 \\ & Br + O_3 \rightarrow BrO + O_2 \\ & Cl + O_3 \rightarrow ClO + O_2 \\ \text{net} \quad & 2O_3 \rightarrow 3O_2 \end{aligned} \tag{18.28}$$

The efficiency of these cycles again may be limited by the formation of acids (hydrogen chloride and bromide, $HCl$ and $HBr$) that can be dissolved in the raindrops. The formation is through the scheme

$$CH_4 + Cl \rightarrow HCl + CH_3 \tag{18.29}$$

Hydrogen chloride can also be reactivated as chlorine by the reaction

$$HCl + OH \rightarrow Cl + H_2O \tag{18.30}$$

Bromine can be used to show how to proceed with the photochemical calculation of the family and its components. The bromine family is given by

$$Br_x = Br + BrO + HOBr + BrONO_2$$

$$Br_y = Br_x + HBr$$

The net chemical production is given by

$$\begin{aligned} d\left[Br_y\right]/dt = & \left\{k_{OH+CH_3B_r}\left[OH\right] + J_{CH_3B_r}\right\}\left[CH_3B_r\right] \\ & + \left\{J_{CF_2ClBr} + k_{O(^1D)+CF_2ClBr}\left[O\left(^1D\right)\right] + k_{OH+CF_2ClBr}\left[OH\right]\right\}\left[CF_2ClBr\right] \\ & + \left\{J_{CF_3B_r} + k_{O(^1D)+CF_3B_r}\left[O\left(^1D\right)\right]\right\}\left[CF_3B_r\right] - rainout \end{aligned} \tag{18.31}$$

And from the $Br_y$ definition, we get

$$\begin{aligned} \left[Br_y\right] &= \left[Br\right]\left(1 + \frac{\left[BrO\right] + \left[HOBr\right] + \left[BrONO_2\right] + \left[HBr\right]}{Br}\right) \\ &= \left[Br\right]\left[1 + \frac{\left[HBr\right]}{\left[Br\right]} + \frac{\left[BrO\right]}{\left[Br\right]}\left(1 + \frac{\left[HOBr\right] + \left[BrONO_2\right]}{\left[BrO\right]}\right)\right] \end{aligned} \tag{18.32}$$

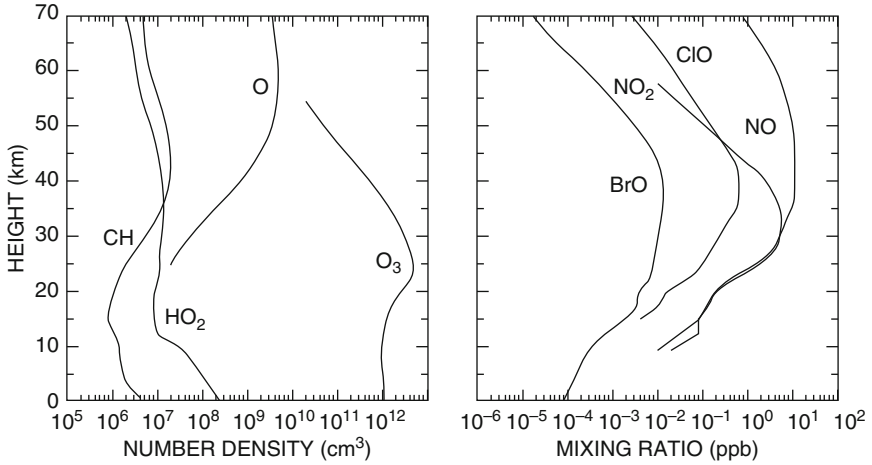



**Fig. 18.8** Concentrations and mixing ratios of some important species for stratospheric chemistry. The figures are based on a two-dimensional model of the stratosphere

Bromine concentration is then

$$[Br] = \left[Br_y\right] / \{1 + C_1 + C_2\,(1 + C_3 + C_4)\} \tag{18.33}$$

where the coefficients are

$$C_1 = [HBr] / [Br]\,; \qquad C_2 = [BrO] / [Br]$$
$$C_3 = [BrONO_2] / [BrO]\,; \quad C_4 = [HOBr] / [BrO]$$

These coefficients are evaluated through iteration, and after that we can obtain the bromine nitrate concentration as

$$[BrNO_3] = \left[Br_y\right] - [Br] - [BrO] - [HOBr] - [HBr] \tag{18.34}$$

It is to notice that like all the radicals, the concentration found with (18.34) depends strongly on the solar zenith angle.

Figure 18.8 shows concentrations and mixing ratios for the different families. These figures are based on two-dimensional model and refer to 45° and the spring season, and they give a fair idea of the number we should expect. Again to have a comparison term we notice that the ozone concentration reported in the figure corresponds to about 8 ppm. The radical concentrations are then really small with respect to ozone. Another important data is that below 40 km the odd oxygen is almost all in the ozone form about 8 ppm. The radical concentrations are then really small with respect to ozone. Another important data is that below 40 km the odd oxygen is almost all in the ozone form.

### 18.2.5 The Effects of the Catalytic Cycles

Figure 18.9 shows the loss rates that can be attributed to the different cycles as calculated by a two-dimensional model of the stratosphere. Although the number may change, this figure gives a good idea of the distribution of loss rates as a

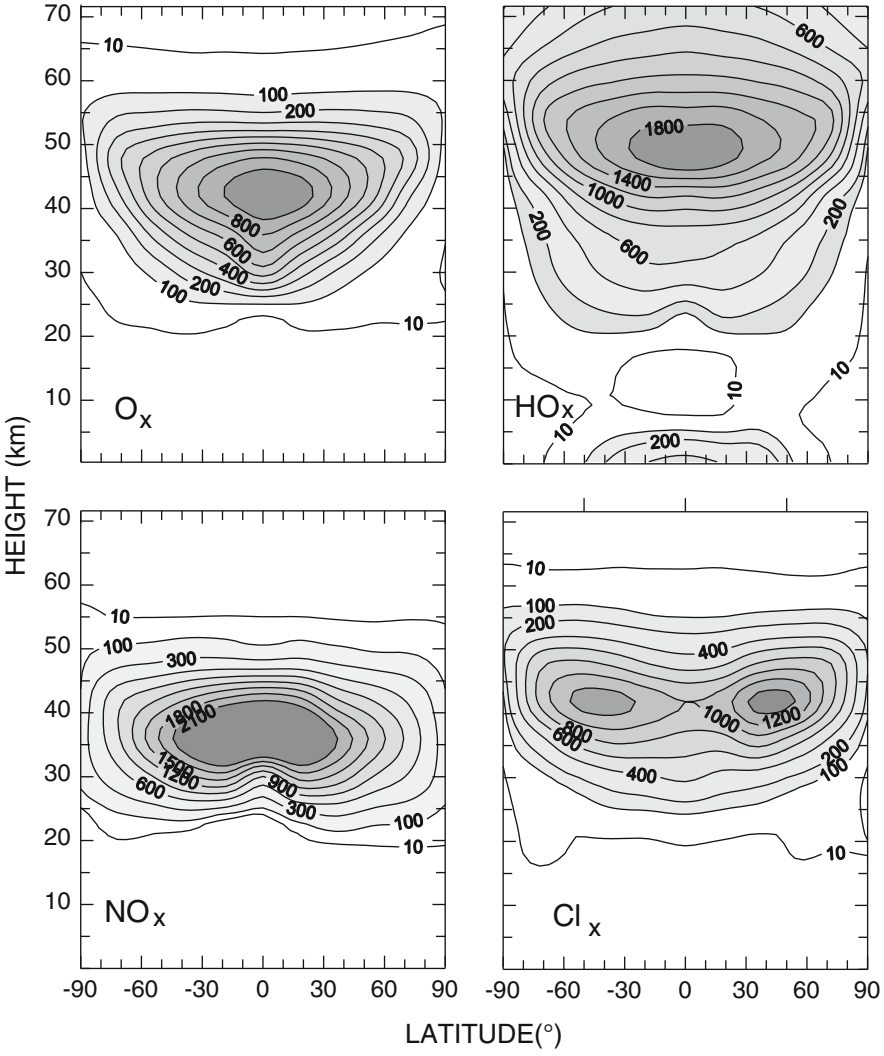



**Fig. 18.9** The destruction rates of the catalytic cycles of odd oxygen ($O_x$), odd hydrogen ($HO_x$), odd nitrogen ($NO_x$), and chlorine ($Cl_x$). Units are $10^3$ odd oxygen molecules $cm^{-3}$ $s^{-1}$ and the data based on a two-dimensional model calculated at the equinox

function of latitude and height. Figure refers to the equinox (this is the reason for the symmetry) and represents the net loss of odd hydrogen in units of 1000 molecules per second and cubic centimeter.

We notice how the cycles that dominate at high altitude are those of oxygen and hydrogen, while at low altitude the nitrogen and chlorine cycles are important. What counts more in the lower stratosphere is the interaction between cycles. Recent results obtained in connection with measurement campaigns have found that in the lower stratosphere between 30 and 50 % of the ozone loss may be attributed to the hydrogen cycle, about 20 % to the nitrogen cycle and the rest to bromine and chlorine. These percentages can change drastically in the presence of large perturbations in the chlorine and nitrogen oxide concentrations.

This is because nitric acid will form while the concentration of $HO_2$ will decrease. The net result is that in the regions where the odd hydrogen cycle is prevalent, the ozone loss will decrease and this will happen mainly in the lower stratosphere. The same thing appears in the region where the loss may be attributed mainly to halogens in the sense that to an odd nitrogen, increase will correspond a decrease in the ozone loss. Increasing further the $NO_x$ concentration, a point is reached where the loss due to the other cycles will be equal to the one due to the odd nitrogen. Beyond this point the loss is dominated by the nitrogen oxides and will increase with their concentration. The ozone loss is then a nonlinear process as shown qualitatively in Fig. 18.10, and this feature will be very important when we will talk about the ozone perturbations.

Measurements made during the campaigns mentioned before put the data in the regions where the ozone loss decreases with increasing nitrogen oxides. The same measurements have shown that for a range of $HO_2$ mixing ratio between 0.5 and 0.25 ppb, the $OH$ concentration remains almost constant, while the $HO_2$ increases in such a way that roughly $[HO_2][NO] \approx$ const. The same thing happens between $NO_2$ and $ClO$ when the mixing ration of the first is between 200 and 400 ppt.

Quite an interesting parameter is the ozone loss frequency that represents the inverse of its lifetime referred to all loss processes. The result of this calculation always based on the two-dimensional model is given in Fig. 18.11 where for practical reasons the residence time is reported. This figure is useful because it gives indications in which regions of the mixing ratio are controlled by the transport with respect to photochemistry. We have mentioned several times that the separation can be put at 1 day because for shorter times the diurnal variation becomes important.

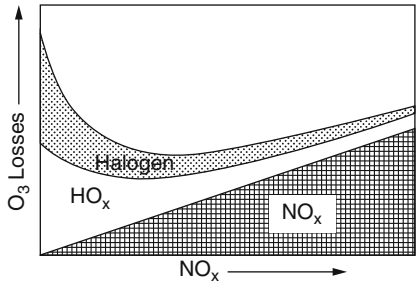



**Fig. 18.10** The qualitative behavior of the ozone loss as a function of $NO_x$ mixing ratio

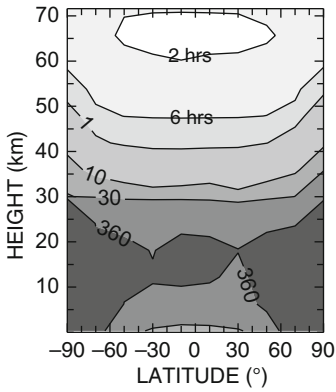



**Fig. 18.11** The ozone lifetime at the equinox. When not indicated the units are days

From this figure it can be argued that below 30 km the ozone does not have diurnal variations and so can be considered as a perfect tracer. Dynamics is important in the upper stratosphere. Another interesting point is that at the equinox the ozone lifetime increases with latitude, and especially in the lower stratosphere, the effect of the autumn can be felt with a longer lifetime. We can see also that in the equatorial troposphere, lifetime decreases again, and this is due mainly to the odd hydrogen and odd nitrogen cycles where the abundance of these compounds increases with increasing solar radiation.

## 18.3 Heterogeneous Chemistry

Until now we have been interested in gaseous phase chemistry. If aerosols are present in the atmosphere, molecules may collide with them and react with the gases on the surface or that are inside the particle. In this case we are dealing with a heterogeneous process. These processes were well known for the troposphere and were introduced for the stratosphere in 1986 to explain the ozone hole by S. Solomon, S. Rowland, and R. Garcia e D. Wuebbles. The idea is that molecules that collide with the aerosol surface may stick on it a little longer so that the probability of reacting increases. A sticking coefficient is defined as

$$\gamma = \frac{\text{number of molecules that reacts } \left(\text{molecules}^{-1}\right)}{\text{number of collisions on the surface } \left(\text{molecules}^{-1}\right)}$$

If the average velocity of the molecules is $\overline{v}$ and the total available surface is $A$, the flux of molecules is given by $n_i\overline{v}$ where $n_i$ is the gas number density. The total number of collisions is then $n_i\overline{v}A/4$. It is convenient to express this number in terms of the total surface per unit volume, $S$. We obtain the total number of collisions per molecule and the equivalent reaction rate is given by

$$k_i = \gamma \ \overline{v}S/4 \tag{18.35}$$

Actually this constant has the unit of frequency so that it is equivalent to the product *reaction constant* $\times$ *molecules* $cm^{-3}$. To obtain the kinetic term for a heterogeneous reaction, it is enough to multiply (18.36) by the concentration of the gaseous reactant. The coefficient $\gamma$ is not easy to evaluate considering that in the lab it is already difficult to measure the rate of a homogeneous reaction.

The aerosol surface density is measured usually in $\mu m^2\ cm^{-3}$ and may change between 1 and $10^2$ with the largest values occurring during catastrophic eruptions and the lowest referring to the background. Knowing the size distribution for the aerosol $n(r)$, the surface density is given by

$$S = \int_0^\infty 4\pi r^2 n(r) dr \tag{18.36}$$

In the lower stratosphere, the reaction rate may be evaluated considering that the average velocity is given by

$$\overline{v} = (8kT/\pi m)^{1/2}$$

For a temperature of 200 K and for a molecule like nitrogen pentoxide, we have $\overline{v} \approx 200\ ms^{-1}$ so that for a surface of 10 $\mu m^2\ cm^{-3}$ and a coefficient $\gamma = 0.01$, we have $k \approx 6 \cdot 10^{-6}\ s^{-1}$ that is a time constant of about 50 h. This is an order of magnitude evaluation because the information about these coefficients are still sparse and they depend on temperature and the composition of the aerosols. However, since the heterogeneous hypothesis has been formulated, considerable progress has been made in the understanding of these processes. The main reactions to be considered are the following:

$$\begin{aligned}
&N_2O_5 + H_2O \quad \text{(aerosol)} \rightarrow 2HNO_3\\
&ClONO_2 + H_2O \quad \text{(aerosol)} \rightarrow HOCl + HNO_3\\
&ClONO_2 + HCl \quad \text{(aerosol)} \rightarrow Cl_2 + HNO_3\\
&N_2O_5 + HCl \quad \text{(aerosol)} \rightarrow ClNO_2 + HNO_3\\
&HOCl + HCl \quad \text{(aerosol)} \rightarrow Cl_2 + H_2O
\end{aligned} \tag{18.37}$$

The first reaction is clearly hydrolysis if the water is in the liquid phase. This may happen, for example, in the volcanic aerosol particles that are essentially a solution of sulfuric acid in water. The same reaction (as well as the second) may occur on the surface of ice particles as in the case of polar stratospheric clouds. In the last three reactions, the hydrogen chloride possibly is adsorbed at the surface of the aerosol particle so that the reaction actually is between the molecules of the two gases although one type is trapped in the solid matrix of the particle. In the chemical reaction scheme (18.37), a notation is added to indicate whether the phase is solid or gaseous as follows:

$$ClONO_2(g) + HCl(s) \rightarrow Cl_2(g) + HNO_3(s)$$

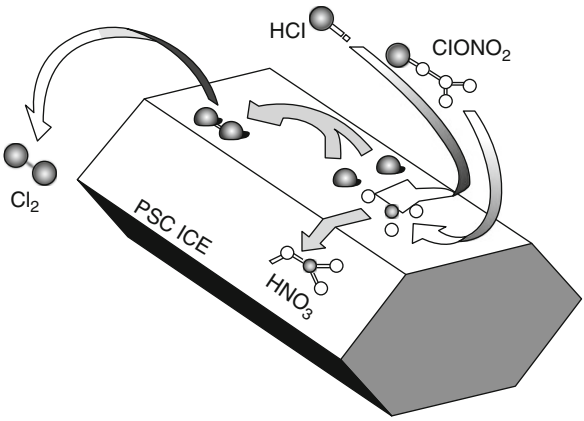



**Fig. 18.12** A cartoon of the heterogenous reaction between chlorine nitrate and hydrogen chloride on a surface of water ice PSC (Adapted from Turco 1997)

where (g) is for gas and (s) for solid. This however is a rather controversial point because we can only say that the reactions are there and their probability is of the same order of that calculated theoretically. It remains hard to know if the nitric acid is formed as solid or is liberated as a gas. Another important point is to establish what is the role of solubility and what is the effective reaction that takes place among the gas components. For example, the first attempts to determine the reaction probability have given very different results because in some case the concentrations used were so high that the saturation pressure was reached on the particle surface so that the gaseous components would sublimate directly.

A possible interpretation of the third of reactions (18.37) is given in Fig. 18.12 (adapted from a beautiful figure by Turco). In this case, the aerosol particle is made of nitric acid trihydrate (NAT) or water ice, and the nitric acid formed remain embedded in the particle. The molecular chlorine $Cl_2$ can then escape the crystal matrix.

Reactions (18.37) have different effects. In general, they transform inert chlorine compounds in active form, that is, chlorine nitrate, hypochlorous acid, and hydrogen chloride in active chlorine. Another effect is to shift the partition inside the $NO_y$ family from $NO_x$ toward nitric acid. This process is called with an ugly word denoxfication and has as consequence the fact that nitric acid embedded in the solid particles is deposited with them causing a depletion of the stratosphere in nitrogen oxides. When the nitrogen oxide mixing ratio decreases, the formation of chlorine nitrate (or bromine) is slowed down, and a larger amount of active halogens will result that will accelerate the ozone destruction.

These processes have a different importance according to the aerosol type involved whether it is of volcanic origin or in a polar stratospheric cloud. The reason is that the probability changes according to the chemical composition of the aerosol surface. A qualitative summary of these effects is given in Fig. 18.13 where the partitions inside the odd chlorine and odd nitrogen families are compared. We see that the sulfate aerosol will reduce the ratio of nitrogen oxide/nitric acid mainly

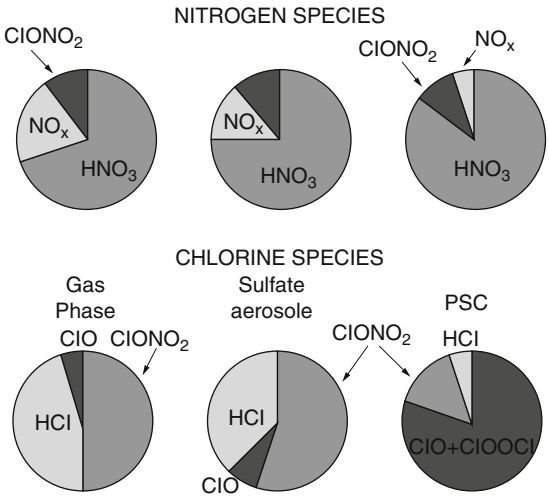



**Fig. 18.13** The partition of nitrogen and chlorine compounds as influence by heterogenous chemistry on sulfate aerosols and PSC (Adapted from Brune 1992)

through the first of reactions (18.37). The reduced availability of nitrogen oxide will increase slightly the abundance of chlorine nitrate.

The chlorine nitrate increase is due to the fourth of reactions (18.37) so that the fraction of hydrogen chloride will decrease. In the presence of polar stratospheric clouds, the heterogeneous reactions are activated and particularly the last three because the hydrogen chloride is more easily soluble in the PSC. A very large quantity of nitric acid may form with a consequent reduction in the nitrogen oxides and a conversion of chlorine in chlorine oxide and dichloroperoxide (*ClOOCl*). Among other things, the first of (18.37) decrease drastically the efficiency of the odd nitrogen catalytic cycle while enhancing the hydrogen and chlorine cycles.

The proofs about the role of heterogeneous chemistry are rather indirect and gathered during dedicated measurement campaign like Airborne Arctic Stratospheric Expedition II (AASE II) in the 1991–1992 winter and Stratospheric Photochemistry Aerosol and Dynamics Experiment (SPADE) at middle latitude in 1992–1993. The two campaigns came after the Pinatubo eruption and have shown through contemporary measures of the compounds of the different families the importance of the heterogeneous chemistry in determining the lower stratosphere composition. We recommend the interested reader to go back to Chap. 9 where we have encountered the stratospheric particles.

## 18.4 The Perturbations to the Ozone Layer

The comprehension of the natural behavior of the ozone layer is really fundamental to recognize the role of some chemical species that determine its concentration. Some of these compounds are a product of human activities so that in principle their abundance may change with time and in turn the ozone amount could also be affected.

Sydney Chapman historically was the first theoretician of the ozone layer and also the one who suggested the possibility to artificially make a hole in it although with good intentions. He noticed that astronomical observations in the ultraviolet are somewhat impeded by the ozone absorption in that spectral region. He suggested that to create an "ozone hole" (his words) using "... a catalyst that without further transformations could have produced the destruction of a large number of ozone molecules...." About 50 years later, this compound is finally found, but it does not facilitate much the ultraviolet astronomical observations because after all it is much less efficient than the one supposed by Chapman. At the beginning of the 1970s, J. Lovelock (the author Gaia) discovers for the first time the presence of chlorofluorocarbons in the atmosphere, and almost at the same time Paul Crutzen and Harold Johnson suggest that the construction of a civil fleet of supersonic transport (SST) could harm the ozone layer because of the large quantities of nitrogen oxide produced by their engines. In 1975, the Climatic Impact Assessment Program (CIAP) report is published, and during the development of this program in 1974, Sherry Rowland and Mario Molina discover that chlorofluorocarbons can be photodissociated in the stratosphere and produce chlorine that at that time had all the numbers to look like the compound requested by Chapman. At this point as we said today "history becomes chronicle." However, nobody even in the ponderous WMO report on 1985 (the Blue Bible) makes any guess about the possibility of an ozone hole. It takes two scientists of the British Antarctic Survey (J. Farman and J. Shanklin) to discover that the seasonal minimum on the Antarctic continent has deepened between 1955 and 1985, but the real stuff has manifested itself at the end of the 1970s. What was promptly called ozone hole is a phenomenon of such proportions that the world governments promoted a politics of emission control that produced the Montreal Protocol. Since the discovery of the ozone hole, the available records were carefully checked to discover that the ozone was decreasing also at the global level. Perturbations can actually be reduced to those introduced by the nitrogen oxides and chlorine. In Fig. 18.14 the relative importance of the

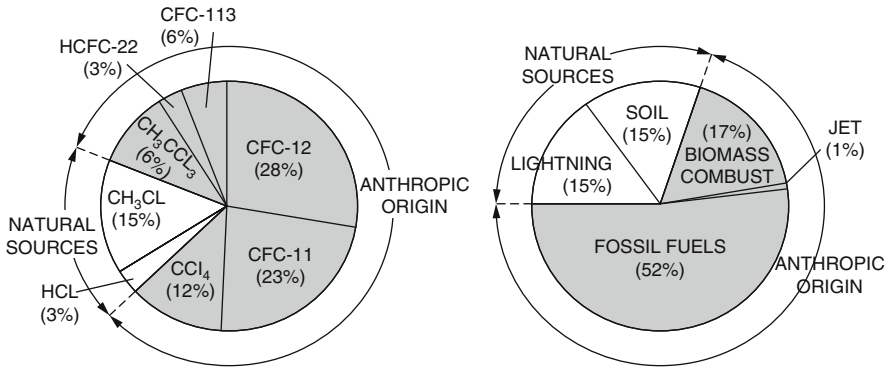



**Fig. 18.14** The sources of chlorine and nitrogen oxide in the atmosphere. Notice the presence of difluorochloromethane (HCFC-22)

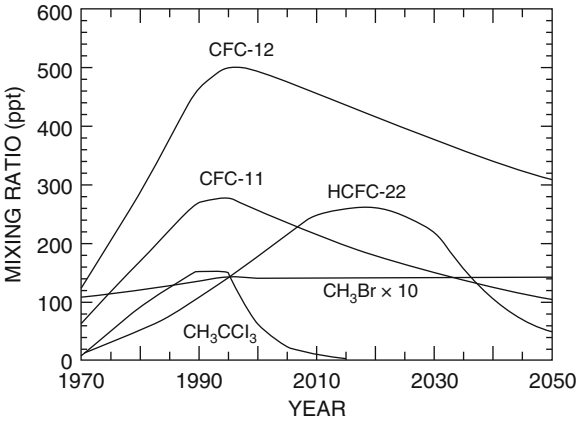



**Fig. 18.15** A possible scenario for the substitution of some chlorine compounds. Notice the different units with methyl chloride multiplied by 10

different sources for such gases is shown. In the case of chlorine, the anthropogenic sources amount to about 82 % of all the chlorine production. However, it is to notice that not all the compounds shown in the figure have the same capacity to release chlorine and so have the same effects on ozone. Among the compounds listed, we found the difluoromethane (CFC-22) that has the chemical formula $CHClF_2$ that is part of the hydrochlorofluoromethanes (HCF) that are the substitutes for CFCs. For what concern the $NO_x$ also in this case a more accurate calculation is necessary to assess the importance of the different sources for the stratosphere. The international agreements that we have mentioned before imply the substitution of some products with others that are less dangerous for the ozone layer. The substitution program actually is expressed in terms of prescribing the mixing ratio of the different species at the surface as it is shown in Fig. 18.15. We see that by the middle of the 1990s, the mixing ratio of the main chlorofluorocarbons (CF-11, CF-12) and methyl chloroform ($CH_3CCl_3$) decreases as they are gradually substituted by the HCF family. This implies that the total chlorine concentration for that period will reach a value $3[\text{CFC-11}] + 2[\text{CFC-12}] + 3[CH_3CCl_3] \approx 2.3$. This is an order of magnitude evaluation because we have neglected some minor gas, and also the total mixing ratio will depend on latitude and altitude. In the preindustrial epoch, the only compound that could produce chlorine was the methyl chloride ($CH_3Cl$) with a concentration of about 0.6 ppb.

In the following paragraphs, we will try to understand what is happening (or already happened) in the stratosphere from a chemical point of view. Some of them will be really qualitative, but we need simply to fix the nature of the problem.

### *18.4.1 The Global Ozone Trend*

The most dated ozone measurements are taken with optical instruments that are essentially spectrometers. Since the 1960s, such measurements have been integrated

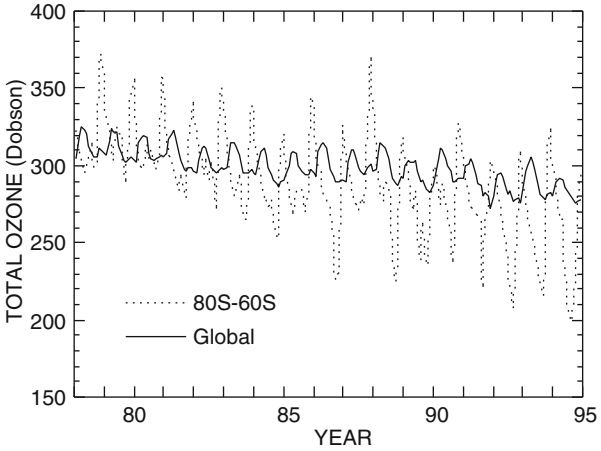



**Fig. 18.16** The global ozone amount between 1978 and 1995 and the average between 80S and 90S latitude band

by satellite data with the result of a quite large record that spans several decades. This long record refers mainly to the total ozone (the columnar density) and is reliable enough that on it we can base the estimation of the ozone trend.

A first example of the data to which we refer is given in Fig. 18.16 where the total ozone is shown and the total content between the latitudinal band 80S–60S that include the depletion of the Antarctic ozone. It is to note that the data are only indicative because they do not take into account the different extension of the latitude bands. Ozone shows a characteristic oscillation that is dominated by the annual cycle. It is noted that at the global level, the ozone has a maximum in spring (of the northern hemisphere), a minimum in summer, and a secondary maximum in winter.

Because we are dealing with global quantities, such variations can only be attributed to the chemistry and not to the transport. A possible reason for the summer minimum is that the catalytic cycles need the radicals that are more abundant in the summer when there is plenty of sun. The maxima are not the same because even if the destruction process is only chemical in nature, the transport modalities between the production and destruction regions are not the same for the two hemispheres so that winters are not symmetrical.

The analysis of Fig. 18.16 shows clearly a decreasing trend that not necessarily should be attributed to an anthropogenic cause. There are a few natural forcing that need to be removed when analyzing the trend. One is the 11-year solar cycle that changes the amount of solar UV radiation and the ozone. Another semi-periodic forcing is the quasi biennial oscillation (QBO) that in stratosphere may change the transport on a time scale of about 2 years. When all these influences are removed from the record, it is possible to obtain data such that shown in Fig. 18.17. This figure illustrates the decadal trend as a function of latitude for the different season.

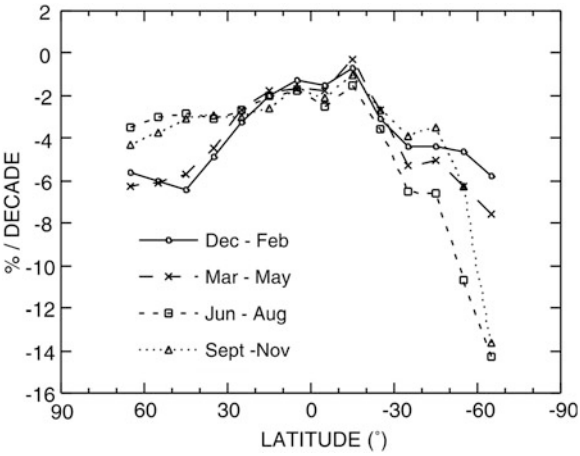



**Fig. 18.17** The ozone decadal trend (in %) as a function of latitude for the years 1979–1994

The data shows that ozone decreases at all latitudes with a tendency for the depletion to be more severe in the winter season. It is quite evident that the southern latitudes are those with the greater weight and the "ozone hole" effect shows up clearly. This was also clear in Fig. 18.16 where the depletion in the polar region is much larger that the global mean.

This ozone behavior can be simply modeled if we assume that the total density $N_{\mathrm{oz}}$ has a total loss $L_0$ expressed in year$^{-1}$ and a production $P$. We have for the trend

$$dN_{oz}/dt = P - LN_{oz} \tag{18.38}$$

If we indicate with $L' = L - L_0$, the perturbation to the loss due to the chlorine increase and with $P_0$ the unperturbed production, we get

$$d\tilde{N}_{oz}/d\tilde{t} = (P/P_0) - \left(1 + L'/L_0\right)\tilde{N}_{oz} \tag{18.39}$$

where

$$\tilde{N}_{oz} = N_{oz}/N_{oz,0} \quad \text{and} \quad \tilde{t} = L_0 t \quad \text{with} \quad N_{oz,0} = P_0/L_0$$

The perturbation to the ozone loss can be related linearly with the fractional change in the chlorine content

$$L'/L_0 \propto \Delta Cl_y/Cl_{y,0}$$

If we use the previous scenario, it is possible to estimate the ozone depletion as shown in Fig. 18.18 where the percentage change for ozone is shown together with the total chlorine change. This very simple calculation is only indicative of the ozone trend although it gives roughly the same values obtained with more

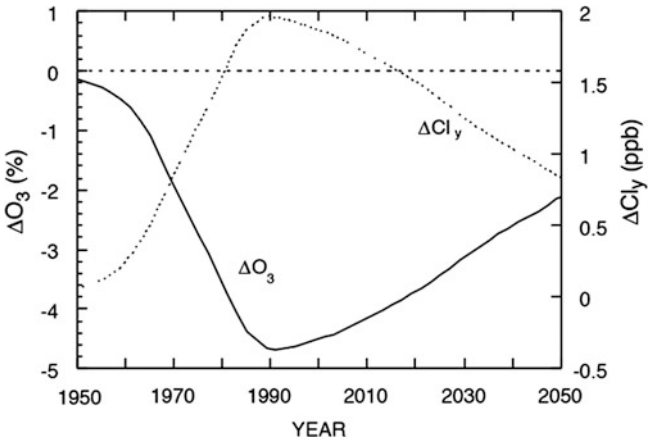



**Fig. 18.18** The total ozone percent change as a function of the absolute change in chlorine content

complex models. It shows also the ozone response without delay to changes in the chlorine mixing ratio. On the other hand, if the ozone changes at the global level are attributed chemical effects, its response time must be of the order of the month that is consistent with the values given previously for the residence time. Times of the same order of magnitude are found in the lower stratosphere where the ozone depletion is more severe.

It is meaningful to ask about the role of the heterogeneous chemistry for the depletion at a global level. With the aerosol at the background level, heterogeneous processes do not have much influence on the ozone depletion at the middle latitude. On the other hand, either we consider perturbation in the $NO_x$ mixing ratio or the stratospheric aerosol load the heterogeneous chemistry may play a very important role. Perturbations of the first kind may be introduced by jet aircraft traffic, and the second may be important in case of catastrophic volcanic eruptions.

### *18.4.2 Natural and Anthropic Perturbations: Volcanic Eruptions*

The hypothesis about heterogeneous chemistry, first formulated for PSC, has received a first indirect confirmation from volcanic eruption and stratospheric sulfate aerosols. Since the measurements taken by the satellite instrument Limb Infrared Monitor of the Stratosphere (LIMS), there was an apparent contradiction between the data on nitric acid and the model predictions with the latter giving values much lower than the experimental data.

After the eruption of the volcano El Chichon in 1982, S. Solomon and D. Hoffman made the assumption that the hydrolysis of nitrogen pentoxide could take place at the surface of sulfate stratospheric aerosols originating from the volcanic eruption. The first of (18.37) convert nitrogen oxides in nitric acid

$$N_2O_5 + H_2O \rightarrow 2HNO_3$$

In this way nitrogen oxide is subtracted to the formation of chlorine nitrate ClONO2 so that more free chlorine is available to destroy the ozone. This hypothesis may imply two things. On one side reconciles the data with the model prediction and on the other hand predicts that in case of catastrophic eruption, we may expect an ozone reduction. This data was around since the Agung eruption in 1963, but at the time the very sparse measurements taken would not be enough to give a solid proof. In 1991, there was another large eruption (Pinatubo) that was followed closely this time even by a satellite (Upper Atmosphere Research Satellite (UARS)), and the data similar to those shown in Fig. 18.16 would confirm that after the eruption the ozone was reduced for a year or so.

The volcanic eruptions are a very good example of the interactions between catalytic cycles. Besides the pentoxide hydrolysis, chlorine nitrate may also react on the sulfate aerosols producing active forms of chlorine according to the heterogeneous reaction

$$ClONO_2 + H_2O \rightarrow HOCl + HNO_3$$

The result is an enhancement of the chlorine cycle with respect to the odd hydrogen cycle. However some attention should be used when comparing the effects of volcanic eruptions. An eruption of the same characteristics of Pinatubo about 100 years ago could have caused an increase in ozone. At that time the chlorine mixing ratio was about 0.6 ppb and was too low to compensate for the loss of nitrogen oxide. Today (about 10 years later) the same eruption would cause a net ozone loss because the chlorine concentration is now about 3 ppb.

This topic of the heterogeneous chemistry on sulfate aerosol can be pursued further to explore the possibility that for an eruption large enough, an ozone hole could be produced anywhere in the Earth's atmosphere and not just in the polar stratosphere. Michael Prather suggested this idea a few years ago.

The starting point is again the heterogeneous reactions (18.37) and in particular those that convert the nitrogen oxides in nitric acid. The effects of these reactions are different. The first effect is the conversion of $NO_x$ ($NO + NO_2 + N_2O_5 + ClONO_2$) in $NO_y$ ($NO_x + HNO_3$), and because less $NO_2$ is available, less $ClONO_2$ will form, and this will result in an enhancement of the chlorine monoxide $ClO$. In the odd chlorine family, the main specie will be $ClO$ rather than $Cl$ because the formation of the latter is somewhat inhibited by the slowdown of the reaction

$$ClO + NO \rightarrow Cl + NO_2$$

However, less chlorine will also decrease its destruction that is due to the reaction

$$Cl + CH_4 \rightarrow HCl + CH_3$$

The increase in the nitric acid concentration will also produce an increase in the hydroxyl abundance through the reaction

$$HNO_3 + h\nu \rightarrow OH + NO_2$$

that is a very important source for these radicals.

