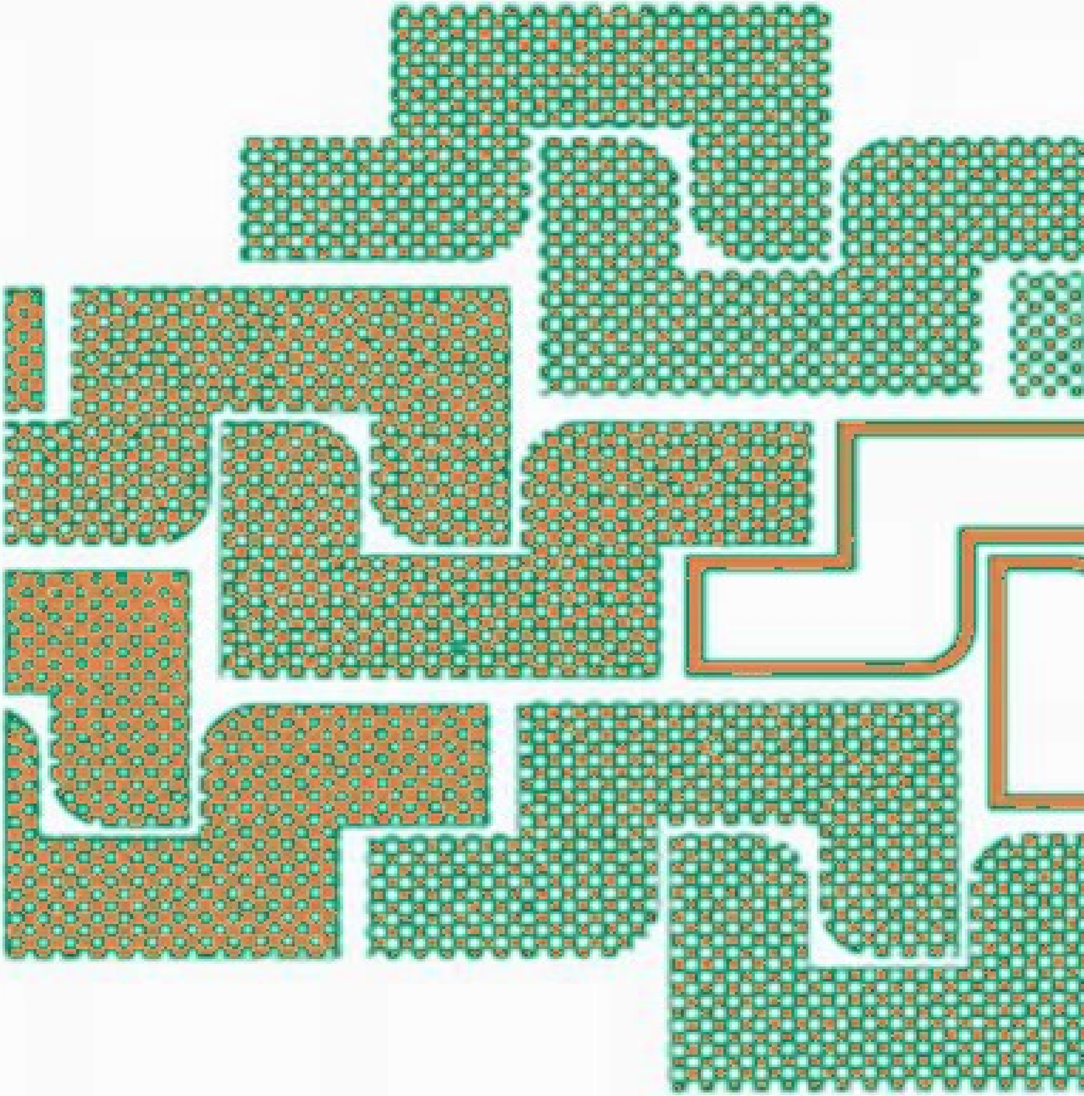

Seasonal Climate: Forecasting and Managing Risk

NATO Science Series

A Series presenting the results of scientific meetings supported under the NATO Science Programme.

The Series is published by IOS Press, Amsterdam, and Springer (formerly Kluwer Academic Publishers) in conjunction with the NATO Public Diplomacy Division.

Sub-Series

I. Life and Behavioural Sciences	IOS Press
II. Mathematics, Physics and Chemistry	Springer (formerly Kluwer Academic Publishers)
III. Computer and Systems Science	IOS Press
IV. Earth and Environmental Sciences	Springer (formerly Kluwer Academic Publishers)

The NATO Science Series continues the series of books published formerly as the NATO ASI Series.

The NATO Science Programme offers support for collaboration in civil science between scientists of countries of the Euro-Atlantic Partnership Council. The types of scientific meeting generally supported are “Advanced Study Institutes” and “Advanced Research Workshops”, and the NATO Science Series collects together the results of these meetings. The meetings are co-organized by scientists from NATO countries and scientists from NATO’s Partner countries – countries of the CIS and Central and Eastern Europe.

Advanced Study Institutes are high-level tutorial courses offering in-depth study of latest advances in a field.
Advanced Research Workshops are expert meetings aimed at critical assessment of a field, and identification of directions for future action.

As a consequence of the restructuring of the NATO Science Programme in 1999, the NATO Science Series was re-organized to the Four Sub-series noted above. Please consult the following web sites for information on previous volumes published in the Series.

http://www.nato.int/science
http://www.springeronline.com
http://www.iospress.nl

Series IV: Earth and Environmental Sciences – Vol. 82

Seasonal Climate: Forecasting and Managing Risk

edited by

Alberto Troccoli

European Centre for Medium Range Weather Forecasts
Reading, UK

Mike Harrison

Independent Consultant
Wolverton Common, UK

David L.T. Anderson

European Centre for Medium Range Weather Forecasts
Reading, UK

and

Simon J. Mason

International Research Institute for Climate and Society
The Earth Institute at Columbia University, NY, USA

Springer

Published in cooperation with NATO Public Diplomacy Division

Book derived from the NATO Advanced Study Institute on
Seasonal to Interannual Climate Variability: its Prediction and Impact on Society
Gallipoli, Italy
23 May–3 June 2005

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-6991-8 (PB)
ISBN 978-1-4020-6990-1 (HB)
ISBN 978-1-4020-6992-5 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

Contents

Preface

Originally formed around a set of lectures presented at a NATO Advanced Study Institute (ASI), this book has grown since then and it has been organised and presented more like a textbook than the standard "collection of proceedings". The lack of a unified reference textbook in *seasonal to interannual climate predictions* that covers both the science of the predictions and the real-world uses of the forecasts was the main driver for the considerable effort placed into producing an amalgamated introductory book. Throughout, our objective has been to present a textbook for people of many disciplines interested in this fascinating and fast emerging sector. An additional novelty for a NATO ASI series book is that all the chapters have been thoroughly peer reviewed: each chapter has received the attention of three or more experts. We believe this reviewing process has considerably raised the level of the book and the extra time (and pain) needed to complete the oeuvre has been entirely justified.

The book is targeted at the intelligent reader at postgraduate level, but who does not need to be an expert in all the fields discussed. The reader may well be coming from only one of the many disciplines that contribute to the fields of seasonal climate forecasting and risk management: this book aims to provide him/her with a general overview of all the major issues related to these fields. A summary at the beginning of each chapter, except for the first, will help all readers select only those chapters that are relevant or of interest to them while still being able to grasp the essentials of every chapter.

The fascination of seasonal climate forecasting, of which El Niño forecasting is the prime example, comes from its multi-faceted character. Not only does it pose interesting new challenges for the climate scientific community but also it is naturally linked to a great variety of practical applications, from security related issues, such as water resource management, food security, and disaster forecasts and prevention, to health planning, agriculture management, energy supply and tourism, to name but a few. Seasonal to interannual climate forecasts are indeed becoming a most important element in some policy/decision making systems, especially within the context of climate change adaptation. Seriously considering the management of risks posed by climate variability and of development in general on the seasonal to interannual scale is key to achieving the longer terms goals of climate change adaptation strategy.

The NATO ASI *Seasonal to Interannual Climate Variability: its Prediction and Impact on Society* was held in the beautiful setting of Gallipoli (Italy) between 23 May and 3 June 2005. This "summer school" attracted applications from a large number of postgraduate students and professionals. Unfortunately places were limited but 62 participants from 27 countries could be accommodated.

It would have not been possible to organise this ASI without the collaboration and support of many people: the team at the NATO Environmental and Earth Science & Technology (EST) Programme with *Mrs. Lynne Nolan* (Secretary) and *Dr. Alain Jubier* first and *Dr. Deniz Beten* later (Programme Directors), who assisted in securing a smooth development of the ASI; *Mrs. Elena Bertocco* (ASI Secretary) assisted with the copious queries from participants, herself assisted by little Edward; the members of the Organising Committee (i.e. the editors of this book plus *Mr. Omar Baddour*, Direction de la Météorologie Nationale of Morocco and World Meteorological Organization, WMO); *Mr. Rob Hine* (European Centre for Medium Range Weather Forecasts, ECMWF, graphic creator) for producing high quality promotional material; *Mr. Nando Micaletto* (ECMWF, technical & local expert) for ensuring the smooth running of the ASI; *Ing Antonio Rizzo* and *Dr. Antonio Tommasi* (Province of Lecce) for the supremely well planned, varied and thoroughly enjoyable social and cultural programme; *Mrs. Annamaria Caputo*, *Mr. Renato Renna* and all the staff at the Ecoresort Le Sirené (Gallipoli) for the warm and professional hospitality.

We are particularly grateful to the various organisations that supported this ASI and the preparation of the book financially: NATO *in primis*, National Oceanic and Atmospheric Administration Office of Global Programs (NOAA OGP), ECMWF, World Meteorological Organization (WMO), the US National Science Foundation (NSF) and the Province of Lecce. In addition, Troccoli was partly supported by the European Union projects ENACT (EVK2-2001-00077) and MERSEA (AIP3-CT-2003-502885) and Mason's contribution was funded by Cooperative Agreement AN07GP0213 from the National Oceanic and Atmospheric Administration (NOAA) and supported by a grant from the NCAR CSL program to the IRI.

It has been a privilege to have so many worldwide experts in the field of seasonal to interannual climate predictions as lecturers at the ASI and as contributors to this book: their contribution made the ASI particularly illuminating and challenging. Likewise, we were fortunate to have so many talented participants who actively and enthusiastically participated in the ASI[1]. Their keen involvement made the school a very stimulating and educational experience for us all. The location, a few metres from the beach, along with the many social and cultural activities no doubt also helped to form an amalgamated group.

[1] For detailed information on the ASI, see: http://www.ecmwf.int/staff/alberto_troccoli/nato_asi/asi_programme/index.html

We would like to thank very much the numerous reviewers who dedicated their time to considerably improving this book: Oscar Alves, Christof Appenzeller, Walter Baethgen, Tony Barnston, Rasmus Benestad, Pierre Bessemoulin, Čedo Branković, Barbara Brown, Dick Dee, Michel Deque, Dave DeWitt, Normand Gagnon, Brad Garanganga, Lisa Goddard, Xiaofeng Gong, Renate Hagedorn, Jim Hansen, Peter Hayman, Jaakko Helminen, Ian Jolliffe, Thomas Jung, Slava Kharin, Ben Kirtman, Willem Landman, Andrew Lorenc, Sabine Marx, Glenn McGregor, Holger Meinke, Saji Njarackalazhikam Hameed, Warwick Norton, Laban Ogallo, Tomoaki Ose, Anders Persson, Michele Rienecker, John Roads, Sandra Robles-Gil, Tim Stockdale, Rowan Sutton, Madeleine Thomson, Coleen Vogel, Richard Washington, Dan Wilks, Toshio Yamagata.