We assume the reaction probability for the first two reactions (18.37) $\gamma_1$ and $\gamma_2$, respectively, and we also consider a limited chemical system that is made by few reactions. This represents an obvious simplification but we will get some reasonable results anyway. We consider the following scheme:

$$\begin{array}{ll} NO_2 + NO_3 + M \rightarrow N_2O_5 + M & k_1 \\ N_2O_5 \rightarrow NO_2 + NO_3 & k_2 \\ N_2O_5 + h\nu \rightarrow NO_2 + NO_3 & J_5 \\ NO_2 + O_3 \rightarrow NO_3 + O_2 & k_3 \end{array} \tag{18.40}$$

The $NO_3$ concentration is important only at night because during the day it is very rapidly photodissociated, and from the scheme (18.40), we can calculate its average concentration during the night:

$$[NO_3] = k_3 [O_3] / k_1$$

where $k_1$ includes also the pressure dependence. Assuming for simplicity the same duration of the day and night, we get

$$[N_2O_5] = k_1 [NO_2] [NO_3] / J_5$$

In the presence of heterogeneous processes, the nitrogen pentoxide will be influenced by the system (18.37) so that

$$[N_2O_5] = k_1 [NO_2] [NO_3] / \left(J_5 + \gamma_1' S\right)$$

where $\gamma_1^{'}$ takes into account the collision velocity and $S$ is the surface density. Substituting for $NO_3$, we obtain

$$[N_2O_5] = \frac{0.5\, k_3 [NO_2] [O_3]}{0.5\, J_5 + \gamma_1' S} \tag{18.41}$$

The factor 0.5 has been added to take into account the fact that both the formation and destruction of $N_2O_5$ take place during the daytime. We notice how the

concentration decreases with increase surface density. To complete the interaction with nitric acid, we need to consider two more reactions

$$\begin{aligned} &NO_2 + OH + M \rightarrow HNO_3 + M \qquad k_4 \\ &HNO_3 + OH \rightarrow NO_3 + H_2O \qquad\qquad k_5 \end{aligned}$$

We get for the nitric acid concentration

$$[HNO_3] = \frac{k_4\,[NO_2]\,[OH] + \gamma_1' S\,[N_2O_5]}{k_5\,[OH] + J_{AC}}$$

where $J_{AC}$ is the nitric acid photodissociation coefficient. Expressing everything as a function of $[NO_2]$, we get

$$[HNO_3] = \frac{[NO_2]\left\{k_4\,[OH] + \gamma_1' S \dfrac{0.5\,k_3\,[O_3]}{0.5\,J_5 + \gamma_1' S}\right\}}{k_5\,[OH] + J_{AC}} \tag{18.42}$$

From this expression, we see that with the increase of the surface density, the nitric acid mixing ratio will increase to the asymptotic value

$$[HNO_3] = [NO_2]\,\{k_4\,[OH] + 0.5\,k_3\,[O_3]\} / \{k_5\,[OH] + J_{AC}\}$$

While for small values of the surface

$$[HNO_3] = k_4\,[NO_2]\,[OH] / \{k_5\,[OH] + J_{AC}\}$$

The saturation effect is shown in Fig. 18.19 where actually we have represented the ratio between the nitric acid abundance in the case of the heterogeneous reaction with respect to the case in which the acid is formed only in the homogeneous phase. It is quite evident that as the aerosol surface density increases, the nitric acid mixing ratio saturates as shown in Fig. 18.19. Calculations refer to an altitude of 24 km, a temperature of 220 K at an average latitude of 45° with $\gamma_1 = 0.01$. In practice, heterogeneous chemistry increases by 50 % the abundance of $HNO_3$ and depletes the total concentration of $NO_x$. It is to notice that the nitric acid saturation mixing ratio is not influenced by the value of the reaction probability. In order to make things simpler, we have neglected the second reaction in (18.37) although at this point it should be clear that its effect should be to increase further the nitric acid mixing ratio. The system must be completed by some other reactions that can be grouped in photodissociation reactions

$$\begin{aligned} &HNO_3 + h\nu \quad \rightarrow \quad NO_2 + OH && J_{AC} \\ &NO_2 + h\nu \quad \rightarrow \quad NO + O && J_N \\ &HCl + h\nu \quad \rightarrow \quad H + Cl && J_H \\ &ClONO_2 + h\nu \quad \rightarrow \quad \text{products} && J_{CN} \end{aligned} \tag{18.43}$$

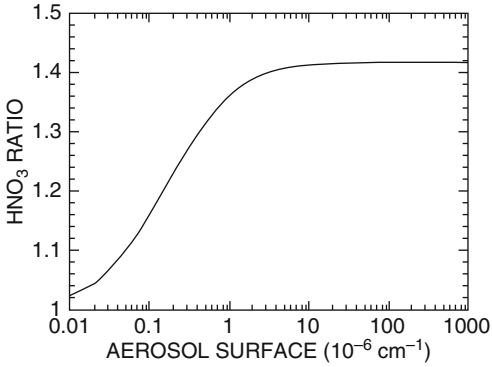



**Fig. 18.19** The ratio between the $HNO_3$ mixing ratio with heterogeneous chemistry with respect the homogeneous phase

and in the homogeneous phase between other minor constituents

$$\begin{array}{lll} HCl + OH \rightarrow H_2O + Cl & & k_6 \\ CH_4 + Cl \rightarrow CH_3 + HCl & & k_7 \\ Cl + O_3 \rightarrow ClO + O_2 & & k_8 \\ ClO + NO \rightarrow Cl + NO_2 & & k_9 \\ ClO + NO_2 + M \rightarrow ClONO_2 + M & & k_{10} \\ NO + O_3 \rightarrow NO_2 + O_2 & & k_{11} \end{array} \tag{18.44}$$

From these additional reactions, the ratio $[NO_2]/[HNO_3]$ can be obtained by assuming the photochemical equilibrium. In particular when all the nitrogen pentoxide is consumed

$$\frac{[NO_2]}{[HNO_3]} = \frac{J_{AC} + k_5\,[OH]}{k_7\,[OH] + k_{10}\,[ClO]} \tag{18.45}$$

The ratio depends then other than the chlorine oxide concentration also on the hydroxyl abundance. We assume initially $[OH] = 2.1 \cdot 10^6$ cm$^{-3}$ and $[ClO] = 8.7 \cdot 10^7$ cm$^{-3}$ so that we have $[NO_2]/[HNO_3] = 0.035$. To have a more precise idea of the value of this ratio for large surface density, we need to have a more realistic value for chlorine oxide. The reaction that converts chlorine in a reservoir in active chlorine is characterized by $k_6$. At photochemical equilibrium, we have

$$[HCl]\,\{J_H + k_6\,[OH]\} = k_7\,[CH_4]\,[Cl] \tag{18.46}$$

Assuming $[CH_4] = 1.12 \cdot 10^{12}$ cm$^{-3}$ (1.1 ppm) and $[Cl] = 0.07$ ppt, we have $[HCl] = 1.2$ ppb. We may expect actually as is also shown in Fig. 18.19 that nitric acid may double so that the same thing happen for the hydroxyl while the $HCl$ abundance will be halved at 0.06 ppb. We may then assume that the decrease in

hydrogen chloride mixing ratio 1.76−0.6 ppb = 1.26 ppb implies a correspondent increase of chlorine monoxide with the result that the ratio given by (18.45) is now 0.004.

To evaluate the nitric acid mixing ratio, we need to know the nitrogen oxide that for a large surface density reduces to the sum $NO + NO_2$. Again at photochemical equilibrium, we get

$$\frac{[NO]}{[NO_2]} = \frac{J_N}{k_{11}\,[O_3] + k_9\,[ClO]} \tag{18.47}$$

Considering that $[O_3] = 4.3 \cdot 10^{12}$ cm$^{-3}$, we have $[NO_x] = 1.1\ [NO_2]$. From the definition of $NO_y$, we express everything as a function of nitrogen dioxide mixing ratio with $[NO_2] = 4.1 \cdot 10^{-2}$ ppb and $[NO] = 4 \cdot 10^{-3}$ ppb. We have $[HNO_3] = 10.6{-}0.045 = 10.55$ ppb. To obtain these results, we need to make some assumptions on the hydroxyl and chlorine abundance. However, the same conclusion can be obtained with a number of iterations.

Concerning the chlorine nitrate concentration, we can make the safe assumption that for large surface densities, we have the total conversion of chlorine nitrate into hypochlorous acid according to the second of (18.37) so that we have for the heterogeneous case

$$[ClONO_2] = k_{10}\,[ClO]\,[NO_2] \,/\, \left\{J_{CN} + 5.4 \cdot 10^{-7} S\right\} \tag{18.48}$$

while in the homogeneous case we had

$$[ClNO_2] = k_{10}\,[ClO]\,[NO_2]\,/J_{CN}$$

Again for large values of the surface density, the heterogeneous case dominates so that for $S = 10^{-5}$ cm$^{-1}$, we have $[ClONO_2] = 0.035$ ppb. The final result is that we have a corresponding increase in $HOCl$ mixing ratio to the chlorine nitrate decrease. Starting from an initial value $[ClONO_2] = 0.9$ ppb, we get for large surfaces values for $HOCl$ that are of the order of 1 ppb. Both the hypochlorous acid and the chlorine monoxide may contribute to the ozone loss through the processes

$$\begin{array}{l} HOCl + h\nu \\ Cl_2O_2 + h\nu \\ ClO + O \end{array} \tag{18.49}$$

The first of these reactions produces only chlorine, while the second one, negligible when the chlorine abundance is low, becomes one of the main destruction mechanisms for ozone when the abundance is large.

However, when dealing with ozone loss, we should not limit our attention to chlorine because heterogeneous chemistry causes an increase in the hydrogen

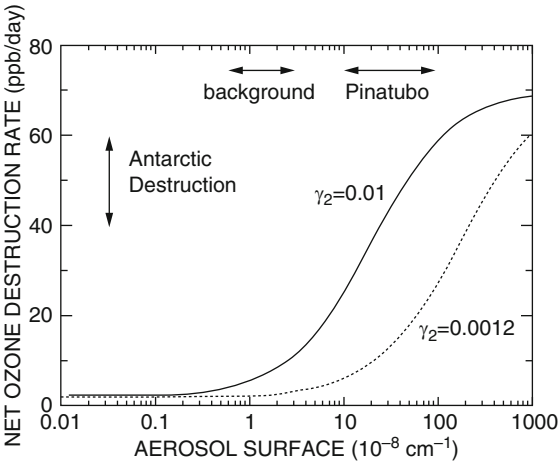



**Fig. 18.20** The ozone loss rate as a function of the aerosol surface density. For comparison typical values found in Antarctica are shown. Also shown are the background and Pinatubo values for the surface density

radicals and a decrease in nitrogen oxides. The hydrogen radicals may increase the ozone production through $HO_2 + NO$ or the destruction through $HO_2 + O_3$. The nitrogen oxides may increase the loss mainly through the reaction $NO_2 + O$. The net effect is that the ozone destruction increases by a factor of 35 with respect to the case of homogeneous chemistry. This is shown in Fig. 18.20 where the loss rate for ozone is shown in ppb per day as a function of the aerosol surface density. We notice that also in this case we have some kind of saturation that corresponds to the fact that not all the chlorine may be converted in reactive form. Two cases are shown for different values of the reaction probabilities.

A few conclusions can be drawn by this exercise. First of all, we have seen that it is possible to slide through a quite complex chemistry. Instead of solving systematically a complex kinetic system, it is much more amusing and instructive to learn the process that counts. It should also be clear that a volcanic eruption of the proportions similar to the Pinatubo should produce ozone loss rates that are comparable to those observed for the ozone hole. We see that while the loss rate may depend linearly on the aerosol surface, its dependence is very nonlinear on chlorine abundance. For example, a decrease of 25 % in the total amount of chlorine gives a reduction of the order of 45 % in the loss rate. This means how the effect of the volcanic eruption of El Chichon in 1983 may have been negligible considering that chlorine was much lower (2.1 ppb vs. 2.8 ppb), and the surface density was 2–3 times smaller.

The paper by Prather was published about a year later the Pinatubo eruption but nothing serious happened to the ozone layer. This may be due to the fact that the large surface requirements were not reached even in the tropical stratosphere. We know however that until the chlorine level remains above 2 ppb, important volcanic eruptions may represent a real ranger for the ozone layer.

### 18.4.3 Natural and Anthropic Perturbations: The Effect of Aviation

The assessment of the effects of the aviation on atmospheric ozone actually initiated the entire ozone debate. In the early 1970s, a specific study of the US Department of Commerce addressed the issue of the effect of a projected fleet of supersonic transport (SST). The CIAP conclusions remained rather uncertain, but in the process the chlorine chemistry was discovered and that appeared as a more immediate danger to the ozone layer considering that in any case the SST construction never materialized. Since then the progress in the understanding of the atmospheric chemistry has been impressive with the most important being the introduction of the heterogeneous processes initially in connection with the polar stratospheric clouds and later with the sulfate stratospheric aerosols. The possibility of hydrolysis of nitrogen pentoxide and chlorine nitrate on sulfate aerosol has drastically changed the perspectives of the aviation effects. The primary cause affecting ozone amount in the lower stratosphere is the emission of nitrogen oxides ($NO_x$) from the jet engines. Whether this results in ozone increase or decrease depends on the local atmospheric composition. On the other hand, heterogeneous processes introduce an additional sink for $NO_x$ with a consequent decrease of the importance of these compounds on the ozone chemistry. Through heterogeneous processes part of the injected $NO_x$ is converted into nitric acid thus affecting also the odd hydrogen and chlorine catalytic cycles. Because of the enhancement in the $NO_y$ ($NO_x + HNO_3$) abundance, less nitrogen is available for the formation of chlorine nitrate, and as a consequence more free chlorine is available for ozone destruction. The lower $NO_x$ amount also implies a reduction in the odd hydrogen sink and this results in an enhancement in the efficiency of the relative catalytic cycle. It can be shown that in the presence of the heterogeneous chemistry, the effect of additional injection of $NO_x$ as those due to air traffic will be strongly damped.

At the beginning only the impact of a fleet of supersonic transport was considered. The reasons being that for fuel consumption optimization, these aircraft must fly in the lower stratosphere, and the injection of nitrogen oxides results directly in the region which is more chemically sensitive. Most recently however attention has been paid to the effect of subsonic aircraft that flies mainly in the upper troposphere. In this case the effects should have been observable because it is estimated that the air traffic may contribute as much as 20 % in the global $NO_x$ production in the troposphere, and this contribution started with the beginning of commercial jet aviation dated almost 40 years ago. In this case, the chemistry is very similar to the one encountered in the urban air with the result that the $NO_x$ injection corresponds normally an ozone increase. The opposing effects, namely, a decrease in the stratosphere and increase in the troposphere, give in most of the published simulations an almost null result on the ozone column, that is, the total ozone amount integrated on a vertical column.

Attention has been paid also to the climatic effects related to the ozone and other environmental changes introduced by air traffic. The change in the atmospheric

composition refers, other than to ozone, also to carbon dioxide and water vapor, both products of the engine combustion. In the last two cases, the changes are so small that they can be really neglected, while a few problems remain in connection with ozone. In this case there is a local radiative forcing related to both the infrared and the absorption of solar radiation. Also the climatic effect (i.e., forcing of surface temperature) can be related to changes in the column content because ozone is a minor greenhouse gas.

Another more subtle climatic effect may be due to changes in cloudiness that again may be influenced in different way by the aircraft. The frequency and extension of the condensation trails (contrails) may have been increasing with time and associated to that an increase in global cloudiness could be envisaged. The knowledge of the physical and chemical mechanisms that produce contrails is so poor that a specific measurement campaign has been recently.

The problem of the effects of aviation on the atmosphere and climate is rather complex, but the research effort which is being dedicated to it seems rather disproportionate especially after some extensive work has appeared in the literature. We show that most of these effects can be treated in the framework of a simple perturbation approach and that there should be no justification for a large-scale research work unless new and rather unexpected data or processes are discovered. The aim is to reevaluate a more physical approach to environmental problems that may show that some or all the impacts are really minor. If this is the case, there should be no need to produce more detailed studies to assess for local or seasonal changes because these effects are in any case negligible. Also in most of the cases, these detailed studies are conducted with tools which are rather primitive and with built-in uncertainties which have never been assessed properly. A typical example is the use of two-dimensional or three-dimensional numerical models of the atmosphere. At the present time, the model which is used most is known as chemical transport model (CTM). These are numerical models that use a fixed dynamics derived either by general circulation models (GCMs) or observations and a simplified chemistry. Most of the time the parameterizations included in CTMs are adjusted to reproduce the current situation, and the degree of reliability of these models is not known exactly although they are used to predict changes that in many cases are a fraction of percent.

A simple evaluation of the expected effects may be useful to decide whether the same effects can be observed experimentally and if the degree of precision of the modeling tools is appropriate to predict the perturbation.

We will start considering the upper troposphere. In this region the ozone net production rate can be approximated as a difference between production of atomic oxygen via photodissociation of $NO_2$ and destruction through the reaction with $NO$:

$$\begin{aligned} NO_2 + h\nu &\rightarrow NO + O \qquad J_{NO} \\ NO + O_3 &\rightarrow NO_2 + O_2 \qquad k_1 \end{aligned}$$

where $J_{NO}$ and $k_1$ are the appropriate rate constant. The net production for ozone is then $J_{\mathrm{NO}}[NO_{\mathrm{x}}] - k_1[NO][O_3]$. It can be easily shown that this term is related to the

$NO_x$ photochemical equilibrium which contains the additional reactions

$$\begin{array}{lll} NO_2 + O & \rightarrow NO + O_2 & k_2 \\ NO + HO_2 & \rightarrow NO_2 + O_2 & k_3 \end{array}$$

The production rate can then be written simply in terms of total $NO_x$

$$P_{O_3}/J_{NO} = [NO_x]\{Rk_3[HO_2] - k_2[O]/J_1\}/(1+R) \tag{18.50}$$

In this equation $R = [NO]/[NO_2]$ remains rather constant for different levels of $NO_x$ in the upper troposphere. In order to evaluate properly (18.50), we need to know also the concentrations of atomic oxygen and of the hydroperoxy radical. This can be done with enough accuracy using about twenty reactions that include the main terms of the odd hydrogen and odd nitrogen cycle. Then the production rate can be evaluated for different levels of total $NO_x$, and it is shown that $P_{O_3}$ has a maximum around 100–200 ppt of $NO_x$. The reason for that is quite simple because for low $NO_x$ mixing ratio, the production is roughly proportional to the $NO_2$ amount. However, as the odd nitrogen increases, the concentration of $HO_2$ is very much affected and actually for large value of $NO_2$ decrease as the inverse of the square of odd nitrogen mixing ratio. There are different degrees of approximation in calculating (18.50) as we have shown in the previous paragraph. However the maximum production is going to be around the local upper tropospheric mixing ratio of odd nitrogen. This is rather important because it facilitates the analysis of the perturbation. Actually in a very general way, it is possible to show that if the concentration of $NO_x$ is perturbed around the maximum by an amount $\Delta NO_x$, the resulting perturbation in the ozone production is approximated simply by

$$\Delta P_{oz}/P_{oz} \approx (\Delta NO_x/NO_x)^2 \tag{18.51}$$

It has been shown already that local perturbation in the odd nitrogen for the subsonic traffic may be around 15–20 % and as a consequence $\Delta P_{oz}/P_{oz} \approx 4$–5 %. Notice that the figure for the $NO_x$ can be obtained as an upper limit simply because the present contribution of subsonic aircraft to the global odd nitrogen production is around 20 %. It can be argued that such small changes are very hard to detect mainly because of the intrinsic variability of ozone, and in any case their impact on the climate and the UV radiation absorption are negligible as we will show later.

We now consider what will happen in the lower stratosphere. A very clear account of the effect of the heterogeneous chemistry and the effluent of the aircraft is given using two-dimensional models.

Although some refinements have been introduced since the publication of these papers, their basic conclusions are still valid. In the case of unperturbed atmosphere and gas phase chemistry in the region between 14 and 23 km, the average contribution of the different catalytic cycles to the ozone destruction in winter is 58 % for $NO_x$, 8 % for $Cl_x$, 18 % for $HO_x$, and 12 % for $O_x$, and the rest is due to bromine chemistry [9]. When the heterogeneous chemistry is introduced limited

to the hydrolysis reaction for nitrogen pentoxide, the role of the different cycles becomes in the same order 8 %, 41.5 %, 36.5 %, 10 %, and 4 %, but the overall destruction rate increases by roughly 26 %. This effect can very easily explained because the hydrolysis reaction is an additional sink for the *NO* besides the dominant loss process $OH + NO_2 + M \rightarrow HNO_3 + M$. The hydrolysis rate is comparable to the rate of formation of nitric acid in the gas phase but really depends on the aerosol surface density available. The additional sink means a lower $NO_x$ mixing ratio and as a consequence a decrease in the rate of formation of chlorine nitrate, that is, $NO_2 + ClO + M \rightarrow ClONO_2 + M$. As seen from the above percentages, the increase in the efficiency of chlorine cycle is not exactly proportional to the decrease of $NO_x$ cycle. The reason is that less nitrogen oxide slows down the conversion of *ClO* to *Cl*. The gas phase formation of nitric acid is also one of the main sinks for odd hydrogen in the stratosphere so we expect for a $NO_x$ reduction an increase in the efficiency of the $HO_x$ cycle.

The key to the relatively small impact, which supersonic aircraft flying the stratosphere will introduce, is the heterogeneous processes. Most of the worst scenarios show a maximum local change of $NO_y$ around 60 % in the lower stratosphere due to the injection of a fleet of supersonic aircraft. If we assume in gas phase chemistry a corresponding increase in the odd nitrogen cycle, we end up with a 34 % increase in overall destruction rate of which now the $NO_x$ cycle is responsible for about 65 %. Actually, the number will be a little different considering that the odd hydrogen will be affected by the change in odd nitrogen, but the change in ozone destruction is quite large. When the hydrolysis reaction is included, we can still assume conservatively that the $NO_x$ will be enhanced, but due to the reduced importance of the odd nitrogen cycle, the total destruction will only be slightly affected, that is, 4 %. In this case however, the increase in nitrogen oxide will affect through heterogeneous chemistry both the chlorine cycle (which will decrease its effects) and the $HO_x$ that will increase slightly.

These simple considerations are confirmed by the complex calculations made with two-dimensional and three-dimensional models which show even smaller numbers. The ozone decrease in the stratosphere however is somewhat compensated by the increase in the troposphere due to subsonic traffic with the net results showing increases or decreases less that 1 % on the column. The expected changes in the total ozone (but also local) are small, and however the published results still show an effort to calculate climatic and other radiative effects. It is possible to demonstrate not only that these effects are small, but again they may be calculated in a very simple way.

Nitrogen oxides injected in the atmosphere may change the ozone content. However, the jet exhaust also contains carbon dioxide and water vapor that are greenhouse gases. The water cycle is so poorly understood that there is no way to estimate the impact of the aircraft emission that must be negligible in any case considering the entity of this source. Carbon dioxide emitted by aircraft is about 2–3 % of the global emission, and its impact on climate forcing may be neglected.

## 18.5 Polar Ozone

A paragraph on its own must be reserved to polar ozone and in particular to the Antarctic ozone. The reason for this interest is that polar ozone has dramatically revealed the human influence on our atmosphere. The reason for this quite different behavior of the Antarctic polar regions is partly found in Chap. 17 where we have shown how the polar vortex actually is responsible of a physical isolation between the polar and the middle latitude regions. This has as consequence the lowering of polar temperatures to such a degree to allow the formation of polar stratospheric clouds (PSC). On their surface heterogeneous processes may take place that are responsible of the rapid ozone loss during the Antarctic spring.

The marked difference between the two hemispheres is again attributable to the different dynamic styles. The stability of the two polar vortexes is quite different because the wave activity in the northern hemisphere is more vigorous so that during the winter the vortex may break a few times. This implies that air can be exchanged between polar and middle regions, and temperature is not low enough for the PSC to form.

The first example of polar ozone depletion is then found in Antarctica as shown in Fig. 18.21 averaged at the measuring station of Halley Bay (76°S). We notice a constant decrease of the seasonal ozone minimum that this halved with respect to its value in 1955. It is quite interesting to know what is the geographical region where the depletion is observed as shown in Fig. 18.22. From the last two figures, we notice how not only the ozone minimum decreases constantly but that the region where this depletion occurs also expand.

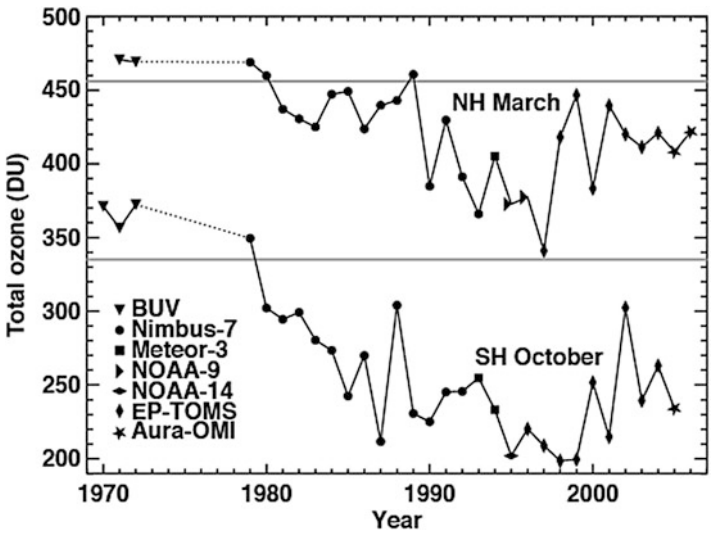



**Fig. 18.21** The ozone behavior (in Dobson unit) in the northern and southern hemispheres

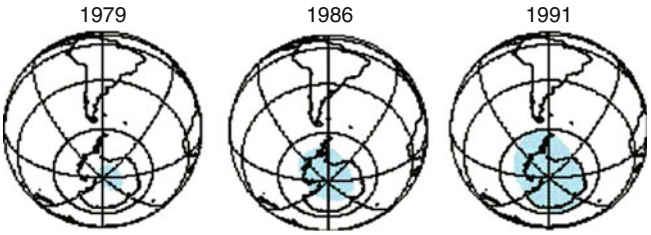



**Fig. 18.22** The extension of the region interested to the Antarctic ozone depletion since 1979

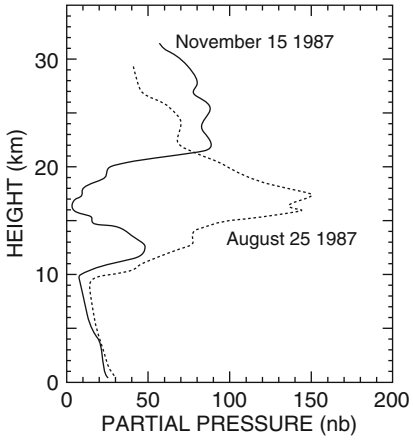



**Fig. 18.23** The ozone partial pressure in 2 days before and after the Antarctic spring

Another impressive data refer to the ozone loss as a function of altitude. A typical example of these phenomena is shown in Fig. 18.23 that shows the results of ozone sounding in 1987. August 25 refers to a period in which large part of the Antarctic stratosphere is still in the dark, while November 15 is within the period in which the minimum in the ozone amount is recorded. During the following years, this behavior has been confirmed so that the altitude region where the ozone depletion occurs is between 14 and 20 km and where most of the PSC clouds are observed.

What we have shown in a sense summarize the observation data on the Antarctic ozone depletion. The theory that still lacks some important detail requires a vision of both the dynamics and the chemistry.

### *18.5.1 The Theory on the Polar Ozone*

Some detail on the explanation of the ozone hole is still missing, but the theory must be right because it has been confirmed by so many data gathered during several measurement campaigns that have involved airborne, ground-based, and satellite instrumentation. Some of these measurements have given results as those shown

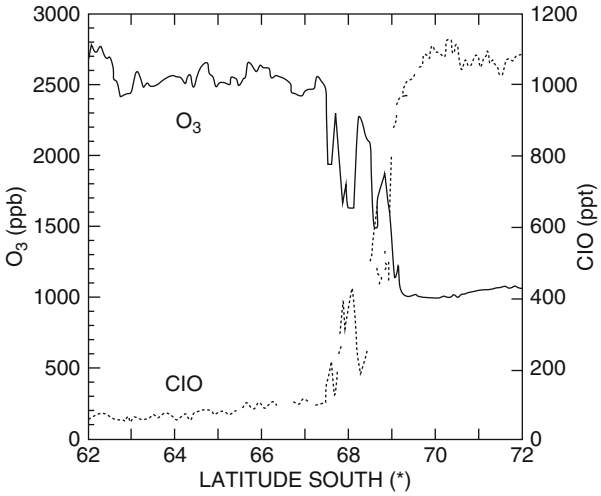



**Fig. 18.24** The mixing ratios for chlorine monoxide and ozone as measured by the airborne instrumentation on a ER-2 plane

in Fig. 18.24 obtained with an instrument flown on the ER-2 that has participated to the 1987 campaign and invented by James Anderson. The data compare ozone and chlorine monoxide measurements, while the aircraft was flying through the polar vortex at about 20 km altitude. We notice that when passing the vortex at latitude of about 68S, there is a sudden increase of the mixing ratio of chlorine monoxide and a corresponding ozone decrease. It is quite surprising that even the smaller fluctuations are perfectly anticorrelated: to a decrease in one concentration corresponds the increase in the other. The most interesting data is the sudden increase in the chlorine concentration that indicates how within the polar vortex all the chlorine is converted into chlorine monoxide to the point that while outside the vortex the mixing ratio is a few ppt inside the vortex it reaches mixing ratios of the order of 1 ppb. This kind of data may be interpreted as a sequence of events like the one shown in Fig. 18.25 where the evolution of the different chemical species is correlated to the temperature evolution.

Once the vortex is formed in late autumn, the temperature starts to decline as the altitude of the vortex. The temperature decreases to the point that PSC form. At this point heterogeneous processes on their surface transform inert chlorine forms (hydrogen chloride and chlorine nitrate) in active chlorine like hypochlorous acid and molecular chlorine:

$$\begin{aligned} &ClONO_2 + HCl \rightarrow Cl_2 + HNO_3 \\ &ClONO_2 + H_2O \rightarrow HOCl + HNO_3 \\ &HOCl + HCl \rightarrow Cl_2 + H_2O \end{aligned} \tag{18.52}$$

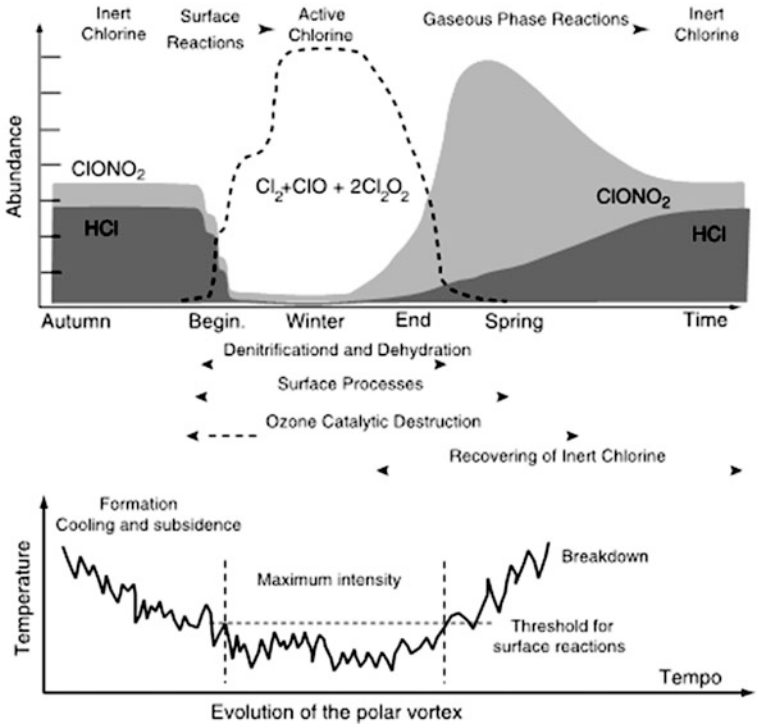



**Fig. 18.25** The evolution of the chemical composition in the Antarctic region and of the stratospheric temperature (Adapted from Jacob 1999)

These reactions may take place in the polar night. At the spring sunrise, the molecular chlorine is photodissociated in chlorine that reacts rapidly with ozone to form chlorine monoxide that starts a cycle that works without atomic oxygen and is based on the dimer *ClOOCl* that is called dichloroperoxide (also known as $Cl_2O_2$):

$$\begin{aligned} & ClO + ClO \rightarrow ClOOCl \\ & ClOOCl + h\nu \rightarrow ClOO + Cl \\ & ClOO \rightarrow Cl + O_2 \\ & 2\,(Cl + O_3 \rightarrow ClO + O_2) \\ \text{net} \quad & 2O_3 \rightarrow 3O_2 \end{aligned} \tag{18.53}$$

For the first of reactions (18.53) to start, we need very high chlorine monoxide concentrations. A similar cycle may also start with one chlorine monoxide substituted by bromine monoxide. The ozone loss due to this cycle is of the order of 1 % per day, and the calculations are in reasonable agreement with the experimental data.

If we go back to Fig. 18.25, we see that almost all hydrogen chloride is converted into active chlorine. The mixing ratio of chlorine nitrate on the other hand is higher

with respect to initial conditions because after the breakdown of the polar vortex, both nitrogen oxides originating from nitric acid and that coming from middle latitude air become available. The result is that not all the chlorine nitrate can be converted in acid also because with the vortex breakdown also the PSC of water ice disappear. The photodissociation of $ClONO_2$ in this phase may contribute to the formation of active chlorine and then to ozone destruction.

Once the polar region "warms" and the air can mix again, the "chemical reactor" as it was formed is destroyed and the gas concentrations rapidly return to the background values.

The ozone hole even if it has been cured by the international agreement has not gone away, and it will take several years (the middle of this century) before it will be a thing to remember.

We have not very much updated this section because in the intervening year since the first edition, there have been very few changes in the ozone hole theory. Some additional interesting detail will be treated in the Examples.

## E.18 Examples

### *E.18.1 The Equivalent Effective Stratospheric Chlorine (EESC)*

The equivalent effective stratospheric chlorine (EESC) is a convenient parameter to quantify the effects of chlorine and bromine on ozone depletion in the stratosphere. We will refer here to the most recent formulation by Newman et al. (2007). The sources of these halogen gases (chlorine and bromine) are mainly in the troposphere and they are released in the stratosphere. EESC was developed to relate this halogen evolution to the tropospheric sources in a simple matter.

EESC, as a function of time t is defined as

$$EESC(t) = a\left(\sum_{Cl} n_i f_i \rho_i + \alpha \sum_{Br} n_i f_i \rho_i\right) \tag{E.18.1}$$

where n is the number of chlorine or bromine atoms of a particular source gas $i$, $f$ represents the efficiency of stratospheric halogen release of the source gas, and $\rho$ is the source gas mixing ratio in the stratosphere. The summations extend to all chlorine- and bromine-containing halocarbons. $a$ can be an arbitrary constant or could be the fractional release for CFC-11 so that EESC represents accurately the amount of inorganic chlorine ($Cl_y$) and bromine ($Br_y$) in some region of the stratosphere. In the classic EESC, $\rho_i$ is calculated assuming a time lag $\Gamma$ from the surface observations

$$\rho_i = \rho_{i,\text{entry}}\,(t - \Gamma) \tag{E.18.2}$$

where $\rho_{i,\text{entry}}(t)$ is the surface observation at time t. Studies assume a value for $\Gamma$ of 3–5 years which represent the time it takes for the compound to reach the stratosphere. The factor a appearing in (E.18.1) takes into account that bromine is much more efficient in destroying ozone. Although this factor should be a function of the specific location, its value is assumed to be roughly 60. Finally in (E.18.1), f represents the fraction of the species that has been dissociated during its movement through the stratosphere. It is defined as

$$f_i = \frac{\rho_i - \rho_{i,\phi,\theta}}{\rho_i} \tag{E.18.3}$$

where $\phi$ is the latitude and $\theta$ represents the potential temperature (and then altitude). To apply now the relations (E.18.1), (E.18.2), and (E.18.3), it is necessary to know the mean age of the air. Details of these calculations can be found in the paper by Newman et al. (2007) from which we report some example. In a previous paper (Newman et al. 2007), $\rho_i$ was taken time dependent according to the relation

$$\rho_i(t) = \int_{-\infty}^{t} \rho_{i,\text{entry}}\left(t'\right) G\left(t - t'\right) dt'$$

In this case, G(t) is the so-called aged spectrum and the fraction release is time dependent, $f_i = f_i(\Gamma)$. This formulation reduces to the classical formulation if $G(t) = \delta\,(t - \Gamma)$ and $\Gamma = 3$ years. Figure E.18.1 shows the results of some calculation made for CF-11. The highest solid line curve refers to the ground concentration as a function of time according to the scenario selected. The two dashed lines refer to the same quantity but with two different age spectrum (3 and 5.5 years), and the lower curve represents the released chlorine after applying a 47 % fractional release. The figure on the right represents the EESC calculation for all the halogens with the old formulation (dashed lines) and the new formulation. Largest values of EESC refer to 5.5 and 3 years mean age of air.

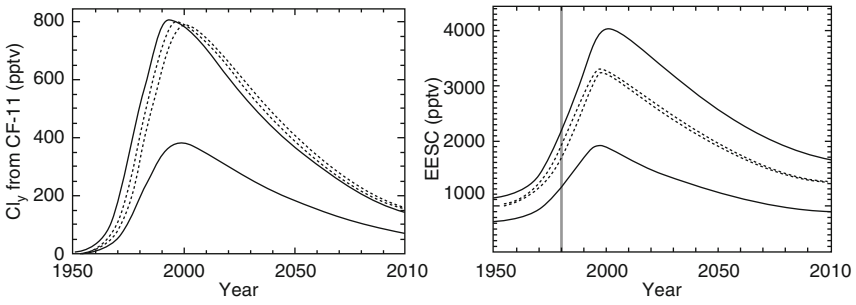



**Fig. E.18.1** The $Cl_y$ released from CF-11 for some scenario (*left*) and EESC from all halogens (See text for explanations)

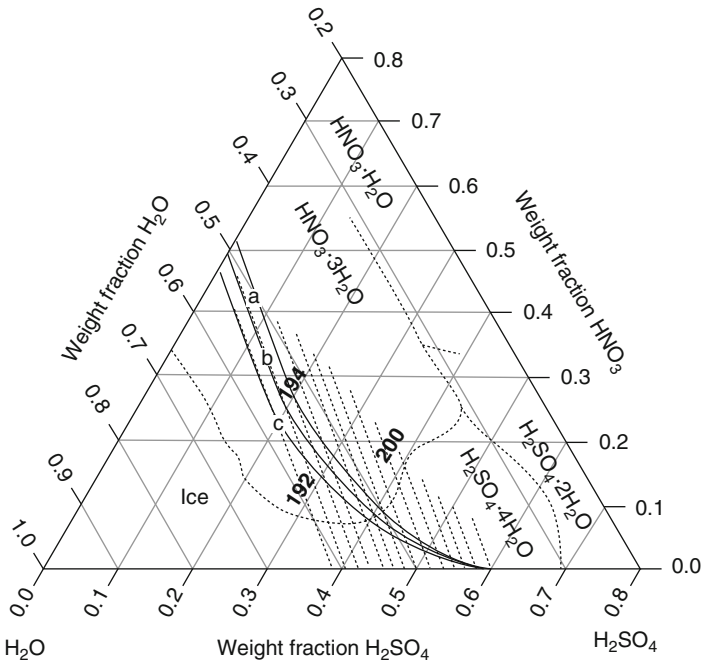



**Fig. E.18.2** Ternary diagram for the system nitric acid/sulfuric acid/water. See text for explanation (Adapted from Molina et al. 1993)

## *E.18.2 Few More Things About Polar Stratospheric Clouds*

We have already treated some points on polar stratospheric clouds (PSC) in the first jter with phase diagram and in paragraph 9.5.2. It is the case then to complete a little bit the topic discussing what physical chemists call a ternary diagram. This is given in Fig. E.18.2 for the system $H_2SO_4/HNO_3/H_2O$. Now with respect to binary systems, we already know the axis is three and on each of them, the weight fraction of the condensable gas is shown. Starting from the water axis, we go from pure water (weight fraction equal to unity) to a weigh fraction of 0.2. The same thing can be noted for the other axis. The heavy dashed curves are the eutectic separating the identities of the various solids that at equilibrium crystallizes first upon cooling liquids with composition bounded by the eutectic lines.