A special thank you to Rob Hine who helped enormously in preparing this book by processing most of the figures and assisting in the final editing. We are grateful also to Anabel Bowen for processing several of the remaining figures. Thank you also to Els Kooij-Connally and Verusca Bertocco for their valuable assistance in editing the book.

Lastly, it should be appreciated that there have been many difficulties in producing such a multi-authored "textbook", hence some gaps and jumps are unavoidable and we hope you will take this into consideration when reading the book. Despite what we like to think are minor drawbacks, we believe this book will provide a very useful reference for all those who would like to venture into the world of climate variability, its prediction and its adaptation strategies. Enjoy reading this book!

Alberto Troccoli
Mike Harrison
David L.T. Anderson
Simon J. Mason

September 2007

Contributors

Dr Yahya Abawi
Department of Natural Resources and Water, 203 Tor Street (PO Box 318), Toowoomba QLD 4350, Australia, E-mail: yahya.abawi@climatechange.qld.gov.au

Prof. David L. T. Anderson
European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom, E-mail: d.anderson@btinternet.com

Mr Omar Baddour
World Climate Data and Monitoring Program, World Meteorological Organization, 7 bis, avenue de la Paix, CH-1211 Geneve 2, Switzerland, E-mail: Obaddour @wmo.int

Dr John Bellow
Florida State University, c/o 3325 Fender Dr., Unit B, Lynnwood, WA 98087, United States of America, E-mail: bellow@comcast.net

Dr Michael Coughlan
Bureau of Meteorology, National Climate Centre, PO Box 1289K, Melbourne, Victoria, 3001, Australia, E-mail: M.Coughlan@bom.gov.au

Dr Mike Harrison
Independent Consultant, Glenmore, Wolverton Common, Tadley Hants RG26 5RY, United Kingdom, E-mail: mikeharrison26@btinternet.com

Prof. Sir Brian Hoskins
Department of Meteorology, University of Reading, PO Box 243, Earley Gate, Reading RG6 6BB, United Kingdom, E-mail: b.j.hoskins@reading.ac.uk

Mr Paul Llanso
ex-Chief, World Climate Applications and Services Programme, WMO, 56 Cedar Summit Rd., Asheville, NC 28803, United States of America, E-mail: pdllanso @charter.net

Dr Simon J. Mason
International Research Institute for Climate and Society, PO Box 1000, Palisades, NY 10964–8000, United States of America, E-mail: simon@iri.columbia.edu

Mr Abdalah Mokssit
Direction de la Météorologie Nationale, Face Préfecture Hay Hassani, B.P 8106 Casablanca-Oasis, Casablanca, Morocco, E-mail: mokssit.abdalah@yahoo.fr

Dr Antonio Navarra
Istituto Nazionale di Geofisica e Vulcanologia, Centro Euro-Mediterraneo per i Cambiamenti Climatici, Viale Aldo Moro, 44, 40127 Bologna, Italy, E-mail: antonio.navarra@cmcc.it

Prof. Jim J. O'Brien
Florida State University, 2035 E. Paul Dirac Drive, R.M. Johnson Bldg., Suite 200, Tallahassee, FL 32310, United States of America, E-mail: jim.obrien @coaps.fsu.edu

Prof. Paul S. Schopf
Department of Climate Dynamics, College of Science, George Mason University, Fairfax, VA 22030, United States of America, E-mail: pschopf@gmu.edu

Mr Rachid Sebbari
Direction de la Météorologie Nationale, Face Préfecture Hay Hassani, B.P 8106 Casablanca-Oasis, Casablanca, Morocco, E-mail: sebbari@marocmeteo.ma

Prof. David B. Stephenson
University of Exeter, North Park Road, Exeter EX4 4QF, United Kingdom, E-mail: d.b.stephenson@exeter.ac.uk

Dr Joe Tribbia
National Center for Atmospheric Research, 1850 Table Mesa Drive, Boulder, CO 80307, United States of America, E-mail: tribbia@ucar.edu

Dr Alberto Troccoli
European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom, E-mail: a.troccoli@ecmwf.int

Dr Jim B. Williams
Consultant with the NRGroup, Mount Ararat, Cave Hill, Maidstone ME15 6DX, Kent, United Kingdom, E-mail: mountararat@btinternet.com

List of Participants

Nina Aleksovska	Pascal Mailier
Vesselin Alexandrov	Constantin Mares
Maura Amici	Ileana Mares
Dmytro Basharin	Cheikh Mbodj
John-Glenn Bellow	Paul-Antoine Michelangeli
Anca-Marina Brookshaw	Ernesto Munoz
Robert Burgman	Warwick Norton
Natalia Calvo	Mirta Patarcic
Andrea Carril	Kathy Pegion
Jessie Cherry	Malaquias Pena-Mendez
Michelle Colley	Sjoukje Philip
Katie Coughlin	Anna Pirani
Mohamed Dawod	Oleg Pokrovsky
Jens Debernard	Antonio-David Pozo-Vazquez
Elia Diez	Cristina Primo-Ramos
Juan-Sebastian Fontecilla	Christel Prudhomme
Maria-Dolores Frias	Angelo Riccio
Alexandre Gagnon	Irina Rudeva
Yoshie Goto	Reija Ruuhela
Annette Guldberg	Serhat Sensoy
Murat Gunduz	Laila Shahin
Mosaad Hassanein	Gregor Skok
Richard Hewston	Lidija Srnec
Antoni Jordi	Gintautas Stankunavicius
Alicia Karspeck	Matthew Swann

Valentina Khan	Qiaobin Teng
Morten-Ødegaard Køltzow	Caroline Ummenhofer
Marina Kordzakhia	Liliana-Florentina Velea
Narasimhan Larkin	Stefania Vergari
Lea-Marjatta Leskinen	Andreas-Philipp Weigel
Paulo-Sérgio Lucio	Xiuhua Zhu

Part I
Seasonal Climate Forecasts in Context

Chapter 1
Introduction

Mike Harrison, Alberto Troccoli, David L.T. Anderson, and Simon J. Mason

Humanity recognised millennia ago the importance of climate variability to the sustenance of life, whether that variability was expressed in the form of droughts, floods, heat, cold, or wind. Coping strategies, developed to handle the consequences of climate variability, helped ensure mankind's survival, although the historic record indicates that not all societies successfully overcame past challenges imposed by long-term droughts, extensive flooding, and the like. Early coping strategies included migration, invasion, appropriation and storage. In addition many, probably most, perhaps all, societies developed indigenous knowledge or belief systems that they felt enabled them to foresee or control those elements of the climate that are so critical for maintaining water and food supplies.

Much has changed for modern societies, with coping strategies such as migration, invasion and appropriation frequently constrained by international boundaries and laws. Indigenous knowledge still plays a major role in many societies, while new structures, often under the umbrellas of the United Nations or national Aid Agencies and Non-Governmental Organizations (NGOs), provide safety nets for those countries currently unable to manage the consequences of climate variability without support. In the developed world, numerous technological advances, including new crop cultivars, integrated approaches to water management, improved drugs and disease control methods, such as for malaria, have introduced major new components in the management of climate risks, although not to the extent that any country has become fully shielded. Nevertheless climate variability in the developed world is more often an irritant than a hazard to life; in fact at times it is viewed as a business opportunity. In many countries, however, climate variability may still threaten life, and, if not, might at the least pose difficult challenges in regards to economic development, individual climate events occasionally resulting

---

Mike Harrison
Independent Consultant

Alberto Troccoli and David L.T. Anderson
European Centre for Medium Range Weather Forecasts

Simon J. Mason
International Research Institute for Climate and Society

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

in economic consequences of magnitudes comparable to individual countries' Gross Domestic Products (GDPs), with several years of re-development often necessary in such instances.

Included amongst the technological advances that have led to increased resilience against climate variability are remarkable achievements in the understanding, monitoring and prediction of climate variability itself, in tandem with developments that significantly aid planning and management, including improved cultivars and cropping methods, new water storage and distribution methodologies, facilitation of international food transportation and storage, and so on. Technology has become an important instrument in protecting against, mitigating, planning for, as well as in the direct management of climate risks, and will continue to be so in the light of future natural and anthropogenically forced climate change. It has been suggested that while management of the risks of climate variability might be managed with current technology, and while these technologies themselves will make substantial contributions to preparations for climate change, new technologies will be required for the full future management of climate variability under a changed climate. At this time, however, for many countries the more immediate challenge is to manage current climate risks both as one key input to sustainable development and as a significant contribution to preparations for a future modified climate.

Within this book we will be focusing on one of the new technologies emerging in the search for improved management of the risks associated with climate variability, namely seasonal to interannual prediction. Prediction, used as one input to preparing for and managing the risks of climate variability, is in itself not a new concept; indigenous methods, normally based on the behaviour of local flora and/or fauna, and/or on belief systems, have flourished around the world and have provided societies with foresights over numerous centuries. Modern systems of prediction, whether based on straightforward empirical links between climate and certain slowly varying aspects of the geosystem, more often than not sea surface temperatures in tropical ocean basins, or on advanced numerical, computer-based models of the geosystem itself, are, however, relatively new, although the genesis of these models may be traced back over the past 100 years.

In principle, modern seasonal to interannual predictions are an answer to the needs of many whose activities are influenced in some manner by climate variability, whether this is in terms of creating profit through the marketing of an appropriate range of goods, or is in terms of critical decisions regarding agriculture and food security. Much of the later body of this book is devoted to exploration of the extent to which current state-of-the-art predictions address the requirements of those who have responsibilities for taking decisions in regard to climate-linked activities, to the impediments, and to the opportunities available. Various examples are provided of the way in which the systems that deliver climate prediction information have been set up and of the benefits achieved.

Earlier chapters of the book are devoted to the science and technology behind the predictions. For the science of seasonal to interannual prediction 1997 was perhaps one of the milestone years. During 1997, amongst other pertinent events, long-term operational support for the Tropical Atmosphere Ocean (TAO) array[1] was authorised by the US Congress, many prediction models of different types became available to take advantage of the information provided by the array, and one of the most significant recorded El Niño events developed to bring its particular signature of climate variability to many parts of the globe. But to understand the significance of 1997 we need to wind back a little, and to consider the lives of communities along the equatorial west coast of South America, particularly around Ecuador, Peru and northern Chile, in previous centuries.

Much of the equatorial west coast of South America is dry in most years, with fishing, particularly for anchovies, providing major sustenance during past eons. Nowadays the story is well known of how the anchovy fisherman around the Gulf of Guayaquil noticed every few years that the fish stocks appeared to disappear for several months at a time, with resultant deleterious impacts on food reserves. At the same times heavy rainfall would strike the area, leading to flooding and wash-aways of crops and mud-built houses. Because these events typically began around Christmas, the fishermen named them ‘El Niño’, after the Christ child. But the fishermen were not the first. At least the Incas, who had never heard of El Niño, recognised its consequences for their food security. Consequently they farmed diverse stocks at different altitudes in the Andes, experience having indicated that rarely was there simultaneous failure of all stocks.

For many years the concept of El Niño was little more than a scientific novelty, studied by few. Even when in the earlier years of the 20th century Gilbert Walker undertook his ground-breaking research into the causes and prediction of the Indian monsoon, and in doing so uncovered the great ‘atmospheric see-saw’ of the Southern Oscillation, the significance of these discoveries, and their relationship to El Niño, was not appreciated. Probably the first El Niño event that drew wider attention was that of 1972/73, which was followed by several scientists building on earlier pioneering work to begin suggesting in the wider literature that El Niño was not something that just affected Ecuadorian and Peruvian anchovy-fishing communities, but was part of a much larger occasional climate anomaly that affected communities in many parts of the world. By the time the large-amplitude 1982/83 event occurred, far greater numbers of scientists were recognising that a breakthrough was being made in regard to understanding and predicting the climate system, and from then on a new ‘industry’ was born: an industry that covers the physical understanding, the consequences for predictability and prediction, and the onward use, including the politics, of the predictions, all of which are inherent in the slow changes in the planetary surfaces underlying the atmosphere.

[1] A network of moored buoys across the tropical Pacific Ocean that delivers via satellites the monitoring information of both the atmosphere and the ocean (to 500 m depth) on which the models and predictions depend.

The basis of this burgeoning industry is that slowly varying components of the geosystem, most significantly sea surface temperatures across tropical ocean basins, can impart a 'memory' to the atmosphere in the vicinity of any such long-lived anomalies. And further that the atmosphere works in such a way that this 'memory' can be transmitted to parts of the globe remote from the originating sea surface temperature anomalies – meteorologists refer to this phenomenon as 'teleconnections'. Thus, for instance, El Niño events are *typically* (see caveats later) associated not only with heavy coastal Ecuadorian and Peruvian rainfall, but with above-average rainfall also in northern Argentina, in East Africa, and in California. Equally, contemporaneous drought can occur in north-east Brazil, in southern Africa and over much of Australia. Climate forcing of this type is not restricted just to changes in the tropical Pacific basin, although as far as is known these are the most important; the other two ocean basins play their own, more limited, roles, as do other slowly varying aspects of the geosystem underlying the atmosphere, such as soil moisture anomalies over various continents and snow extent over Eurasia.

El Niño, and its related cousin La Niña, represent major changes in the distribution of sea surface temperatures across the tropical Pacific basin, with warmer waters spreading eastward towards South America from their usual position in the west of the basin during an El Niño. Anchovies thrive in the cold current running northwards along the west coast of South America, but during an El Niño this cold current becomes overlaid by the warmer waters, and the anchovy descend towards the colder nutrient-rich waters below.[2] For the fishermen the anchovies have disappeared; in practice they are thriving deeper within the ocean than usual, beyond the reach of any netting system.

Once scientists began to recognise the significance of events in the Pacific basin, the next stages were to understand the mechanisms involved, to model the pertinent aspects of the geosystem, and to determine if prediction might be possible based on this new knowledge. Arguments still exist over the precise mechanisms involved in El Niño events, but the basics are understood, as is demonstrated within this book. Many models of varying complexity have been built to understand the system. And many of these same models have been used to provide predictions. The advances in this field over the past 30 years are spectacular. These advances benefited enormously from the TAO array and other observing systems, both in situ and satellite-based.

Building on developments that have resulted from the recognition of the importance of, and the growing understanding of the dynamics of El Niño events, in this book we cover: overviews of the climate system and the manner in which it works; current capabilities to model and predict the climate system out to several months

[2] During La Niña events waters along the western South American coast become colder than usual and in the eastern tropical Pacific warmer than usual. During La Niña events climate anomalies worldwide tend to be amplified in a canonical pattern roughly the reverse of that for an El Niño event, but in this case the anchovies remain near the surface.

based on the ability to simulate ocean circulations in the Pacific basin and elsewhere; the manner in which the information produced by the models is treated and delivered; and finally the ways in which this information is used in decision making in numerous activities. It is a story of success, but it is also a story of complexity in several senses, complexities that need further resolution if the full benefits of the scientific advances are to be obtained.

Complexities emerge in several ways. First, the geosystem itself is complex in the manner it works, including in the ways in which the various components interact with one another. One prime example of this complexity is that while El Niño is the major forcing known on timescales of a few seasons, it is irregular (events being separated by anything from 2 to 7 years), and, not being alone as a forcing mechanism, its influence might be overcome by other sources of forcing. Many around the Indian Ocean basin, for example, recall the 1997/98 El Niño event, sometimes referred to by meteorologists as the strongest on record, not for the canonical response expected (perhaps as during 1982/83) but for the deviations from that response. For example many areas were braced for droughts – southern Africa, India, parts of Australia – but rainfall was perfectly adequate in all of these despite the strength of the event. Equivalently in East Africa above average rainfall is the canonical response, but there was no expectation of the devastating amount of rain that fell at that time (Fig. 1.1). These differences from the best-wisdom canonical response were attributed to unusual and strong sea surface temperature anomalies across the tropical Indian Ocean, anomalies not always fully incorporated by the prediction models then available. Assumption of canonical responses with regard to climate variability is unlikely to represent the safest available approach.

Scientists have not unravelled the complexity of the geosystem in full, and models remain relatively simplified approximations of the real world. Hence any predictions from these models cannot be perfect as the models themselves are not perfect, but there is a further crucial aspect of complexity here in that the models are sensitive to various small changes in values of observations used in the initialization stages, and to aspects of their own formulation in detail, sensitivities that can lead to entirely different predictions when brought into play. Scientifically sensitivity to small differences in starting positions is known as ‘chaos’; chaos, which strictly refers to the characteristic of non-linear systems at certain (but not all) times to be markedly dependent on various relatively small differences, results in the inherent impossibility to predict the future in a deterministic sense at some, and in general for seasonal predictions at all, times – only probabilistic predictions are appropriate for chaotic systems. Most modern prediction approaches acknowledge chaos and produce probabilistic forecasts, but the delivery and interpretation of probabilistic forecasts introduces further issues. Ultimately the information produced by the models is incorporated into decision processes relevant to managed systems which themselves often have chaotic or uncontrolled aspects. The entire system is one of complexity throughout, complexities that as yet are not fully understood nor managed.

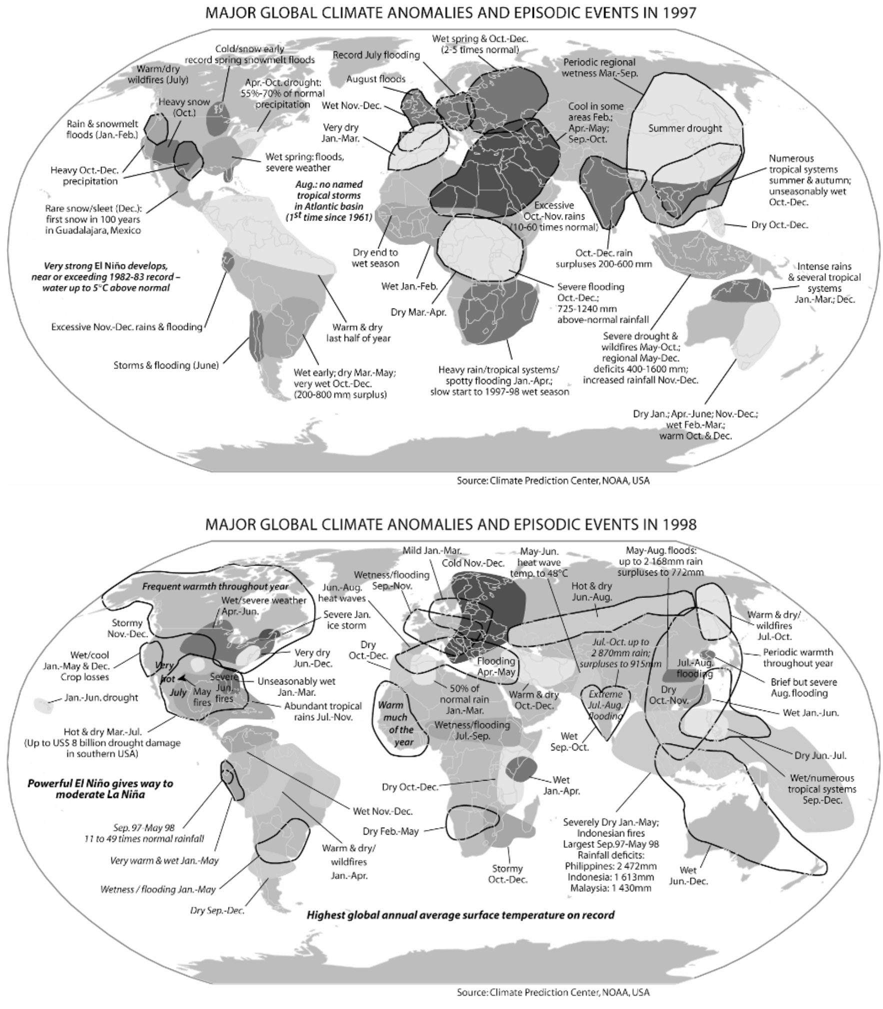

Fig. 1.1 Effects of climate variability during the years 1997 (top) and 1998 (bottom) which include (but extend beyond) the major 1997/98 El Niño event. Compare these effects with those during a 'canonical' El Niño year in Fig. 6.10. Careful comparison indicates that there were differences during the 1997/98 from those of the canonical expression, particularly around the Indian Ocean basin, including: more rainfall than typically occurs over parts of south-eastern Africa, a wet monsoon, and again more rainfall than typically occurs over northern Australia. Additionally rainfall over East Africa was far more intense than might have been expected. Strong anomalies of sea surface temperatures over the tropical Indian Ocean, contemporaneous with and perhaps related to those in the Pacific Ocean, have been identified as a possible cause (Adapted from WMO 1999, report No. 905)

Complexity is not assisted by the fact that the degree to which predictions can be made with success, even in a probabilistic sense and whether from statistical or from numerical models, varies geographically, it varies by seasons, it varies by forecast timescale, it varies by the variable being predicted, and it may exist only during specific 'windows of opportunity'. Thus in general terms the highest predictability of atmospheric temperatures and rainfall exists across the tropical ocean basins, in particular that of the Pacific, and over certain land areas within or immediately adjacent to those basins. Predictability tends to decrease further away from the Equator and from the oceans, although some areas, such as North America, are favoured in certain seasons through enjoying higher predictability than similar regions at the same latitudes because of the manner in which teleconnections work in those areas. There is evidence that predictability in the global sense is higher during El Niño and La Niña events than otherwise, and that in some regions, such as Europe, it may not exist at times other than during these 'window of opportunity' events [but equally may not necessarily be high during specific individual events]. Temperature tends to be more predictable than rainfall. But even for the most predictable variable at the location with the highest overall predictability it is always necessary to provide probabilistic predictions. And with that comes the challenge of interpretation and of translation into effective decisions.