The figure and the example we discuss are taken from the classical paper by Molina et al. appearing on science in 1993. Superimposed to the ternary diagram are the dilution curves for a parcel of air at 100 hPa ( $\approx$ 16 km) with 5 ppm of water vapor and with (a) 10 ppbv of $HNO_3$, (b) 5 ppbv $HNO_3$, and (c) 2.5 ppbv of $HNO_3$. The dotted lines show the equilibrium temperature.

Several things can be studied from this diagram. Following one of the cooling curves, we notice that changing the temperature from 196.5 K to 193.5 K increases

the weight fraction of $HNO_3$ from roughly 10 % to more than 40 % increasing the supersaturation for nitric acid trihydrate.

### *E.18.3 How to Calculate the Loss Rate of Ozone Over Antarctica*

We start out remembering the main reactions of the chlorine cycle:

$$\begin{aligned} &ClO + ClO \xrightarrow{M} ClOOCl \\ &ClOOCl + h\nu \rightarrow ClOO + Cl \\ &ClOO \xrightarrow{M} Cl + O_2 \\ &2\,(Cl + O_3 \rightarrow ClO + O_2) \\ &\text{Net}: \quad 2O_3 \rightarrow 3O_2 \end{aligned} \tag{E.18.4}$$

At this point, we may evaluate the destruction rate for ozone making some simplifying assumptions. First we neglect the interaction with $NO_x$ ($ClO + NO \rightarrow Cl + NO_2$). This can be justified because of the almost complete denitrification of the Antarctic atmosphere. Then we neglect the bromine cycle and assume because of the very rapid reaction of $Cl + O_3$ that all the reactive chlorine $[(ClO + 2(ClOOCl))]$ is partitioned between chlorine peroxide and ClO. We also assume that all the inorganic chorine has been converted into reactive form ($[Cl_x] = [Cl_y]$) considering the very low concentrations of HCl and $ClNO_3$. We finally assume that the air is cold enough for the decomposition of ClOOCl to be neglected. Then the equilibrium between the first two reactions (E.18.4) gives

$$k[ClO]^2 = J\,[ClOOCl] \tag{E.18.5}$$

we the assume that

$$[Cl_x] = [ClO] + 2\,[ClOOCl] \tag{E.18.6}$$

we can then substitute in (E.18.5) to obtain

$$k\{[Cl_x] - 2\,[ClOOCl]\}^2 = J\,[ClOOCl]$$

and solve for [ClOOCl]

$$[ClOOCl] = \frac{\alpha}{8}\left[\left(1 + 4\frac{[Cl_x]}{\alpha}\right) - \left(1 + 8\frac{[Cl_x]}{\alpha}\right)^{1/2}\right] \tag{E.18.7}$$

where $\alpha = J/k\,[M]$. If we refer to October 1 at 75S latitude, we have for this a coefficient value of 8.5 ppbv. Then for the second and fourth equations (E.18.4), we have for the ozone loss

$$\frac{\partial\,[O_3]}{\partial t} = -2J\,[ClOOCl] = -2k[ClO]^2$$

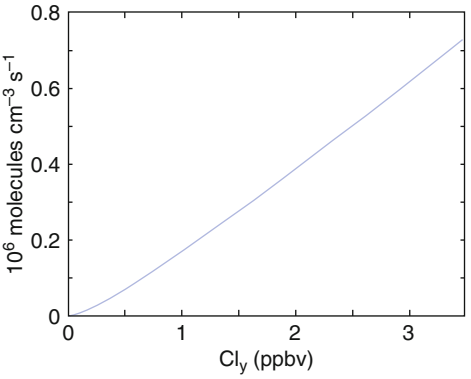



**Fig. E.18.3** Destruction rate of ozone as a function of total chlorine content. Roughly 1 ppbv corresponds to 1960 while 3 ppbv to mid-1990s

This destruction rate can be shown as a function of the total chlorine content as in Fig. E.18.3.

## References[1]

### *Books*

Baird C (1995) Environmental chemistry. Freeman, New York
Scientific Assessment of Ozone Depletion (2014) WMO, Report N.55, Geneva
Turco RP (1997) Earth under siege. Oxford University Press, Oxford/New York

### *Articles*

Anderson JG et al (1989) Kinetics of $O_3$ destruction by ClO and BrO within the Antarctic vortex: an analysis based on in situ ER-2 data. J Geophys Res 94:11480
Brune W (1992) The use of EOS for studies of atmospheric physics. Elsevier, Amsterdam/New York
Cicerone RJ, Elliot S, Turco RP (1991) Reduced Antarctic ozone depletions in a model with hydrocarbon injections. Science 254:1191
Elliott S, Cicerone RJ, Turco RP, Tabazadeh KDA (1994) Influence of the heterogeneous reaction HCl + HOCl on an ozone hole model with hydrocarbon additions. J Geophys Res 29:3497
Fahey et al (1993) In situ measurements constraining the role of sulfate aerosols in mid-latitude ozone depletion. Nature 363:509
Graedel TE, Crutzen PJ (1993) Atmospheric change: an earth system perspective. Freeman, New York

[1]Again this is a chapter containing topics that evolve fast. The NASA/WMO documents are very useful and the books by Baird and Turco provide some basic understanding. The single papers are mostly instructive.

Jacob D (1999) Introduction to atmospheric chemistry. Princeton University Press, Princeton
Molina MJ et al (1993) Physical chemistry of the $H_2S0_4/HNO_3/H_20$ system: implications for polar stratospheric clouds. Science 261:1418
Newman PA (2013) Chemistry and dynamics of the Antarctic ozone hole. In: Polvani LM, Sobel AH, Waugh DW (eds) Chemistry and dynamics of the Antarctic ozone hole. American Geophysical Union, Washington
Newman PA, Daniel JS, Waugh DW, Nash ER (2007) A new formulation of equivalent effective stratospheric chlorine (EESC). Atmos Chem Phys 7:4537
Pitari G, Rizi V (1993) An estimate of the chemical and radiative perturbation of the stratospheric ozone following the eruption of Mt. Pinatubo. J Atmos Sci 50:326
Pitari et al (1993) High speed civil transport impact: role of sulfate, nitric acid trihydrate and ice aerosols studied with two dimensional model including aerosol physics. J Geophys Res 98:23141
Plumb RA (1996) A tropical pipe model for stratospheric transport. J Geophys Res 101:3957
Plumb RA, Ko MKW (1992) Interrelationships between mixing ratios of long-lived stratospheric constituents. J Geophys Res 97:1014
Stolarski R et al (1992) Measured trends in stratospheric ozone. Science 256:342
Wennberg PO et al (1994) The removal of the lower stratospheric O3 by free radical catalysis: in situ measurements of OH, HO2, NO, NO2, ClO, and BrO. Science 266:398
Wooldridge PJ, Zhang R, Molina MJ (1995) Phase equilibria of H2SO4, HNO3, and HCl hydrates and the composition of polar stratospheric clouds. J Geophys Res 100:1389

# Chapter 19
# Chaos and Nonlinearities

The conclusion of this book cannot escape the aspect of atmospheric physics that is most debated nowadays, that is, the chaotic character of the climatic system and the atmosphere. In the previous chapters we have not questioned the deterministic character of this system, and most of the time we have worked with equations for which analytical solutions existed that enabled us to predict at each instant of time the status of the system. These were relatively simple systems that implied a total predictability. Sometimes more complex systems do not give rise to the same conclusions or to regular solutions. The atmosphere could be such a system and, according to Edward Lorenz, it may be intrinsically unpredictable. You may think about a cloud: today there is no theory that could predict the evolution of a cloud in the presence of updraft, wind, humidity, advection, etc. There are no two identical clouds; nonetheless, we know that at the base of cloud evolution, there are processes that could be described and understood. If we think a little, this is not a novel situation in physics. In a completely different context, the kinetic theory of gases solves another impossible problem because it avoids the question of how to describe the exact position of each molecule in a gas. Instead it gives their collective properties, describing their statistical behavior.

For complex dynamic systems, in recent years, a method has been developed that is known as chaos theory which tries to find a logic behind apparently irregular phenomena. It is not surprising that such a theory could be originated in the realm of meteorology with the genius of Edward Lorenz. At the beginning of the 1960s, studying convecting fluids, he discovered that sometimes they could give rise to very irregular behavior. After Lorenz, the developments have been so fast that they cannot be summarized in one chapter only even for those aspects relative to the

**Electronic supplementary material** The online version of this chapter (doi: 10.1007/978-3-319-29449-0_19) contains supplementary material, which is available to authorized users.



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_19

atmosphere and climate. On the other hand, it would be a pity to omit from the book a topic so important. We will thus start giving the most important definitions.

## 19.1 Simple Examples from the Theory of Dynamic Systems

The evolution of a dynamic system in general can be described by a number of rules that can be written as differential equations. The most obvious dynamic system is a pendulum. By referring to Fig. 19.1 and indicating with $\theta$ the position and with $\dot{\theta}$ its angular velocity, we have the equations

$$d\theta/dt = \dot{\theta}; \quad d\dot{\theta}/dt = -(g/l)\,\theta - \beta\dot{\theta} \tag{19.1}$$

And with respect to the simple pendulum we have added a damping term. The description of the motion of such a system can be made in phase space corresponding to represent the function $\dot{\theta} = f(\theta)$, which in this case assumes the form of a spiral that goes toward the origin of the axis in the phase space $\theta = \dot{\theta} = 0$. In Fig. 19.1 these two trajectories are qualitatively shown. The solution of the dynamic system is such that, whatever the initial conditions, the system ends up either at the origin or is "stranded" in the orbit shown at right. These two geometrical objects in the phase space are called *attractors*. An attractor that assumes the form of a closed curve is representative of a periodic motion and indicates a *limit cycle*.

Another form of the attractor is the torus, as shown in Fig. 19.2. This attractor is characteristic of quasi-periodic motions. An example is a pendulum in which the amplitude of the oscillation is modulated at a different frequency. In the phase space this mode is confined on the surface of the torus. A property of such a system is that if the ratio of the characteristic frequencies is not commensurable (i.e., 1 and $2^{1/2}$), then two points that enter the surface at a distance arbitrarily small will stay near. These attractors are also called "well behaved" and are a particular case of systems that are predictable. Often they are also called *non-chaotic attractor*s and mathematically speaking are characterized by an integer dimension that belongs to a submanifold of the relative phase space. For example, the attractor for the simple

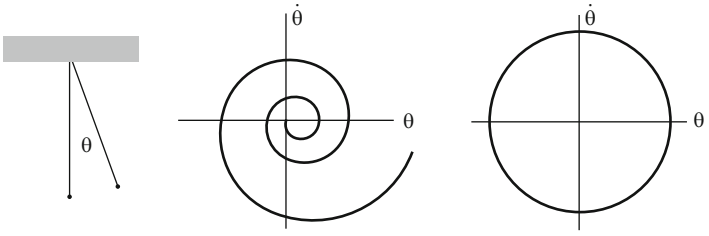



**Fig. 19.1** The phase space trajectories for a simple pendulum (*right*) and a damped one (*center*)

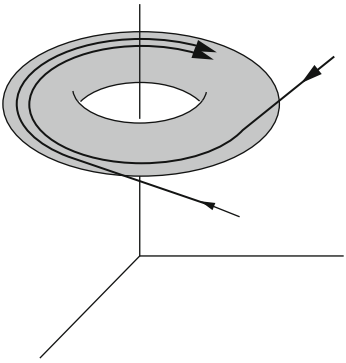


**Fig. 19.2** The surface of the torus as an attractor. Trajectories entering the surface arbitrarily close will stay near to each other (Tsonis and Elsner 1989)

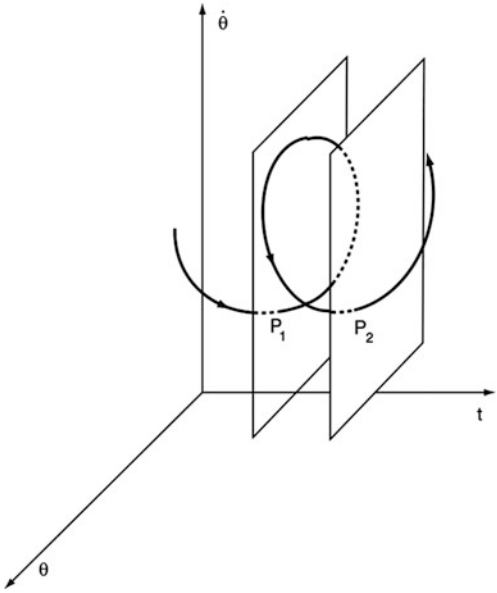



**Fig. 19.3** Illustration of the Poincarè section. The phase space trajectory intersects at point $P_1$ and $P_2$ planes parallel at the $\theta$ and $d\theta/dt$ plane at time intervals equal to the period of the motion (Tsonis and Elsner 1989)

pendulum is a point (zero dimensions) with respect to two-dimensional phase space. The attractor of the damped force pendulum has one dimension with respect to a two-dimensional or three-dimensional phase space. In order to deal with more complex attractors we need to introduce two other concepts, the Poincarè section and the fractal dimension.

### *19.1.1 The Poincarè Section*

The system of Eq. (19.1) can be modified with the introduction of a third coordinate $\varphi = \omega t$ that corresponds to the equation $\dot{\varphi} = \omega$. In this way the forced pendulum in a three-dimensional phase space ($\theta$, $\dot{\theta}$, $\varphi$) will have a trajectory very similar to a helix, as shown in Fig. 19.3. If we make a cross section of this trajectory at a fixed

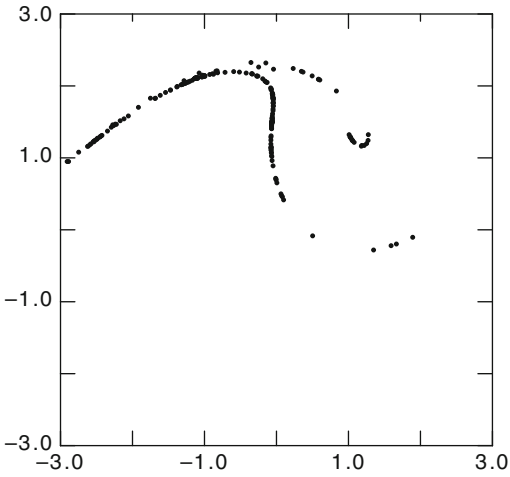



**Fig. 19.4** The Poincarè section in the plane angle; the angular velocity for a nonlinear pendulum. Abscissa is the angle (Baker and Gollub 1990)

time interval corresponding to a period $T = 2\pi/\omega$, we will have in the $\theta$, $\dot{\theta}$ plane a point with the same coordinates. On the other hand, if we consider a nonlinear pendulum (for which we do not use the small amplitude approximation), it will be governed by the equations

$$d\theta/dt = \dot{\theta}$$
$$d\dot{\theta}/dt = -\sin\theta - \dot{\theta}/q + F\cos\omega t$$
$$d\varphi/dt = \omega$$

where the symbols are self-explanatory. It is possible to show that such a system does not always give periodic solutions while the parameter that controls mainly the results is the forcing term $F$. We now calculate the Poincarè section for this nonlinear case and obtain a rather surprising result, shown in Fig. 19.4. The pattern in the figure is composed of about 200 intersections of the kind shown in Fig. 19.3, with the pattern more and more defined as the number of points increases. All these points taken together form the attractor, to which we may now give a different interpretation.

We actually could think of solving the nonlinear system starting from different initial conditions (chosen, e.g., with a random number generator). After a certain time we would find that the system is actually in one of the points that make up the attractor or in any case in the general area occupied by the attractor. It is quite interesting at this point to pose the question of the dimension for the attractor which in this particular case is not an integer number. The attractor dimension is important because it is somewhat related to the minimum number of variables required to describe the dynamic system. The evaluation of these dimensions can only be made through the introduction of the fractal concept.

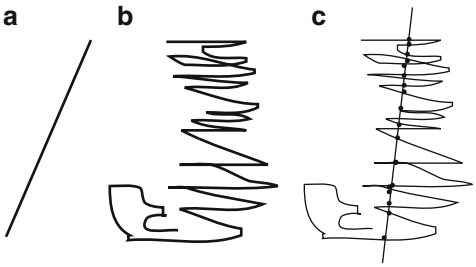



**Fig. 19.5** Examples of different fractal dimensions. In (**a**) the line has dimension 1. In (**b**) the scrawl has dimensions between 1 and 2. In (**c**) the fractal dimension has been reduced by one unit with respect to (**b**). Points have dimensions between 0 and 1 (Stull 2000)

### *19.1.2 Fractal Dimension*

Fractals can be defined as patterns made up from the superposition of similar shapes but on different scales. Typical examples are the dendrites, snowflakes, or tree branches. In meteorology, lightning, clouds, and turbulence (to name a few) may exhibit a fractal behavior. The fractal dimension is a measure of the space filling capability of a particular object. In Euclidean geometry the dimensions are expressed by integer: a point has $D = 0$ dimensions, a line $D = 1$, and so on. Fractal geometry allows also decimal dimensions. Consider, for example, drawings made

by children, as shown in Fig. 19.5. In part (a) of this figure the simple line has an Euclidean dimension $D = 1$. In part (b) the wiggly line has a fractal dimension between 1 and 2 with the larger dimension corresponding only to that the case the line should completely fill the plane. There are different methods to determine the fractal dimension. The simplest one is based on the fact that when we add a spatial dimension the fractal dimension increases by one unit. As a matter of fact, we may imagine the line in Fig. 19.5 as a cross section of a newspaper sheet that has been crinkled. When the sheet is flat on the table, it has dimension $D = 2$. When it is folded or wadded, its dimension is between 2 and 3 and may approach $D = 3$ when it is well compressed into a perfect ball. If the folded or crinkled newspaper is cut along a plane, the cross section of the cut has a dimension between 1 and 2. For example, if its dimension in three-dimensional space was 2.3, now the cut will have a dimension $2.3 - 1 = 1.3$, and its aspect could be that of Fig. 19.5b. If this section is now cut with a line, as in Fig. 19.5c, the intersection points will generate a pattern with dimension $1.3 - 1 = 0.3$.

Reducing the dimension makes it easier to measure it. As an example we could take the "cloud" of points shown in Fig. 19.6. In this case the plane is divided in elementary squares each of size $\varepsilon$ and the number of squares $N(\varepsilon)$ necessary to cover all the points is counted. Now the dimension of the square size is decreased and the count is repeated so that a plot can be made of the number $N(\varepsilon)$ as a function of $1/\varepsilon$. The slope of the resulting line will be defined as fractal dimension $d_c$ ($c$ is for "capacity"), that is,

$$d_c = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{1/\varepsilon} \tag{19.2}$$

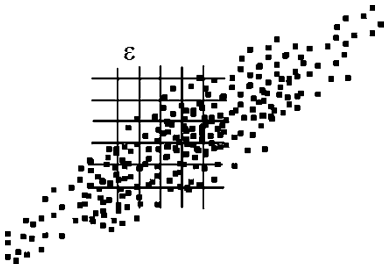



**Fig. 19.6** The measurement of the attractor dimension using the box system (Alligood and Sauer 1997)

If the attractor has been obtained from Poincarè sections, as those of the pendulum, its fractal dimension must be increased by one unit because we reduced one dimension at the beginning. There are several other methods to determine the fractal dimension, and in the specific case of the pendulum, it can be found to be about 2.3. In three-dimensional phase space, the volume of the attractor is zero because it can be shown that trajectories never intersect so they actually constitute surfaces that do not fill out all the space. If the dimension of the attractor is not an integer we are in the presence of a *strange attractor*. This is a rather limiting definition because the most important characteristic is that the strange attractors, to repeat Lorenz's definition, are "an infinite number of curves, surfaces or higher dimension manifolds – generalization of surfaces to multidimensional space – often occurring in parallel set, with a gap between two members of the set."

## 19.2 The Climate

Again, to repeat Lorenz's words, "... for a very complicated chaotic system – the global weather- the attractor is simply the climate, that is the set of weather patterns that at least have some chance of occasionally occurring."

In Chap. 14 we illustrated some important features of the climatic system and we also discussed that the possibility to detect anthropogenically induced climatic changes is based in large part on the capacity to recognize natural from forced variations. In practice we have been asking how deterministic is the climatic system.

As we said earlier, the equations that describe the climate are the same describing the weather. However they are used in a very different way. The forecast is made starting from a set of initial conditions on a number of variables (wind, humidity, and so on). The equations are then integrated for a period of several days so that a forecast can be released within the same time frame. When the equations are used to predict climate, the boundary conditions are fixed instead (solar constant, carbon dioxide content of the atmosphere, and so on) and the equations are integrated until a steady state is reached.

Slowly changing variables like sea surface temperature or ice coverage are of very little interest in the case of weather forecasting yet they could be very important in determining the climate.

The equations that describe either the weather or the climatic system are nonlinear. This implies that, starting from different initial conditions, after some time the same solution is reached: in this case the system is called *transitive*. More correctly, we could say that after an infinite time, the system tends to a set of solutions that have the same statistical properties. On the other hand, it may happen that different initial conditions imply at least two sets of different solutions, and in this case we are in the presence of an *intransitive system*. Finally, the system can be such that it goes from one state to the other after a finite time and we will call it *quasi-intransitive*.

To understand such definitions Edward Lorenz derived a system of equations from the simplest form of the equation of motion. The starting point is the vorticity equation written in a slightly different form with the introduction of the streamfunction $\psi$ such that $u = -\partial\psi/\partial y$ and $v = -\partial\psi/\partial x$

$$\frac{\partial}{\partial t}\nabla^2\psi + \mathbf{k}\cdot\nabla\psi\times\nabla\left(\nabla^2\psi\right) = 0 \tag{19.3}$$

The solution of this barotropic vorticity equation can be found with a two-dimensional Fourier series

$$\nabla^2\psi = X\cos ly + Y\cos kx + 2Z\sin ly\sin kx \tag{19.4}$$

Integrating we find the streamfunction to be

$$\psi = -\frac{X}{l^2}\cos ly - \frac{Y}{k^2}\cos kx - \frac{2Z}{k^2+l^2}\sin ly\sin kx \tag{19.5}$$

Substituting in the Eq. (19.3) we obtain the equations for the amplitudes

$$\begin{aligned}
\frac{dX}{dt} &= -\left(\frac{1}{k^2} - \frac{1}{k^2+l^2}\right)klYZ \\
\frac{dY}{dt} &= \left(\frac{1}{l^2} - \frac{1}{k^2+l^2}\right)klXZ \\
\frac{dZ}{dt} &= -\frac{1}{2}\left(\frac{1}{l^2} - \frac{1}{k^2}\right)klXY
\end{aligned} \tag{19.6}$$

A possible estimation of the coefficients can be made by assuming $2\pi/l = 10^4$ km, $2\pi/k = 5\cdot 10^3$ km so that

$$\dot{X} = -0.1YZ; \quad \dot{Y} = 1.6XZ; \quad \dot{Z} = -0.75XY \tag{19.7}$$

In Eq. (19.5) the first term is independent of the longitude and represents then the zonal velocity. The other two terms are the eddies and are superimposed on the zonal motion. The system Eq. (19.6) is nonlinear, but to have something that resembles

the atmospheric circulation, it must be completed by a friction term (proportional to -*X*, -*Y*, and -*Z*, like damping terms). Also, the eddies represented in Eq. (19.5) are advective and we need to add terms that result from the direct interaction between eddies. The forcing of the system could be assured, for example, by a temperature difference between pole and the equator and can be represented by an explicit forcing term. For the eddies only one component will be explicitly forced. More reasonable, phenomenological equations could be

$$\begin{aligned}\frac{dX}{dt} &= -Y^2 - Z^2 - aX + aF \\ \frac{dY}{dt} &= XY - bXZ - Y + G \\ \frac{dZ}{dt} &= bXY + XZ - Z\end{aligned} \tag{19.8}$$

where $aF$ and $G$ are the forcing terms for the zonal wind and the eddies, respectively. The terms $bXZ$ and $bXY$ represent the eddy advection, while $XZ$ and $XY$ represent the amplification of the zonal wind due to the interaction of the eddies with the zonal wind. From these equations it is clear that the eddies grow at the expense of the zonal wind. The value of $b$ establishes the relative role of the advection with respect to the eddy amplification: if it is greater than 1, advection is faster than amplification. The genesis of Eq. (19.8) is still controversial, and actually we think the best trail for it is a work that Lorenz published in the *Journal of the Atmospheric Sciences* in 1981.

From this simple system of equations, we obtain the total energy

$$\frac{1}{2}\frac{d}{dt}\left(X^2 + Y^2 + Z^2\right) = aX\left(F - aX\right) + Y\left(G - Y\right) - Z^2$$

that is conserved in the absence of eddy terms and damping. A steady-state solution for Eq. (19.8) is easily obtained and is given by

$$\begin{aligned}Y &= \left(1 - X\right)G/\left(1 - 2X + \left(1 + b^2\right)X^2\right) \\ Z &= bXG/\left(1 - 2X + \left(1 + b^2\right)X^2\right)\end{aligned} \tag{19.9}$$

in the absence of eddies and the forcing $G$ this solution gives

$$Y = Z = 0 \text{ and } X = F$$

This solution simply says that the mean zonal flux is established as the equilibrium between the Coriolis term and the zonal friction so that we are speaking about a Hadley-type circulation. It is possible to analyze the stability of such a solution

using the perturbation method so that each variable is perturbed for a small quantity around the steady-state solution $X = F + x, Y = y, Z = z$. These must be substituted in the last two equations of Eq. (19.8) to find

$$\frac{1}{2}\frac{d}{dt}\left(y^2 + z^2\right) = (F - 1)\left(y^2 + z^2\right)$$

which means a growth for the eddies if $F > 1$ and vice versa.

To integrate Eq. (19.8), we need to assign numerical values to the quantities $a$, $b$, $F$, and $G$. We choose to normalize the amplitudes to the value of the vorticity at middle latitude ($10^{-4}$ s$^{-1}$) because this should be the order of magnitude for the quantities $X$, $Y$, and $Z$. The damping time for the eddies could be 5 days (remember the damping time in the boundary layer for spin down). For the damping time of the zonal current we use 20 days. As zonal mean velocity we assume 10 m s$^{-1}$ with the forcing $F$ being of the same order of magnitude. The normalized value would then be 10 m s$^{-1}$ x 5 d/$10^7 = 0.4$. Distances are then normalized to $10^4$ km.

The study of the steady-state solutions shows that other solutions can be obtained with $G$ being between $-2$ and 2. Our first integration will use the values $a = 0.25$, $b = 4$, $F = 8$, and $G = 0.2$. The results of this integration are shown in Fig. 19.7 where the three variables exhibit an almost perfect periodic behavior. It should be noted however that the amplitude of the zonal wind is small with respect to the amplitude of the eddies and the period is of the order of 10 days, while there is a $\pi/2$ phase difference between the two eddy components. If G is interpreted as a parameter that introduces an asymmetry between the X and Y eddy components, we can conclude that this particular solution in its regularity reproduces quite well the Hadley regime.

If the asymmetry factor grows we notice a number of things as a result of Fig. 19.8. In this case $G = 0.8$ and the oscillation is much less regular while the current and the eddy amplitudes are comparable. The minima and maxima however have very different amplitudes.

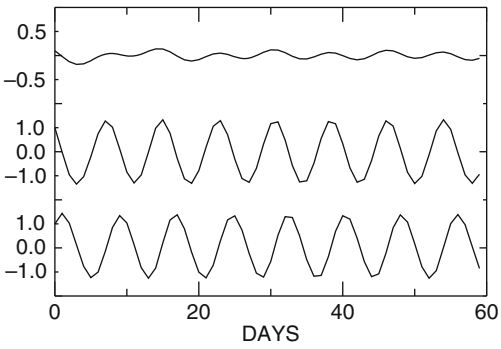



**Fig. 19.7** The results of the numerical integration for the system of Eq. (19.8) with $a = 0.25$, $b = 4$, $F = 8$, and $G = 0.2$

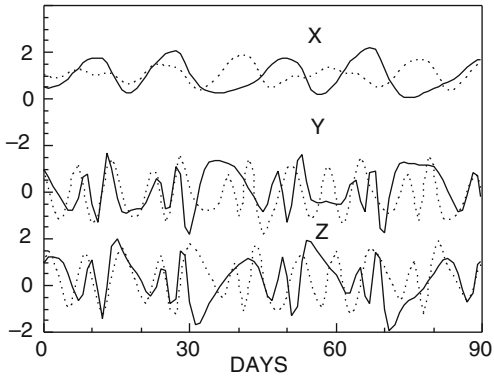



**Fig. 19.8** The same as Fig. 19.7 with $G = 0.8$. The *dashed line* refers to different initial conditions for X

We notice in any case that both in the zonal current and the eddies, two different regimes seem to be present: in $X$ there are two waves of different amplitudes that are not the same, while in the other two variables, there are periods of time in which either the low frequency or the high frequency dominates. High frequency is more common in the periods in which the zonal flux is high. These two distinct regimes suggest that we are in the presence of a case of intransitivity, that is, in climatic terms we could say that the system, without any external influence, gives rise to two completely different types of climate. A confirmation of this characteristic is that by changing only the initial condition of the first variable, the system gives completely different solutions which in the figure are represented by the dashed line. In any case the existence of the two regimes does not mean that we have identical variations in them and it is more like a season in different years (e.g., two winters): there may be snow and low temperatures, but they are far from identical. Even more interesting is the case shown in Fig. 19.9, where the asymmetry has been increased further. In this case the only zonal component is shown and we notice that the very long periods of time dominated by large differences between minima and maxima correspond to periods with smaller differences.

Calculating the attractor of the system (that turns out to be strange) enables us to make a more objective analysis of this behavior. This is shown in Fig. 19.10, which is composed of about 1250 points that represent the intersections of the trajectories with the plane $Y = 0$. In a three-dimensional space, the trajectories enter the plane of the sheet on the right, then exit, and reenter from behind on the left side. Lorenz made an extensive analysis of this attractor.

Some important considerations can be made starting with the examination of Fig. 19.7. We may assume that a weather forecast model shows a similar sensitivity to the initial conditions. In this case the initial values for the variables have a finite precision (like 5 m $s^{-1}$ for the wind and 1 °C for the temperature). These initial errors double about every 5 days so that after some time we are no longer sure of the forecast's reliability, because its present status could result from data different with respect to those used. This suggests the conclusion that the intrinsic nonlinearity of the system implies a limitation in the forecast capability that at the

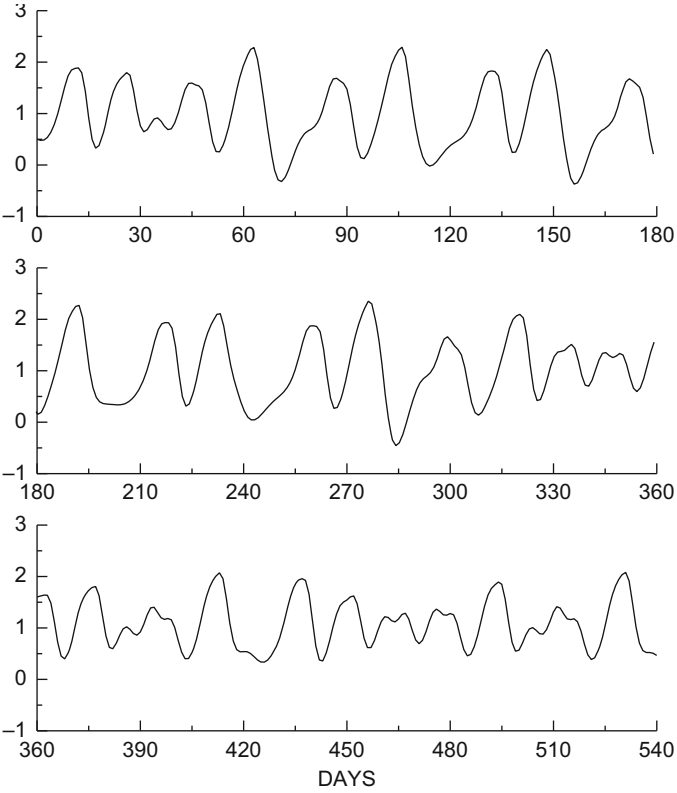
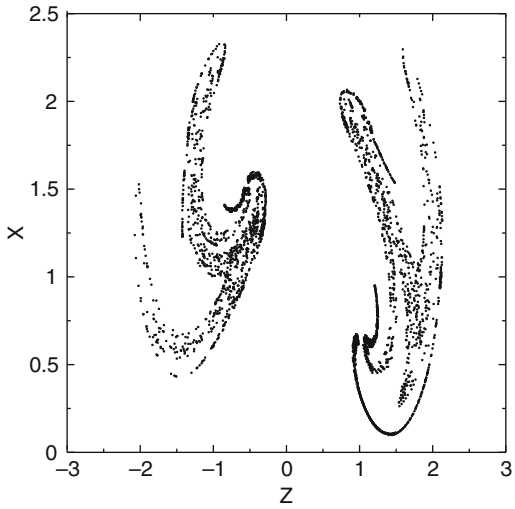



**Fig. 19.9** The $X$ component for three consecutive periods of 6 months each. The parameters are the same as in Fig. 19.7 except for $G = 1$

**Fig. 19.10** The attractor for the system of Eq. 19.8 includes about 1250 points



present time is of the order of 5 days. In order to improve this limit, the precision of the measurements must be improved by various orders of magnitude, with the obvious financial implications. It is possible to show that the doubling time for the error is linear with the logarithm of the precision, so that a significant increase in the forecast time implies considerable investment in the improvement of the instrumental techniques. This would not guarantee a better precision because we are not sure whether the forecast models include all the relevant nonlinear processes in the correct way.

## 19.3 Is El Niño Chaotic?

Each epoch invents its culprits and ours seem to have found one in El Niño. To this still quite unknown phenomena are attributed most of the weather troubles and anomalies. El Niño involves a very complex interaction between the tropical ocean and the atmosphere; this for that reason is also called ENSO (El Niño Southern Oscillation). The southern oscillation is a phenomenon referring mainly to the pressure behavior over the ocean. El Niño begins with an anomalous warming of the tropical Pacific Ocean that starts from the Peruvian coast and moves westward. The warming of the ocean increases the local convection so that the warmer regions correspond to intense precipitation while regions with subsidence have conditions of pronounced drought.

The theory of El Niño is still fundamentally the one suggested by Bjerknes which assumed that the trigger for El Niño is a weakening of the trade winds. These usually blow from east to west and force the ocean water westward, thus forcing a rising of cold water to the east of the Peruvian coast. When the trade winds weaken, the cold water updraft is blocked and the ocean starts to warm to the east. This warming reinforces the convection and consequently the trade winds so that the cycle starts again. Actually the most interesting data on the El Niño is the fact that the cycle does not have a fixed period and the intensity is not related to the frequency. It is these data that make El Niño a perfect candidate to be explained as chaotic phenomena. Another important consideration is that the ocean and the atmosphere have very different response times, with the atmosphere reacting in a few weeks and the ocean in months.

Considering the very unpleasant effects produced by El Niño on the regions that are on the tropical oceans, the predictability and the possible forecast of ENSO has considerable importance. However for a chaotic system the predictability has a quite different meaning.

As an exercise we can illustrate some of these points using a very simplified model, suggested by Geoffrey Vallis and shown in Fig. 19.11. In this model the ocean is divided in two regions, the upper and deep water. The latter is assumed to have a constant temperature $\overline{T}$, while the upper ocean has a horizontal temperature gradient resulting from the difference between the temperature on the eastern side

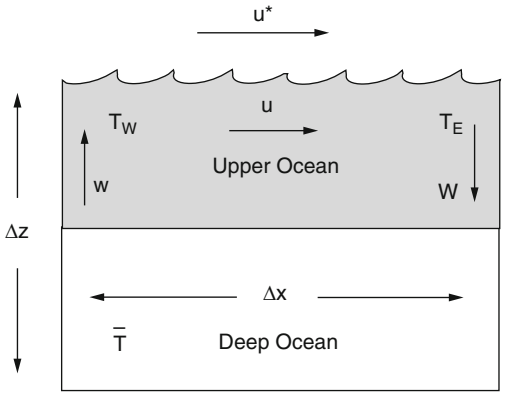



**Fig. 19.11** The model for El Niño. The ocean is assumed to be divided in two regions, one at constant temperature and the other in which the temperature is determined by the dynamics

$T_E$ and the western side $T_W$. The ocean surface current is determined largely by the wind $u^*$ whose direction is determined by the temperature gradient

$$(T_E - T_w)/\Delta x$$

The acceleration of the oceanic current $u$ will be

$$du/dt = B(T_E - T_W)/2\Delta x - C\left(u - u^*\right) \tag{19.10}$$

In practice it is assumed that the ocean current is due to the wind stress and to the damping term $-C(u-u^*)$. For the ocean temperature we assume that there is an advective term due to the oceanic current

$$u\left(\overline{T} - T_E\right)/\Delta x$$

and again a relaxation term proportional to the difference $(T_W - T^*)$. We have then

$$dT_W/dt = u\left(\overline{T} - T_E\right)/2\Delta x - A\left(T_W - T^*\right) \tag{19.11}$$

$$dT_E/dt = u\left(T_W - \overline{T}\right)/2\Delta x - A\left(T_E - T^*\right) \tag{19.12}$$

The temperature $T^*$ is the relaxation value in the absence of currents.