Many centres now generate predictions up to seasonal, and in some cases on longer scales, using dynamical models on either an operational or a regular research basis; many of these products are placed on either open or password-protected web sites. Dynamical models, being expensive to develop, maintain and run, are mainly the preserve of a relatively small number of meteorological organisations and universities. Broadcasting and distribution of these forecasts comes, in general but not universally, under the overview of the UN Specialised Organization, the World Meteorological Organization (WMO). WMO is coordinating the establishment of recognised Global Producing Centres as well as of Regional Climate Centres as centres of excellence to support climate services.

By comparison with dynamical models, developing and distributing predictions based on statistical approaches is relatively straightforward. Thus many national meteorological services, particularly most within Africa, that do not possess the resource to run dynamical models have created statistical modelling capabilities, either just for their own country or for wider areas, which form important bases for national prediction services. Most current evidence suggests that the qualities of predictions from statistical and numerical sources are competitive. It is possible also to combine statistical and numerical approaches, either in the prediction stage where one component is achieved through statistical means, or through the creation of a consensus of predictions from individual sources.

While there is a relatively small number of forecast producers, those interested in taking advantage of the predictions are globally widespread. Given that prediction skill tends to be highest overall at lower latitudes, with active advantage of that fact taken in Australia, the greatest concentration of users (Australia excepted) might be expected in developing countries, users with responsibilities ranging from international management of development, including issues such as food and water security, through all levels down to those taking decisions in the field. Climate-sensitive commercial interests are growing in the developing world, including from businesses based in the developed world. The three classic areas of interest (but numerous others exist) are agriculture, water resources and health, all of which are covered in this book in some detail. At higher latitudes, where skill levels tend to be lower, the greatest number of users are probably those with commercial interests, with government planners a second important interested group. In all cases the available evidence suggests that the costs of developing and maintaining the forecasts are significantly outweighed by the benefits produced.

The book is laid out in five parts. In Part 1, a background to the science and to the use of the predictions in decision making is provided, in part through this introduction chapter. The scientific core is discussed in Part 2, in which focus is given to the workings of the climate system and to approaches to prediction, both dynamical and statistical. Methodologies for adjusting the prediction information that emerges from the various models so that that information is better tuned for later decision making, is covered in Part 3. Decision making and some specific uses of the prediction information are discussed in Part 4, while loose ends and views to the future are drawn together in Part 5.

To an extent the structure of this book is reminiscent of an end-to-end approach to the production, delivery and use of the prediction information. In other words it might be viewed as outlining a unidirectional system in which predictions are fed through necessary delivery stages for ultimate use in applications. There is nothing new in such an end-to-end approach, this having been the principal model for delivery of weather forecasts over many decades. The end-to-end principle was assumed in first attempts to deliver seasonal predictions in the 1990s and the early 2000s, it was the underlying paradigm for the creation of WMO's Climate Information and Prediction Services (CLIPS) and the US-based International Research Institute for Climate and Society (IRI), and it remains the assumed principle for a large body of forecasters and service providers. Experience has indicated, however, that because of the complexities of the systems involved throughout, the end-to-end approach is non-optimal, and new approaches/paradigms are being sought.

These new approaches are based on steadily improving understanding of the decision processes involved in the use of climate information. Decision processes vary significantly to the extent that a simple one-size-fits-all, end-to-end, approach to the delivery of climate services is frequently, in practice, unsatisfactory. From the most broad-brushed perspective, decision processes, and therefore the manner in which climate information should be delivered, vary between the developed world and the developing world, between commercial and development contexts, between sectors (agriculture, water, health and so on), and between the various levels at which decisions are made (from intergovernmental down through to the field level). End-to-end delivery of information might be appropriate in, say, commercial contexts, whereas different approaches are necessary for social and economic development contexts within the developing world.

The necessity for climate information providers to be sensitive to the specific decision needs within each context places an onus on those providers for customisation of services, an onus that requires close cooperation with those taking specific decisions. The IRI has changed its strategy to approach this challenge through integrated assessment of all information needs (not limited simply to climate information) within each context, with the expectation that lessons learned will ultimately lead to greater facility in optimisation of information delivery across countries, sectors, and so on. But this raises the question of identification, and nomenclature, of these decision makers. From the perspective of the end-to-end model the concept was simply one of delivery to 'end users' for use in their 'applications'. The new paradigm, covering intermediaries/recipients/decision makers/decision takers/stakeholders/end users, at the full range of levels, with responsibility for numerous decisions that often do not conform to the straightforward concept of 'application', has not yet generated an appropriate nomenclature that places all involved and their actions into clear context. Within this book the nomenclature used is variable as a result, although we try to be as consistent as possible, but should throughout be considered within the context of the new, evolving, paradigm. As will be seen, the learning process in service delivery is still at an early stage and is not covered in full within this book; the examples provided give insight, nonetheless, into contexts within which climate information is being provided and used. Undoubtedly service delivery is one area demanding active and creative consideration from those engaged within it.

The potential readership of this book is broad, covering numerous disciplines and levels of expertise. Climatologists with interests specific to atmospheric dynamics and numerical modelling cannot be expected to be expert in issues of communication nor of the behaviour of *Anophe les* mosquitoes and its links to climate and malaria. Equally agriculturalists may not be interested in the detailed structure of climate models. In order to assist those with the limited expertise in the contents of specific chapters, each chapter begins with a summary of its contents written in such a way as to be accessible to all readers. A list of references is provided at the end of the book, including a separate list for further reading of interest to both specialists and non-specialists. Also, two glossaries have been included to assist all readers, the first dealing with acronyms and the second with terminology.

Acknowledgements The authors would like to thank Mmes. Cynthia Cudjoe and Leslie Malone (WMO) for kindly providing the original figure used to generate Fig. 1.1.

Chapter 2
Seasonal Forecasts in Decision Making

Mike Harrison, Alberto Troccoli, Michael Coughlan, and Jim B. Williams

A new and developing vibrant science has been born capable of providing significant benefits to humankind, from development work aimed at sustaining and enhancing the quality of life to increasing the profits of commercial activities. At the heart of this science lies an improved understanding of the climate system, of its predictability, and of its links with natural and social systems. An overview of the integrated structures of these non-independent systems within the context of the new capabilities in seasonal to interannual prediction is provided in this chapter, including the fundamental interactions between the various systems, their natural complexity, the confusion that often arises between the terms 'climate variability' and 'climate change', and the essential role climate information, including predictions, plays in the management of risks associated with climate variability and change. There follows an introduction to decision making in which climate information is involved, including discussions on decision processes and communication, a brief history of relevant climate science, and an overview of political and social issues directly linked to climate. Finally, two perspectives are provided of activities that might benefit from decision making that takes advantage of climate information: first, a predominantly end-to-end perspective in which climate information is delivered directly to a particular application; second, a perspective where the challenge is to integrate climate information into the broader context of sustainable development. These two positions, direct delivery into specific decisions for 'private' benefit and information provision for the 'public good', perhaps represent the two ends of the broad spectrum within which this new science can contribute.

Mike Harrison
Independent Consultant

Alberto Troccoli
European Centre for Medium Range Weather Forecasts

Michael Coughlan
Australian Bureau of Meteorology

Jim B. Williams
Consultant with the NRGroup

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

2.1 Climate Variability and Change: The Overlaps and the Differences

2.1.1 About Systems

The process that starts with the generation of a seasonal to interannual prediction and ends with someone making use of the prediction is a road that takes us from the application of pure science in physical systems to the pragmatism of real-world uncertainties, via the practicalities of operational forecasting frameworks. The latter systems are arguably more complex and unpredictable than the physical systems from which we started.

In trying to deliver on the promise offered by scientific knowledge of climate, we must deal with several 'systems' – scientific, environmental, social and economic – not only how each functions in its own right, but also how they interface, overlap and interact with each other. From the pure scientific perspective some systems are simple and driven by a single dominant force or set of independent linear forces. Such systems are generally highly predictable, e.g. planetary motion where gravity is by far the dominant force.[1] On earth, forces are rarely independent of each other and are often non-linear. Sometimes there are only a few dominant forces that give rise to chaotic outcomes; such systems exhibit some level of predictability but also often have inherent and unpredictable instabilities. At the far end of the scale there are systems with many roughly equal forces at work, which lead to random outcomes. In random systems the predictability of any individual outcome within the system is virtually impossible to assess but statistics may still tell us quite a lot about how the system will behave as a whole.

Meteorologists, ever the pragmatists, have long recognised the uncertainty in their science and that there are good reasons for limits to the predictability of explicit outcomes of the non-linear systems that generate our weather and climate (Lorenz 1963). Yet by capturing the essence of the physics, dynamics and chemistry of the system and by exploiting the 'laws' of large numbers, meteorologists and climatologists have become adept both heuristically and mathematically in stretching the levels of useful skill towards the outer limits of predictability.

[1] However, when two nearly similar gravitational pulls act on a single body then the system can become unpredictable.

2.1.2 Climate and Weather

Climate is traditionally viewed as the integration 'upwards' of the characteristics of discrete weather events and variables over time and to some extent space; occasionally climate is described as 'the statistics of weather'. The corollary is that the components of global climate change should be manifest 'downwards' on all time and space scales. This critically important concept (Fig. 2.1) has only recently been recognised by those concerned with appropriate responses to climate change. Successful adaptation to climate change will not simply be a case of adding another row of bricks to a sea wall to stem sea level rise, for example, or building another dam to catch more water in a drier climate. The consequences of 'global warming' will not just appear as an inexorably rising graph of global temperature but will also be evident through a set of complex changes in the global circulations of the atmosphere and ocean that will arise, in part, because it is expected that the warming will be greater over the land than over the sea. In turn, this means that some areas will become drier or wetter than others, but not every year – just more frequently than before. It follows that in any given year the mix of weather patterns that a decision maker will have to deal with will also change.

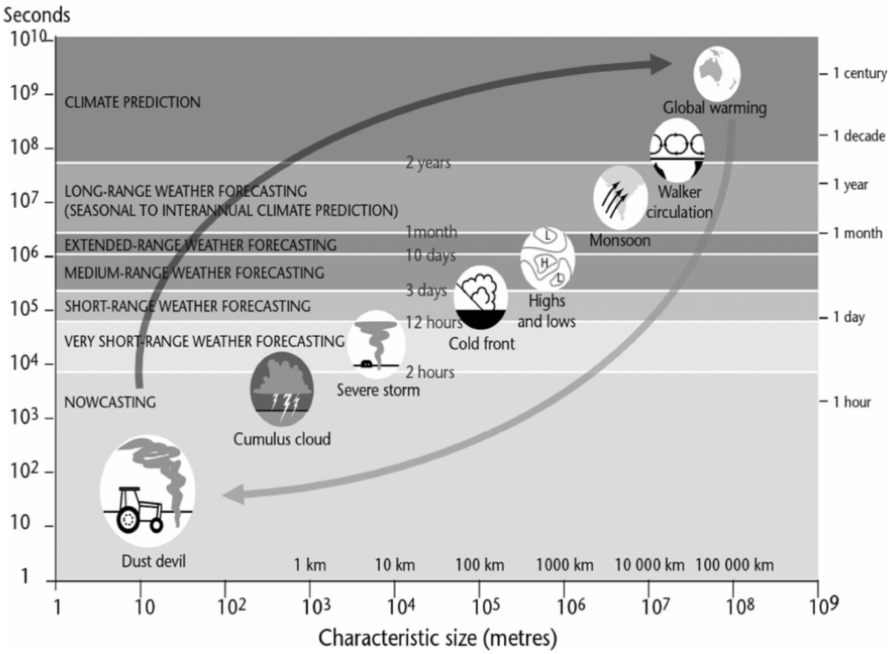

Fig. 2.1 Climate is traditionally viewed as the integration of discrete weather events and variables over time and space. The corollary is that the components of global climate change should be manifest 'downwards' on all time and space scales

Clearly then there is scope for adaptation to climate change on all time and space scales. Using the information that a seasonal to interannual forecast offers is as practical a response to climate change as it is to varying seasonal conditions.

The difficulty of distinguishing between ‘climate variability’ and ‘climate change’ has been addressed within the United Nations Framework Convention on Climate Change (UNFCCC) by constraining *climate change* to mean only that component which is directly the consequence of human activities, in particular the emissions of greenhouse gases, but also including land use transformations. All other components of change in the climate are referred to within the UNFCCC as natural *climate variability*. Note that these definitions are both independent of timescale, and thus change and variability according to the UNFCCC definition cover all scales from the very shortest to those acting over extended periods of centuries and beyond, the only difference being one of attribution, i.e. between natural and anthropogenic forcing.

This separation of change and variability is logical when viewed from a UNFCCC perspective, not least that natural climate variability cannot be ‘managed’ in the UNFCCC sense whereas management is possible to an extent for climate change as it is by definition human-induced. Two approaches to the management of climate change are envisaged within the UNFCCC: mitigation of emissions and adaptation to a changed climate. Within the UNFCCC context actions and funding regarding mitigation, with emissions taken as the main driver of change, become self-defining, and it is this perspective that provides the foundation for the UNFCCC definitions of change and variability. There is less clarity, however, when it comes to actions and funding for adaptation activities, which in the strictest UNFCCC sense should apply only to adaptation to whatever modulations on whatever timescales result purely from anthropogenic causes. In reality, such a partitioning is highly, if not totally, impractical as making a clear separation between weather and shorter scale climate fluctuations that are naturally forced from those that are anthropogenically forced cannot be made. Any adaptation responses, whether managed or endogenous, will need to factor in the integrated totality of fluctuations that have resulted from the combination of all sources. Management of the risks of climate variability on timescales of a season to a year are thus an inherent aspect of adapting to the consequences of climate change whatever the timescale. The contribution that management of short-term climate risk can make to the overall response to long-term climate change has generally been undervalued during the formative years of the UNFCCC. The broadening in recent years of the UNFCCC process beyond mitigation to embrace adaptation to a growing extent has led to a greater appreciation of the need to manage climate risks over all timescales including the vital contribution that seasonal predictions can make. Some of the tools that will assist in understanding and managing the consequences of the totality of climate variability and change, whatever the cause, are covered in this book.

2.1.3 Adaptation, Climate Variability and Change

Even if separating adaptation to climate variability from adaptation to climate change becomes problematic, as it will be in many practical instances, what are the main pathways for adaptive responses? Figure 2.2 suggests that climate science can tell us how 'forcing' within the climate system will produce or induce changes in weather and longer term climate patterns. Such outcomes will have their consequences or 'impacts', the severity of which will be determined by the level of vulnerability of a society or ecosystem that is sensitive to weather and climate. If the impact is sufficiently strong to elicit a response within the community, that response may take several forms. In the case of a serious or severe event that leads to a disaster, for example, the normal human response will be one of providing emergency relief to affected communities as quickly as possible. Experiencing an impact might lead one to attempt to do something about future levels of the undesired forcing. Experimenting with cloud-seeding to prevent damaging hail is one example of such a response on the shortest timescale. Efforts at mitigation or abatement of greenhouse gas emissions to forestall further global warming lie at the other end of the time spectrum. A further "lesson learned" response is to take adaptive measures that build resilience to future occurrences of similar events. Such responses would include building sturdier houses to withstand storm-force winds or even adding that extra row of bricks on the sea wall.

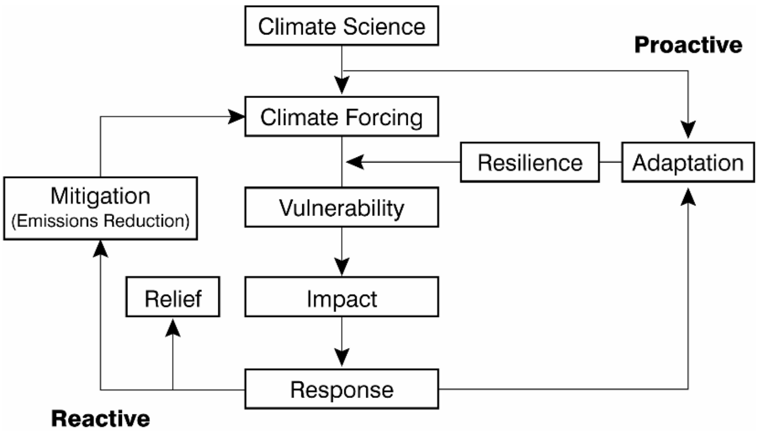

Fig. 2.2 Pathways for responding to climate variability and change. The central axis represents the sequence of a climatically forced event. The side arms provide optional actions to reduce the negative impacts of such events, either proactively with the application of scientific understanding or, in a more reactive sense, when the consequences of an event have already been experienced

So far, however, all the actions or responses discussed have been ‘reactive’, and follow once an event has occurred or begun to occur. Figure 2.2 suggests that climate science has the potential for providing a more proactive pathway to adaptation. Such a pathway provides opportunities for building resilience and hence reducing vulnerability to an event before it occurs. It is important to recognise that while reliable and useful prediction is a highly desirable tool to have at one’s disposal on this pathway it is not always necessary for deriving effective adaptation strategies. Even in the absence of any predictive capacities, statistical information about how the climate varies in time and space can be a powerful planning tool, at least so long as one is confident that the past climate is a good model for future climate.

2.1.4 Forecasts, Predictions, Projections and Scenarios

The rapidly growing societal awareness of climate change highlights a degree of terminological confusion within the broader climate community, not only among those interested in response and adaptation measures but also among climate scientists. Not entirely at one on how best to define the term ‘climate’ as it relates to the past and to the present, climatologists are faced with the need to describe what it means when one is talking a bout climate and climatic events into the future. The term ‘climate forecast’ seemed to suggest an extension of explicit weather type forecasts out to climate timescales, something that, as we have seen, is clearly not possible; the addition of the pre-fix ‘long-range’, as in “long-range weather forecasts”, did little to resolve the confusion on the shorter climate timescales. In fact the use of synonymic terms to define a range of very different concepts has left many scrambling to sort out the details, e.g. ‘projection’ as something distinct from a ‘forecast’ or a ‘prediction’, along with the now almost hackneyed term ‘scenario’. Figure 2.1 provides one attempt at a rational nomenclature, but the inclusion of climate projections and scenarios on this figure would probably require a third axis. Those with a sceptical bent on the climate change issue rose quickly to exploit some of this terminological confusion, despite the best efforts of the Intergovernmental Panel on Climate Change to have everyone reading from the same glossary (IPCC 2001).

In essence, all expressions of what the future may hold, whether they are called forecasts, predictions, projections or scenarios, embody degrees of uncertainty. Consequently, from a practical or even a basic conceptual point of view, it is the level of uncertainty that matters and not so much the exact meaning of the term being used.

2.2 A History and Status of Seasonal to Interannual Predictions in Decision Making

2.2.1 Introduction

From a practical perspective, there is only one reason for undertaking research and development to advance seasonal to interannual predictions and for investing in the infrastructure to produce and deliver them. That reason is to assist whatever decision processes are of concern to those who might make use of them. To be of real and measurable value, prediction information must be readily assimilable into the decision processes of recipients. In practice this goal may represent the ideal more than the complex reality, but it implies nevertheless that coordination between supplier and recipient is essential for the derivation of optimal benefit from the prediction information. Such optimal benefit is difficult to achieve in seasonal prediction:

- When information is couched in language that recipients find difficult to interpret – jargon such as "chaos", "probabilities", "terciles" and so on
- When the provider does not have a clear view of the needs of the recipient
- When the recipient does not have a clear view of the uncertainties inherent in the information
- Without adequate and ongoing coordination and dialogue between provider and recipient

The process of dialogue and coordination has been building for many years, but there remains much to be done in order to achieve optimal support for decision making.

Following the global emergence of seasonal forecasting after the commissioning of the Tropical Atmosphere Ocean (TAO) array, the first approach taken was to disseminate seasonal to interannual predictions in an "end-to-end" way. This process generally involves one or more forecast producers delivering predictions to one or a group of recipients within a specific sector, an approach adopted initially by both the World Meteorological Organization's Climate Information and Prediction Services (CLIPS) initiative and by the International Research Institute for Climate and Society (IRI).[2] This end-to-end process has been the traditional approach taken in the delivery of short-range weather forecasts, and therefore seemed a logical way forward. In practice end-to-end has proven often to be sub-optimal for seasonal to interannual predictions because of the intrinsic difficulties in linking the probabilistically framed predictions to many practical decision processes. The outcome to date, by and large, has been a mosaic of small projects,

[2] Originally named the International Research Institute for Climate Prediction.

with few that can be regarded as seminal to a more generalised approach. Translation of results between projects/sectors/geographical areas has proved to be difficult.

Recognising these difficulties, some organisations have developed strategies built around the concept of focussed solutions within particular sectors. In this approach attention is placed on the coordination of all activities within the delivery and application chain in order to develop a comprehensive decision making package that will benefit the stakeholders within a specific sector. For example, one pilot IRI project covered the management of water resources in two dams in Ceará (Brazil). Water from these dams was used for hydropower generation, for irrigation (during the greater part of the year when rain is not expected) and for general purposes, including industrial and personal, consumption. The project involved the generation of predictions on various timescales, the convening of several committees of users and water managers, the application of market forces, involvement of the insurance sector, and the creation and delivery of a tailored information package to all stakeholders. While not yet commissioned operationally, this project provides a cogent example of the type of innovative solution that might be applied elsewhere.

However, even this approach, which is still essentially end-to-end in concept, and similar such approaches, may not be sufficient to tackle the larger issues. As already mentioned, prediction, when available, is just a single, albeit important, tool in the management of climate risks. In a broader context the potential contribution of predictions lies in the need to manage climate risks on all timescales. This broader context includes the management of risks arising from climate change and desertification and, in a more political/social framework, the achievement of objectives such as the Millennium Development Goals (MDGs). It covers additionally incorporation and melding of sources of risk other than climate *per se*, and aspects of management of the totality of those risks, including development and administration of appropriate policies. All approaches require an outcome-oriented perspective of interaction of all involved disciplines with all users.

2.2.2 Decision Making

The decision is everything: without serving as a basis for decisions, seasonal to interannual prediction would be little more than a stimulating intellectual challenge. Yet providing information for possible use in making a decision is not of itself enough; that information should enlighten a new decision, confirm the validity of a decision already made, or cause the recipient to adjust a previous decision, if it is to have *value*. Without providing value, even the stimulating intellectual challenge is at risk.