The system of Eqs. (19.10, 19.11, and 19.12) has a nonlinear character, and before going to its solution, it is interesting to find the equilibrium values. These may be obtained by putting for simplicity $\overline{T} = 0$ and $u^* = 0$ and making Eqs. (19.10, 19.11, and 19.12) zero. It is convenient to make a change in the variables according to

$$u = u/2A\Delta x; \quad t = At; \quad y = (T_E - T_W)/2T^*; \quad y = (T_E + T_W)/2T^*$$

to obtain

$$du/dt = by - cu; \quad dy/dt = uz - y; \quad dz/dt = -uy - (z - 1) \tag{19.13}$$

where we have put

$$b = T^*B/A^2(2\Delta x)^2 \text{ and } c = C/A$$

and also $u^* = 0$ and $\overline{T} = 0$. The system of Eq. (19.13) at the steady state has real solutions only if

$$c/b = AC(2\Delta x)^2/T^*B < 1$$

In some sense this solution is valid until $c \approx b$ so that $u = 0$ and $T_E = T_W = T'$. The limit is due to the fact that when $c \leq b$, the possible solutions for the variables become two that are stationary until $b$ does not reach the value

$$b_c = c^2\,(4 + c)\,/\,(c - 2)$$

Beyond this value the behavior of the system is chaotic. An equilibrium or stable solution means that small perturbations around the equilibrium points decay with time, while in a chaotic solution the initial perturbation will be amplified. In these conditions the behavior may be exemplified by looking at Fig. 19.12 which shows the behavior of the velocity $u$ and the temperature difference for the values of the different parameters of the system $C = 0.25$ month$^{-1}$, $A = 1$ year$^{-1}$, $u^* = -0.45$ m s$^{-1}$, and $T^* = 12$ °C. $B$ in this case has a value 2 m$^2$ s$^{-2}$ and the chaotic behavior occurs when this parameter is large enough to influence the wind.

The study of the chaotic behavior is much easier if we observe the trajectories in the phase space $T_E - T_W$, $u$, shown in Fig. 19.12. The thing we notice is that the system stays much longer in the region where the wind velocity is negative so that this represents the "normal" situation that is stationary but unstable. The El Niño state on the other hand is much less frequent.

## 19.4 Dimensions of Weather and Climate Attractors

One of the most ambitious ideas of the chaos theorist is to find how many variables are needed to describe complex nonlinear systems like the one that regulates weather and climate.

At the end of the 1980s, a number of papers were published in which people claimed with some surprise that the number of variables could be as low as 3 and no larger than 8. This was obtained by analyzing experimental data on climate or meteorological variables. The method used was quite ingenious and circumvented

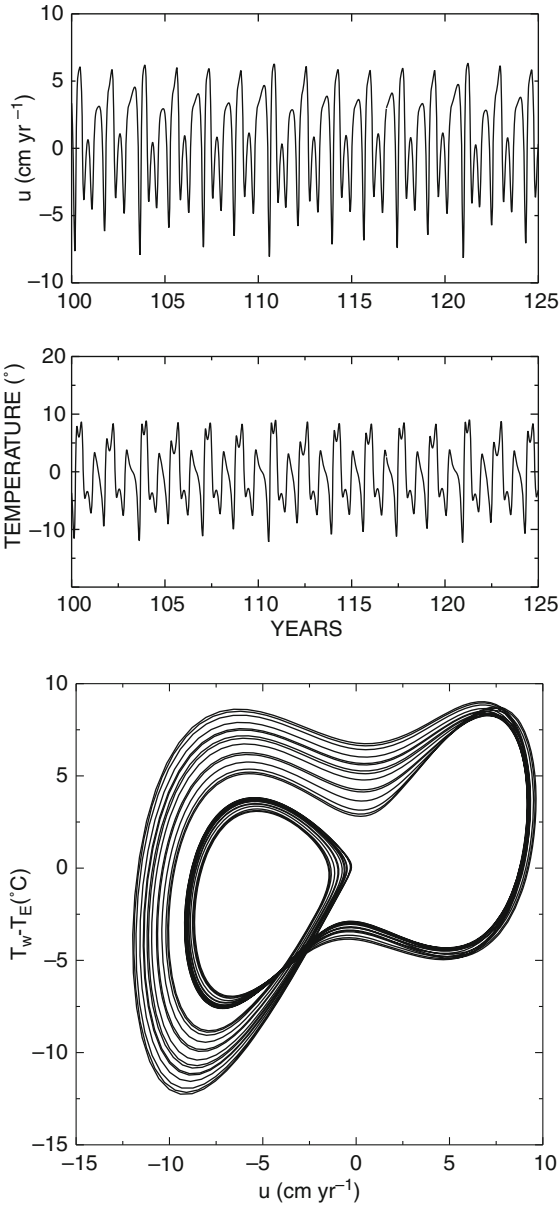



**Fig. 19.12** At the *top* is shown the behavior of the velocity $u$ and the temperature difference $T_E - T_W$ between 100 and 125 years of integration time of the system. The *bottom* shows the trajectory of the system on the phase space

the difficulty that nobody knows all the equations (or the right equations?) that govern weather and climate. The idea is that if we have a single record $x(t)$ of some variable $x$ as a function of time, from this we should be able to extract the information about the other variables. The physics behind the approach is that $x(t)$ should contain this information because it is the result of all interacting variables.

We may try to use some mathematics to explain the method. The evolution of the system may be described by a set of first-order differential equations:

$$\begin{aligned}\dot{x}_1 &= f_1\,(x_1, x_2, .....x_n)\\ \dot{x}_2 &= f_2\,(x_1, x_2, .....x_n)\\ &.\\ &.\\ \dot{x}_n &= f_n\,(x_1, x_2, .....x_n)\end{aligned}$$

It is well known that such a system can be reduced to a differential equation of the n$^{th}$ order that is

$$x_1^{(n)} = f\left(x_1, \dot{x}_1, ....., x^{(n-1)}\right)$$

This corresponds to substituting the state space coordinates ($x_1$, $x_2$, ….$x_n$) with the first $\left(x_1, \dot{x}_1, \ddot{x}_1, .....x_1{}^{(n-1)}\right)$ or second $\left(\dot{x}_1, \ddot{x}_1, .....x_1{}^{(n)}\right)$ derivatives with respect to time. In a rather famous paper, Ruelle, one of the pioneers of nonlinear physics, suggested in 1981 that rather than using the continuous variable $x(t)$ and its derivative, the discrete time series $x(t)$ and its successive shift by a delay parameter $\tau$ could be used.

In practice it should be possible to reconstruct a trajectory in a p-dimensional phase space by taking as coordinates $x(t), x(t+\tau), x(t+2\tau)\ldots x(t+(p-1)\tau)$. The shifting in time is actually equivalent to differentiation of a continuous-time series and it is quite evident for a periodic function like sine or cosine. The procedure is shown in Fig. 19.11 and is drawn for a three-dimensional space. In a chaotic regime the points along the same trajectory but far apart in time are uncorrelated. We can characterize however the correlation between points by introducing a function $C(r)$ such that

$$C(r) = \lim_{m\to\infty} \frac{1}{m^2}\left[\text{number of pairs } i,j \text{ whose distance } \left|\mathbf{x}_\mathrm{i} - \mathbf{x}_\mathrm{j}\right| < \mathbf{r}\right] \quad (19.14)$$

where $i$ and $j$, as shown again in Fig. 19.13, are the indices ordering the points along a trajectory containing a total of $m$ points. The form of $C(r)$ could be a power law of the form

$$C(r) \propto r^{d(m)}$$

where $d(m)$ may well be considered the correlation dimension for the attractor. As $m$ increases $d$ will approach a limit $d_G$ that should be the correlation dimension for the attractor of the reconstructed system. A very simple example will clarify the approach. We start with the generation of a white noise signal using a random

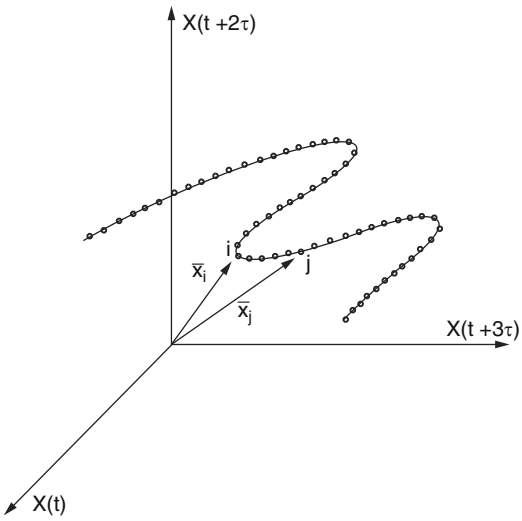



**Fig. 19.13** The reconstruction of a trajectory in phase space starting from a simple variable *x(t)* using the method of delays. In this case a 3D space is used with coordinates $x(t)$, $x(t+\tau)$, and $x(t+2\tau)$ (Bergè et al. 1984)

number generator. We then apply, starting from the first value, the second coordinate $x(t+\tau)$, the third $x(t+2\tau)$, and so on. We can calculate the distance between two points in the $p$-dimensional space by using

$$\left|\mathbf{r_{ij}}\right| = \left[\left(x_{1i} - x_{1j}\right)^2 + \left(x_{2i} - x_{2j}\right)^2 + \cdots \left(x_{(p)_i} - x_{(p)j}\right)^2\right]^{1/2} \tag{19.15}$$

where $x_1 = x(t)$, $x_2 = x(t+\tau)$, ... : At this point we can calculate the number of "pairs" for which the distance is less than $r$, and the result can be plotted as shown in Fig. 19.12. For each value of $p$ the slope of the curve can be determined and from it the exponent $d$. It is noted that as the number of dimension increases the slope also increases so that for large $p$ we find $d \;\; p$. The same procedure could be applied in principle to any signal with the difference that, while for white noise there is no saturation, that is, the slope increases with the number of dimensions, it is possible that for the signal we will observe a saturation. This would indicate that the process from which the signal results has an attractor embedded in a space with a dimension corresponding to that of the saturation.

This procedure was applied, for the first time, by Nicolis to climatic data that refer to the oxygen isotopes we studied in connection with the ice ages. He found, surprisingly, a dimension for the attractor around 3.1. Again the procedure was applied to weather data (geopotential, pressure, and wind), and the dimension again was found to be within 3 and 8.

Edward Lorenz became a little suspicious about these estimates and used a model similar to the one describing the atmospheric circulation but for which the variables could be expanded according to the rule $(3x(2^L - 1))$ with $L = 1$ or $L = 3$. The complete system looks like this

$$
\begin{aligned}
dX_j/dt &= -q_j\left(Y_j^2 + Z_j^2\right) - a_j\left(X_j - F_j\right) + U_j \\
dY_j/dt &= X_j\left(q_j Y_j - b_j Z_j\right) - p_j\left(Y_j - G_j\right) + V_j \\
dZ_j/dt &= X_j\left(b_j Y_j + q_j Z_j\right) - p_j Z_j + W_j
\end{aligned} \tag{19.16}
$$

for $j = 1, \ldots, 2^L - 1$ and where

$$
\begin{aligned}
U_{2j} &= -cp_j Y_j \\
U_{2j+1} &= -cp_j Z_j \\
V_j &= cp_j\left(X_{2j} - h_{2j}\right) \\
W_j &= cp_j\left(X_{2j+1} - h_{2j+1}\right)
\end{aligned} \tag{19.17}
$$

for $j = 1, \ldots, (2^{L-1} - 1)$.

The only difference between Eqs. (19.16) and (19.8) is the coupling terms $U$, $V$, and $W$. The values used for the other quantities were $p_j = q_j = 1$, $a_j = 0.25$, $b_j = 4.0$, $F_j = 8.0$, $G_j = 1.0$, and $h_j = 1.0$ for all $j$. The real coupling is through $c$ which is taken as 0.1 for a weak coupling and 1 for a strong coupling. The procedure adopted by Lorenz was to estimate the attractor dimension $d_c$ from the box counting procedure we have illustrated earlier and also from reconstructing the trajectory and calculating $C(r)$. For the low-order system, the result found by Lorenz was that for $L = 1$, when the new model reduces to Eq. (19.8), the dimensions $d_c$ and $d_G$ roughly coincide. For $L = 3$ there are 21 variables and in this case the difference between the two dimensions depends on both the coupling coefficient $c$ and the variables chosen for the evaluation. For weak coupling ($c = 0.1$) the capacity dimension is 17 while the correlation dimension $d_G$ is roughly 5. For strong coupling ($c = 1$) the capacity remains around 17 with $d_G$ about half of that.

The conclusion is that the procedure of delays gives a systematic underestimation of the attractor dimension so that when dealing with a natural signal, one could get the wrong impression of a low number of variables. The reason found by Lorenz is that, as shown in Fig. 19.14, the power law should be particularly good for a large number of points and "small enough $r$." However although it is not particularly difficult to find a large number of data, there are no indications about the smallness of $r$. Lorenz suggested that the atmosphere–climate system may consist of a number of low-order subsystems loosely coupled. The real data analysis reporting the low dimension for the attractor was probably sampling one of these low-order systems.

Few doubt that in considering the complexity of the Earth system, chaos theory is the way to go, but the road is still quite unclear, especially because the models used until now are quite far from reality.

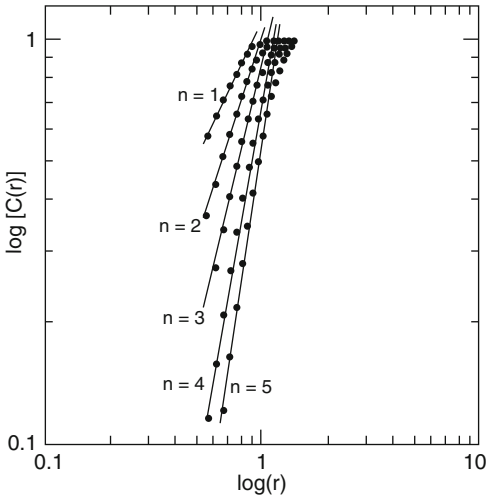



**Fig. 19.14** The function $C(r)$ for white noise. Notice the increasing slope with increasing embedding dimension $n$

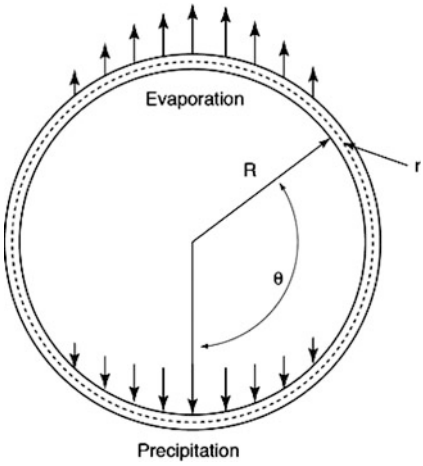



**Fig. 19.15** A loop oscillator for the haline circulation: a tube filled with salty water. The freshwater is passed through the skin of the tube (Huang 2009)

## 19.5 A Bridge to Nonlinearities: The Loop Oscillator

This is an example due to Dewar and Huang. We assume to have a loop (as in Fig. 19.15) where the difference between precipitation (p) and evaporation (q) is given by

$$p - e = E \cos \theta \tag{19.18}$$

We use a cylindrical coordinate system with tangential components denoted by subscript $\theta$, so we write the continuity equation as divergence equal to freshwater flux:

$$\partial u/\partial\theta = 2RE\cos\theta/r \tag{19.19}$$

define the angular velocity as

$$\omega = u/R \tag{19.20}$$

Integrating (19.19) we have

$$u = 2RE\sin\theta/r + C \Rightarrow \omega = 2E\sin\theta/r + C$$

Defining the average angular velocity as

$$\Omega = \int_0^{2\pi} \omega d\theta/2\pi$$

we have $C = \Omega$ and then

$$\omega = \Omega + 2E\sin\theta/r \tag{19.21}$$

For the salt concentration $S$, we can write the continuity equation

$$\frac{\partial S}{\partial t} + \frac{\partial(\omega S)}{\partial\theta} = \frac{K}{R^2}\frac{\partial^2 S}{\partial\theta^2} \tag{19.22}$$

where $K$ is the diffusivity coefficient for salt.

The momentum balance for an infinitely small sector of the water tube is given by

$$\rho_0\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial l}\right) = -\frac{\partial P}{\partial l} - \rho g\sin\theta - \varepsilon\rho u$$

where l is the direction along the tube, $P$ is the pressure, and the first term is just a viscous term. Again integrating the momentum equation along the entire loop,

$$\frac{\partial\Omega}{\partial t} = -\varepsilon\Omega - \frac{g}{2\pi R}\int_0^{2\pi}\frac{\rho}{\rho_0}\sin\theta d\theta \tag{19.23}$$

Using the density relation $\rho = \rho_0(1 + \beta S)$, with β salinity coefficient the equation becomes

$$\frac{\partial\Omega}{\partial t} = -\varepsilon\Omega - \frac{g\beta}{R}\langle S\sin\theta\rangle \tag{19.24}$$

Equations (19.22) and (19.24) can be normalized according to

$$S = \overline{S}S'; \quad t = Tt'; \quad \omega = \omega'/T$$

where $\overline{S}$ is the average salinity and $T = \sqrt{R/g\beta\overline{S}}$. Equation (19.24) becomes

$$\frac{\partial \Omega'}{\partial t'} = -\varepsilon T\Omega' - \frac{g\beta T^2}{R}\overline{S}\langle S' \sin\theta\rangle$$

T can be determined by imposing that the coefficient of the second term on the right equal to unity. We have after dropping the primes

$$\dot{\Omega} = -\alpha\,\Omega - \langle S \sin\theta\rangle \tag{19.25}$$

where $\alpha = \varepsilon T = \varepsilon\sqrt{R/g\beta\overline{S}}$. Equation (19.22) becomes after being normalized

$$\dot{S} = kS_{\theta\theta} - [(\Omega + \lambda \sin\theta)\, S]_\theta \tag{19.26}$$

where

$$\lambda = 2ET/r \quad \kappa = KT/R^2$$

In the most general case, solution to the set (19.25 and 19.26) can be obtained by expanding the solution for the salinity in a Fourier series:

$$S = 1 + \sum_1^\infty (2a_n \sin n\theta + 2b_n \cos n\theta) \tag{19.27}$$

The simplest case is to retain only the $\sin\theta$ and $\cos\theta$ mode so that the set of equation reduces to

$$\dot{\Omega} = -\alpha\,\Omega - a_1$$

$$\dot{a}_1 = \Omega b_1 - \kappa a_1$$

$$\dot{b}_1 = -\Omega a_1 - \kappa b_1 - \lambda/2$$

This system can be reduced to the classical Lorenz equation with the linear transformation

$$x = \Omega; \quad y = -\frac{a_1}{\alpha}; \quad z = \frac{b1}{\alpha} + \frac{\lambda}{2\kappa\alpha} \tag{19.28}$$

and the system becomes

$$\begin{aligned}\dot{x} &= \alpha\,(x-y)\\ \dot{y} &= x\,(r-z)-\kappa y\\ \dot{z} &= xy-\kappa z\end{aligned} \tag{19.29}$$

where $r=\lambda/2\alpha\kappa$. This system coincides with the Lorenz system with $b=1$ when $\kappa=1$.

## 19.6 The Thermohaline Circulation According to Stommel

### *19.6.1 The Model*

To illustrate further some properties of the nonlinear system, we leave for a moment the atmosphere and again go into the ocean. The oceanic circulation is largely determined by the interaction with atmospheric wind (we mention that for El Niño) but also by buoyancy effects due to different salinity and temperature. Salinity is expressed as the grams of salt dissolved in a kg of water. In the ocean this number changes between 32 and 37. The equation of state for ocean water in its simplest form is written as

$$\rho=\rho_0\left[1-\alpha\,(T-T_0)+\beta\,(S-S_0)\right] \tag{19.30}$$

where $\alpha$ and $\beta$ are two constants with value of $5.26\cdot10^{-5}$ K$^{-1}$ and $7.86\cdot10^{-4}$ psu$^{-1}$. The other constants are $\rho_0=1028.1$ Kgm$^{-3}$ and $S_0=35$ psu (practical salinity units). The simplest model of the oceanic circulation due to difference in density can be the one illustrated in Fig. 19.15. The two boxes represent the high-latitude ocean at temperature $T_1$ and salinity $S_1$ which loses salinity because of the precipitation, while the low-latitude ocean at temperature $T_2$ and salinity $S_2$ gains salinity because of evaporation. The two boxes exchange water at the surface and at the bottom with a flux $q$ (measured in s-1) with following convention. At the surface poleward flow is when q > 0 that implies equatorward bottom flow. Conversely at the surface q < 0 means equatorward flow and poleward bottom flow. We simply assume that the flux is given by

$$q=k\,(\rho_1-\rho_2)\,/\rho_0 \tag{19.31}$$

So that according to (19.30) we have

$$q=k\left[\alpha\,(T_2-T_1)-\beta\,(S_2-S_1)\right] \tag{19.32}$$

We further assume that the temperature is fixed, so we need to worry only about salinities with conservation equations:

$$\begin{aligned}\dot{S}_1 &= -H_S + |q|\,(S_2 - S_1)\\ \dot{S}_2 &= +H_S - |q|\,(S_2 - S_1)\end{aligned} \tag{19.33}$$

where $H_S$ is the virtual salinity flux defined as $S_0E/D$ with $S_0$ reference salinity, $E$ freshwater flux (measured in $\mathrm{ms}^{-1}$), and $D$ depth. We further define

$$\Delta T = (T_2 - T_1) \quad \Delta S = (S_2 - S_1) \quad \rho = (\rho_1 - \rho_2)$$

Implying that

$$q = k\,(\rho/\rho_0) = k\,[\alpha\Delta T - \beta\Delta S]$$

Under normal conditions we expect salinity and temperature to be both higher at low latitudes and both lower at high latitude. When the temperature difference dominates the salinity difference, high-latitude density will be higher than low-latitude density. Therefore $q>0$ and the surface flow is poleward. In this case

$$q > 0: \quad |q| = q = k\,[\alpha\Delta T - \beta\Delta S]$$

So that in this case temperature difference drives the circulation and salinity breaks it. The opposite happens when salinity difference dominates temperature difference. In this case $q<0$ the surface flux is equatorward

$$q < 0: \quad |q| = -q = k\,[\beta\Delta S - \alpha\Delta T]$$

To solve the system we add and subtract (19.33) to obtain

$$\dot{S}_1 + \dot{S}_2 = 0 \tag{19.34}$$

That is the conservation of salinity. And

$$\dot{S} = 2H_S - 2k\,[\alpha\Delta T - \beta\Delta S]\,\Delta S \tag{19.35}$$

The equilibrium solutions can be found by putting

$$H_S - k\,|\alpha\Delta T - \beta\Delta S|\,\Delta S = 0$$

We must distinguish between the two cases in which the argument of the modulus is positive or negative. In the first case

$$\overline{q} > 0; \quad \alpha\Delta T > \beta\Delta\overline{S}$$

We then have

$$H_s - k\left(\alpha\Delta T - \beta\Delta\overline{S}\right)\Delta\overline{S} \Rightarrow \left(\beta\Delta\overline{S}\right)^2 - \beta\Delta\overline{S}\left(\alpha\Delta T\right) + \beta H_S/k = 0$$

which has the roots

$$2\beta\Delta\overline{S} = \alpha\Delta T \pm \left[(\alpha\Delta T)^2 - 4\beta H_S/k\right]^{1/2} \tag{19.36}$$

Then for poleward surface flow when

$$(\alpha\Delta T)^2 > 4\beta H_S/k \Rightarrow \beta H_S/k(\alpha\Delta T)^2 < 1/4$$

we have two solutions. In the other case it is easy to show that

$$\overline{q} < 0; \quad \alpha\Delta T < \beta\Delta\overline{S}; \left(\beta\Delta\overline{S}\right)^2 - \left(\beta\Delta\overline{S}\right)\left(\alpha\Delta T\right) - \beta H_S/k = 0$$

with the single root

$$\beta\Delta\overline{S} = \frac{\alpha\Delta T}{2} + \frac{1}{2}\left[(\alpha\Delta T)^2 + 4\frac{\beta H_S}{k}\right]^{1/2} \tag{19.37}$$

In this case we have to discard the negative root because in this case $\beta\Delta\overline{S} < 0$ in contradiction with the condition $\alpha\Delta T < \beta\Delta\overline{S}$. Of the three solutions we can show that only two are stable. We just perturb the quantities

$$\Delta S = \Delta\overline{S} + S'; \quad q = \overline{q} + q'$$

where $\overline{q} = k\left[\alpha\Delta T - \beta\Delta\overline{S}\right]$. We can easily obtain

$$\dot{S'} = 2\left|\overline{q}\right| S' \pm k\beta S'\left(\Delta\overline{S} + S'\right); \quad + : \overline{q} > 0; \quad - : \overline{q} < 0$$

This expression cab be linearized if we assume $S' << \Delta\overline{S}; \quad q' << \overline{q}$ and in that case

$$\dot{S'} = -2\left|\overline{q}\right| S' \pm k\beta S'\Delta\overline{S}; \quad + : \overline{q} > 0; \quad - : \overline{q} < 0$$

Now assuming the perturbation evolves as $S' = S_0 e^{\lambda t}$ we obtain

$$\lambda = -2\left|\overline{q}\right| \pm 2k\beta\Delta\overline{S} = -2k\left|\alpha\Delta T - \beta\Delta\overline{S}\right| \pm 2k\beta\Delta\overline{S}$$

It is convenient at this point to introduce the dimensionless salinity difference defined as

$$\delta = \beta\Delta\overline{S}/\alpha\Delta T$$

so that the normalized flux becomes

$$q = \frac{\overline{q}}{k\alpha\Delta T} = \frac{k\left(\alpha\Delta T - \beta\Delta\overline{S}\right)}{k\alpha\Delta T} = 1 - \delta \tag{19.38}$$

The condition for instability is when $\lambda > 0$ so when we use the plus sign ($\overline{q} > 0$) the condition for instability is satisfied if

$$2\beta\Delta\overline{S} > \alpha\Delta T \Rightarrow \beta\Delta\overline{S}/\alpha\Delta T = \delta > 1/2$$

That coupled with the condition $q > 1$ gives the instability region when $0.5 < \delta < 1$ while in case $\overline{q} < 0$ we have stability for $\delta > 1$. In terms of $\delta$, we have the solutions

$$\begin{aligned}
\delta_1 &= \frac{1}{2}\left(1 - \sqrt{1 - 4E}\right) \quad 0 < \delta < 0.5 \\
\delta_2 &= \frac{1}{2}\left(1 + \sqrt{1 - 4E}\right) \quad 0.5 < \delta < 1 \\
\delta_3 &= \frac{1}{2}\left(1 + \sqrt{1 + 4E}\right) \quad \delta > 1
\end{aligned} \tag{19.39}$$

The first two solutions are further constrained by the fact that $E < 0.25$. These conditions are summarized in Fig. 19.16.

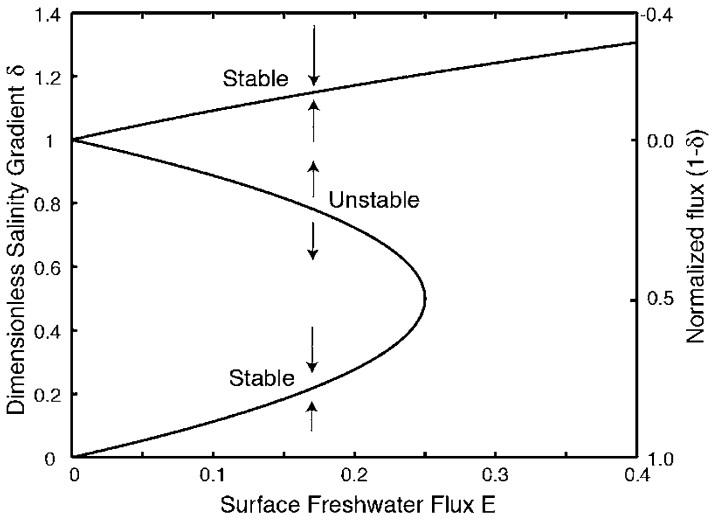



**Fig. 19.16** Graphical representation of the box model. Curves show the equilibrium solutions, while the arrows show the tendencies. Notice that for $E < 0.25$ there are three possible equilibrium states

### 19.6.2 Stability of the Solutions

We start out from Eq. (19.35) normalized:

$$\frac{1}{2k\alpha\Delta T}\frac{d}{dt}\left(\frac{\beta\Delta S}{\alpha\Delta T}\right)=\frac{\beta HS}{k(\alpha\Delta T)^2}-\left|1-\left(\frac{\beta\Delta S}{\alpha\Delta T}\right)\right|\left(\frac{\beta\Delta S}{\alpha\Delta T}\right)$$

At this point we need to introduce a normalized time:

$$t'=t\,(2k\alpha\Delta T)$$

So that the previous equation becomes

$$\dot{\delta}=E-|1-\delta|\,\delta \tag{19.40}$$

This equation tells that on the equilibrium curve the time derivative is zero. However on the left of that curve $E$ is less than that required by equilibrium so that $\dot{\delta}$ is negative and the salinity difference decreases with time. This is indicated by the arrows. The opposite happens on the right of the equilibrium curve where now $\dot{\delta}$ is positive and the salinity difference increase with time. This means that if we start from a point at $\delta < 0.5$, the system will return to the equilibrium curve as indicated by the arrows. The same happens if we start from a point with $\delta > 1$. However, if we start from a point where $0.5 < \delta < 1$, then the system will not return to the initial point but will move toward the stable region with $\delta > 1$. Notice that based on the definition of $\delta$, when this parameter is larger than unity $\beta\Delta S > \alpha\Delta T$ and the regime will be dominated by salinity difference. On the other hand when $\delta < 0.5$, then $\beta\Delta S < 0.5\,(\alpha\Delta T)$ and the regime will be determined by thermally strong flow. This implies that the stable region $\delta > 1$ will be dominated by salinity differences. In particular the point $E = 0.25$, $\delta = 1$ beyond which no thermally direct steady state is possible in language of dynamical systems is called a saddle node bifurcation.

These instabilities can be interpreted in real terms referring to Fig. 19.17. Starting from a point that would correspond to the present climate if the freshwater flux is increased (e.g., by melting the ice polar caps), the system will move toward the bifurcation point and then will "fall" to the southward-directed surface flux. The system will not recover by simply decreasing the freshwater flux because the new transition to the present climate will occur when the freshwater flux is stopped. This simple model has been used by many to make some interesting observation on the occurrence of glacial cycles. Reaching the bifurcation point would correspond to the collapse of the North Atlantic Deep Water (NADW) formation. Considering the catastrophic character of the collapse, the transition could happen only in a few decades, something like the movie "The day after tomorrow."

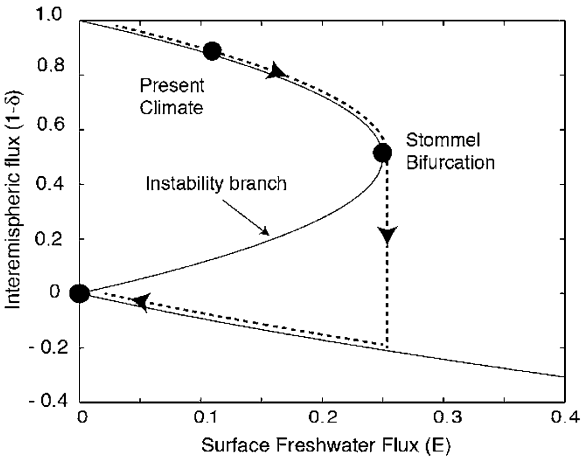



**Fig. 19.17** The hysteresis cycle produced by the Stommel 2 box model. This figure is the same as the previous one with the scales changed. Notice that the flux is southward for the lowest branch

## 19.7 The Difference Equations

As usual it was Edward Lorenz that started the business in a paper published in 1964. As we have seen in many instances, the resolution of a differential equation (or a system of differential equation) must go through a transformation in the difference equation. We report exactly the words of Lorenz:

> When the original equations are nonlinear, the equivalent difference equations generally cannot be written in finite form in term of familiar analytic functions. The existence of the difference equations is assured, however, by the existence of the solutions of the differential equations.
>
> We therefore lose no generality in choosing an arbitrary system of equation to illustrate the problem of deducing the climate, if we choose a system of difference equations instead of differential equations.

In that paper Lorenz was interested in the definition of climate (something we have done before) as "long-term statistical properties of the atmosphere." He was mostly concerned with climatic changes which "are of necessity changes in statistics taken over finite intervals" and refer mainly to intervals from decade to centuries. The interest is concentrated on the fact that even if the environmental conditions do not change, the statistics may change. If the fluctuations of the atmosphere are of short period (i.e., the weather), the separate sample may be hardly distinguishable from one another. Sometime it could happen that the fluctuations have a duration comparable to that of the sample and the separate sample may have different statistics: this is more like climatic change. The interest is then to find very long period fluctuations when the environmental conditions do not change.

It is relatively easy to explain the weather fluctuations (periods from 1 day to 1 year) in terms of instability, while it is much more complex to explain

low-frequency variability in the presence of a constant environment. In mathematical terms climate may be expressed in a set of differential equations with the property that the long-term statistics of certain time-dependent solutions are not the same as those of other time-dependent solutions. In practice some systems have the properties that there are two or more sets of statistics, any one of which could constitute the climate of the system. The particular climate which prevails depends upon the particular conditions that prevail when the system first was established. Such systems are called *intransitive* according to Lorenz. Systems in which only one climate is physically possible are called *transitive*. We do not know if our climatic system is either transitive or intransitive. In particular for the intransitive system, the transition from one state to another could result from a change in the environmental conditions (i.e., changes in the solar constant), and the return to the initial state could happen when the previous conditions are reestablished. Two particular time-dependent solutions of the system may appear to have different statistics if the solutions are considered only in a moderate time span and the system may appear to be intransitive. However, if the time interval is sufficiently extended, the statistics may look quite similar. This means that a solution will exhibit different statistical properties within different intervals of a long time span. This system is called *almost intransitive.* Figure 19.18 exemplifies such a definition. In a transitive system two initial states A and A' evolve into the single equilibrium state B. An intransitive

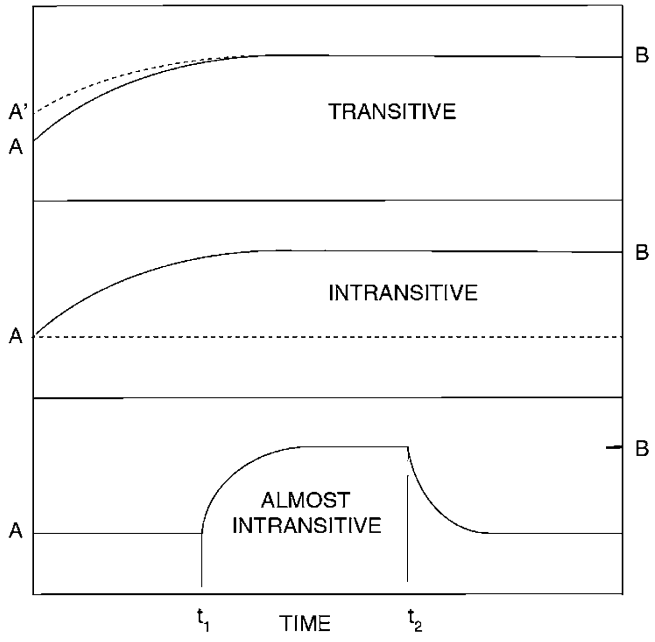



**Fig. 19.18** Diagram showing a transitive (*upper*), intransitive (*middle*), and almost intransitive (*bottom*) climate system from an initial state A (Adapted from Peixoto and Oort 1992)

system may have two or more equilibrium state A and B for the same boundary conditions. An almost intransitive system may behave as if were intransitive up to a time $t_1$ shifting to an alternative state B where it could remain up to a time $t_2$.

### *19.7.1 Examples for Transitive and Intransitive System*

Following the previous considerations, we lose no generality if we consider a system of nonlinear difference equations capable of generating a stable climate. The simplest one could be of the form

$$X_{n+1} = f(X_n) \tag{19.41}$$

where $f$ is continuous and single valued in $X$. If $X_0$ is an arbitrary chosen initial condition, Eq. (19.41) will generate a series $\{X\} = \{X_0, X_1, X_2, .....\}$, and the long-term statistics of this series constitute the climate determined by Eq. (19.41). Lorenz observed that if (19.41) has not a steady-state solution, that is, $X_0 = X_1 = X_2 \ldots$, then in view of the continuity, either $X_0 < X_1 < ..$, or $X_0 > X_1 > \ldots,$. This means that $X_n \to \infty$ or $X_n \to -\infty$ as $n \to \infty$ since any finite limit would correspond to a steady state. The series $\{X\}$ then possesses no climate and we shall require that $f(X) = X$ for at the least a finite value of $X$.

The simplest continuous nonlinear function would appear to be a quadratic function. The most general quadratic equation with at least a steady-state solution can written as

$$X_{n+1} = aX_n - X_n^2 \tag{19.42}$$

with $a \geq 0$. It can be easily verified that if $a > 4$, the choice $X_0 = \frac{1}{2}a$ implies $X_1 > a$ and $X_2 < 0$ after which $X_n \to -\infty$ as we shall then require that $0 \leq a \leq 4$ and that $0 \leq X_0 \leq a$.

Using Eq. (19.42) we will derive a "climate" defined with the mean

$$\overline{X} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{n=N-1} X_n \tag{19.43}$$

Noting in particular how $\overline{X}$ varies with $a$.