The value obtained in practice can be determined in numerous ways, including through (but not restricted to) the form and quality of individual decision processes and the degree to which predictions are customised to those decision processes. High order predictions, such as ones for total agricultural production in a region, in principle offer the greater potential value as compared to those at lower orders, say the number of growing degree days during a season, or those at lower orders still, such as mean temperature and rainfall anomalies over a period. Most seasonal predictions currently offer only the lowest order of climate predictions, typically of mean temperature and rainfall anomalies, although there is expanding activity to support higher order predictions in certain geographical regions and sectors.

While predictions tend to possess a relatively monochromatic character, decisions in a complex environment come in a vibrant spectrum of forms and approaches. Few decisions are independent of others, and most are based on a range of information streams. Climate will generally be only one factor under consideration – see the example of Food Security in Fig. 2.3 – and may be perceived as not even particularly important. Predictions of first order variables such as rainfall and temperature, unless perceived, or ideally proven, to be of a quality sufficient to provide value, may receive less attention than basic climatic data, and less attention than other data streams informing a decision. Yet, in practice, relevant climate and other data are only infrequently supplied alongside the predictions themselves as part of a climate service. Similar arguments apply to predictions of second and third order variables.

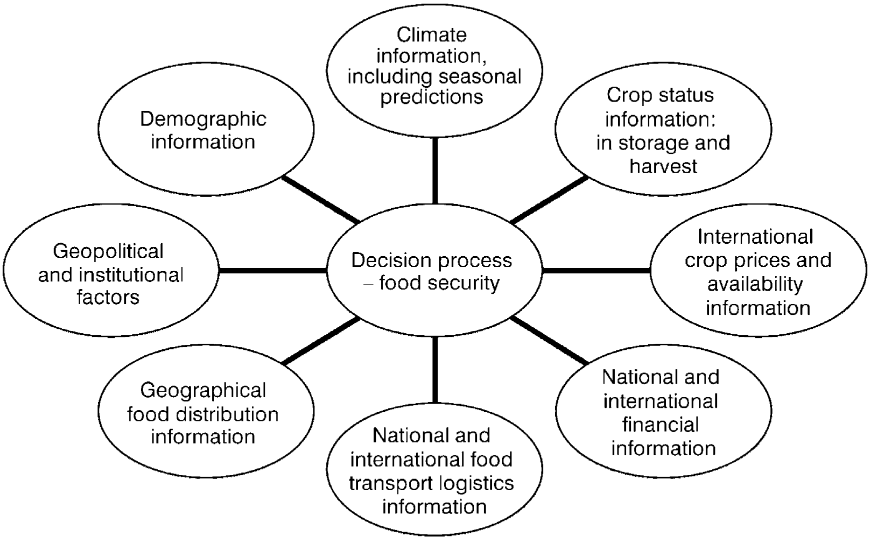

Fig. 2.3 A simplified example of information streams that might be used in a single decision process related to food security

Decisions are made within a rich continuum of overlapping domains involving sectors, cultures, economics and politics, as well as timescales. Numerous sectors are affected by climate variability, and indirect interactions can extend those affected into surprising areas. For example, most of the eight MDGs, even those not explicitly related to climate, can be detrimentally or beneficially influenced respectively by climate variability that undermines or supports the economic and/or political and/or physical infrastructure of a country. MDG No. 8, Develop a Global Partnership for Development, which covers mainly international trade and finance issues, is but one example. Most cultures approach decision making through their own time-honoured traditions, with many continuing to use indigenous knowledge developed over centuries to guide their day-to-day decisions. The economic, political and statutory backdrop to any decision can influence both the manner and the outcome of specific decisions. All such issues should be considered in order to deliver seasonal to interannual predictions tuned according to pertinent decision processes.

Even the matter of timing proves complex, as decisions are made on a wide variety of timescales, with a variety of lead-times, few of which will correspond neatly to the scales and lead-times common to contemporary prediction capabilities. There is often a tension between the window of opportunity for seasonal prediction that comes from sea surface temperature anomalies and the real requirements of the decision maker. It is this tension that organisations such as the IRI, the Australian Bureau of Meteorology (BoM) and the Queensland Department of Primary Industries (QDPI), for example, are attempting to address.

2.2.3 Communication

Effective communication between provider and recipient is an essential prerequisite for maximising the benefits from short-range climate predictions. Good communication, in both verbal and visual forms, needs to be appropriate to all stages of the process, starting with the initial introduction of predictions in specific decision making contexts, continuing through the period of forecast use, and then extending to the support necessary for further development in their application. Like most scientists, climatologists tend to use the jargon peculiar to their field. Recipients also tend to belong to particular disciplines or sectors, each with its own vernacular. The inevitable consequence is scientist-recipient communication at a sub-optimal level. This language problem certainly is not restricted to climate, but attention within the climate context could break down some of the perceptual barriers that cause predictions to be discounted or used ineffectively. Hence climate scientists have a fundamental responsibility to understand how their information is to be used, and to communicate their information in the language of that use. It helps if the recipients also have some understanding of climate jargon, but in practice this may not be necessary provided there is confidence in the information being received, a confidence more likely to grow given communication primarily in the vernacular of the recipient. Confidence is linked in context to credibility, credibility being gained through numerous processes including extensive experience of the quality of the predictions or of receipt of persuasive information confirming that quality, a requirement again demanding communication in a form suitable to the recipient. Unfortunately most training activities to date have focused on teaching recipients climate jargon, rather than on teaching climate scientists the essential language of recipients and the nature of their decision making processes.

Sitting alongside verbal communication is the powerful tool of visual communication. Data visualisation techniques have developed rapidly in recent years, particularly through the use of computers, and have made substantial contributions to the advancement of all sciences. Data visualisation can also be a potent means of communicating science and scientific information to the layperson. It is regrettable, therefore, that novel methods of communicating visual information on seasonal to interannual climate predictions that are readily accessible to recipients have been slow to develop and that, in general, visual presentations remain tied to the perceived communication needs of the climate scientist rather than to the actual needs of recipients.

Well-designed visualisations could play a vital role within the framework of specific decision processes:

- To help explain the science
- To provide climatological and other information
- To provide the predictions themselves of whatever type
- To provide information on the quality of the predictions (i.e. verification)
- To place climate information within the context of other information required for a decision

As yet, many predictions, together with any accompanying verifications, are made available in formats that do little to assist decision makers. Frequently complementary explanations are written in the jargon of the scientist rather than the language of the recipient. While it may be a difficult and slow process to improve the quality of the predictions themselves, much could be done now to improve the communication of them and their current levels of skill in ways that facilitate their incorporation into decision processes, with consequent rapid gains in the value of the predictions. Equally, well designed visualisations can be used to communicate to the climate scientist how decisions in recipient communities are made. Communication through an effective mix of enhanced verbal and visualisation techniques offers outstanding potential for major advances in targeting and improving the value of the forecasts.

2.2.4 A Brief History

Climate prediction, at least that covering the next few seasons, is one of the oldest professions, with known examples stretching back millennia. Life depends on climate, and decision making to sustain life requires methods of foreseeing climate aberrations that threaten life. The extant wealth of indigenous knowledge, built over many generations, has resulted in a complex of information still frequently used and implicitly trusted in many parts of the world. Not surprisingly this indigenous knowledge universally tends to be derived around seasonal changes in local flora and fauna, plus astronomical observations. Religion and other belief structures (including dictums and maxims) are added to the mix in many countries. Wherever indigenous knowledge is considered fundamental then the usefulness of any new information source naturally will be first judged against this; it is conceivable, of course, that such comparisons may be biased. Nevertheless the existence of a culture of indigenous knowledge provides an opportunity for the climate scientist to introduce new techniques in a sympathetic and synergistic manner.

While a rich global history exists of attempts to predict weather in coming seasons, it is generally agreed that modern seasonal to interannual prediction originated in the work of Sir Gilbert Walker, tasked while Director of the Indian Meteorological Service in the early years of the 20th century with predicting the monsoon in order to bolster food security for the subcontinent. Indian food security in practice has come through the coordinated planning of resources over a few years, rather than through Walker's work. However Walker's legacy lives on through both the world's longest-running statistical seasonal prediction system, as maintained by the Indian Meteorological Department (IMD), and his identification of the Southern Oscillation, the great "see-saw" in atmospheric pressure differences between the South Pacific and the Indonesian region. It was to be several decades before the relevance of Walker's work was to be recognised in full, but his work provides the observational foundation for most modern approaches.

By the 1970s a few scientists were beginning to recognise the relationship between the Southern Oscillation and El Niño (to be discussed in Chapter 3), a periodic warming of sea surface temperatures along the equatorial Pacific South American coast, and further to acknowledge the societal impacts of individual El Niño events. With that progress came evidence of the general potential for sea surface temperature anomalies, primarily but not uniquely tropical, to influence remote climates on seasonal timescales. Although the 1972/73 event created a stirring of interest, it was the 1982/83 event, with its "classic global climate anomaly configuration" (also known as teleconnection pattern and shown in Fig. 6.10 later in the book – compare it to the main climate anomalies for the 1997/98 ENSO event in Fig. 1.1), that propelled El Niño into global prominence. That event transformed the agenda of the First International Conference on Southern Hemisphere Meteorology, coincidentally held during August 1983 in a Brazil feeling the full impacts of the event from flood rains in the south to drought in the northeast. The event also set in train an industry building statistical prediction models based on links between rainfall and anomalous sea surface temperature patterns, an industry that continues today alongside the sophistication of the global coupled climate models, the former providing benchmarks for skill assessments of the latter.

Gilbert Walker worked from an entirely pragmatic base, and that same pragmatism has been the main driver for new investment in prediction infrastructure. Certainly influential theoretical work, such as that undertaken in the USA through the 1980s and 1990s by Peter Lamb and associates (Mjelde et al. 1993, and references therein), suggested that the financial returns to be expected from seasonal prediction could be substantial. Practical experience, such as that gained in the 1990s using seasonal predictions in the Nordeste region of Brazil, an area with some of the highest seasonal rainfall predictability anywhere, supported the theory.

A further boost to information delivery was given by the major 1997/98 El Niño event, which happened to coincide with the commissioning of the TAO array of moored buoys straddling the equatorial Pacific, with the maturation of the numerous ocean prediction models using TAO data, and with the first of the Regional Climate Outlook Forums (RCOFs), held to deliver information into tropical countries most influenced by El Niño events.

However progress since has been more constrained than appeared to have been promised by these early successes. A number of dramatic predictions of the consequences of the developing El Niño event openly and widely broadcast in 1997, often taking advantage of the emergence of the Internet, were felt to have been incorrect. The 1997/98 event, although unarguably one of the largest on record in terms of its intensity and effects, failed to impose the 1982/83-style "classic global climate anomaly configuration" on which these predictions were based. Confidence was eroded and questions were raised concerning the free and open distribution of independent and sometimes contradictory predictions. Scientists pressed the need for presenting predictions as probabilities, a concept that immediately raised a barrier to understanding and acceptance for some users. And recipients did not always gain the assurance necessary to incorporate this new prediction information into their decisions; many recognised that a false decision might have long-term effects that might be difficult to reverse. In the worst cases gambler's ruin beckoned. The initial positive results from the Brazilian Nordeste proved difficult to duplicate even in this same region, with later spectacular forecast failures in the region severely denting confidence (Lemos 2003; Meinke et al. 2006).

The science has now entered perhaps a period of consolidation. There is no doubt that the predictions have measurable skill in the technical sense, and experiments such as PROVOST and DEMETER have demonstrated certain levels of technical skill beyond the preliminary expectations of participating scientists (for example over Europe, where earlier research had indicated minimal, if any, predictability). Prediction models continue to be improved, new sources of prediction skill are being examined – in part through the COPES (Coordinated Observation and Prediction of the Earth System) experiment in which research into seasonal predictability originating in land surface soil moisture, ice cover and stratospheric circulations is being assessed – and new activities generated to introduce prediction information to additional user groups and sectors. Yet, as indicated earlier, it remains unclear that maximum value is being extracted from the current skill levels of the predictions. In part the apparent lack of value in seasonal forecasts almost certainly results from non-optimal incorporation of climate information into the decision matrix of climate-sensitive enterprises. Therein lies a key to the delivery of the societal benefits inherent within the science.

2.3 Climate-Related Decision Making Under Uncertainty

The proposition that, from the societal point of view, decision making is the ultimate goal of seasonal to interannual climate prediction has been emphasized already. It has also been highlighted that climate predictions – and climate information in general – will be just one component in most decision making processes (see Fig. 2.3). Most important of all, however, is the fact that climate prediction is inherently probabilistic in nature and probabilities always indicate uncertainty in the final outcome (this fact will be stressed many times throughout the book). Decision makers who make use of such predictions need to factor in this intrinsic uncertainty. Defining a practical framework for taking uncertainty into account in order to assess the level of risk associated with decision making processes is the subject of this section.[3] Such a framework is based on decision analysis, a subject developed under the discipline of decision theory.

Decision theory is a body of knowledge and a related set of analytical methods of different levels of formality designed to assist decision makers in choosing a course of action from among a set of alternatives through a careful consideration of the possible consequences of each alternative. In turn, decision analysis is essentially concerned with breaking complex problems into manageable parts, by adopting the 'divide and conquer' approach. A large body of work has been developed in the field of decision analysis, and only its surface will be scratched here. A good reference for a deeper understanding of the subject is provided by Goodwin and Wright (2003).

Two of the most important tools in decision analysis are decision tree diagrams and influence diagrams. These are two tools that attempt to model the decision making process by illustrating graphically the alternatives, uncertainties, risks and

[3] The concepts of risk and uncertainty, while related, are very different: uncertainty involves variables that are constantly changing, whereas risk involves only the uncertain variables that affect or impact the system's output directly (Mun 2004). Note, however, that not everyone finds uncertainty, and its associated probabilities, easy to incorporate into their decision making processes.

objectives of the problem at hand. By offering a visual representation of the decision problem, these tools are helpful in clarifying the various steps in the decision process in ways that can lead to creative thinking and to the identification of issues not previously considered. These features make these two diagrammatic approaches appealing to decision makers faced with complex decisions. Decision making in general involves multiple objectives as well as multiple stakeholders. In order to simplify the treatment, only a single objective and a single stakeholder are considered here.

The decision tree diagram is a flow diagram that includes the timing of decisions, coverage of uncertainties, and quantification of each possible decision. Once the objective of the decision has been identified, the decision tree analysis requires five steps:

1. Determine all possible options and risks related to the problem
2. Calculate the consequences of all options
3. Determine the uncertainty associated with each option
4. Generate a tree diagram using the information from the first three steps
5. Assess the best course of action

In societies that are driven mainly by economic considerations, the numerical quantity that expresses the objectives of the situation, and summarises the outcomes of all the options, is money. In principle, however, there is no reason why other quantities could not be used; for instance the number of people at risk of starvation due to a possible drought, or measures that are problematic to quantify, such as the effects on the environment of particular management options (e.g. desertification, salination, erosion, etc.).

The graphical representation of a decision tree diagram is made up of activity forks or decision nodes (a square) and event forks or chance nodes (a circle). The use of a triangle to terminate a branch in the tree is customary. An activity fork is used when a definitive decision amongst two or more options is required, whereas an event fork is used when the option is subject to uncertainty. Given the complete tree diagram, the best course of action is determined by considering the implications of each option starting from the right of the diagram and moving to the common start of the tree, towards the left. This process of evaluation of the best action plan decision is referred to as “folding back” or “pruning” the tree.

Referring to the food security example (Fig. 2.3), it is possible to construct a highly simplified decision tree diagram by considering only three information streams: “Crop status information”, “Climate information, including seasonal prediction” and “International crop prices and availability information”. Imagine the following situation: one million people may be at risk of starvation – the risk is dependent on the amount of food in the reserves and on the predicted climate conditions. In order to decide on the best course of action (i.e. to reduce the risk of starvation by providing the population with sufficient food for the coming season) a decision tree diagram might be built, as in Fig. 2.4. If the crop reserves are sufficient, then no action is required.[4] However, if reserves are insufficient available options need to be assessed. It is assumed here that the only two accessible additional pieces of information are a seasonal climate prediction and international crop prices. The climate prediction offers a 30% chance that rainfall will be sufficient to produce enough crops to meet national demands. For crop prices there are two options: one is to buy crop in advance, the other is to buy it after the cropping season has started. In the former case the cost is, say, €10 million, in the latter €30 million. So, for unfavourable predicted climate conditions, the options are to buy now and spend €10 million or buy later and spend three times as much. In the case of favourable conditions, the options are to buy now or to hedge, e.g. by purchasing insurance or by buying part of the crop that might be needed. In the case of hedging, it is assumed that costs of either alternative are €2 million.

The best course of action is then given by the branch with the associated lowest expense or, in the commercial parlance, the largest profit. By “pruning” the branches of the tree, one obtains the monetary values as presented. The only value which needs some explanation is €7.6 million. At each node, the value before that node is calculated by considering the probability of each branch following the node. This probability is multiplied by the amount on the corresponding branch and then summed over the contributions from all branches. In this case there are

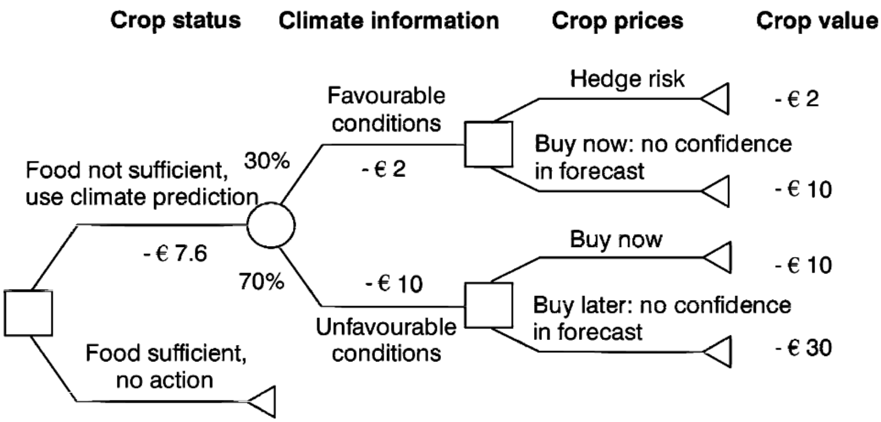

Fig. 2.4 Example of a decision tree diagram with reference to the food security application of Fig. 2.3. Squares represent decision nodes and circles chance nodes. The use of a triangle to terminate a branch is customary. Amounts are in million of euros. This is a highly simplified decision tree, purposely constructed to focus on its mechanics (see text for details)

[4] For simplicity, options such as building national food reserves or generating foreign income through sales have been ignored.

only two branches after the (single) chance node and so the value before that node is (€2 × 0.3 + €10 × 0.7) million = €7.6 million. In evaluating this problem, it has been assumed that the decision maker is risk-neutral. The results generalise however to arbitrary risk attitudes of the decision maker, whether they are risk averse or risk seeking. The attitude to risk may be assessed by eliciting a utility function.

In the present example, it is straightforward to assess what is the most convenient action when food reserves are not sufficient, and this is to spend €2 million to hedge the risk given a favourable prediction, or to buy now otherwise. However, it is also true that only a single estimate of costs was provided. In practice, because uncertainty generally exists in the various options forming the decision tree, sensitivity analyses are conducted with the aim of providing error estimates associated with all possible outcomes. A more meaningful evaluation of the risk associated with the selected course of action would thus be obtained. It is important to note that the use of expert advice or judgment – seasonal prediction in this case – in event forks generally sharpens the uncertainties associated with the options in that particular fork. Probabilities would be equal in the absence of any information, including of historical records,[5] i.e. 50–50% instead of the 30–70% (see Fig. 2.4) coming from the knowledge of climate information. The procedure used to incorporate expert advice in the decision making probability assessment is referred to as the Bayesian approach. A discussion of Bayesian theory is given in Chapter 9.

In practice, situations tend to be rather more complex than that shown in Fig. 2.4, as can be inferred from the number of entries in Fig. 2.3. The number of possible options would grow substantially were the simple decision tree of Fig. 2.4 generalised to take into account all the entries in Fig. 2.3. The rapid growth of complexity represents a drawback of tree diagrams as they can become difficult to follow or to validate.

The decision making problem is further complicated when different entries in the tree are interdependent; for example in the food security case above the act of issuing a public climate forecast may affect crop prices directly. At first glance tree diagrams appear to represent an end-to-end process, in that they flow sequentially from left to right; a closer examination shows, however, that a diagram can become highly interactive due to the interdependence of the various processes.

An alternative approach to decision trees is the use of influence (or relevance) diagrams. The high-level (compact) visual representation of influence diagrams makes them particularly valuable for the structuring phase of problem solving, and for visually representing large, intricate problems. The complexity of the details present in decision trees becomes embedded into the general structure of influence diagrams, structure which clearly calls attention to the relationships between the

[5] In practice, it is virtually impossible not to be able to have access to some additional prior information. Any such information would modify the prior 50–50% probability.

various elements of the problem. As a consequence, influence diagrams can cope with situations in which there is substantial sophistication and complexity. All these features make influence diagrams easier to interpret and overall more powerful than tree diagrams. Indeed, some authors contend that decision trees should only be used as a teaching device for beginners. Note that a 'properly' formed influence diagram can always be converted into a decision tree (Howard 1990, explains the rules to build a 'proper' influence diagram).

The symbols used for influence diagrams are similar to those for decision trees, but with some differences. A decision node, drawn as a rectangle, represents a variable under the control of the decision maker; an uncertainty node, drawn as an oval, represents a variable not directly controlled by the decision maker; a deterministic node, drawn as a double oval, represents an uncertainty that is functionally determined by the variables influencing it; and a value node, drawn in many ways, including a hexagon or a diamond, represents the variable to be optimised by the decision. The nodes are connected to each other via arrowed arcs, which generally indicate 'relevance'. An example of an influence diagram with reference to the food security application of Fig. 2.3 is shown in Fig. 2.5.