For some, the value of a in equation (2) may have shown to be transitive, that is, to determine a stable climate. For all other values between 0 and 4 we will make the hypothesis that Eq. (19.42) is transitive and speak of the value of $\overline{X}$ corresponding to a. There is no proof of transitivity, but the many numerical solutions do not indicate any suggestion for intransitivity.

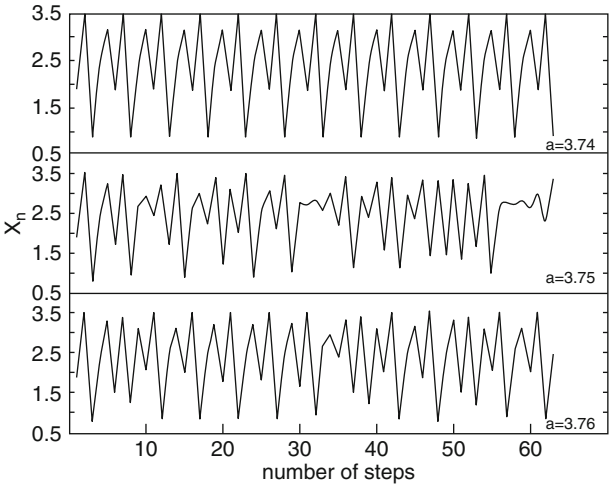



**Fig. 19.19** Numerically determined solutions of Eq. (19.42). Notice the periodicity of the first solution

**Fig. 19.20** The mean $\overline{X}$ for values of a from 3.74 to 4

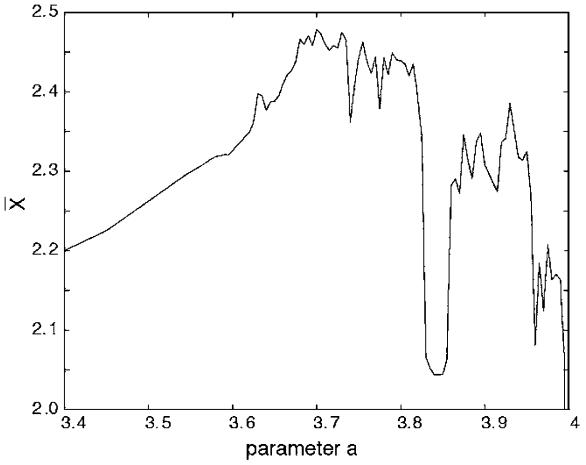



Figure 19.19 shows the results obtained by solving numerically Eq. (19.42) for three different values of $a$. The initial value $X_0 = \mathrm{a}/2$. We can see that for $a = 3.74$, the solution is periodic with period of 5, while for all the other values there is nonapparent periodicity. In any case to obtain a reasonable mean, we may use a large number for n. We have used $\mathrm{n} = 1024$ and have evaluated the mean for values of a going from 3.74 to 4 with steps of 0.005 and the results are shown in Fig. 19.20. The most striking characteristic of this graph is its irregularity so that we can hardly find any possible definition of this climate.

A proposal for a similar equation to (19.42) for an almost intransitivity climate was formulated by Lorenz in 1976. He considered the following single first-order cubic difference equation:

$$X_{n+1} = aX_n\left(3 - 4X_n^2\right) \tag{19.44}$$

This equation as the previous one may not have any connection with reality except for the fact that as we have seen before in a few problems, we have a quadratic term (i.e., Eq. 19.8) and in many other instances the results of transforming a differential equation in a difference equation may give rise to quadratic terms. We will use then Eq. (19.44) to show some mathematical properties of a transitive or intransitive system. If $a > 1$, solution of Eq. (19.44) will blow up. However, if $a \leq 1$ and $|X_n| \leq a$, then $|X_{n+1}| \leq a$. A choice of an initial value $|X_0| \leq a$ will determine an infinite sequence $X_0, X_1, X_2 \ldots$, with each term lying between $-a$ and $a$. Based on the previous considerations, the term $\left(3 - 4X_n^2\right)$ in (19.44) is the one that will change the sign of $X_{n+1}$ with respect to $X_n$, and this will happen when $X_n^2 \leq 3/4$. As a consequence if $a^2 \leq 3/4$, then $X_n$ and $X_{n+1}$ must have the same sign so that a sequence consists entirely of positive or entirely of negative numbers. Since the positive and negative sequences have different means the system is clearly intransitive. If on the other hand $a^2 \geq 3/4$, change of sign within a sequence is possible. For example, if $X_n$ is close enough to 0.5, then $X_n^2 > 3/4$ and $X_{n+2}$ and $X_{n+1}$ have opposite sign. For entrain values of $a^2 \geq 3/4$, the sequence will contain mostly positive terms and other mostly negative terms, and so we have two "climates." However for most of the values of $a^2 \geq 3/4$ the system is transitive. When $a^2$ exceeds ¾ only slightly, the range of value of $X_n$, about ½, is very small. It may be anticipated that most sequences will be characterized by large numbers of successive terms of one sign and transitions will be relatively rare. Examination of a few short segments might not reveal any changes and the system qualify as almost intransitive (Fig. 19.21).

## 19.8 Nonlinearity and Delayed Differential Equations

A simple example will clarify the purpose of this paragraph. We consider the so-called continuous logistic equation that deals with the population growth as a function of the carrying capacity, $K$, and the growth rate $r$:

$$\frac{dN}{dt'} = rN\left(1 - \frac{N}{K}\right) \tag{19.45}$$

In the exercise we will treat the discrete form of this equation, and we will see why it has been used to study chaos. There are two steady-state solutions, one with $N = 0$ and the other with $N = K$. The continuous logistic equation assumes that the birth

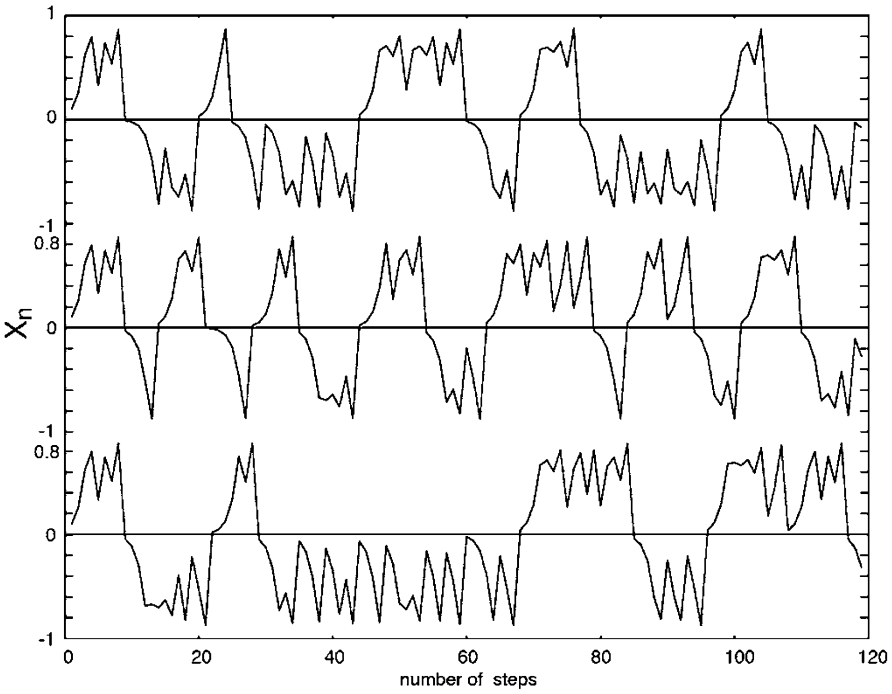



**Fig. 19.21** Solutions for Eq. (19.44) for initial conditions of 0.0999 (*upper*), 0.1000 (*middle*), and 1.001 *lower*

rate depends instantaneously on the population size, while it seems obvious that there should be a delay that can be incorporated in the terms between brackets:

$$\frac{dN}{dt'} = rN\left[1 - \frac{N\left(t' - \tau\right)}{K}\right] \tag{19.46}$$

This means that the loss term is not based on the instantaneous population but on the one at time $(t' - \tau)$. Equation (19.46) can be normalized using the variables

$$y = N/K \qquad t = t'/\tau$$

and we arrive to the nondimensional equation (Shampine et al. 2003)

$$\frac{dy}{dt} = \alpha y\left[1 - y\left(t - 1\right)\right] \tag{19.47}$$

where $\alpha = r\tau$. A particular aspect of the delay equation is that initial conditions need to be specified over the time interval from 0 to 1 (the so-called history). Typically this is assumed to be a constant. Equation (19.47) is clearly nonlinear and

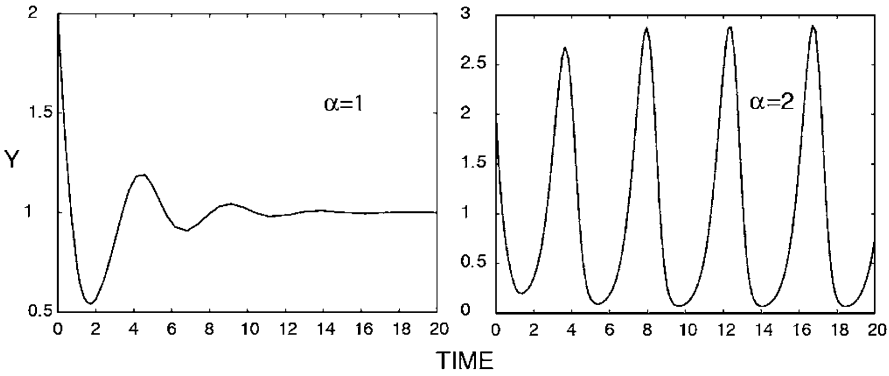



**Fig. 19.22** The solutions of the logistic equation for two values of the parameter $\alpha$

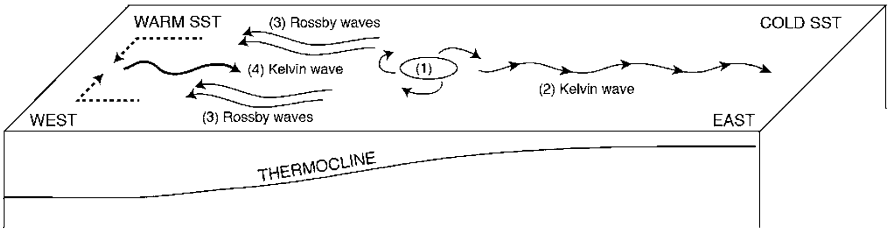



**Fig. 19.23** A schematic picture of the delayed oscillator mechanism (Adapted from Suarez and Schopf (1988) and Tzipermann)

the solution can be obtained with a standard MATLAB program. Solutions depend on the value of $\alpha$ and are depicted in Fig. 19.22, where it is shown that for $\alpha = 1$, the solution is a damped oscillation, while for $\alpha = 2$, the solution is a sustained oscillation. This behavior will be clarified is some of the examples.

Now that we are equipped with this additional knowledge, we can apply the delay differential equation to atmospheric problems.

### *19.8.1 ENSO as a Delay Oscillator*

We have modeled ENSO previously as a possible example of chaotic system. However Suarez and Schopf in a 1988 paper made the hypothesis that ENSO could be treated as a delayed oscillator. The idea is based on the complex interaction between ocean and the atmosphere in which Kelvin and Rossby waves play a fundamental role. We may refer to Fig. 19.23 to fix the ideas. A weakening of the wind (1) creates an equatorial (downwelling) Kelvin wave (2) that travels to the east Pacific within 1–2 months where the thermocline deepening induces an SST heating and starts the El Niño event. The SST heating further weakens the central Pacific

winds, and the event is therefore amplified by the ocean–atmosphere instability. The initial wind weakening also creates off-equatorial cold (upwelling) Rossby waves (3) that are reflected from the western boundary as cold Kelvin waves (4) arrive at the eastern boundary and terminate the event.

To obtain some numbers, we call $\tau_K$ and $\tau_R$ the basin crossing times for Kelvin and Rossby waves. If we call $h_{eq}(x_c)$ the positive deep anomaly at the central Pacific that happens at time $t-\frac{1}{2}\tau_K$ excites an eastward-propagating Kelvin wave that arrives after about $\frac{1}{2}\tau_K$ to the eastern Pacific and deepens the thermocline there within $\tau_K \approx 1-3$ months. Similarly, a negative off-equatorial deep anomaly in the central Pacific ($h_{off-eq}(x_c)$) at time $t-[\frac{1}{2}\tau_R+\tau_K]$ excites a westward propagating Rossby wave at the central Pacific that is reflected off the eastern boundary as an equatorial Kelvin wave arrives at the eastern boundary at time, shallows the thermocline there, and cools the SST. We may the write the change of temperature at the time t:

$$\frac{dT(t)}{dt} = a' h_{\text{eq}}\left(x_c, t-\tfrac{1}{2}\tau_k\right) + b' h_{\text{off-eq}}\left(x_c, t-[\tfrac{1}{2}\tau_R+\tau_K]\right) \tag{19.48}$$

To this equation we must add a dissipative term that stabilizes the system proportional to the cube of the temperature:

$$\frac{dT(t)}{dt} = a' h_{\text{eq}}\left(x_c, t-\tfrac{1}{2}\tau_k\right) - b' h_{\text{off-eq}}\left(x_c, t-[\tfrac{1}{2}\tau_R+\tau_K]\right) - cT(t)^3 \tag{19.48a}$$

If we assume that the wind stress in the central Pacific is a pretty much simultaneous response to the east Pacific SST, we can write

$$\frac{dT(t)}{dt} = aT\left(t-\tfrac{1}{2}\tau_k\right) - bT\left(t-[\tfrac{1}{2}\tau_R+\tau_K]\right) - cT(t)^3 \tag{19.49}$$

where $a$, $b$, and $c$ are positive constants. We have seen that the first term is delayed by roughly 1 month, while the second term has a delay of roughly 6 months. This equation constitutes the delayed oscillator model for El Nino.

Equation (19.49) can be simplified neglecting the delay due to the Kelvin wave and in that case becomes

$$\frac{dT(t)}{dt} = T(t) - \alpha T(t-\delta) - T(t)^3 \tag{19.50}$$

This is however in nondimensional form where the temperature has been scaled by $(a/c)^{1/2}$ and the time by $c^{-1}$. This can be interpreted as measure of the influence of the returning signal relative to that of the local feedback.

Before we proceed to the integration of Eq. (19.50), we may consider some linear stability analysis. Besides the stationary solution $T=0$, the equation has two additional stationary states:

$$T_0 = \pm(1-\alpha)^{1/2} \qquad \text{for} \qquad \alpha < 1$$

If we perturb the solution around $T_0$ an linearize the equation we have for $T = T_0 + T'$

$$\frac{dT'}{dt} = (3\alpha - 2)\, T' - \alpha T'\,(t - \delta) \tag{19.51}$$

And we look for solution of the form

$$T' = T \exp(\sigma t)$$

where $\sigma = \sigma_r + i\sigma_i$ so that substituting in (19.51)

$$(\sigma_r + i\sigma_i) = (3\alpha - 2) - \alpha \exp\left[-(\sigma_r + i\sigma_i)\,\delta\right]$$

From which we obtain

$$\sigma_r = (3\alpha - 2) - \alpha \cos(\sigma_i \delta)\, e^{-\sigma_r \delta} \qquad \sigma_i = \alpha \sin(\sigma_i \delta)\, e^{-\sigma_r \delta}$$

The neutral curves ($\sigma_r = 0$) gives immediately

$$\begin{aligned} \delta &= \cos^{-1}\left\{(3\alpha - 2)/\alpha\right\}/\sigma_i \\ \sigma_i &= \left[\alpha^2 - (2 - 3\alpha)^2\right]^{1/2} \end{aligned} \tag{19.52}$$

For each value of $\alpha$ in the interval $0.5 < \alpha < 1$, there are infinite neutral curves, but for $\delta < 10$ there are only a few neutral curves, and the neutral solutions indicate a period given by $2\pi/\sigma_i$. Solutions to Eq. (19.50) have been obtained with a simple MATLAB script and are illustrated in Fig. 19.24. The behavior of the nonlinear oscillator is depicted as a function of time divided the delay that changes between 2 and 10. The coefficient $\alpha$ is kept constant at 0.75. The solutions indicated that as the delay grows the wave tends to become square with a period twice the delay. For smaller delay the period is much longer and the solutions are more sinusoidal. For example, for the neutral solution, the parameter $\alpha$ gives a $\sigma_i$ of 0.7 which implies a period of about 8.9 rather than 10 as given in the upper part of the figure.

The question now is if this simple model can mimic some characteristic of El Niño. If the period is twice the delay in order to have a period around 5 years, we see that near the neutral curve, we need a delay of 2.5 years. This is roughly twice the time waves employed to go back and forth, but nonetheless we cannot pretend that such a simple model could give such precision. While the delayed oscillator model is useful in providing us with an idea of what could be the mechanism of ENSO, it represents a limit for a fuller dynamics. The simple program for ENSO is given in the example.

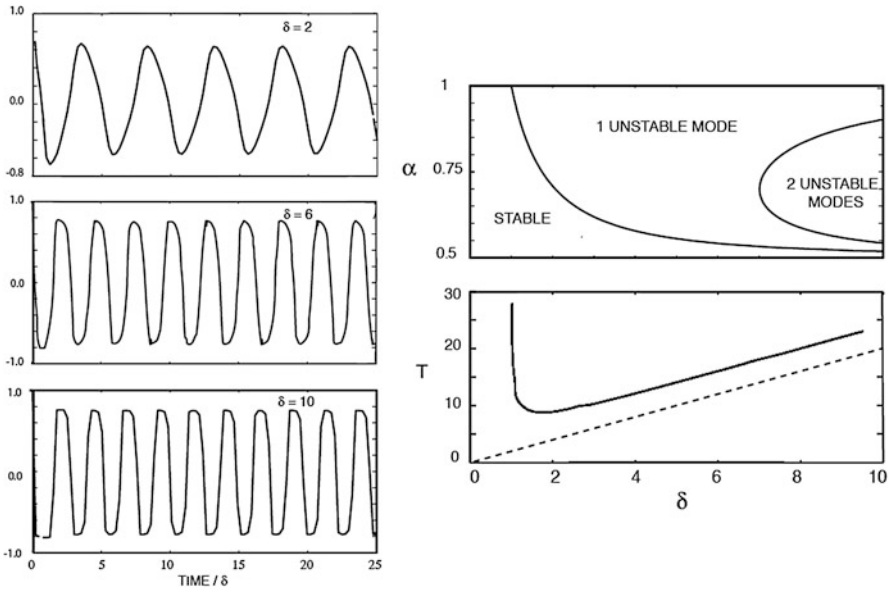



**Fig. 19.24** On the *left* are the oscillations obtained for three different values of the delay. On the right (*top*) the stability curves for the neutral solution with stable modes below the lower curve and unstable modes to the right of the same curves. The *lower figure* on the *right* shows the period for the neutral solution always higher than $2\delta$ shown by the *dotted line*

### *19.8.2 Aerosol–Cloud–Precipitation as the Predator–Prey Problem*

The predator–prey is a classical problem for the application of mathematics to biological systems. If we imagine in a desert island populated only by rabbits and foxes the species numbers will result by the strong interactions between the two. As the foxes hunt rabbits, the latter population will decrease to the point that foxes will start to starve and die and the rabbit population will then increase again to initiate a new cycle. The two populations will follow each other with some phase delay. This system is regulated by the so-called Lotka–Volterra equation:

$$\frac{dC}{dt} = C\,(a - bR)\,; \qquad \frac{dR}{dt} = R\,(eC - f) \tag{19.53}$$

where $C$ is the population of prey (rabbits), $R$ the population of predator (foxes), and $a$, $b$, $e$, and $f$ are system-dependent constants. The $C$ population will grow exponentially in the absence of $R$, but it is reduced by $R$ preying on $C$. As for the predator its population depends on $C$ and will decrease exponentially in the absence of $C$.

For the cloud we can imagine that the role of predator is assumed by the rain, while the prey is the liquid water path, that is, the amount of liquid water per unit area in the cloud (g $cm^{-2}$). The rain will increase at the expense of LWP and as the latter is depleted, the rain will start to decrease until the cloud will reform.

Actually in the model of Koren and Feingold (2011), the two variables are the cloud depth $H$ and the concentration of aerosol $N_d$. Rain is parameterized as a function of aerosol density and cloud depth.

We start defining LWP

$$\text{LWP} = \int_0^H q(z)dz = \frac{c_1}{2}H^2$$

where $q$ is the cloud water content and $c_1$ is a function of cloud-base temperature and pressure; $c_1 \simeq 2 \cdot 10^{-6}$ mm $m^{-2}$. The balance equation for $H$

$$\frac{dH}{dt} = \frac{H_0 - H}{\tau_1} + \dot{H}_r\,(t - T) \tag{19.54}$$

The first term on the right is a relaxation term toward the asymptotic value $H_0$ that would be the depth reached by the cloud in the absence of water sinks. The second term represents the loss of liquid water due to rain, a stochastic process that converts small cloud droplets to raindrops in a relatively short time $T \approx 15$ min. The delay term $(t - T)$ accounts for the time dependence of rain production, which is a function of the state of the cloud some period before the current time step.

To the first order, rain rate $R$ can be diagnosed from cloud depth $H$ and $N_d$

$$R = \alpha H^3 / N_d \tag{19.55}$$

based on theory and observations and where $\alpha = 2$ mm $m^{-6}$ $d^{-1}$. In this way the cloud depth controls the precipitation much more than $N_d$.

The loss term for $H$ can be obtained by

$$\dot{H}_r = \frac{dH}{dt} = \frac{dH}{dLWP}\frac{dLWP}{dt} \tag{19.56}$$

where $dLWP/dt = -R$. Using (19.53) we obtain

$$\dot{H}_r - R\frac{dH}{dLWP} = -\frac{R}{c_1 H} = -\frac{\alpha H^2}{c_1 N_d} \tag{19.57}$$

A similar equation to (19.54) is obtained for $N_d$

$$\frac{dN_d}{dt} = \frac{N_0 - N_d}{\tau_2} + \dot{N}_d\,(t - T) \tag{19.58}$$

In this case $N_0$ corresponds to the background aerosol concentration. The loss term for $N_d$ results from the conversion of cloud water to rain via drop collection and is calculated on a simple expression:

$$\dot{N}_d = -c_2 N_d R \tag{19.59}$$

where $c_2 \simeq 3 \cdot 10^{-1} \mathrm{m}^{-1}$

To simplify the delay term we notice that both $H$ and $N_d$ are continuous functions and assume that there is an equivalent delay time $T'$ that represents the mean value in the interval. $\dot{H}_r$ is expressed as $\alpha H^2 (t - T') / c_1 N_d (t - T')$ and $\mathrm{N_d}$ is calculated at time $(t - T')$. Finally $R$ is written based on (19.55)

$$R(t) = \frac{\alpha H^3 (t - T')}{N_d (t - T')}$$

Now we can rewrite (19.54) and (19.58) as a system using (19.57) and (19.59)

$$\begin{aligned} \frac{dH}{dt} &= \frac{H_0 - H}{\tau_1} - \frac{\alpha H^2 \left(t - T'\right)\right)}{c_1 N_d (t - T')} \\ \frac{dN_d}{dt} &= \frac{N_0 - N_d}{\tau_2} - \alpha c_2 H^3 \left(t - T'\right) \end{aligned} \tag{19.60}$$

Steady-state solutions are easily found to be

$$H = \frac{\left(N_d^2 + 4\gamma \tau_1 N_d H0\right)^{\frac{1}{2}} - N_d}{2\gamma \tau_1} \tag{19.61}$$

and this shows that for low values of $N_d$ (<30 $\mathrm{cm}^{-3}$), H strongly depends on $N_d$ while for greater value depends mainly on $H_0$.

Time-dependent solutions are shown in Fig. 19.25. The data used to draw this diagram are $H_0 = 670$ m, $N_0 = 515$ $\mathrm{cm}^{-3}$, $\tau_1 = 80$ min, $\tau_2 = 84$ min, and $T' = 21.5$ min. The behavior predator–prey is quite evident. The solution in the first 10 h that the amount of rain follows the maxima of the cloud depth by about 5 h while the aerosol density precedes the cloud depth by about one and half hour comparable to the $\tau_1$ and $\tau_2$ times. The "phase diagram" on the other hand shows a behavior very similar to the predator–prey diagram with closed loops that decrease in amplitude with time due to the damping. The latter is not visible in the upper figure because the short integration time (10 h with respect to a week). The MATLAB script to solve this problem is given in the examples.

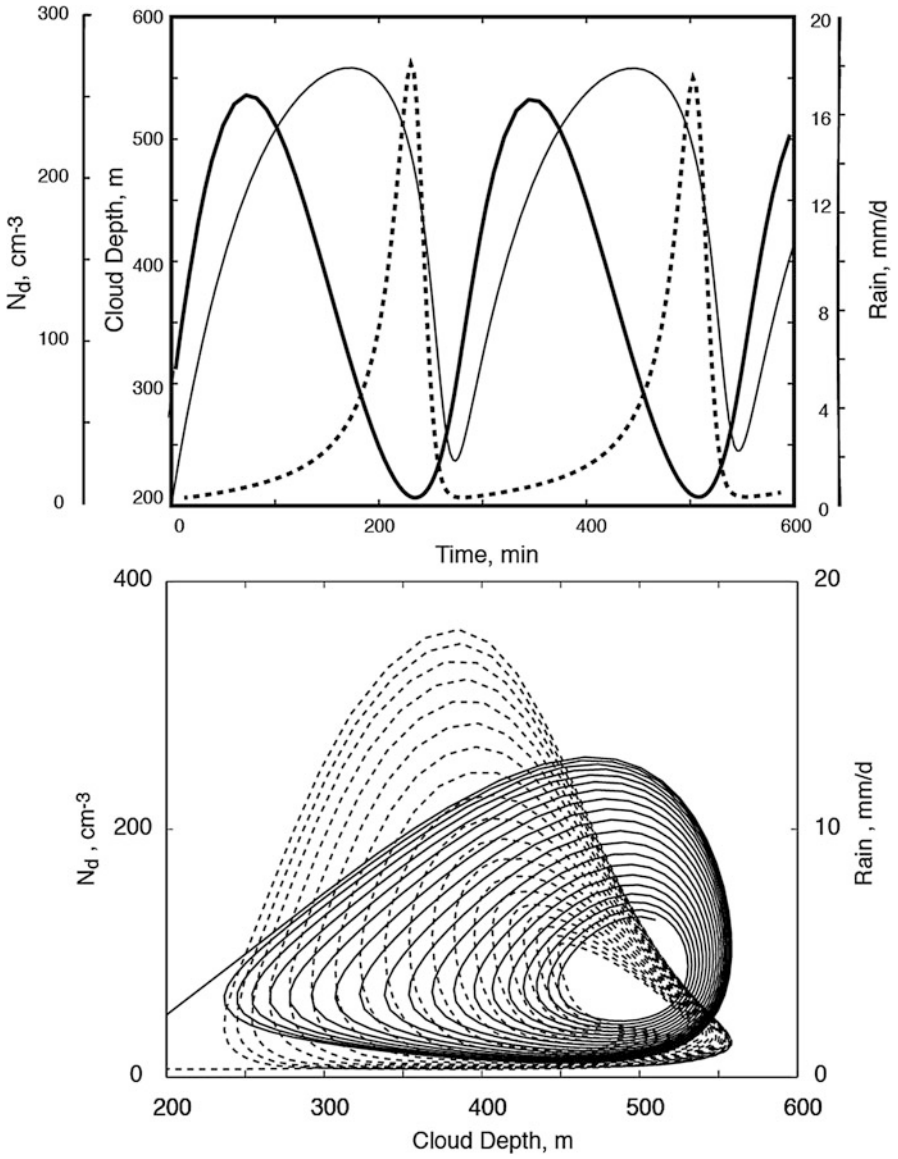



**Fig. 19.25** Oscillating, limit cycle response of the H, R, and Nd system. The *upper figure* shows the first 600 min. *Solid thin line*, Nd; *solid heavy line*, H; *dashed line*, Rain

## E.19 Examples

### *E.19.1 The Lorenz System: The Mother of All Chaotic Systems*

We consider a vertical loop as in Fig. E.19.1 (filled with a fluid) of radius $a$ while the tube radius is negligible. The external temperature changes linearly with height:

$$T_E = T_0 - T_1 z/a \tag{E.19.1}$$

The quantities inside the loop depend only on the angle $\phi$ and time. Then for velocity q and temperature T we have

$$\begin{aligned} q &= q(\phi, t) \\ T &= T(\phi, t) \end{aligned} \tag{E.19.2}$$

We then assume incompressibility ($\partial\rho/\partial t = 0$) so that divergence of velocity is also zero ($\partial q/\partial\phi = 0$). This means that velocity is only a function of time (solid body rotation), while for temperature we must assume a more complex behavior:

$$T - T_0 = T_2 \cos\phi + T_3 \sin\phi \tag{E.19.3}$$

According to this relation the temperature difference between the top ($\phi = \pi$) and the bottom ($\phi = 0$) is 2 $T_2$. The difference between the two sides ($\phi = \pi/2$ and $\phi = 3\pi/2$) is $2T_3$. Both $T_2$ and $T_3$ vary with time.

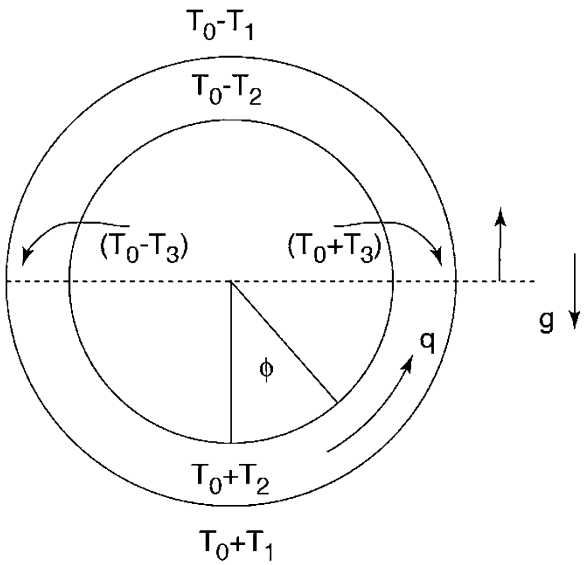



**Fig. E.19.1** The convective loop oscillator

We now have to find the equations of motion starting with the Navier–Stokes equation:

$$\frac{\partial \overline{u}}{\partial t} + \overline{u} \cdot \nabla \overline{u} = -\frac{1}{\rho} \nabla p - g\alpha \Delta T + \nu \nabla^2 T$$

Consider now that the divergence is zero and also we substitute the viscous term with something just proportional to the velocity:

$$\frac{\partial q}{\partial t} = -\frac{1}{\rho a} \frac{\partial p}{\partial \phi} - g\alpha \left(T - T_0\right) \sin \phi - \Gamma q \tag{E.19.4}$$

The only term to interpret here is that the vertical component of the buoyancy force has been projected along the tangent to the loop. Substituting the temperature difference from (E.19.3), we obtain

$$\frac{\partial q}{\partial t} = -\frac{1}{\rho a} \frac{\partial p}{\partial \phi} - g\alpha \left(T_2 \cos \phi + T_3 \sin \phi\right) \sin \phi - \Gamma q$$

To eliminate the pressure term, we integrate along the loop

$$2\pi \frac{\partial q}{\partial t} = g\alpha \int_0^{2\pi} \left(T_2 \cos \phi \sin \phi + T_3 \sin^2 \phi\right) d\phi - 2\pi \Gamma q$$

Evaluating the integral, we get

$$\frac{dq}{dt} = -\Gamma q + \frac{g\alpha T_3}{2} \tag{E.19.5}$$

To obtain the temperature equation we start out with the equation of diffusion

$$\frac{\partial T}{\partial t} + \overline{u} \cdot \nabla T = \kappa \nabla^2 T$$

Then we assume that the diffusion within the tube is negligible and that the heat is transferred through the walls at a rate $K\left(T_E - T\right)$ so that the above equation becomes

$$\frac{\partial T}{\partial t} + \frac{q}{a} \frac{\partial T}{\partial \phi} = K\left(T_E - T\right) \tag{E.19.6}$$

Remembering that the external temperature changes with height we have

$$T_E = T_0 + T_1 \cos \phi; \quad T - T_0 = T_2 \cos \phi + T_3 \sin \phi;$$

$$T_E - T = \left(T_1 - T_2\right) \cos \phi - T_3 \sin \phi$$

When we substitute in the temperature equation, we have

$$\frac{dT_2}{dt}\cos\phi + \frac{dT_3}{dt}\sin\phi - \frac{q}{a}\left(T_2\sin\phi - T_3\cos\phi\right) = K\left[\left(T_1 - T_2\right)\cos\phi - T_3\sin\phi\right]$$

equating the terms containing $\sin\phi$ and $\cos\phi$ we have

$$\frac{dT_3}{dt} - \frac{qT_2}{a} = -KT_3; \quad \frac{dT_2}{dt} - \frac{qT_3}{a} = K\left(T_1 - T_2\right)$$

We introduce the variable $T_4(t) = T_1 - T_2(t)$ to obtain

$$\frac{dT_3}{dt} = -KT_3 + \frac{qT_1}{a} - \frac{qT_4}{a}; \quad \frac{dT_4}{dt} = -KT_4 + \frac{qT_3}{a}$$

These equations together with (E.19.5) define the Lorenz system

$$\begin{aligned}
\frac{dq}{dt} &= -\Gamma q + \frac{g\alpha T_3}{2} \\
\frac{dT_3}{dt} &= -KT_3 + \frac{qT_1}{a} - \frac{qT_4}{a} \\
\frac{dT_4}{dt} &= -KT_4 + \frac{qT_3}{a}
\end{aligned}$$

The variable can be normalized as follows

$$X = \frac{q}{aK}; \quad Y = \frac{g\alpha T_3}{2a\Gamma K}; \quad Z = \frac{g\alpha T_4}{2a\Gamma K}; \quad t' = tK$$

So finally we get

$$\begin{aligned}
\dot{X} &= -PX + PY \\
\dot{Y} &= -Y + rX - XZ \\
\dot{Z} &= -Z + XY
\end{aligned} \tag{E.19.7}$$

That is the Lorenz system and that is clearly a nonlinear system. The parameters $r$ and $P$ are the Rayleigh number and the Prandtl number

$$r = \frac{g\alpha T1}{2a\Gamma K}; \qquad P = \frac{\Gamma}{K} \tag{E.19.8}$$

The real Lorenz system differs from the one above because the damping term of the $Z$ component contains a constant $b$.

We will then discuss the stability of such a system. The steady solutions correspond to put the derivative to zero. We have two possible sets of solutions:

$$X^* = Y^* = Z^* = 0$$
$$X^* = Y^* = \pm\sqrt{r-1}; \quad Z^* = r-1$$

The first solution corresponds to quiet (there is no motion), while the second exists only if r > 1. We called r "Rayleigh number," but actually it can be seen that it is the ratio between the real Rayleigh number and the "critical Rayleigh number" that happens when the convection starts. The stability analysis is carried out by perturbing the solution and linearizing the equations. If the perturbations are $\delta X$, $\delta Y$, and $\delta Z$, then the equations become

$$\delta\dot{X} = -P\delta X + P\delta Y$$
$$\delta\dot{Y} = -\delta Y + r\delta X - X^*\delta Z - Z^*\delta X$$
$$\delta\dot{Z} = -\delta Z + X^*\delta Y + Y^*\delta X$$

Assuming the perturbation of the form $\delta X = \delta Y = \delta Z \propto e^{\sigma t}$, the above system is reduced to a linear system:

$$\delta X_0\,(\sigma + P) - P\delta Y_0 = 0$$
$$\delta X_0\,(r - Z^*) - \delta Y_0\,(1+\sigma) - X^*\delta Z_0 = 0$$
$$Y^*\delta X_0 + X^*\delta Y - \delta Z_0\,(1+\sigma) = 0$$

Nontrivial solutions are obtained when the determinant is zero. For the rest solution we have

$$(\sigma+1)\left[\sigma^2 + \sigma\,(P+1) - P\,(r-1)\right] = 0$$

which has three roots

$$\sigma_1 = -1; \quad \sigma_{2,3} = \frac{-(P+1) \pm \sqrt{(P+1)^2 + 4P\,(r-1)}}{2}$$

And we see that in order to have instability ($\sigma > 1$) $r > 1$. For the second set of solution, we have a cubic equation:

$$\sigma^3 + A\sigma^2 + B\sigma + C = 0$$

where

$$A = (P+2) \quad B = P + r \quad C = 2P\,(r-1)$$

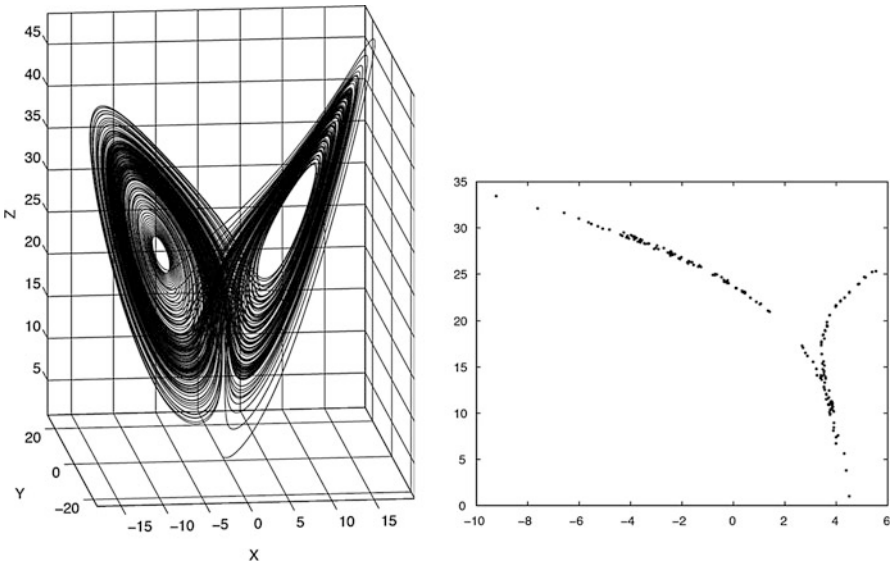



**Fig. E.19.2** The phase diagram of the Lorenz system (*left*) and the Poincarè section on the plane z = 0

It can be shown that instability occurs when

$$r > rc = P\,(P + 4)\,/\,(P - 2)$$

The Lorenz model with $\dot{Z} = -bZ + XY$ gives instability for $r > r_c$ when

$$r > rc = [P\,(P + 3 + b)]\,/\,(P - 1 - b)$$

and with $P = 10$ and $b = 8/3$, $r_c = 24.74$. Figure E.19.2 gives an example of the behavior of the Lorenz system in phase space. The appearance of the famous butterfly is quite evident.

## *E.19.2 The Logistic Map as an Example of Difference Equation*

In the previous paragraph, we have already the equation for the population growth (Eq. 19.45). We can now recast that equation in another form

$$\frac{dN}{dt} = rN\left[1 - \frac{N}{K}\right] = rN - \frac{rN^2}{K} = aN - bN^2 \qquad \text{(E.19.9)}$$

as a difference equation becomes

$$N_{i+1} - N_i = a' N_i - b' N_i N_i \Rightarrow N_{i+1} = N_i \left(a' + 1\right) - b' N_i N_i$$

With appropriate choice of $a'$ and $b'$ this equation becomes

$$N_{i+1} = a N_i \left(1 - N_i\right) \tag{E.19.10}$$

This recursive relation is also called logistic map because it maps out the sequence $N_0, N_1, N_2 \ldots$.

Developing in detail the difference equation we notice that

$$a' = r\Delta t; \quad b' = r\Delta t / K_i$$

And the recursive relation becomes

$$N_{i+1} = N_i \left(1 + r\Delta t\right) \left(1 - \frac{r\Delta t N_i}{K \left(1 + r\Delta t\right)}\right)$$

And in order to satisfy (E.19.10) must be

$$a = \left(1 + r\Delta t\right); \qquad b = \frac{r\Delta t}{K \left(1 + r\Delta t\right)} = 1$$

Which impose some condition on ten variables. Most important is that $N_i$ be always less than 1.

Equation (E.19.10) can be solved for different values of a and some initial results are shown in Fig. E.19.3. We can see that increasing a corresponds to having asymptotic values for x that goes from 1 to 2 to 4 and then for some values we have a sequence of number that represents the so-called route to chaos. This can be seen more clearly in a diagram which gives the solution as a function of a as shown in the same figure.

The points where the solutions has two values or four values, etc. are called bifurcation point. An interested reader could find more on many books on chaotic dynamics.

### *E.19.3 The Lyapunov Exponent*

Chaotic behavior can be quantified in terms of the Lyapunov exponent $\lambda$ which is defined as

$$dx_n = dx_0 2^{\lambda n} \tag{E.19.11}$$

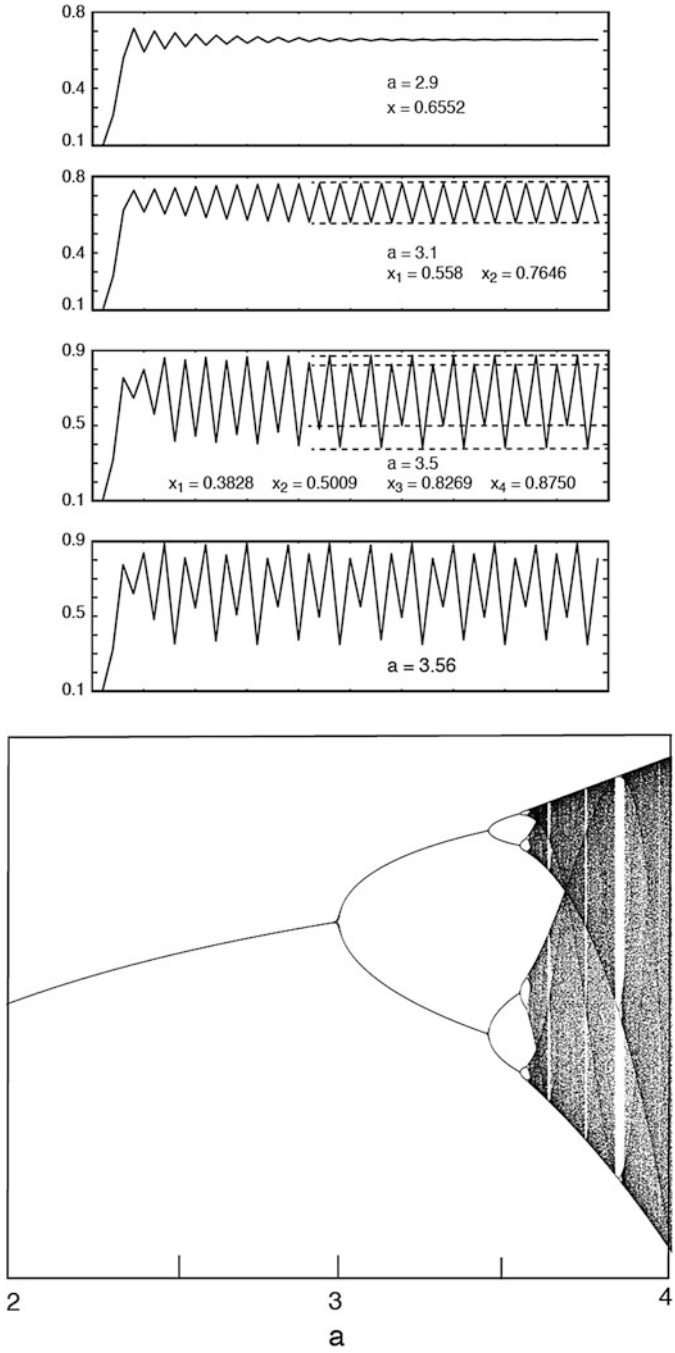



**Fig. E.19.3** The solution to the logistic map (*upper panel*) and the solution as a function of a (*lower panel*)

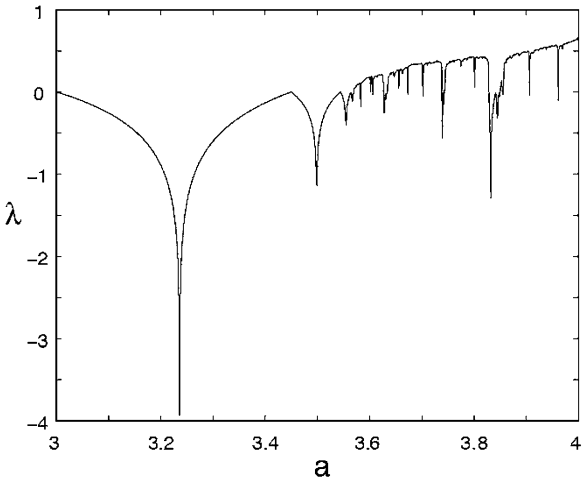



**Fig. E.19.4** Lyapunov exponent for the logistic map as a function of the parameter a

where $dx_n$ is the incremental difference after n iterations and $dx_0$ is the incremental difference in the initial value. According to this definition, if the Lyapunov exponent is negative, adjacent solutions converge and deterministic solutions are obtained. On the other hand if Lyapunov exponent is positive, adjacent solution will diverge exponentially and chaos will follow. In order to calculate the Lyapunov exponent, we consider the incremental difference in a single iteration. So we write based on (E.19.10):

$$x_{n+1} + dx_{n+1} = f(x_n + dx_n) = f(x_n) + \left(\frac{df}{dx_n}\right) dx_n \tag{E.19.12}$$

where for the logistic map we have

$$x_{n+1} = f(x_n) \quad \Rightarrow \quad dx_{n+1} = \left(\frac{df}{dx}\right)_n dx_n \tag{E.19.13}$$

And for the logistic map

$$f(x) = ax(1-x) \quad \Rightarrow \frac{df}{dx} = a(1-2x) \quad \Rightarrow \quad dx_{n+1} = a(1-2x_n)\, dx_n$$

And from the definition (E.19.11) we have

$$\lambda = \lim_{m\to\infty} \frac{1}{m} \sum_{n=0}^{m} \log_2 \left| \left(\frac{df}{dx}\right)_n \right| \tag{E.19.14}$$

The Lyapunov exponent for the logistic map is represented in Fig. E.19.4. It is clearly shown that in the region $3.5699 < a < 4$, the Lyapunov exponent assumes positive values and is then responsible for chaos.

### *E.19.4 MATLAB Program for El Niño Delayed Oscillator*

```
function ensodde
lags=10;tspan = [0 250],
%sol1=dde23(@dde, lags,@history, tspan);
sol1=dde23(@dde, lags,1, tspan);
tplot=linspace(0,250);
T1=deval(sol1,tplot);
%T2=deval(sol2,tplot);
tplot=[-2 tplot];
T1=[1 T1];%T2=[2 T2];
plot(tplot/lags, T1)
%sunfunction-------
function dydt=dde(t,T,Z)
dydt=T-3*T^3-0.75*Z;
function s = history(t)
s=1-t;
```

### *E.19.5 MATLAB Program for the Predator–Prey Problem*

```
TT = 21.5;
len = 600;
%options = ddeset('MaxStep',1e-1)
options = ddeset('RelTol',1e-4);
sol = dde23('num_lv',[TT, TT],[200,50*1000000]',[0,len], options);
R = 2*sol.y(1,:).*sol.y(1,:).*sol.y(1,:)./sol.y(2,:);
%plot
figure(1),clf
plot(sol.x,sol.y(1,:),'g');
title('H')
xlabel('time t');
ylabel('y(t)');
figure(2),clf
plot(sol.x,sol.y(2,:)*1e-6,'r');
title('N in cm^-3')
xlabel('time t');
ylabel('y(t)');
figure(3),clf
plot(sol.x,R)
title('R in mm/day')
xlabel('time t');
ylabel('y(t)');
del = find(sol.x-TT ==min(abs(sol.x-TT)));
```

```
Rnew = R;
for t= del+1:length(R)
   Rnew(t)= R(t-del);
end
R = Rnew;
%size(R)
figure(4),clf
[ax,p1,p2] = plotyy(sol.y(1,:), sol.y(2,:)*1e-6,sol.y(1,:),R);
title('verde = N, Rosso = R')
figure(5),clf
plot(sol.y(1,:),R)
function dydt = num_lv(t,y,Z)
   ylag1 = Z(:,1);
   ylag2 = Z(:,2);
    alfa=2*6.944*0.0000001;
   c1=0.000002*0.001;
   c2=0.3*100000;
    %c1 = 2e-6;
   %alfa = 2/1440.0;
   %c2 = 0.3;
   %tau1=60;
   tau1=80;
   %tau2=60;
   tau2=84;
   %h0=530;
   h0=670;
   %n0=1.80e8;
   n0=5.15e8;
   dydt = [h0,n0]';
   dydt(1) = (h0-y(1))/tau1-alfa*(ylag1(1))^2/(c1*ylag2(2));
   dydt(2) = (n0-y(2))/tau2-alfa*c2*(ylag1(1))^3;
```

# References[1]

## *Books*

Alligood KT, Sauer TD, Yorke JA (1997) Chaos, an introduction to dynamical systems. Springer, New York

Baker GL, Gollub JP (1990) Chaotic dynamics, an introduction. Cambridge University Press, Cambridge/New York

[1]There are many books that introduce chaos theory. I would recommend Baker and Gollub, Korsch and Jodl, Alligood et al., and Bergè et al. The book by the founding father of this theory, E. Lorenz, gives unparalleled insight. The papers by Tsonis are very illuminating.

Bergè P, Pmeau Y, Vodal C (1984) Order within chaos. Wiley, New York
Huang RX (2009) Ocean circulation. Cambridge, Cambridge
Lorenz EN (1993) The essence of chaos. University of Washington, Seattle
Peixoto JP, Oort JH (1992) Physics of climate. AIP, New York
Shampine LF, Gladwell I, ThompsonS (2003) Solving ODE with Matlab. Cambridge, Cambridge University Press
Stull RB (2000) Meteorology for scientists and engineers. Brooks/Cole, Pacific Grove

## *Articles*

Koren I, Feingold G (2011) Aerosol-cloud-precipitation system, as a predator prey problem. Proc Natl Acad Sci U S A 10:1073
Lorenz EN (1964) The problem of deducing the climate from the governing equations. Tellus 16:1
Lorenz EN (1976) Non deterministic theories of climatic change. Quater Res 6:495
Lorenz EN (1980) Attractor sets and quasi-geostrophic equilibrium. J Atmos Sci 37:1685
Lorenz EN (1984) Irregularity: a fundamental property of the atmosphere. Tellus 36A:1984
Lorenz EN (1991) Dimension of weather and climate attractors. Nature 353:241
Suarez MJ, Schopf PS (1988) A delayed action oscillator for ENSO. J Atmos Sci 45:3283
Tsonis AA, Elsner JB (1989) Chaos, strange attractors and weather. Bull Am Meteorol Soc 70:14
Tsonis AA et al (1994) An investigation of the ability of nonlinear methods to infer dynamics from observables. Bull Am Meteorol Soc 75:1623
Vallis GK (1986) El Nino: a chaotic dynamical system? Science 232:243

# Chapter 20
# Geoengineering

In the last few years, atmospheric physics had been invaded by a growing large number of researchers that got very excited about the possibility of changing artificially the environment to correct what the human beings (consciously or not) are doing to the climate. This very arduous task goes with the name geoengineering. It was invented many years ago in the science fiction literature and was known as terraforming. It was Jack Williamson that first mentioned that word in a science fiction novel of 1942. Then in 1995, Martyn Fogg published a book on terraforming where he gave the term geoengineering for planetary engineering applied to the planet Earth. Planetary engineering is the application of technology for the purpose of influencing the global properties of a planet. In 1992, the National Research Council published a report (Public Implications of Greenhouse Warming) that had a chapter on geoengineering and that could be considered as the institutionalization of the science. A very original consideration for the times was the sentence:

> It is important to recognize that we are at present involved in a large project of inadvertent "geoengineering" by altering atmospheric chemistry, and it does not seem inappropriate to inquire if there are countermeasures that might be implemented to address the adverse impacts.

This is a matter of debate other than a good occasion for research money but also a very good opportunity to apply what we have learned before in this book.

First of all, we will describe some of the proposed technologies for geoengineering the planet and then we will describe in detail a few of them. Geoengineering acts on the main cause that presumably is changing the climate, and that is the carbon dioxide. There are a number of techniques suggested to remove and store $CO_2$, and these go under the acronym CCS (carbon capture and sequestration). Other techniques are directed to correct the warming, reducing the amount of solar radiation absorbed by the planet. These techniques refer to the solar radiation management (SRM).



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_20

## 20.1 A Short Inventory of Geoengineering Technologies

We refer to the scheme reported in Fig. 20.1 (redrawn from Lenton and Vaughn 2009) and start with the simplest (in theory) methods to modulate the solar radiation entering the atmosphere. The idea is to have sunshades placed somewhere out of the Lagrange point L1. These would be essentially mirrors that could also be place in Earth's orbit. If the radiation enters the atmosphere, it could be reflected back to space by an artificial stratospheric aerosol layer with composition similar to the volcanic layer produced by violent eruptions, and so particles would be sulfate. Doing the same thing for the troposphere would be dangerous to your health. Another possibility to increase the atmospheric albedo would be the introduction or brightening of clouds. Finally, the albedo of the Earth could be increased by changing the reflectivity of the surface, using a number of different technologies.

The CCS techniques can be roughly classified as those that suck carbon dioxide from the atmosphere and store it somewhere and those that increase the capacity of the Earth system to absorb carbon dioxide. These include then afforestation and

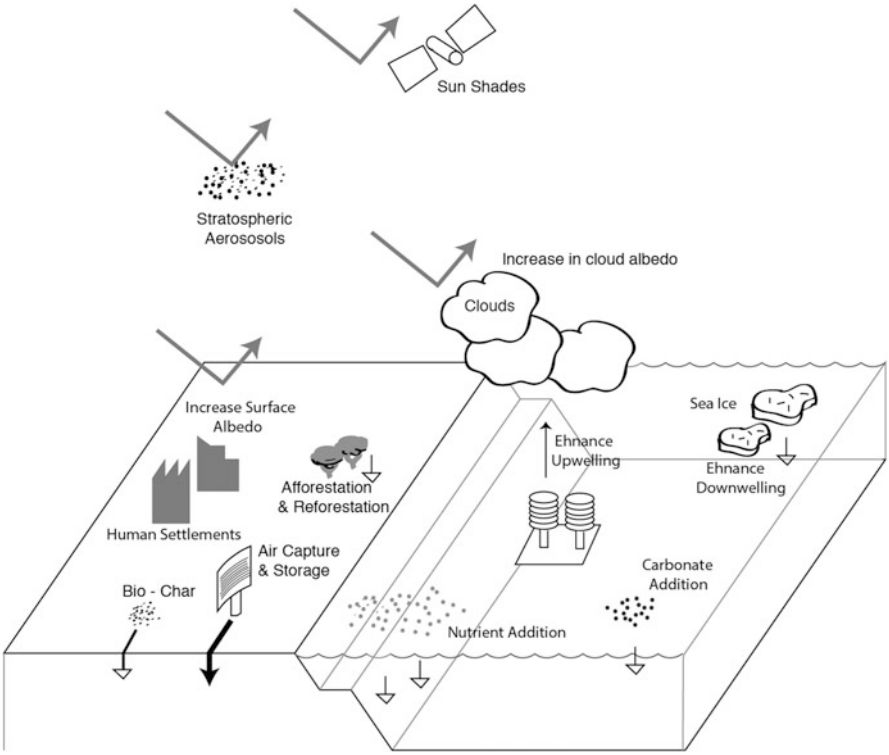



**Fig. 20.1** Overview of the geoengineering proposals. *The gray arrow* indicates the shortwave radiation and the white arrows the carbon disposal (Adapted from Lenton and Vaughan 2009)

reforestation, air capture through some chemical process, and transformation of biological products in charcoal. Also included in this category is the change of the ocean capacity to absorb $CO_2$. The oldest suggestion deals with the iron fertilization where the surface water is added with iron or other nutrients that enhance the carbon fixation. In the opposite way, the enhancement of upwelling of nutrient-rich waters would do the same job (the cylinders depicted in the figure should do the job). Other proposals would increase the deepwater formation or through carbonate addition would increase the alkalinity of the ocean.

Following the review by Lenton and Vaughan (2009), in Fig. 20.1, the black arrows indicate engineered flow of carbon, while the white arrows indicate the enhancement of natural flows of carbon. Following the same paper, we will compare the different geoengineering techniques with respect to the radiative forcing ($RF$) they would produce. $RF$ is linearly related to the global temperature change ($\Delta T$) through the climate sensitivity parameter $\lambda$, that is, $\Delta T = \lambda\ RF$. Because of the linearity, we will not fuss around with strange numbers and assume for a standard RF corresponding roughly with a doubling of $CO_2$ the value of 4 w $m^{-2}$. This would give, with an average value of 0.86 $Cw^{-1}$ $m^{-2}$ for $\lambda$, a change in temperature of 3.44 °C.

## 20.2 Carbon Sequestration and Storage

In this paragraph, we will not deal with the technical details on how it is possible to suck carbon dioxide from the air or capture it directly at the source. Rather, we will make some general observation on the feasibility of such an intervention. We will follow the paper by Lenton and Vaughn (2009) and the very insightful piece by Stocker (2013).

First of all, we ask how much reduction can we obtain in the radiative forcing by subtracting a certain amount of $CO_2$, and this can be simply evaluated using the approximation that the radiative forcing ($RF$) is roughly a logarithm function of the amount of carbon dioxide

$$RF = \beta_1\ \ln\left(\frac{CO_2}{CO_{2,0}}\right) \tag{20.1}$$

where $CO_{2,0}$ is a reference concentration taken as 278 ppm (the value it had in 1800) and $\beta_1 = 5.35$ $wm^{-2}$. This means that at the present time for a mixing ratio of 400 ppm, the radiative forcing amounts to 1.94 $wm^{-2}$. If we derive (20.1), we obtain the sensitivity of the radiative forcing with respect to $CO_2$ mixing ratio

$$\frac{dRF}{dCO_2} = \frac{\beta_1}{CO_2} \tag{20.2}$$

The present sensitivity would then be 0.0133 $\mathrm{wm}^{-2}$ $\mathrm{ppm}^{-1}$ while the preindustrial value would be 0.0193 $\mathrm{wm}^{-2}$ $\mathrm{ppm}^{-1}$. If a certain amount $\Delta C_{\mathrm{atm}}$ of carbon is subtracted to the atmosphere by the geoengineering activity, the effect on the radiative forcing will be

$$RF(t) \approx \frac{\Delta C_{atm}}{k} \frac{\beta_1}{\mathrm{CO}_2(t)} \tag{20.3}$$

If $\Delta C_{\mathrm{atm}}$ is measured in Pg ($1\mathrm{Pg} = 10^{15}$ g), then $k = 2.14$ Pg C $\mathrm{ppm}^{-1}$. This represents the conversion factor between carbon amount and ppm of $CO_2$. The relation (20.3) can be used directly expressing the amount subtracted in ppm. The relation (20.3) assumes that all the emitted $C$ ends up in the atmosphere. If the emitted quantity is $C_e$ and all the carbon ends up in the atmosphere, then the increase in $CO_2$ in ppm would be

$$\Delta \mathrm{CO}_2\,(\mathrm{ppm}) = \frac{C_e\,(m_{atm}/m_C)}{M_{atm}}$$

where $m_{\mathrm{atm}}$ and $m_c$ are the molecular mass of the atmosphere and the carbon and $M_{\mathrm{atm}}$ is the mass of the atmosphere ($5.2 \cdot 10^{18}$ kg). For 1Pg of $C_e$, the expression gives an increase of 0.467 ppm.

However, if we want to base our evaluation on the emitted carbon, we need to make some simplification. In this, we will follow the clear ideas of Wallace Broecker and his $CO_2$ arithmetic. The airborne fraction must be taken into account, and just to make memorizing easier, we assume that if the emitted carbon corresponds to 4 Pg of carbon, this would give 1.85 ppm of carbon dioxide. Then, assuming an airborne fraction of 0.54, we have the proportion that 4 Gt of $C$ corresponds to an increase of 1 ppm in the atmosphere. At the present time, each year 10 Gt of carbon is produced that would imply an increase of 2.5 ppm/year. The Broecker arithmetic is very simple. If we want to double the preindustrial $CO_2$ concentration of $CO_2$ (560 ppm), we must burn up to $4 \times (560–400) = 640$ Gt of carbon and so on. This is what he calls the size of the pie. At the present time, 25 % of the carbon is produced by the USA and Europe and 35 % from India and China. If we assume that the "rich countries" maintain the 30 % share, their slice of the pie (192 Gt) would be consumed in 32 years. A drastic reduction would be obtained if we assume that each country limits their share to 20 %. In this case, their slice would be 128 Gt that would be consumed in just 21 years. At the time Broecker wrote his note (2007), the situation was summarized by a diagram like the one in Fig. 20.2. A scenario is shown in which the pie of $C$ to be consumed amounts to 150 Gt. This wedge is utilized, decreasing linearly the rate of consumption which starts with 6 Pg of $C$ and decreases linearly to zero after 50 years. The hypothesis is that the real consumption is much higher and is depicted with darker gray. The excess is what must be disposed of with carbon sequestration technology. This seems to show that sequestration technology must be a priority and a problem that must be solved urgently.

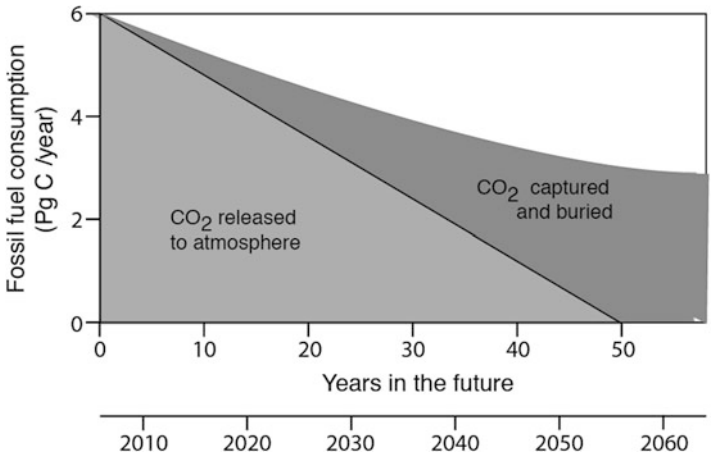



**Fig. 20.2** Hypothetical scenario for a 150 Pg wedge pie. The real production is limited by the curve while the wedge is the straight line. All the excess carbon must be captured (Adaped from Broecker 2007)

The urgency has been shown very clearly in another note by Thomas Stocker (2013). He assume a very simple analytical scenario for the emissions

$$E(t) = \begin{cases} E_0 e^{r(t-t_0)} & t_0 < t \leq t_1 \\ E_0 e^{r(t_1-t_0)} e^{-s(t-t_1)} & t > t_1 \end{cases} \tag{20.4}$$

$E(t)$ is the emission at time t, and $E_0 = 9.3$ Gt C/year is the emission at time $t = 0$ taken at the year 2009, while $r$ is the rate of emission increase until the time $t_1$. The cumulative emission at time $t_0$ is taken as $C_0 = 530$ GtC. Time $t_1$ is when the mitigation starts with emission reduction at the constant rate $s$. It is to notice that there is simply a reduction in the rate of emission but not a "negative" emission that would mean removal of carbon dioxide from the atmosphere. The cumulative emission is given by

$$C_\infty = C_0 + \int_{t_0}^{\infty} E(t)\, dt = C_0 + E_0 \left( \frac{1}{r} + \frac{1}{s} \right) e^{r(t_1-t_0)} - \frac{1}{r} E_0 \tag{20.5}$$

At this point, we assume there is a linear relationship between warming and cumulative emission

$$\Delta T = \beta C_\infty \tag{20.6}$$

To find the minimum peak warming in Eq. (20.5), we use the limit $s \to \infty$ and get

$$\Delta T_{\min} = \beta \left( C_0 + \frac{1}{r} E_0 \left( e^{r(t_1-t_0)} - 1 \right) \right) = \beta C_1 \tag{20.7}$$

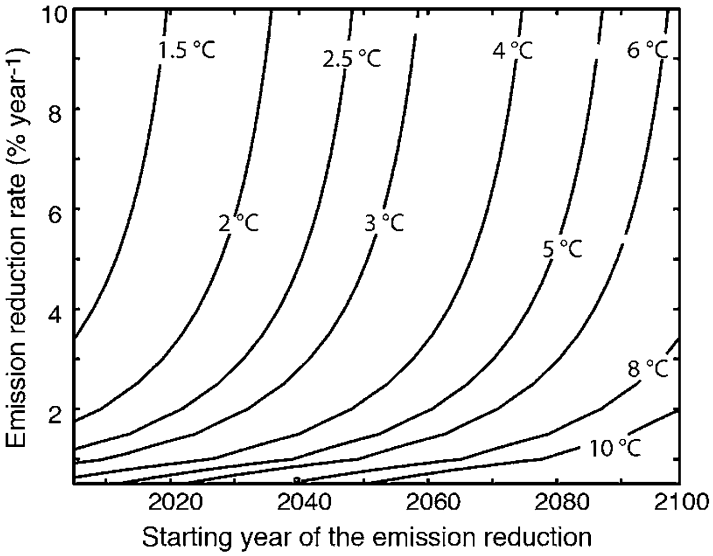



**Fig. 20.3** Contours of peak warming as a function of the starting date of emission reduction and for fixed parameters $C_0$, $E_0$, $\beta$, and $r$ (Adapted from Stocker 2013)

This represents the best case because it refers to zero emissions after time $t_1$ and also the minimum peak warming. Achievable climate target is then determined by the cumulative emission at time $t_1$, $C_1$.

From Eqs. (20.6) and (20.5), it is possible to evaluate the peak warming that will depend (once $\beta$ is assigned) on the current rate of emission increase, the starting time of the mitigation, and the rate of emission reduction. This can be seen from Fig. 20.3 adapted from the paper by Stocker. To draw the figure, we have used $C_0 = 530$GtC, $E_0 = 9.3$ Gt C year $^{-1}$, $\beta = 2$°C (TtC) $^{-1}$, and $r = 1.8$ % per year. The figure shows that a delay in a year of mitigation increases the peak warming, while for a fixed date of mitigation, the peak warming decreases with increasing rate of reduction. The first conclusion is that already we are almost out of business because to contain the peak warming below 2°C, we need the mitigation to start around 2020 with a reduction rate of 3.2 % per year. If the starting date is moved to 2030, the reduction rate jumps to almost 6 % per year.

The starting date could be important for another reason because according to the relation (20.7), the minimum peak warming depends on the quantity of carbon emitted until time $t_1$. We must also remember that the relation has been obtained with the hypothesis of zero emission after time $t_1$. Stocker calls this minimum peak warming *climate target*, and the lowest values are progressively lost as we move the date of reduction. Figure 20.4 shows the possibility to reach the climate target as a function of the starting date and the rate of reduction. In (A), it is shown that there is an area of unachievability of climate target that grows as we move the starting year. On the other hand, for low values of the emission reduction rate $s$, the lowest values

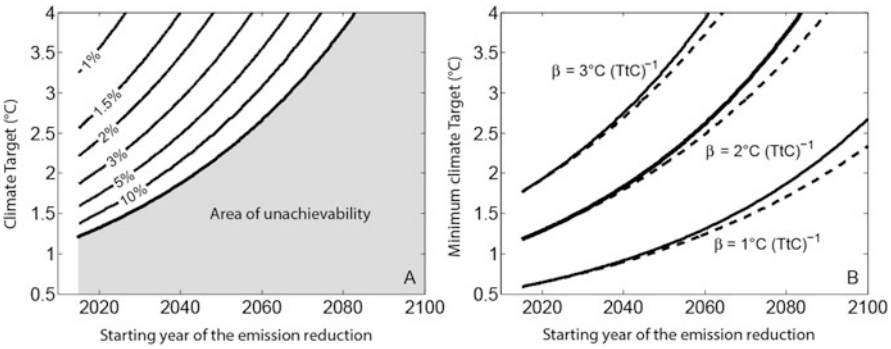



**Fig. 20.4** In (**a**) are shown the contours of required reduction emission rates (%) as a function of the starting date of reduction and the climate target. In (**b**), the achievable minimum climate target is explained in the text (Adapted from Stocker 2013)

of climate target are not achievable yet. In (B) of the same figure, the climate target is shown as a function of three values of the peak response to cumulative emission $\beta$ and for single value of the rate of increase $r = 1.8\,\%$ per year (solid lines) compared to the average rate of increase in the last decade (1.5 % per year). This shows that a larger rate of increase implies greater warming.

Economic models estimate that feasible maximum rate of reduction may not exceed 5 %. Under this assumption, the 1.5 °C target has been lost in 2012, the 2 °C would be lost after 2027, and the 2.5 °C target will be lost after 2040.

## 20.3 What Geoengineering Can Do

The previous section has shown that at the present time, there are very few chances to reduce future concentrations of $CO_2$ by simply reducing the emission. Actually, in Chap. 16, we have shown that using a very simple model, the only way to achieve stabilization and then a slow reduction was to just stop the emission (see Fig. 16.11). This result can be obtained using the same analytical model of Stocker. Integrating the emission (20.4), we can obtain the total carbon content of the atmosphere after a time $t$:

$$\begin{aligned} C(t) &= C_0 + \tfrac{1}{r}E_0\left(e^{r(t-t_0)} - 1\right) \text{ for } t_0 < t < t_1 \\ C(t) &= C_1 + \tfrac{1}{s}E_0 e^{r(t_1-t_0)}\left(1 - e^{-s(t-t_1)}\right) \text{ for } t > t_1 \end{aligned} \tag{20.8}$$

where $C_1$ is evaluated from the first of (20.8) for $t = t_1$. It is clearly seen that the stabilization is already achieved very rapidly with a reduction rate of 10 %. We see clearly that for $s \to \infty$ the total burden stabilizes at $C_1$. Notice that the amount of carbon does not decrease with time because the model is lacking the action of the ocean. Also, the number appearing on the figure may look puzzling. Based on

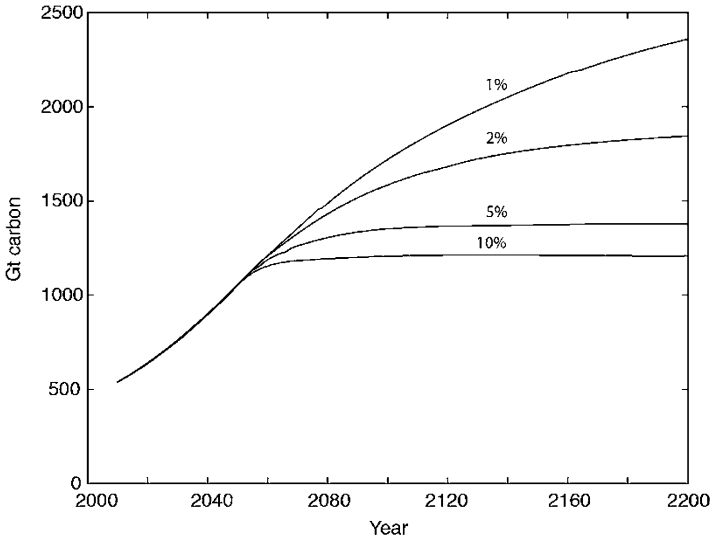



**Fig. 20.5** The total content of carbon in the atmosphere as a function of time according to the simple model by Stocker. The percentage indicated on the *curves* is the parameter s in Eq. 20.8

what we said in the previous paragraph, 530 Gt ($C_0$) would correspond to a mixing ratio of 247 ppm, and for 2009, that would give a mixing ratio in the atmosphere of 527 ppm when added to preindustrial concentration (280 ppm). To get reasonable numbers again, we have to take into account the airborne fraction that would reduce the increase in the atmosphere by roughly 50 %.

A similar calculation can be done with the model used in Chap. 16 with the results shown in Fig. 20.5. The curve labeled "no sequestration" refers to the base case where a grow rate of 1.5 % year$^{-1}$ was used from 2000 to 2040 and then a negative growth rate of 2 % year$^{-1}$ was adopted. The maximum concentration of around 800 ppm is obtained followed by a slow decline. This result could be compared with the 1 % curve of Fig. 20.6 where after 200 years, a concentration of roughly 1200 ppm resulted. Then, we make the hypothesis to absorb carbon dioxide from the atmosphere at a rate of 3.2 and 6.4 Gt $CO_2$ year$^{-1}$. These values correspond roughly to 10 and 20 % of the present rate of emission. We can see that the geoengineering effort decreases the maximum value of the mixing ratio and has the effect of reducing effectively the amount of carbon dioxide in the atmosphere.

It is not our intention to detail the way this could be done, but rather we would like to do some very general consideration. First of all, let us look at the energy requirements if we consider that $CO_2$ must be subtracted from the atmosphere, compressed, and stored geologically. The minimum energy required to extract $CO_2$ from a mixture of gases in which the partial pressure of carbon dioxide is $p_0$ and delivered as a pure $CO_2$ stream at final pressure $p$ is given by net enthalpy of mixing $kT \ln(p/p_0)$, where $k$ is the Boltzmann constant (8.3 J mol$^{-1}$K$^{-1}$) and $T$ is

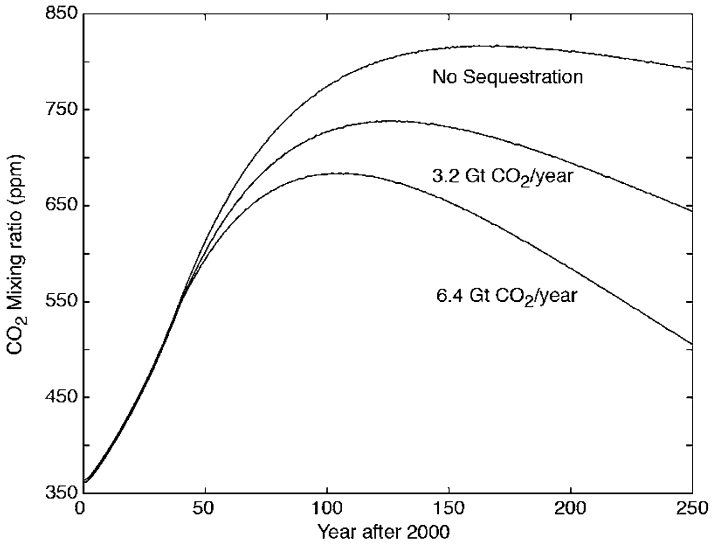



**Fig. 20.6** The change with time of the mixing ratio of carbon dioxide in the presence of pumping from the atmosphere. The *upper curve* has a growth of 1.5 %/year up to 2040 and a negative growth of 2 %/year after that time. In the other two curves, $CO_2$ is subtracted from the atmosphere at the rates indicated

the ambient temperature. If we use $p = 1$ atm, $p_0 = 4 \times 10^{-4}$, and $T = 300$ K, we get about 20 kJ/mol or 1.6 GJ/tC. To this energy, we must add that required to compress $CO_2$ to 100 atm. This adds something like 11 kJ/mol that must be doubled to take into account efficiency. We arrive at a total of 4 GJ/tC. This must be compared with the energy densities of the most common fossil fuels, that is, coal, oil, and natural gas, which are 40, 50, and 70 GJ/tC, respectively. In principle, if the energy for air capture is provided by fossil fuels, then the carbon captured is much larger than the carbon content of the fuels used to capture it.

There is however another requirement for land. An air capture system is limited by the amount of $CO_2$ transported to the absorber by the atmospheric motions. It is possible to estimate (see Examples) that the limit in the boundary later is about 400 GtC/ha-year. We can require then zero net $CO_2$ emission and divide the energy content of the fossil fuels by the area required to capture carbon dioxide. As an example, we can refer to coal so that the energy density would be $40 \times 10^9 \times 400 = 1.6 \times 10^{13}$ J/ha-year $= 50$ Wm$^{-2}$. This is an energy flux that would produce no net $CO_2$ emissions. This result would scale with the energy content of other fuels. As a comparison, biomass-based systems produce roughly 1 Wm$^{-2}$, while wind power reaches something like 20 Wm$^{-2}$. This shows that land requirements are not a major problem. The other way in which $CO_2$ can be removed is captured directly from power plants. This is rather an economic comparison, and it is not within the purpose of this book.

**Table 20.1** The radiative forcing potential for air capture for the two scenarios shown in Fig. 20.5

| Year | $\Delta CO_2$ (ppm) | | $\Delta C$ (Pg C) | | RF ($Wm^{-2}$) | |
|---|---|---|---|---|---|---|
| 2100 | 44 | 89 | 94.2 | 190.6 | −0.12 | −0.62 |
| 2150 | 85 | 161 | 182 | 256 | −0.56 | −1.05 |
| 2200 | 120 | 227 | 256 | 486 | −0.79 | −1.5 |
| 2250 | 147 | 286 | 314 | 612 | −1.98 | −1.92 |

A final consideration has to do with the application of Eq. (20.3) to evaluate the radiative forcing potential of this particular geoengineering option. Table 20.1 reports the radiative forcing potential of geoengineering reduction starting 100 years after 2000. The adoption of the second scenario would imply a considerable change in the radiative forcing.

## 20.4 Shortwave Options

In this paragraph, we will examine some of the suggestions to modify the solar radiation arriving on the Earth's surface and known collectively as solar radiation management (SRM). As usual, the discussion will be quite general, while for the details, the reader may want to get to the original work. In this respect, the work of Lenton and Vaughan (2009) is very much recommended.

### *20.4.1 Increase Albedo*

We will follow in this paragraph the work of Lenton and Vaughan (2009) that is focused essentially in reducing the amount of solar radiation absorbed by the Earth. If the planetary albedo is $\alpha_p$, the change in amount of solar radiation $\Delta S$ necessary to obtain a radiative forcing $RF$ is simply

$$RF = \Delta S \left(1 - \alpha_p\right) \tag{20.9}$$

On the other hand, if we change the planetary albedo by $\Delta\alpha_p$, then the resulting radiative forcing will be

$$RF = -S_0 \Delta\alpha_p \tag{20.10}$$

where $S_0$ is the solar constant. To simplify the problem, $S_0$ is not constant in our case but differs if we refer to the intervention to change the land albedo and in that case is 330 $Wm^{-2}$ or the oceans 345 $Wm^{-2}$. It is not so simple to relate the change in the albedo of clouds or land $\Delta\alpha$ to the change in planetary albedo. A possible relationship is

$$\Delta\alpha_p = f_a f_{\text{Earth}} \Delta\alpha \tag{20.11}$$

where $f_{\text{Earth}}$ is the fraction of the Earth's surface affected by the intervention and $f_{\text{a}}$ is a factor which takes into account the effects of absorption and reflection of the other layers.

In the previous chapters, we have discussed at length some of these questions and found, for example, that for clear sky conditions, $f_a$ could be intended as a two-way transmittance of the atmosphere $f_a = \mathrm{T}_a^2$ where $\mathrm{T}_a$ is the transmittance of the atmosphere.

Cloudy skies are the most interesting case if we want to find simple ways to treat them and also because one of the most popular geoengineering proposals deals with the whitening of clouds. This could be obtained by adding aerosols to the clouds and increasing the number density of the cloud droplets, which follows the so-called Twomey effect. Aerosol number concentration $N_a$ can be related to cloud droplet number density, $N_d$, by

$$N_d = N_a^{\gamma} \tag{20.12}$$

where $\gamma$ ranges between 0.06 and 0.48. An empirical relation is used to link the number density to the effective radius $r_e$

$$r_e = k(N_d)^{-1/3} \tag{20.13}$$

where $k$ is a constant derived from observation. As shown by Twomey (see Examples), the cloud optical thickness can be related to the liquid water content $l$, the cloud height $h$, and the water density $\rho_w$

$$\tau = \frac{3lh}{2\rho_w r_e} \tag{20.14}$$

If we assume that the Twomey effect requires that the product $l \cdot h$ stay constant, then using (20.13) we have

$$\tau = K(N_a)^{\gamma/3} \tag{20.15}$$

The cloud albedo can be calculated using the Eddington approximation (see Examples) as

$$\alpha_c = \frac{0.866\,(1-g)\,\tau_c}{1 + 0.866\,(1-g)\,\tau_c} \tag{20.16}$$

where $g$ is the asymmetry factor 0.85. The numerical factor depends on the diffusivity factor used. We stick to the expression given by Lacis and Hansen (1974). Differentiating (20.16), we have the change in cloud albedo

$$\Delta\alpha_c = \frac{0.866\,(1-g)}{\left[1 + 0.866\,(1-g)\,\tau_c\right]^2}\Delta\tau_c \tag{20.17}$$

This expression can be used to evaluate $N_d$ once $\alpha$ and $\tau$ have been fixed.

The effect of the clouds on the planetary albedo is taken into account in a simple way. We assume a transmittance to the top of the clouds $T_a = 0.925$ and correct this value, taking into account the reflection from the ocean surface to the clouds and the transmittance to and through the clouds $T_c \sim 0.5$ and then we have

$$f_a = T_a^2 \left(1 - T_c^2 \alpha_s\right) \tag{20.18}$$

where $\alpha_s$ is ocean albedo. Using $\alpha_s = 0.09$, $f_a$ without cloud $= 0.856$, we got $f_a = 0.84$.

These simple calculations can be used to evaluate some basic requirement for geoengineering. For example, suppose we want to offset the radiative forcing for $CO_2$ doubling (3.7 $Wm^{-2}$), increasing the cloud albedo. From (20.10), we calculate the change in planetary albedo $\Delta\alpha_p$ with $S_0 = 345$ $Wm^{-2}$ and we find $\Delta\alpha_p = 0.01$. Then from (20.11), we evaluate the change in cloud top albedo to obtain such a change in planetary albedo once we have fixed the fraction of interested area $f_{\text{Earth}}$. Taking 17.5 %, we obtain a change in cloud albedo of 0.075, and this value can be used through (20.15) and (20.17) to get the change in aerosol number density

$$\frac{\Delta\tau}{\tau} = \frac{\gamma}{3}\frac{\Delta N_a}{N_a} \tag{20.19}$$

where $\Delta\tau$ is given by (20.17). We get a change in number density of the order of 30 % well within reasonable limit. There are many other possibilities to change the albedo of the Earth that go from changing the color of cropland to modifying the reflectivity of buildings, and all of these can be worked out with some detail.

### *20.4.2 Stratospheric Aerosol or How to Create a Volcanic Eruption*

We have mentioned in the previous chapters that volcanic eruptions may somewhat cool the Earth. The massive amount of sulfur compounds emitted by catastrophic volcanic eruptions directly in the stratosphere condenses in sulfate aerosol with very high reflectivity and so may reduce the amount of solar radiation reaching the surface of the Earth. Hansen (2005) claims to have found a relationship between optical thickness measured at 0.55 μm and radiative forcing, and that is

$$RF\left(Wm^{-2}\right) \approx 25\tau \tag{20.20}$$

This is an "empirical" relation obtained using general circulation models. However, following a very old paper by Harshvardan and Cess (1976), we can model the influence of the aerosol layer as a reflecting layer with reflectivity R over the surface

of the Earth with reflectivity a. We just neglect all the absorption and we get that the net radiation absorbed at the ground is

$$F = (Q/4)\,(1-a)\,(1-R)\left[1 + (aR) + (aR)^2 + \ldots\right]$$

and that is

$$F = F_0\,(1-R)\,/\,(1-aR) \tag{20.21}$$

where $F_0 = (Q/4)\,(1-\mathrm{a})$ is the absorbed flux in the absence of aerosol. The radiative forcing is then

$$F_0 - F = \Delta F \approx F_0 R\,(1-a)\,(1+aR) \approx F_0 R\,(1-a) \tag{20.22}$$

where we have assumed $(1+aR) \sim 1$. Then, we get $\Delta F \sim 324\,R$ $(\mathrm{Wm}^{-2})$, and following the estimate of Lacis et al. (1992), we may get a relation between $\tau$ and the reflectivity $\mathrm{R} \sim 4.5 \cdot 10^{-2}\ \tau$, that is, $\Delta F \sim 14.5\ \tau(\mathrm{Wm}^{-2})$. This is less than the estimate (20.20) but gives an idea where the proportionality between optical thickness and radiative forcing may originate.

Once we have the above relation, we may adopt a climate sensitivity and get the expected change in temperature. Optical thickness for large volcanic eruption may be slightly larger than 0.1 and may get a radiative forcing around 2.5 $\mathrm{Wm}^{-2}$ to which corresponds a temperature change of the same order or larger.

We have then the example from nature, and the idea is then why we do not reinforce the aerosol's presence in the stratosphere artificially. We may inject some Gt of sulfur in the stratosphere and create an artificial volcanic eruption. Calculations are easily made using models, and it turns out that 1.5–5 Tg of *S* per year would do the job.

The central problem of this geoengineering effort is how to pump such large amount of sulfur in the stratosphere, considering that the lifetime is around 1 or 2 years. This means that there must be ways to inject such large amount of sulfur on a continuous basis. At present, there is no aircraft capable of transporting this mass of sulfur. The only option left is to build new planes or devise new ways to refurbish the stratosphere with sulfur. The development of a new aircraft capable of transporting 1 Mt/year could be around 1–2 billion dollars that would increase to 5–10 billion dollars for 5 Mt/year. The development time is another problem as shown by experience in the last few years; it could take between 10 and 20 years. The operative costs could correspond to that of a small airline company (1 billion dollar), while the acquisition cost of the new aircrafts could be around 10 billion. These are very rough estimations and could be in error by a factor 5 or 10.

Such large expenses could be justified on the ground that the mitigation cost for global warming could be in the range of 0.2–2.5 of the gross global product that is between 20 and 250 billion dollars.

It must be clear that the direct injection in the stratosphere is essential for the generation of a stratospheric cloud. The simple cycle outlined by Rasch (2008) has been used to evaluate the sulfate stratospheric burden, its stratospheric deposition, and the tropospheric scavenging. The total tropospheric source has been 65 Mt/year, while the natural quiescent volcanic source is 0.1 Mt/year. The reference quantities are a scavenging of 50 Mt/year and a sedimentation of 0.24 Mt/year with a sulfate burden in the stratosphere of 0.24 Mt. If the tropospheric source is perturbed by 1 Mt/year, the scavenging flux increases on the same amount and the sedimentation rises to 0.25 Mt/year. For a perturbation of 5 Mt/year, the scavenging increases up to 54.8 and the sedimentation to 0.26 Mt/year, and as expected, the perturbation is completely absorbed in the troposphere.

On the other hand, if the same perturbation is operated in the stratosphere, then for 1 Mt/year, the scavenging does not change, while the sedimentation arrives to 1.14 Mt/year. If the perturbation is for 5 Mt/year, the scavenging flux stays practically the same, while the sedimentation is up to 5.14 Mt/year, and the sulfate burden has the same value. With respect to the basic tropospheric flux of 65 Mt/year, 5 Mt/year could be negligible, but if the perturbation is applied directly to the stratosphere, the effects could be very important.

The claim usually is that if compared with the natural flux (roughly 100 Mt/S), the perturbation is only 5 %, but this is true only if the perturbation is performed in the troposphere but not when the sulfur is injected directly into the stratosphere.

The other topics usually neglected in this kind of proposal are the effects of deposition. A recent paper by Sigl et al. (2014) reports some data on the sulfate deposition. For the Samalas eruption (1257) for a total emission of 170 Mt/s, we have a deposition of 73 k/m$^2$ in the Antarctic region. For the 1458 Kuwae (Vanuatu island in the Coral Sea) with a possible emission of 100 Mt, the deposition amounted to 64 kg/km$^2$. Finally, for the Tambora, the emission was 40 Mt and the deposition of 46 kg/km$^2$.

Based on the study on a number of eruptions, it is possible to establish a relation between the total hemispheric mass of aerosol and the sulfate flux in the atmosphere. The relation is linear

$$M \simeq 1.25S$$

where $M$ is the mass in Mt and $S$ is the injected $SO_2$ mass. For the arctic area, we found a slightly nonlinear relation

$$F \simeq 0.3S^{1.3}$$

These relations say that if we double the injection of sulfur, the sulfate mass deposition increases by a factor 2.5. These results confirm what other models have to say. For a volcanic source ranging between 10 and 14 %, the wet sulfate deposition is not proportional but changes between 30 and 36 %.

Only a few papers detail the data on the deposition, and typically for a 5 Mt annual source, we have a production of 7.5 Mt of $SO_2$ that if distributed uniformly

would give a flux of $1.45 \cdot 10^{-5}$ kg $SO_4/m^2$ and using the nonlinear relation could be as low as a $1.20 \cdot 10^{-5}$ kg $SO_4/m^2$. This flux would impact on the acidity of the soil and ocean.

The most interesting data in this case are those published by the National Atmospheric Deposition Program (NADP) (http://nadp.isws.illinois.edu) which show that in 2012, most of the fluxes over the USA were lower than $8 \cdot 10^{-4}$ kg $SO_4/m^2$, with the Middle West regions having larger fluxes up to $20 \cdot 10^{-4}$ kg $SO_4/m^2$, while northern regions have fluxes of the order of 1 and $3 \cdot 10^{-4}$ kg $SO_4/m^2$. The geoengineering approach would then produce acidic fluxes at the least comparable or higher than the natural fluxes.

Another possible negative effect of an artificial volcanic eruption is the ozone depletion. The formation of a stable stratospheric aerosol layer may favor the heterogeneous chemistry over the $H_2SO_4$ particle as we have seen in the previous chapters.

## 20.5 Space Shields

The most futuristic approach to geoengineering is the proposal to limit the amount of solar radiation reaching the Earth by putting "sunshades" directly in space. The original idea was formulated by Early in 1989 but made popular by Roger Angel in 2006. We repeat here some of the simple considerations done with energy balance climate models where the equilibrium temperature results from the equivalence between the incoming solar radiation and the outgoing longwave radiation

$$\overline{T} = \frac{Q(1-\alpha) - A}{B}$$

where $A$ and $B$ are the usual coefficient relating the IR flux to temperature

$$I = A + BT$$

If the solar constant $Q$ is perturbed by an amount $\delta Q$, then the resulting temperature change is

$$T(t) = \overline{T} + \frac{\delta Q}{B}(1-\alpha)\left[1 - \exp\left(-\frac{t}{\tau}\right)\right] \tag{20.23}$$

where $t = C/B$ with $C$ as the heat capacity of the ocean mixed layer. The change in temperature for a change $\delta Q$ in the solar constant will then be $\delta T = \delta Q(1-\alpha)/B$.

Assuming for $B$ a value of 2.17 W/m$^2$, we have that for a change of 2 K, we need a change of 1.7 % in solar constant in the case of artificial volcanic eruption that was obtained with the volcanic cloud.

The idea beyond the space shield is to put a blind directly on the sun. In principle, that could be obtained by putting a disk in a place in the solar system, which requires

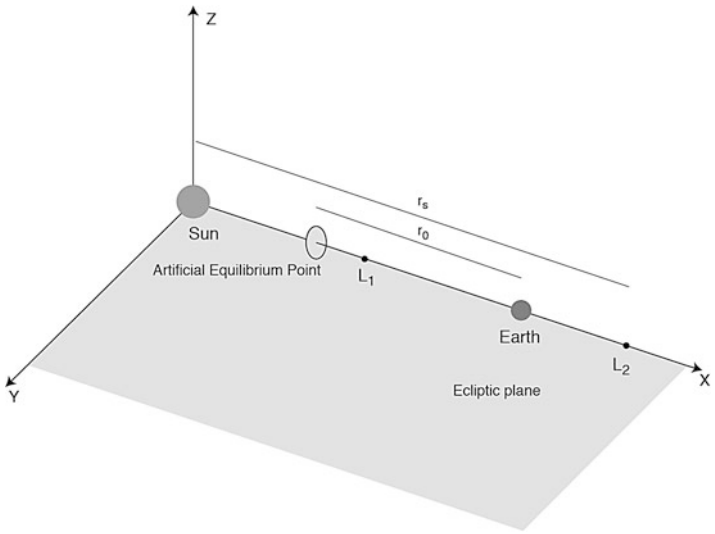



**Fig. 20.7** Occulting solar disk positioned at an artificial equilibrium point. Points $L_1$ and $L_2$ are the classical Lagrange points (Adapted from McInnes 2010)

the least energy to control. The most obvious place could be one of the equilibrium points in the solar system called the Lagrange points. We can refer to Fig. 20.7 and suppose to have a disk of radius $R_s$ at a distance $r_s$ from the Earth. The sun on the other hand will have a radius $R_0$ at a distance $r_0$. The solid angles subtended by the two are $\Omega_S = \pi R_S^2/r_S^2$ and $\Omega_O = \pi R_O^2/r_O^2$, respectively, and the fractional reduction of the solar constant would just be the ratio of the two solid angles

$$\frac{\delta Q}{Q} = \frac{\Omega_S}{\Omega_O} = \left(\frac{R_S}{R_O}\right)^2 \left(\frac{r_O}{r_S}\right)^2 \tag{20.24}$$

From this relation, we obtain the disk radius once its distance from Earth and the solar constant reduction are specified. We get

$$R_S = R_O \left(\frac{r_S}{r_O}\right) \left(\frac{\delta Q}{Q}\right)^{1/2} \tag{20.25}$$

Actually, the disk radius is important because the disk mass depends on it. On the other hand, the distance $r_s$ is now determined not only by the gravitational force but also by the radiation pressure. We have at the equilibrium

$$\frac{GM_E}{r_S^2} - \frac{GM_O}{(r_O - r_S)^2} + \omega^2 (r_O - r_S) + a_S = 0 \tag{20.26}$$

where $\omega = \left(GM_O/r_O^3\right)^{1/2}$, that is, the angular velocity of the Earth relative to the sun, and $a_S$ is the acceleration due to the radiation pressure. If $P_E$ is the radiation pressure at the distance of 1 astronomical unit, the acceleration $a_S$ would be

$$a_S = \frac{2kP_E A_S}{M_S}\left(\frac{r_O}{r_O - r_S}\right)^2 \tag{20.27}$$

where $A_S$ is the disk area, $M_S$ its mass, and $k$ a coefficient that takes into account the optical properties of the disk. The coefficient $k$ is a very complicated function of the specular reflectivity $\eta$ and the emissivity of the sun facing side $\varepsilon_F$ and the Earth facing side of the disk $\varepsilon_B$

$$k = \frac{1}{2}\left[(1+\eta) + \frac{2}{3}(1-\eta)\frac{\varepsilon_F - \varepsilon_B}{\varepsilon_F + \varepsilon_B}\right] \tag{20.28}$$

Now, we get $M_S$ from (20.27) and substitute for $A_S$ the expression we get from (20.24). We have for the mass of the disk

$$M_S(r_S) = \frac{2\pi kP_E R_O^2}{a_S(r_S)}\left(\frac{\delta Q}{Q}\right)\left(\frac{r_O}{r_O - r_S}\right)^2 \tag{20.29}$$

The mass of the disk is a function of the distance and so can be minimized by finding the minimum of the function

$$f(r_S) = \frac{1}{a_S(r_S)}\left(\frac{r_O}{r_O - r_S}\right)^2$$

which gives a disk located at $r_S = 2.56 \cdot 10^6$ km from Earth a little closer to the sun than the classical Lagrange point. The results for the radius that would reduce the solar constant by 1.7 % are about 1450 km, and its mass would be in the range $2.6 \cdot 10^8$ tons. This would be quite a fit if you consider that the concrete used to build the Chinese Three Gorges Dam is around $6 \cdot 10^7$ tons. You can only imagine how many rockets you need to build such a disk and how much it would cost!

## 20.6 Can Solar Radiation Management Work?

We have assumed so far that reducing the absorbed solar radiation has the opposite sign of increasing the greenhouse effect. There are few people that think differently based on some elementary thermodynamic constraints. That does not mean that a lowering of the absorbed solar radiation does not decrease the temperature but simply that the amount of reduction must be calculated differently, and besides, there could some additional unpleasant consequences.

Bala et al. (2008) have used a GCM to evaluate the sensitivity of the climate system to a $CO_2$ doubling (from 315 to 630 ppm) and a reduction in solar flux of 1.8 %. They have found that for the $CO_2$ doubling, the temperature increased by 2.42 K while decreased by 2.4 K for solar reduction. The same happens for the water vapor content that rises by 15 % in the first case and decreases of the same amount in the second case. The surprise comes from the global precipitation that increases by 3.7 % for a $CO_2$ doubling and decreases by 5.8 % for the solar radiation reduction. Apparently then, while the geoengineering could "cure" the warming, it would then produce a tendency to make the world drier. The same tendency was noted with similar experiments.

This effect can be explained qualitatively in the following way: at the surface of the Earth, the flux differences between the control and perturbed case must be zero, that is,

$$\Delta R + \Delta S - \Delta L - \Delta H = 0 \tag{20.30}$$

where $\Delta R, \Delta S, \Delta L, \Delta H$ are the differences in longwave radiation, shortwave radiation, latent heat flux, and sensible heath flux. It is important to distinguish the radiative flux between "response" and "forcing" where the forcing is the instantaneous impact of some perturbation on radiation. Then, (20.30) becomes

$$F + \Delta R_r + \Delta S_r - \Delta L - \Delta H = 0 \tag{20.31}$$

where $F$ is the sum of shortwave and longwave radiative forcing and the subscript $r$ represents the response component of the change in radiation.

In the model used, the surface forcing is roughly $-3$ $\mathrm{Wm}^{-2}$ for the solar experiment, but it is only a few tens of a $\mathrm{Wm}^{-2}$ for the $CO_2$ doubling. Now, the response is independent of the forcing mechanism, and $\Delta R_r$ has then the same magnitude but opposite sign in the two experiments (solar and $2 \times CO_2$). The same happens for the solar radiative response $\Delta S_r$. The radiative response on (20.31) can be neglected so that the forcing must be equal to the sensible and latent heat

$$F = \Delta L + \Delta H \tag{20.32}$$

Now, the solar forcing causes a negative radiative forcing at the surface, and considering that latent heat is about five times the sensible heat, it will dominate the response. This is the primary cause of the large decrease in precipitation for the solar engineering.

This qualitative argument has been substantiated, recently, using a simple energy balance model by Kleidon and Renner. The model detailed in the example assumes the energy balance at the surface given by

$$R_l - S - \lambda E - H = 0 \tag{20.33}$$

Dove $R_l$ is the longwave contribution, $S$ is the absorbed solar radiation, $H$ the sensible heat, and $\lambda E$ the latent heat flux with $\lambda$ the latent heat of vaporization. All the fluxes are linearized and written as

$$\begin{aligned} R_l &= k_r \left(T_S - T_a\right) \\ H &= c_p \rho w \left(T_S - T_a\right) \\ E &= \rho w \left(q_{\text{sat}}\left(T_S\right) - q_{\text{sat}}\left(T_a\right)\right) \end{aligned} \tag{20.34}$$

where $T_s$ and $T_a$ are the surface and atmospheric temperatures, $q_{\text{sat}}$ is the saturation pressure at the indicated temperature, $w$ is the velocity which describes the vertical mass exchange, and $k_r$ is an exchange coefficient. The dependence of $q_{\text{sat}}$ from temperature is also linearized so that the latent heat flux is written as

$$\lambda E = c_p \rho w \frac{s}{\gamma} \left(T_S - T_a\right) \tag{20.35}$$

where $\gamma = c_p p / \left(0.622\lambda\right) \approx 65\ \text{PaK}^{-1}$ is the psychrometric constant and s is the slope of the saturation curve $s = \left(\partial e_{\text{sat}} / \partial T\right)$. It is assumed that the system works as a Carnot engine in which the net source of heat is given by the sensible heat flux utilized according to the efficiency of the Carnot engine $H\left(T_S - Ta\right) / T_S$. This energy is used to generate the convective motions, and when is maximized, the velocity is obtained as

$$w = \frac{\gamma}{\gamma + s} \frac{S}{2 c_p \rho \left(T_S - T_a\right)} \tag{20.36}$$

At this point, we have all the elements (almost) to evaluate the hydrologic sensitivity defined as follows:

$$\frac{1}{E}\frac{dE}{dT_S} = \frac{1}{E}\frac{\partial E}{\partial s}\frac{ds}{dT_S} + \frac{1}{E}\frac{\partial E}{\partial S}\frac{dS}{dT_S} \tag{20.37}$$

This can also be written as

$$\frac{1}{E}\frac{dE}{dT_S} = \frac{1}{E}\frac{\partial E}{\partial s}\frac{ds}{dT_S} + \frac{1}{E}\frac{\partial E}{\partial S}\left(\frac{dT_S}{dS}\right)^{-1} \tag{20.38}$$

The two terms on the right side represent the change in evaporation ($E$) to the surface temperature for a change in $s$ and can be written (see Examples) as

$$\frac{1}{E}\frac{\partial E}{\partial s}\frac{ds}{dT_S} = \frac{\gamma}{\gamma + s}\frac{1}{s}\frac{ds}{dT_S} \tag{20.39}$$

This term can be evaluated and it can be shown as

$$\frac{1}{E}\frac{\partial E}{\partial s}\frac{ds}{dT_S} \approx 2.2\% K^{-1} \tag{20.40}$$

The second term is the sensitivity to change in solar radiation ($S$)

$$\frac{1}{E}\frac{\partial E}{\partial S}\left(\frac{dT_S}{dS}\right)^{-1} = \frac{4k_r\sigma^{1/4}}{2\sigma^{1/4}S + k_r S^{1/4}} \approx 1\%K^{-1} \tag{20.41}$$

This means that the total sensitivity to a change in solar radiation is $1\,\% + 2.2\,\% = 3.2\,\%$. Now, suppose we change the greenhouse forcing so that the surface temperature increases by 2 K. Then the percentage change in evaporation would be $\Delta E/E = 4.4\%$ and positive. If this increase in temperature is counteracted using SRM, the change would be $\Delta E/E = -6.4\%$ and negative. The net effect would be a reduction in the evaporation by $2\,\%$ and consequently on the precipitation.

As shown by the simulation with GCM and by these simpler models, the solar radiation management is somewhat unsymmetrical, and while compensating, the warming would produce a reduction in precipitation. A similar conclusion can be reached for the mass exchange.

## 20.7 A Cure for the Ozone Hole with Geoengineering

Many years ago, well before geoengineering became popular, there was a proposal to do something about the ozone hole. The paper appearing on science in 1991 proposed to inject several thousands of ethane (or propane) into the Antarctic stratosphere to reduce the quantity of free chlorine and then to favor the ozone recover. We will follow here the approach proposed by Jacob as exercise and then we will correct the proposal and show that it was quite wrong. The interesting thing is that it was corrected basically by the same people who had proposed it. This is a rare example of the power of scientists: you make a mistake and correct it and you get published twice. We will follow an excercise by Jacob (1999).

We start out with the reactions taking place in the Antarctic atmosphere

$$\begin{aligned}
&\mathrm{HCl} + \mathrm{ClNO_3} \overset{\text{aerosol}}{\rightarrow} \mathrm{Cl_2} + \mathrm{HNO_3}\\
&\mathrm{HCl} + \mathrm{N_2O_5} \overset{\text{aerosol}}{\rightarrow} \mathrm{ClNO_2} + \mathrm{HNO_3}\\
&\mathrm{N_2O_5} + \mathrm{H_2O} \overset{\text{aerosol}}{\rightarrow} 2\mathrm{HNO_3}\\
&\mathrm{Cl_2} + \mathrm{hv} \rightarrow 2\mathrm{Cl}\\
&\mathrm{ClNO_2} + \mathrm{hv} \rightarrow \mathrm{Cl} + \mathrm{NO_2}\\
&\mathrm{Cl} + \mathrm{O_3} \rightarrow \mathrm{ClO} + \mathrm{O_2}\\
&\mathrm{ClO} + \mathrm{NO_2} + \mathrm{M} \rightarrow \mathrm{ClNO_3} + \mathrm{M}\\
&\mathrm{ClO} + \mathrm{ClO} \rightarrow 2\mathrm{Cl} + \mathrm{O_2}
\end{aligned} \tag{20.42}$$

If we assume initial values of 1.5 ppbv for HCl, 0.3 ppbv for $ClNO_3$, and 1.8 ppbv for $N_2O_5$, it is easy to show that after completion of the above reaction, we end up with zero HCl, 0.6 ppbv of (Cl + ClO), and 1.2 ppbv of $ClNO_3$. This can be as follows. From the first reaction, only 0.3 ppbv of $Cl_2$ can be formed because of the limited quantity of $ClNO_3$. The remaining 1.2 ppbv of HCl may form the same amount of $ClNO_2$ from the second reaction, consuming only 1.2 ppbv of $N_2O_5$. Then, the photodissociation reactions will produce 0.6 ppbv of Cl and 1.2 ppbv of Cl and the same amount of $NO_2$. The Cl will be consumed in the reaction with ozone forming 1.8 ppbv of ClO. Of these, 1.2 ppbv will be consumed in the reaction with $NO_2$ producing the same amount of $ClNO_3$.

When ethane is added we must consider the reaction

$$Cl + C_2H_6 \rightarrow HCl + C_2H_5 \tag{20.43}$$

followed by

$$HCl + ClNO_3 \overset{\text{aerosol}}{\rightarrow} Cl_2 + HNO_3$$

$$Cl_2 + hv \rightarrow 2Cl; \qquad Cl + C_2H_6 \rightarrow HCl + C_2H_5$$

$$HCl + ClNO_3 \overset{\text{aerosol}}{\rightarrow} Cl_2 + HNO_3; \qquad Cl_2 + hv \rightarrow 2Cl$$

Adding 3 ppbv of ethane, we have

$$0.6\ Cl + 3\ C_2H_6 \rightarrow 0.6\ HCl + 0.6\ C_2H_5 + 2.4\ C_2H_6$$

where the numbers are the mixing ratios in ppbv. These will be followed by

$$0.6\ HCl + 1.2\ ClNO_3 \rightarrow 0.6\ Cl_2 + 0.6\ ClNO_3 + 0.6\ HNO_3$$

$$0.6\ Cl_2 + hv \rightarrow 1.2\ Cl;$$

$$1.2\ Cl + 2.4\ C_2H_6 \rightarrow 1.2\ HCl + 1.2\ C_2H_5 + 1.2\ C_2H_6$$

$$1.2\ HCl + 0.6\ ClNO_3 \rightarrow 0.6\ Cl_2 + 0.6\ HNO_3 + 0.6\ HCl;$$

$$0.6\ Cl_2 + hv \rightarrow 1.2\ Cl$$

$$1.2\ Cl + 1.2\ C_2H_6 \rightarrow 1.2\ HCl + 1.2\ C_2H_5$$

It can be easily seen that any lesser amount of 3 ppbv of $C_2H_6$ would bring an increase in (Cl + ClO) and consequently an increase in ozone destruction.

Before going on examining the pitfalls of the above proposal, let us see how much ethane we should have used to damp the ozone hole. We assume a total surface of the ozone hole of about $20 \times 10^6$ km$^2$, that is, $2 \times 10^{13}$ m$^2$. This area extends from 15 to 20 km with a pressure difference of 150 hPa. We get a total mass of about

$$\Delta m = S\Delta p/g = 2 \cdot 10^{13} \times 1.5 \cdot 10^4/9.8 \approx 2 \cdot 10^{16} kg$$

Given the molecular weight of ethane (30) being very similar to the one of the atmosphere, we can use the volume mixing ratio to evaluate the mass to be injected

$$C_2H_6\,(\text{mass}) \approx 2 \cdot 10^{16}\text{kg} \times 3 \cdot 10^{-9} = 6 \cdot 10^{7}\text{kg} = 60,000\ \text{tons}$$

This amount of ethane should be injected above 15 km. At the present time, there is no aircraft capable of such a feat. Even the most modern of aerial tankers (Boeing KC 46) would carry about 90 tons of fuel per flight which means you need roughly 1000 flight of a nonexistent plane to do the job. However, even considering this "minor problem" 3 years later, the same authors published a paper that admitted they had neglected something so important to make the entire proposal useless if not dangerous.

The main point of the consideration of the heterogeneous reaction is

$$\text{HCl} + \text{HOCl} \overset{\text{aerosol}}{\rightarrow} \text{Cl}_2 + \text{H}_2\text{O} \tag{20.44}$$

with the main source for hypochlorous acid being the reaction

$$\text{ClO} + \text{HO}_2 \rightarrow \text{HOCl} + \text{O}_2$$

Actually, you need some hydroxyl that in polar stratosphere can be produced by methane oxidation and then similarly by oxidation of ethane. The paper showed that considering only ethane the chlorine cycle would sum to the reaction

$$\text{C}_2\text{H}_6 + 2\text{O}_3 \rightarrow \text{C}_2\text{H}_4\text{O} + \text{H}_2\text{O} + \text{O}_2$$

or starting with the hydroxyl

$$\text{C}_2\text{H}_6 + \text{OH} + \text{HCl} + 2\text{O}_3 \rightarrow \text{C}_2\text{H}_4\text{O} + 2\text{H}_2\text{O} + \text{Cl} + 2\text{O}_2$$

The production of acetaldehyde ($C_2H_4O$) means that the subsequent reactions would imply further ozone reduction. Again, with the chlorine

$$\text{C}_2\text{H}_4\text{O} + 3\text{O}_3 \rightarrow \text{CH}_2\text{O} + \text{CO}_2 + \text{H}_2\text{O} + 3\text{O}_2$$

or the hydroxyl

$$\text{C}_2\text{H}_4\text{O} + \text{OH} + \text{HCl} + 3\text{O}_3 \rightarrow \text{CH}_2\text{O} + \text{CO}_2 + 2\text{H}_2\text{O} + 3\text{O}_2$$

and finally, with photodissociation

$$\text{C}_2\text{H}_4\text{O} + 2\text{HCl} + 3\text{O}_3 \rightarrow \text{CH}_2\text{O} + \text{CO} + 2\text{H}_2\text{O} + 2\text{Cl} + 3\text{O}_2$$

In each of these cases, the ozone destruction is increased.

## E.20 Examples

### *E.20.1 Back to Radiative Transfer*

We simplify somewhat what we have done in Chap. 13 and try to obtain directly an expression for the downward and upward intensity of radiation. We start by defining the intensity-only function of the zenith angle

$$I(\mu) = \frac{1}{2\pi}\int_0^{2\pi} I(\mu, \varphi)\, d\varphi \tag{E.20.1}$$

Then, the upward and downward intensities are defined by

$$I(\mu) = \begin{cases} I^{\uparrow} & \mu > 0 \\ I^{\downarrow} & \mu < 0 \end{cases} \tag{E.20.2}$$

Then, we may write directly the expression for the change of $I^{\uparrow}$

$$\frac{1}{2}\frac{dI^{\uparrow}}{d\tau} = I^{\uparrow} - \omega I^{\uparrow} + \omega b I^{\uparrow} - \omega b I^{\downarrow} \tag{E.20.3}$$

On the rhs, the first term is attenuation, the second is the intensity lost because of backscattering, and the third term is the fraction of the scattered radiation diffused in the same direction of the incident radiation. The last term is the same but for the downward intensity. In this relation, $\omega$ is the single-scattering albedo and is the backscattering fraction. Similar expressions can be written for the forward intensity

$$-\frac{1}{2}\frac{dI^{\downarrow}}{d\tau} = (1-\omega) I^{\downarrow} - \omega b \left(I^{\uparrow} - I^{\downarrow}\right) \tag{E.20.4}$$

We can find a simple relation between the backscattered fraction and the asymmetry factor g considering that if the scattering is isotropic, then $g = 0$, so $b = 1/2$. If $g = 1$, then all the radiation is scattered in the same incident direction and $b = 0$. Finally, if $g = -1$, all the radiation is scattered back and $b = 1$. Then, we have the relation

$$b = (1-g)/2 \tag{E.20.5}$$

Substitution in (E.20.3) and (E.20.4) gives

$$\begin{aligned} \frac{1}{2}\frac{dI^{\uparrow}}{d\tau} &= (1-\omega) I^{\uparrow} + \frac{\omega(1-g)}{2}\left(I^{\uparrow} - I^{\downarrow}\right) \\ -\frac{1}{2}\frac{dI^{\downarrow}}{d\tau} &= (1-\omega) I^{\downarrow} - \frac{\omega(1-g)}{2}\left(I^{\uparrow} - I^{\downarrow}\right) \end{aligned} \tag{E.20.6}$$

If we consider the case of conservative scattering $\omega = 1$, then the sun of the last two equations becomes

$$\frac{1}{2}\frac{d}{d\tau}\left(I^{\uparrow} - I^{\downarrow}\right) = 0 \tag{E.20.7}$$

then corresponds to

$$\left(I^{\uparrow} - I^{\downarrow}\right) = \text{constant} = \frac{F_{\text{net}}}{\pi} \tag{E.20.8}$$

If we subtract Eq. (E.20.6), we get

$$\frac{d}{d\tau}\left(I^{\uparrow} + I^{\downarrow}\right) = 2\,(1-g)\left(I^{\uparrow} - I^{\downarrow}\right) = \frac{2F_{\text{net}}}{\pi}\,(1-g)$$

which integrates to

$$\left(I^{\uparrow} + I^{\downarrow}\right) = \frac{2F_{\text{net}}\,\tau}{\pi}\,(1-g) + \text{cost} \tag{E.20.9}$$

This can be solved with the help of (E.20.8) to give

$$\begin{aligned} I^{\uparrow} &= \tfrac{F_{\text{net}}}{2\pi}\left[1 + 2\tau\,(1-g)\right] + \tfrac{1}{2}\cos \text{t} \\ I^{\downarrow} &= -\tfrac{F_{\text{net}}}{2\pi}\left[1 - 2\tau\,(1-g)\right] + \tfrac{1}{2}\cos \text{t} \end{aligned} \tag{E.20.10}$$

The boundary conditions for a cloud of optical thickness $\tau^*$ are $I^{\uparrow}(\tau^*) = 0$ and $I^{\downarrow}(0) = I_0$. The solution becomes

$$I^{\uparrow} = \frac{I_0\,(1-g)\,(\tau^* - \tau)}{1 + (1-g)\,\tau^*}; \qquad I^{\downarrow} = \frac{I_0\left[1 + (1-g)\,(\tau^* - \tau)\right]}{1 + (1-g)\,\tau^*} \tag{E.20.11}$$

The albedo and the transmittance are then

$$r = \frac{I^{\uparrow}(0)}{I^{\downarrow}(0)} = \frac{(1-g)\,\tau^*}{1 + (1-g)\,\tau^*}; \quad t = \frac{I^{\downarrow}(\tau^*)}{I^{\downarrow}(0)} = \frac{1}{1 + (1-g)\,\tau^*} \tag{E.20.12}$$

## *E.20.2 The Twomey Effect*

We would like to explain the relation (20.14) following the paper by Brenguier et al. (2011). However, we will start from the simple and short note by Sean Twomey that gave origin to all. The optical thickness is defined as

$$\tau = \pi \int_0^H \int_0^{\infty} Q_E(x) n\,(r,z)\,r^2 dr dz \tag{E.20.13}$$

where n is the size distribution of water droplets with radius r having extinction coefficient $Q_E$ for the size parameter $x$. $H$ is the depth of the cloud. Assuming a constant value for the extinction $\approx 2$, we arrive at the simple formula

$$\tau = 2\pi N\bar{r}^2 H \tag{E.20.14}$$

where $N$ is now the number of droplets per unit volume. On the other hand, the water content is given by in unit of mass per unit area

$$W = 4/3\pi\rho_w N\bar{r}^3 H \tag{E.20.15}$$

Keeping the $W$ constant, we can obtain $\bar{r}$ from the last relation so that the optical thickness will scale as $N^{1/3}$. We have just seen (E.20.12) that the albedo increases with increasing optical thickness and so with the number densities of particles. This approach can be detailed a little bit. Again, start by calculating the optical thickness of a cloud following the relation

$$\begin{aligned}\tau &= \pi\int_0^H\int_0^\infty Q_E(x)n\,(r,z)\,r^2 dr dz \\ &= \int_0^H \pi Q_E\,(\bar{x})\,N(z)r_2^2(z)dz = \int_0^H \pi Q_{ext}\,(\bar{x})\,M_2(z)dz\end{aligned} \tag{E.20.16}$$

In this case,

$$r_2^2 = (1/N)\int_0^\infty n\,(r,z)\,r^2 dr$$

is the mean surface radius, and $M_2(z) = N(z)r_2^2(z)$ is the second moment of the droplet spectrum. In the same manner, the water content can be defined as

$$\begin{aligned}W &= \int_0^H q_c(z)dz = 4/3\pi\rho_w\int_0^H N(z)r_3^3(z)dz \\ &= 4/3\pi\rho_w\int_0^H M_3(z)dz\end{aligned} \tag{E.20.17}$$

where $q_c = 4/3\pi\rho_w N r_3^3$ is the liquid water content, $r^3$ the mean volume droplet radius, and $M_3$ the third momentum. (E.20.16) and (E.20.17) can be treated as in Twomey, so we get

$$\tau = \pi Q_E M_2 H = \pi Q_E N^{1/3} M_3^{2/3} H = A(NH)^{1/3} W^{2/3} \tag{E.20.18}$$

where $A = \pi Q_E/(4/3\pi\rho_w)^{2/3}$. This relation can be obtained considering that $M_2 = r^2 NH$, $M_3 = r^3 NH$. We get $r = (M_3/N)^{1/3}$ and $M_2 = N^{1/3}M_3^{2/3}$. This expression

is valid for a monodisperse droplet spectrum and in this case, $r_2 = r_3$, while in the most general case, the bias between $r_3$ and $r_2$ could be taken into account with a correction factor

$$k = \left(\frac{r_3}{r_e}\right)^3 = \left(\frac{r_2}{r_3}\right)^6$$

where $r_e$ is the droplet effective radius $r_e = r_3^3/r_2^2$. In this case, $M_2 = (kN)^{1/3} M_3^{2/3}$ and (E.20.18) becomes

$$\tau = A(kNH)^{1/3} W^{2/3} = A\left(\frac{r_3}{r_e}\right)[NH]^{1/3} W^{2/3} \tag{E.20.19}$$

where

$$(NH)^{1/3} = \frac{W^{1/3}}{\left(\frac{4}{3}\pi\rho_w r_3^3\right)^{1/3}}$$

Substituting in (E.20.19), we get

$$\tau = \frac{3}{2\rho_w}\frac{W}{r_e} \tag{E.20.20}$$

## *E.20.3 Energy Balance Model*

For the illustration of this model, the interested reader may go back to the original paper by Kleidon and Renner (2013), and here, we will give only some useful detail.

The atmospheric temperature is obtained as a balance between the solar radiation and the emitted infrared radiation

$$T_a = (S/\sigma)^{1/4} \tag{E.20.21}$$

To obtain the surface temperature, we assume that solar radiation is split between atmosphere and surface, so we get

$$k_r\,(T_S - T_a) = S/2 \qquad \Rightarrow \qquad T_S = T_a + S/2k_r \tag{E.20.22}$$

The evaluation of the optimum values for the variables must be found starting by linearizing all the fluxes as a function of the temperature difference $(T_S - T_a)$, and this must be done mainly for the latent heat flux

$$E = \rho w \left(q_{\text{sat}}\left(T_S\right) - q_{\text{sat}}\left(T_a\right)\right)$$

Notice that the $q$'s are the specific humidity and we may want to express them as a function of the saturation pressure. We can write

$$q_{\text{sat}}(T_S) \approx q_{\text{sat}}(T_a) + \left(\frac{dq_{\text{sat}}}{dT}\right)_{T=Ta} (T_S - Ta)$$

Then $E$ can be written as

$$E = \rho w \left(\frac{dq_{\text{sat}}}{dT}\right)_{T=Ta} (T_S - Ta)$$

However,

$$q = \frac{\rho_v}{\rho} = \frac{e}{p}\frac{M_v}{M_a} = 0.622\frac{e}{p}$$

and the derivative

$$\frac{dq}{dT} = \frac{0.622}{p}\frac{de}{dT}$$

We easily obtain for the latent heat flux expression (20.35)

$$\lambda E = c_p \rho w s (T_S - T_a)/\gamma$$

with $s$ and $\gamma$ defined in the text. At this point, we require the equilibrium

$$\lambda E + H + k_r (T_S - T_a) = (T_S - T_a)\left[c_p \rho w \left(1 + \frac{s}{\gamma}\right) + k_r\right] = S$$

And the temperature difference

$$(T_S - T_a) = S\left[c_p \rho w (1 + s/\gamma) + k_r\right]^{-1} \tag{E.20.23}$$

To obtain the maximum strength of the hydrologic cycling drive by the sensible heat, we simply calculate the power of a thermodynamic engine

$$G = H (T_S - T_a)/T_S \tag{E.20.24}$$

substituting (E.20.23) and the expression for $H$, we have

$$G = \frac{S^2}{T_S}\frac{c_p \rho w}{\left[c_p \rho w (1 + s/\gamma) + k_r\right]^2} \tag{E.20.25}$$

This expression can be maximized with respect to $w$ and can obtain the optimum values

$$w_{opt} = \frac{\gamma}{\gamma + s}\frac{k_r}{c_p\rho}; \qquad H_{opt} = \frac{\gamma}{\gamma + s}\frac{S}{2} \qquad \lambda E_{opt} = \frac{s}{\gamma + s}\frac{S}{2} \tag{E.20.26}$$

At the maximum power, the latent heat flux

$$E = \frac{s}{s + \gamma}\frac{S}{2\lambda} \tag{E.20.27}$$

that can be derived to obtain

$$\frac{1}{E}\frac{dE}{ds} = \frac{\gamma}{s + \gamma}\frac{1}{s} \tag{E.20.28}$$

The same expression (E.20.26) can be derived with respect to $S$ to obtain

$$\frac{dE}{ds} = \frac{1}{2\lambda}\frac{s}{s + \gamma} \tag{E.20.29}$$

## References[1]

### *Books*

Hester RE, Harrison RM (eds) (2014) Geoengineering of the climate system. Royal Society of Chemistry, Cambridge

Jacob D (1999) An introduction to atmospheric chemistry. Princeton University Press, Princeton

National Research Council (U.S.) (2015a) Climate intervention: reflecting sunlight to cool earth. National Academy Press, Washington, DC

National Research Council (U.S.) (2015b) Climate intervention: carbon dioxide removal and reliable sequestration. National Academy Press, Washington, DC

Royal Society (Great Britain) (2009) Geoengineering the climate: science, governance, and uncertainty. The Royal Society, London

[1]This is a new chapter and on geoengineering there is already plenty of material. However, except for popular books, there are no textbooks on this matter. There are reports that could be assimilated to books.

## *Articles*

Angel R (2006) Feasibility of cooling the Earth with a cloud of small spacecraft near the inner Lagrange point (L1). Proc Natl Acad Sci U S A 103:17184
Bala G, Duffy PB, Taylor KE (2008) Impact of geoengineering schemes on the globalhydrological cycle. Proc Natl Acad Sci U S A 105:7664
Brenguier JL (2011) Cloud optical thickness and liquid water path-does the k coefficient vary with droplet concentration. Atmos Chem Phys 11:9771
Broecker WS (2007) $CO_2$ Arithmetic. Science 315:1371
Caldeira K et al (2013) The science of geoengineering. Annu Rev Earth Planet Sci 41:231
Hansen JE (2005) Efficiency of climate forcing. J Geophys Res 110, D18104
Harshvardan, Cess RD (1976) Stratospheric Aerosols: effect upon atmospherc temperatures and global climate, Tellus, 28:1
Kleidon K, Renner M (2013) Thermodynamic limits of hydrologic cycling within the Earth system: concepts, estimates and implications. Hydrol Earth Syst Sci 17:2873
Kleidon A, Kravitz B, Renner M (2015) The hydrological sensitivity to global warming and solar geoengineering derived from thermodynamic constraints. Geophys Res Letters 42:138
Lacis AA, Hansen JE (1974) A parametrization for the absorption of solar radiation in the Earth Atmosphre. J Atmos Sci 31:118
Lacis AA et al (1992) Climate forcing by stratospheric aerosols. Geophys Res Lett 19:1607
Lenton TM, Vaughan NE (2009) The radiative forcing potential of different climate geoengineering options. Atmos Chem Phys Discuss 9:2559
McInnes CR (2010) Space-based geoengineering: challenges and requirements. J Mech Eng Sci 224:571
Rasch PJ et al (2008a) An overview of geoengineering of climate using stratospheric sulphate aerosols. Phil Trans R Soc A 366:4007–4037
Sigl M et al (2014) Insight from Antarctica on volcanic forcing during common era. Nat Clim Chang 4:493
Stocker T (2013) The closing door of climate targets. Science 339:280

**ERRATUM**

# Fundamentals of Physics and Chemistry of the Atmospheres

**Guido Visconti**


G. Visconti, *Fundamentals of Physics and Chemistry of the Atmospheres*,
DOI 10.1007/978-3-319-29449-0

---

**DOI 10.1007/978-3-319-29449-0_21**

The title of the second edition did not match that of the first edition. The title has been corrected to "Fundamentals of Physics and Chemistry of the Atmosphere".

---

The updated original online version for this book can be found at
http://dx.doi.org/10.1007/978-3-319-29449-0



G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0_21

# Index


**A**
Abel transform, 170, 178–180
Absorptance, 164, 449, 450, 458
Absorption
  direct and diffuse, 420, 421
  ozone, 40, 64–66, 146, 418–421, 423, 486, 690
  ozone, parameterization, 421
  solar radiation, 43–52, 165, 339, 399, 423, 485, 702
  spectral lines, 443, 448
  water vapour, 49, 59, 182, 417, 418, 421–424
Absorptivity
  $CO_2$, $H_2O$, $O_3$ parameterization, 421, 422, 443
Adiabatic, gradient
  dry, 6, 7, 290
  saturated, 13, 290
Aerosol
  coagulation, 276
  microphysics, 281
  nucleation, 266–276, 283
  size distribution, 263–265, 275, 276, 280, 283, 284, 290–292
  stratospheric, 11, 276, 281–286, 452, 600, 694, 695, 701, 766, 776–779
Agung, eruption, 505, 695
Airborne Arctic Stratospheric Expedition II (AASE II), 689
Albedo, 43, 58, 60, 141–142, 342–344, 418, 437, 439, 440, 456–457, 464, 466, 469,472, 477, 478, 490, 507, 509–511, 525–526, 541, 585, 766, 774–776, 787–789
Albedo, single scattering, 434–435, 440, 453, 457, 787
Alkalinity, 580–582, 586, 767
Anderson, J., 707
Andrews, D.G., 317, 336, 452, 620, 624, 668
Angle
  hour, 41–43
  scattering, 63
  zenith, 42, 43, 50, 54, 64, 418–420, 433, 435, 439, 457, 787
Angular momentum
  conservation and Hadley circulation, 341, 412
Asymmetry factor, 436, 437, 440, 723, 775
Atmospheric
  composition, 146–148, 526, 675
  properties, 168–171
Atmospheric chlorine
  scenario, 583
  sources, 574, 577, 578
Atmospheric particulate (aerosol)
  diffusion and absorption, 421, 426
  log-normal distribution, 264, 265, 284, 290
  optical properties, 425–432
  size distribution, 263–265
Atmospheric transmission, infrared, 49
Attractor
  dimension, 718, 720, 732
  strange, 720, 724
  weather and climate, 728–733




G. Visconti, *Fundamentals of Physics and Chemistry of the Atmosphere*,
DOI 10.1007/978-3-319-29449-0

**B**
Band
  intensity, 298
  parameters, 450, 458
Baroclinic, fluid, 113, 126
Baroclinic instability
  Charney problem, 387–389
  Eady problem, 383–387
  qualitative treatment, 380–383
  steering level, 387
  two levels model, 389–396
Barotropic, fluid, 88, 92, 113, 188, 216
Bayesian statistics, 542, 566
Biogeochemical cycle
  for carbon, 576–588
  for nitrogen, 577, 581
Biological productivity, 584, 586–588
Biological pump, 581
Biosphere
  numerical program for, 198
  survival, 584–588
Boltzmann
  constant, 3, 38, 50, 267, 268, 294, 772–773
  distribution, 444
Boundary layer
  planetary, 225–260
  and secondary circulation, 242–246
  spin down, 244, 246
  surface, 225, 228, 233–237, 239–241
Bowen, ratio, 236, 254
Broecker, W., 24, 578, 768
Brunt, V., frequency, 8, 205, 297, 303, 383, 384, 408, 619
Budyko, M., 463, 469, 475, 477

**C**
Carbon
  biogeochemical cycle, 576–588
  biosphere content, 148
Carbonates chemistry, 570, 578–584
Carbon dioxide
  biogeochemical cycle for carbon, 576–587
  lifetime, 571, 583
  partial pressure in the ocean, 155, 587, 588
Catalytic cycle
  bromine and chlorine, 681–683
  effects, 684–686
  nitrogen, 679–681, 689
  odd hydrogen, 677–679
Centrifugal force, 82, 94–97, 124, 137, 244
Cess, R.
  the band absorptivities, 450–451
Chaos
  climate, 720–726, 728–733, 740–745
  El Niño, 726–728, 736, 747, 748, 762
Chapman, S.
  ozone model, 155–157
Chappuis band, 64, 420
Charney, J.
  the problem, 387–389
Chemical composition
  atmospheres planets, 142, 144
  kinetics 134, 148–150
Chemistry and transport, 151–154
Chemistry in the troposphere, 569–612
Chlorofluorocarbons (CFC)
  lifetime, 571, 574, 575, 657
  numerical program for, 690
CIAP. *See* Climatic impact assessment program (CIAP)
Circulation
  absence of viscosity, 376
  Bjerknes, 87
  definition, 71–72, 85
  diabatic, 621, 630
  equatorial, 365–378
  mean zonal, 352–357
  meridional, 341, 350, 351, 355, 356, 366, 368, 372, 620
  middle latitude, 378–395
  non linear symmetric, 370–376
  residual, 621, 622, 626, 630, 633
  secondary, 242–246
  symmetric, 366–370
  theorem, 87
  transformed eulerian mean, 620–622
Clausius Clapeyron equation, 10, 12, 14, 23, 26, 286, 478
Climate
  ensemble average, 504, 549
  stability, 507
Climatic impact assessment program (CIAP), 690, 701
Climatic models
  one-dimensional, 468–483
  potential, 467, 468, 473, 476
  stability, 473–475
  zero-dimensional models, 462–468, 472, 473, 476
Climatic system
  components, 503–505, 512, 578
  external and internal influences, 504
  transitivity, intransitivity and almost-intransitivity, 507
Clouds in planetary atmospheres, 286–290

Columnar density
atmosphere, on, 673
definition, 49
Conditional convective instability, 20–21
Conservation of the angular momentum, 112, 246, 404
Conservation of the vorticity
Ertel, 192, 194
potential, 194
Continuity equation
isobaric coordinates, 115–116
Coordinate
conservative, 651–657
continuity equation on, 75–77, 205
isentropic, 184–188, 638–657
vorticity equation on, 90, 186–188
Coordinate system
Cartesian, 74, 628
isentropic, 184–188, 638, 640, 643
isobaric, 628
natural, 122–126
sigma, 533
Coriolis
acceleration, 73, 80, 82, 86, 94–97, 101, 107, 114, 143, 238, 244, 352, 624
parameter, 73, 96, 106, 109, 110, 118, 119, 130, 131, 138, 199–201, 203, 206, 211, 212, 218, 308, 367, 383, 388, 390, 409, 659
Critical level, 312, 322
Crutzen, P., 690
Cyclostrophic, equilibrium, 124, 125

**D**
Delayed differential equations
aerosol cloud, 750–753
ENSO, 747–750
Derivative
convective, 74
in isentropic coordinate, 185, 187
local, 74, 89
total, 74–75, 94, 154, 182, 183, 185, 205, 315, 328, 347, 533
Dew point temperature, 15, 17, 19
Diabatic, warming, 182
DIC. *See* Dissolved inorganic carbon (DIC)
Difference equations, 222, 391, 492, 507, 534, 551, 558, 665, 741–745, 759
Diffusion
in the atmosphere, scattering, 43
climate models, 470, 477, 527–529, 558
eddy coefficient, 158, 250
Mie theory, 425
Rayleigh, 423
solar radiation, 43
stationary functions, 474
turbulent, 228, 246–251, 316
Diffusion coefficient, 158, 226, 228, 231, 232, 236, 241, 250, 270, 271, 277, 321, 401, 469, 471, 473, 484, 497, 498, 527, 533, 627–630, 632, 633, 639–644, 664
Diffusion equation, 160, 226, 255, 277, 321, 628
Dissolved inorganic carbon (DIC), 580–582
DNA and ultraviolet radiation, 165, 504, 602, 671, 703
Dobson unit, 419, 705
Downward control, 635, 637, 640, 641
Dynamical systems, 740
Dynamic forcing, 631

**E**
Eady, E. T.
problem, 383–387
Earth
atmosphere composition, 142, 144, 146, 148, 504
temperature, 22, 464, 469
volatile elements, 523
Earth radiation budget experiment (ERBE)
and greenhouse effect, 487–489
Eccentricity of the orbit, variations, 508, 510
Eddies
stationary, 350
transient, 350, 351
turbulence, 229
Eddington, approximation, 437–440, 451, 456, 499, 775
δ Eddington, approximation
application, 438–440
in the infrared, 451–452
Efficiency
absorption, 426
extinction, 426, 427, 453
reaction, 689
scattering, 176, 177
Ekman
layer, 237–243, 253
spiral, 238, 241, 252
El Chichon, eruption, 695
Electric field
incident, 44, 62, 429
Electronic transition
rotational, 443
vibrational, 443

Eliassen-Palm, flux, 312–315, 326, 619, 621, 631
El Niño, 726–728, 736, 747, 748
Emagram, 16, 18–20, 29, 33, 34
Emissivity, 52, 56, 68, 403, 407, 449, 451, 483, 487
Energetic of waves
  baroclinic, 395–402
  gravity, 315–316
Energy
  available potential, 34, 395, 397
  conservation, 51, 226, 311, 345
  forms of, 346–348
  internal, 5, 346
  kinetic, 3, 102, 232, 310, 311, 318, 324, 327, 346, 347, 351, 362, 380, 387, 396, 399–402
  potential, 3, 34, 101, 102, 318, 324, 325, 346, 347, 351, 395–397, 399, 400, 403, 408, 496
  rotational, 443
  total potential energy (TPE), 346, 396, 397
  in two levels model, 398–400
  vibrational, 443
Energy-balance, models, 462, 468–483, 488, 491, 497–501, 510, 515, 779, 790–792
Energy budget
  atmospheric, 342–352
  earth-atmosphere, 343, 344, 350, 462
  zonal, 342, 343, 347, 350
Equal area requirement, 31–32
Equilibrium
  cyclostrophic, 124, 125
  hydrostatic, equation, 4, 7, 22, 59, 107, 113, 116, 119, 171, 185, 192, 202, 204, 308, 309, 318, 325, 330, 366, 422, 531, 533, 618
  photochemical, 594, 604, 633, 679, 698, 699, 703
Equinox
  definition, 40
Equivalent density, 184, 638
Ertel vorticity, 190, 192, 640, 652–654
Eutectics, 10
Exit time 421, 468

**F**

Farman, J., 690
Feedback
  factor, 525, 526, 539–541, 544, 558
  ice-albedo, 462, 464, 483, 490, 511, 523, 526, 585
  mechanisms, 402, 462, 464, 483, 486, 523, 525, 541
  water vapour, 59, 483, 487–491, 526, 541, 585
Flux
  diffusive in ocean, 497, 527
  geopotential mean zonal, 347, 348, 351
  latent heat, 59, 236, 253, 256, 257, 351, 468, 469, 782, 783, 790–792
  over an obstacle, 222–223
  in presence of vegetation, 255–258
  sensible heat (SH), 236, 256, 348, 351, 469, 476, 478, 498, 782, 783
  zonal, 232, 313, 314, 348, 355, 358, 619, 624, 625, 722, 724
Fractal dimension, 717, 719–720
Friction
  force, 78–80, 82, 91, 105, 130, 239
  Rayleigh, 366, 631
  turbulent, 229–233
Friction velocity, 235, 237

**G**

Garcia, R., 269, 630, 631, 668, 686
GCM. *See* General circulation model (GCM)
General circulation model (GCM)
  equations, 540
  global warming, 503
  physical processes, 531
General circulation models
  Bayesian, 567
  performances, 535–549
  Taylor diagrams, 535–539
Geoengineering
  energy balance models, 790–792
  extinction, 789
  ozone hole, 784–786
  space shields, 779–781
  technologies, 766–767
Geometrical cross section
  extinction, 177, 432, 452
Geopotential
  definition, 6
  height, 4, 5, 194, 413, 414
Geostrophic
  approximation, 105, 106, 109, 116, 207, 213
  degeneracy, 111
  stream function, 341
Geostrophic motion
  definition, 105
Geostrophy, quasi, 111, 202–206, 215, 307, 308, 311–312, 325, 383, 389, 636

Gibbs, free energy, 267
Glaciers
  ablation, 515
  accumulation, 515
  limit cycle, 518
Global warming, 25, 72, 503, 523–531, 777
Gravity waves
  absorption, 322
  breaking, 317–322
  diffusion coefficient, 321
  energetics, 315–316
  internal, 304–307
  orographic, 302–304
  physics, 311–316
  saturation, 317–322
  shallow water, 220, 300–302
  transport, 307, 315, 324, 327
Greenhouse effects
  runaway, 59–61
Group velocity, 211, 297–299, 303, 304, 306, 307, 315, 316, 319, 321, 324, 328, 333, 358, 359
Growth rate
  baroclinic instability, 387

**H**
Hadley
  cell, 340, 341, 351, 356, 365, 373, 374, 377, 378, 403, 404, 406–408, 412, 413
  circulation, 357, 378, 395, 399, 409–413
Halley Bay, 705
Hansen, J., 420, 439, 440, 459, 485, 775, 776
Hartley band, 64
HCFC. *See* Hydrochlorofluoromethanes (HCFC)
Heterogeneous chemistry, 12, 23, 671, 681, 686–689, 694, 695, 697–699, 701, 703–704, 779
Hoffman, D., 695
Holton, J., 103, 190, 201, 223, 336, 626, 657, 658, 664, 665, 668
Huggins band, 64, 65
Hydrocarbon and urban air, 593
Hydrochlorofluoromethanes (HCFC), 691
Hydroxyl (OH radical), 589

**I**
Ice ages
  $CO_2$ content, 504
  lithosphere-atmosphere coupling, 513–519
  oscillation of the climate systems, 520
  stochastic resonance, 513, 519–524
  temperature and $CO_2$ correlations, 506, 513
  theories, 519
Ice latitude, 464, 469, 472, 473, 490
Inertial
  instability, 120–122, 405–406
  oscillation, 96–98
Infrared radiation
  absorbance, 53, 55, 58, 61
  transport, 441–443
Intergovernmental Panel for Climatic Change (IPCC), 523, 530, 539
IPCC. *See* Intergovernmental Panel for Climatic Change (IPCC)
Isallobaric, wind, 111–112
Isallobars, 111
Isotopic, ratio for O, 512

**J**
Jetstream, 626

**K**
Karman, von
  constant, 235, 237
Kasting, J., 584–586
Kelvin waves, 660, 664, 747, 748

**L**
Lacis, A., 42, 439, 440, 459, 485, 775, 777
Laminar layer, 226
Latent heat flux, 59, 236, 253, 256, 257, 468, 469, 782, 783, 790–792
Legendre, polynomials
  albedo expansion, 464, 478
  in North's model, 469–473, 478
Lidar, 161–163, 175–177, 281
Lifetime
  definition, 150
  ozone, 673, 686
Limb infrared monitor of the stratosphere (LIMS), 694
Limb measurement, 167
LIMS. *See* Limb infrared monitor of the stratosphere (LIMS)
Log-pressure, coordinate system, 204, 311, 315, 325, 627
Longitude of the perihelion, variation, 509
Loop oscillator, 733–736, 754
Lorentz, E.
  system, 445, 446
Los Angeles atmosphere, 593
Lyapunov exponent, 759–761

**M**
Margules formula, 118–120
Marine chemistry
  $CO_2$ concentration, 578–584
Mars
  atmosphere, 72, 141–144, 287
  composition, 143, 144
  dust storms, 142, 407
  radiative constant, 60
  temperature, 14, 60
McIntyre, M., 620, 635, 668
Mean
  deviation, 232, 292, 348, 353, 381, 536
  temporal, 350
  transformed eulerian mean, 620–622
  zonal, 350, 352–357, 361, 625, 663, 722
Methane (CH4)
  lifetime, 571
  oxidation, 572
Mie, G.
  scattering, 417, 425, 432–433
Minor gas inventory, 570–576
Mixing length, 231–233, 235, 239, 400
Mixing ratio, 14–17, 19, 20, 22–25, 28, 30, 31, 33, 35, 50, 59, 130, 143–145, (BAL)
Molecular spectra, 443–446
Molina, M., 690, 711
Montreal, protocol, 690
Motion equation
  primitive, 533

**N**
NAT. *See* Nitric acid trihydrate (NAT)
Newtonian cooling, 371, 410, 616, 622, 631
Nitric acid ($HNO_3$), 147, 151, 153, 283–286, 590, 592, 596, 599, 671–672, 676, 678–681, 685, 688–689, 694–699, 701, 704, 708–709, 711–714, 784, 785
Nitric acid trihydrate (NAT), 284–286, 688, 711–712
Nitrogen compound, partition
  chlorine, partition, 675, 676, 680, 685, 689, 690
Nitrogen oxide (NOx)
  sources, 144, 572
Noise
  climatic, 523
  white, 480, 520–522, 730–731, 733
North, G.
  model, 469–473, 476–479, 491–495, 497
Nucleation
  heterogeneous, 266, 268, 283
  heteromolecular, 266, 283
  homogeneous, 268, 276, 283
  homomolecular, 266

**O**
Ocean, thermal capacity, 479, 529
Odd oxygen ($O_x$), 156, 157, 594, 612, 675, 677, 683, 684, 703
Oerlemans, J., 515
Oeschger, H., 513
Optical thickness, 47–50, 53–57, 59–61, 64, 163, 168, 417, 435–440, 442, 443, 448–449, 451, 456–457, 499, 775–777, 788–789
Orbital parameters of the earth
  and solar radiation, 142, 508–511
Orbital variation, 142, 505
Ozone
  catalytic cycles, 677, 678, 681, 682
  destruction rate, 677, 684, 700, 704, 712, 713
  global ozone trend, 691, 694
  hole, 23, 52, 53, 71, 155–156, 310, 569, 657, 677, 686, 690, 693, 695, 700, 706, 709, 784–786
  homogeneous chemistry, 674–686, 700
  lifetime, 672, 673, 686
  net chemical production, 682
  perturbations to the ozone layer, 689–704
  polar, 705–709
Ozone measurement from satellite, 164–168, 691–692

**P**
Phase function in scattering, 46–47, 432–433, 437–438
Phase velocity, 99, 133, 196, 211, 220, 297, 299, 302–304, 320, 322–325, 357, 380, 384, 392–394, 633, 660–662
Photodissociation
  coefficient, 149, 156, 595, 596, 649, 697
Photolysis, 149, 150, 153, 157–158, 676
Pinatubo, eruption, 281, 282, 505, 689, 695, 700
Planets
  chemical composition, 142–145, 286–290, 569
  clouds on, 288, 776
Poincaré, H.
  section, 717–720, 758
Polarizability, 44–46, 62–64, 174, 177–178, 427–430, 433

Polar stratospheric clouds (PSC)
one and two, 284
size distribution, 284
Polar vortex, 625, 638, 656, 705, 707–709
Potential temperature
equivalent, 16, 19, 26–28, 31, 33
Prandtl
laminar boundary layer, 226
Precipitable water, 421, 422
PSC. *See* Polar stratospheric clouds (PSC)

**Q**
QBO. *See* Quasi-biennal oscillation (QBO)
Quasi-biennal oscillation (QBO), 659, 663–665, 667–668, 692

**R**
Radar 156, 161–163, 171–175, 281
Radiation
flux, 505
infrared, 37, 39, 52–61, 67, 68, 116, 145, 146, 342, 343, 417, 441–443, 462, 477, 483, 486, 487, 498, 790
intensity, 37–38, 47, 53, 63, 163, 167–168, 435, 442, 787
solar, 37, 39–53, 67, 68, 116, 142, 145, 146, 162, 165, 257, 339, 342–343, 366, 399, 417, 419–420, 422–433, 440, 461, 462, 464, 472, 485, 486, 504–512, 517–518, 523, 525–526, 530, 541, 585, 616, 674–675, 686, 702, 765, 766, 774, 776, 779, 781–784, 790
Radiative
equilibrium, 25, 55–61, 146, 461, 483–485, 615–618, 621–623, 625, 637, 642
forcing, 25, 67, 526, 527, 530, 562, 702, 767–768, 774, 776, 777, 782
Radiative convective
atmosphere, 57–59
Radiative-convective models
$CO_2$ increase, 487–488
convective adjustment, 487
energy balance, 462, 491
greenhouse effect, 487–490
heating rate, 484–487
ice-albedo feedback, 483, 491
net flux, 484, 485
Radiative, time constant, 66–67, 404
Radiative transfer
equation of, 53–57
in presence of scattering, 433–440
Radical
alkyl, 590
methyl, 591
OH, 589
and tropospheric chemistry, 589, 671
Radiometric, variables definition, 37–39
Radio occultation (RO), 162, 168–171
Radius parameter, 452
Rainbow, 424
Ramanathan, absorbance, 450–451
Rate constant, 148–150, 153, 702
Rayleigh
friction, 631
natural light, 46, 62–64
scattering, 44–48, 62–64, 167, 168, 417, 418, 426, 428, 432, 436
Reactive chemical species
reservoirs, 675, 676, 698
stratospheric, 615
Redfield, ocean, 581
Reference system, equation of motion
inertial, 73, 81
rotating, 244
Reflectivity
of the atmosphere, 418, 423, 438
total, 418
Refractive, index, 162, 170–171, 173, 177–178, 426, 430, 431, 452
Remote sensing, 161–180
Reynolds, number of, 226–228, 272, 293, 294
RO. *See* Radio occultation (RO)
Rossby
dispersion relation, 133, 211, 661
number, 72, 73, 93, 105, 108, 112, 133, 142, 143, 207, 212, 219, 221, 397, 405, 407, 408
waves, 130–134, 190–192, 194–196, 209–211, 214, 219–221, 297, 307–312, 357, 359, 388, 747, 748
Rowland, S., 686, 690

**S**
SAT. *See* Sulphuric acid tetrahydrate (SAT)
Saturation, ratio, 14, 267–270, 276, 292, 293
Scale height, 4–5, 22, 23, 57, 114, 158, 159, 204, 289, 290, 381, 397, 406, 412, 618
Scattering, angle, 63, 433, 435
Sea breeze, 126–127, 135–136
Sellers, W.D.
fluxes, 468, 479
model, 477–479, 483
Sensible heat
flux, 236, 348, 350, 351, 401, 461, 468, 477, 478, 484, 485, 498, 531, 782, 783
zonal flux, 348

Shallow water vorticity
Rossby waves, 219–221
Solar constant, 39, 60, 66, 141, 142, 463, 465, 466, 471–473, 479, 504, 505, 508, 587, 588, 720, 742, 774, 779–781
Solomon, S., 629, 686, 695
SPADE. *See* Stratospheric Photochemistry Aerosol and Dynamics Experiment (SPADE)
Spectral line shape
Doppler, 445–447
Lorentzian, 445–447
Spherical harmonics, 534
Spin down in a tea cup, 244–246
Static stability, 7–8, 20, 182, 390, 402, 408, 627
Stratosphere
eddy flux, 353, 355–357, 360, 620, 627, 633, 641
radiative equilibrium, 25, 55–61, 146, 461, 483, 485, 615–618, 621–623, 625, 637, 642
thermal structure, 51, 72, 122, 162, 164, 233, 616–618
troposphere-stratosphere exchange, 67–659
zonal wind, 117–118, 139, 184, 230, 310, 315, 319, 320, 322, 341–342, 353, 354, 356, 368, 370, 371, 375, 377, 407, 408, 617, 618, 620, 624, 631, 640, 656, 683, 722, 723
Stratospheric Photochemistry Aerosol and Dynamics Experiment (SPADE), 689
Stream function
geostrophic, 109–111
meridional, 341
Montgomery, 185, 641
stochastic resonance, 513, 519–524
Sudden stratospheric warming, 310, 623–625
Sulphur cycle
and acid rain, 599–600
Sulphuric Acid
clouds, 711
solution, 11
Sulphuric acid tetrahydrate (SAT), 284–286
Surface density, 204, 284, 687, 696–700, 704
Sverdrup, unit, 582

**T**
Taylor Proudman theorem, 113, 115
Themohaline circulation
stability, 740–741
Stommel model, 736–739, 741
Thermal emission measurement, 163–164
Thermal wind, 93, 112–118
Thermodynamic
atmospheric, 1–35
equation, 181–183, 305, 315, 319, 325, 366, 371, 383, 389, 391, 408–409, 533, 619
water vapour, 8–13
Tracers
slope, 642–645
transport in the stratosphere, 626–637
Trade winds, 340–342, 350, 368, 726
Transport
chemical compounds, 626
and diffusion coefficient, 401
in isentropic coordinates, 184–188, 638–656
parameterization, 400–402
tracers in the stratosphere, 626–637
Two levels model, 534
Twomey effect, 775, 788
Two stream, approximation, 54, 452

**U**
Upward control, 635
Urban air, chemistry, 593–595

**V**
Van de Hulst, H.C., 431
Venus
atmosphere, 67, 72, 125, 143, 144, 287
composition, 143, 144, 282
cyclostrophic equilibrium, 124, 125
greenhouse effect, 52, 53, 59, 67–68, 463, 464, 481, 487–490, 506, 513, 526, 541, 585, 781
Hadley cell, 340, 341, 351, 356, 373, 374, 377, 378, 403, 407–408, 412, 413
radiative constant, 60
temperature, 124, 145
volatile elements, 523
Vibrational motion of a molecule, 442
Virtual temperature, 32–33
Viscosity
cinematic, 80, 226, 229, 244
dynamic, 78, 272, 293
Volatile elements, 523
Volcanic eruptions and climate, 262, 671, 694–700, 766, 776
Vortex lines, 88, 92
Vorticity
absolute, 83, 90, 92, 93, 130, 134, 187, 190–192, 208, 216, 297, 634
application, 83, 126–134, 190–192

and circulation, 83–94
definition, 84, 85, 125, 132, 186, 187, 203, 206, 212
equation, 88–94
Ertel potential, 188–196
geostrophic, 112, 121, 202–206, 307, 308, 325, 389
horizontal divergence, 93, 100, 112, 307, 634
interpretation, 252, 507, 542, 619, 630, 653
inversion, 206–211, 218
modified potential, 652, 655
potential, 99–101, 186, 190–194, 198, 199, 203, 206, 208–210, 218, 220, 222, 309, 311, 313, 314, 325, 383, 389, 625, 632, 633, 640–642, 651–655
quasi geostrophic, 204–206
Vorticity equation
and gravity waves, 317
in isentropic coordinates, 652

**W**
Wallace, J.M., 17, 35, 279, 352, 379, 394, 399, 424, 425
Water vapour
distribution planetary atmospheres, 22–24
effective quantity, 13–21
thermodynamics effects, 8–13
Wave action, 313–316, 319
Waves
atmospheric, 302
baroclinic, 385, 387
dispersion relation, 133, 211, 303, 305, 306, 310, 311, 333, 358, 661, 662
orographic, 302–304, 306
over an obstacle, 196, 222–223
properties, 297–299
Rossby, 130–134, 190, 192, 194–196, 209–211, 214, 219–221, 297, 307–312, 357, 359, 389, 747, 748
surface, 304
three dimensional Rossby waves, 307–312
Weathering, rate, 587–588
Wet bulb temperature, 15, 17, 19

**Y**
Yamamoto, G., 421, 422