This example is just one of the many possible combinations for linking the various variables or options of Fig. 2.3. It is important to recognise that decision tree diagrams and influence diagrams are never unique in the sense that they aim to 'model' the experts' natural thought processes. It is therefore crucial to elicit information for the specific problem at hand from as many experts as possible, also to try to ensure that wider economic, social and environmental considerations (i.e. the three pillars of sustainability) are taken into account. For example, in Fig. 2.5

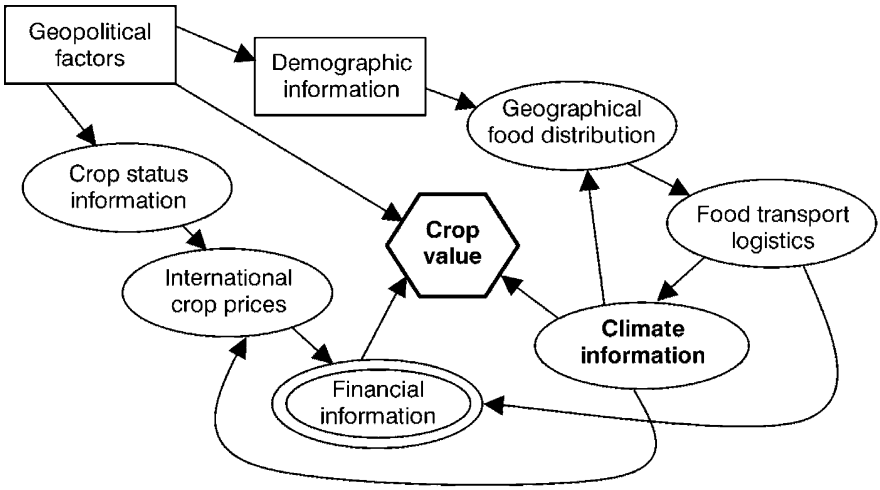

Fig. 2.5 Example of an influence diagram with reference to the food security application (see Fig. 2.3). For the sake of clarity not all the dependencies have been represented in this diagram (e.g. the link between 'crop status information' and 'climate information')

the customary choice to assign certainty to the "Geopolitical factors" and "Demographic information" variables has been made, and the variable to be optimised is "Crop value", as in the decision tree of Fig. 2.4, but this problem could be stated in other ways.

To summarise, the two most relevant concepts highlighted by Fig. 2.5 are:

1. Climate information is just one of the components in the decision making process
2. Climate information enters the decision process under several facets through its interdependency with other information streams

2.3.1 A Holistic Approach to Seasonal Climate Prediction

A direct corollary to the decision analysis presented in the previous section is that climate information has to be considered within its broader context, especially the social and economic aspects. This holistic approach, defined as *Climate Affairs* by Glantz (2003), is helpful not only in understanding and managing the many ways in which climate variability influences human activities and environmental processes, but also in identifying how societal and environmental issues not related to climate may act as confounding factors to the climate information in the decision making process. Indeed, the concept of Climate Affairs was developed with the aim of placing climate and climate-related factors on the list of items that decision makers should take into consideration. Climate Affairs consists of the following component fields:

- *Climate science*: the description of the components of the physical climate system, including the role of human activities as forcing factors to the system
- *Climate impacts*: the impacts of the climate on both societies and ecosystems
- *Climate politics*: the process needed to produce climate-related regulations and laws
- *Climate policy and law*: the legal and regulatory aspects of climate–society–environment interactions
- *Climate economics*: the financial aspects of the climate, including cost-assessments carried out in order to assist in the decision making process
- *Climate ethics and equity*: the set of principles of right conduct and the state of being just, impartial, and fair in the context of climate-related impacts (e.g. the poor generally have fewer options than the rich in tackling climate-related harmful events)

Food security, as discussed in the previous section, is one of the sectors that would greatly benefit from a holistic approach, especially in regions with large interannual climate variability. By definition, climate-related issues are crucial in such areas, but many problems also arise because of the pressure to exploit these areas, as well as through other human choices and *perceptions* of acceptable risks.

It is worth noting that the concept of Climate Affairs goes beyond mere applications to decision making. Its other purposes are to encourage education and communication on climate-society-environmental issues, and to improve understanding of how climate variability affects society and the environment. Despite the feasibility of assessing risks attributable to climate in a physical sense (e.g. probability of a drought, a quantity directly usable in decision analysis), there are also many societal aspects which are difficult to quantify, and thus subjective judgment often plays a significant role in decision making. It is through focussing on education and by developing appropriate communication approaches that the level of arbitrary subjectivity of decisions may be reduced.

The subjective role of climate information was emphasised by the use of the word *perception* above. The perception of the climate (and of its prediction) is often distinct from its physically measured characteristics. A prediction for a colder winter than normal, for example, can be interpreted in many different ways, depending on the person to whom this prediction is addressed. A perception is often related to the association a person makes to his/her most memorable cold winter. Thus perception relates more to the psychology of the decision maker rather than to ignorance or lack of information[6] (Weber 2001; Loewenstein et al. 2001). The notion of perception is critical to any decision making process is further explored in Chapter 11.

In the following specific perceptions of climate by society are discussed briefly. There are three non-exclusive ways in which climate may be perceived by society: as a hazard, as a resource and as a constraint (Glantz 2003).

The hazard component is probably the most common way in which society tends to view climate, especially when high impact events, such as devastating floods or persistent droughts, hit the headlines. The hazard perception is particularly strong for governments, since climate-related harm to a population may have repercussions on a government's duration in power or on its likelihood of re-election. For governments, it may be more relevant therefore to be concerned with climate-related disasters than with enhancing climate as a resource.

Despite the perceived presence of this sword of Damocles, societies around the world view climate as a resource too. In fact people's lives and activities, including commerce, are adjusted in general to the expected flow of the seasons in order to take advantage of local climate conditions.

There are also environments in parts of the globe to which humans are less able to adapt. These environments are characterised by conditions where climate is seen as a constraint, an impediment to productivity or even to survival. Such is the case in marginal agricultural areas, for example, where annual rainfall averages are low and interannual climate fluctuations in precipitation are so large that production may be meagre in some years, at which times it may be accompanied by

[6] A linked issue is that of legitimacy, which concerns the perception that the system is being provided in the interests of the stakeholders and those of the providers.

economic losses or even starvation. When such constraints are occasionally relaxed, e.g. in the rare ‘good’ years of the semi-arid tropics, decision makers are presented with opportunities that need to be exploited: in the rain-fed wheat systems of Australia, for instance, 70% of profits are made in 30% of the years.

Certain global phenomena, such as the seasons or ENSO, may lead to all three perceptions of the climate across diverse locations. ENSO, for instance, is perceived as a hazard in some regions, as its occurrence is associated variously with droughts and floods. Similarly, ENSO may be seen as a resource in those regions where its outcomes are beneficial (e.g. warmer winters in Florida during La Niña events; see Chapter 12). Finally, ENSO is a constraint on productivity and/or security in places where resources are not sufficient to cope with its consequences. The objective of factoring seasonal to interannual predictions into decision making processes is then one of modifying perceptions towards reducing losses related to the hazard and constraint components, and to increasing gains related to the resource component.

2.4 Identifying the Users and the Uses

For seasonal to interannual forecasts to be of benefit to society it is imperative to identify clearly the users as well as the context of each use. Numerous users are likely to be interested in decision-making processes for which seasonal forecasts might be relevant, sometimes beyond those directly affected by climate variability. One example is that of crop switching (e.g. planting sorghum instead of cotton) or by using superior drought-adapted varieties when water-deficient conditions are expected. Here we give a brief overview of some potential direct uses of predictions; in the following section a more focused discussion from the perspective of the developing world is provided. Where possible examples are provided, firstly, of ground-level decision processes, and secondly of decision processes at national and international level. Note that while the list given below may offer the impression of independence, some of the sectorial examples nevertheless may be interconnected: thus there is a need for appropriate interaction amongst sectors if optimal decisions are to be attained. A second factor to recognise is that, while all examples in the following are quoted within positive contexts, seasonal forecasts may be used, say by traders, to the disadvantage of those unfortunate to live and work in negatively affected (e.g. drought) areas.

2.4.1 Agriculture

Agriculture, including both plant cultivation and livestock production, is a sector heavily dependent on climate, such as in the amount and timing of rainfall, the occurrence of damaging frosts, the length of the growing season, and the number of growing degree-days. Seasonal forecasts would therefore assist pre-emptive actions, such as the use of varied crop species, or the altered composition and/or allocation of browsing herds for more effective exploitation of marginal areas. Thus improved use of climate information in agriculture could increase profitability and sustainability by allowing farmers to match cropping decisions to expected climatic conditions (Stern and Easterling 1999). At the national, regional and international levels, matters of food security, of international crop yield estimation and food flows, and of food marketing can be informed through climate services. Practical examples of the uses of seasonal forecasts in agriculture are discussed in Chapter 12.

2.4.2 Disaster Forecasts and Prevention

Natural disasters associated with extreme climate events, resulting in the loss of life, destruction of shelters and food reserves, disruption of food production and transportation systems, and health risks are situations faced by large parts of the world population. General systems of emergency preparedness and response, such as early warning systems, might benefit and avoid costly damages (see Chapter13). Seasonal forecasts could play a role in warning systems in cases in which their skill was judged to be at levels sufficient for alerts. International preparations for disaster response might also take advantage of climate information.

2.4.3 Energy

Most forms of energy production (e.g. gas and hydropower) and the level of energy consumption are, to varying degrees, affected by climate conditions. Using seasonal forecasts as input for load-balance models could potentially decrease the overhead necessary to maintain the agreed baseline energy availability, thus helping to optimize the matching of supply and demand. They might also be used for planning international energy transfers.

2.4.4 Finance and Insurance

Climate information can be used in the financial sector to optimize capital requirements, and to hedge the risk of financial losses due to climate-related events. For example, seasonal forecasts can be used by an energy company to optimize the use of climate-linked financial products designed to reduce the potential impact of adverse weather conditions on the company's balance sheet, or by an insurance company to assess its exposure to climate-related risk. Insurance is now being tested in response to climate risk management in the developing world.

2.4.5 Fisheries

Fish population fluctuations, whether due to climatic factors or to harvesting or to other reasons, are by their nature more difficult to analyse. As a consequence, for fishing management, which normally aims at constraining both biological and economic overfishing, it might be more challenging to use forecast information effectively. One notable exception might be that of Peruvian anchovies whose population is highly influenced by El Niño events.

2.4.6 Food Security

Food security is naturally related to the agricultural examples given earlier. Droughts, floods and cyclones are some of the essential factors in determining the quantity and quality of food supply, also referred to as food security. Food security is particularly an issue in regions where the interannual climate variability, especially in rainfall, is large and local production is the main food supply. For such regions, rainfall forecasts could help alleviate problems in low rainfall years. It must be noted, however, that climate is only part of the story in food-hardship periods; confounding factors such as political situation or locust infestation may contribute to exacerbate the problem (e.g. the 2005 food crisis in Niger).

2.4.7 Health Management

Human health is sensitive to several types of climatic variations. For some diseases close direct and indirect links with climate conditions exist (e.g. malaria epidemics). In such cases, climate forecasts might give public health systems early warning of the likelihood of epidemics. For instance, tropical disease risk management is an application in which the use of climate information is receiving increasing attention. Health planners need information on the predicted level of risk for malaria, meningitis, or cholera epidemics to develop. International strategies for improving health and for relevant pandemic responses would benefit from an enhanced use of climate information.

2.4.8 Hydrology and Water Resource Management

Water managers may benefit from rainfall forecasts for the planning of irrigation systems, surface water storage, groundwater pumping capacity and trans-basin diversion. Such forecasts might also contribute to a more effective deployment of emergency flood management and relief operations. Information on climate variability, including predictions, can form an important knowledge source in decisions on water security, facility development, and cross-border basin management.

2.4.9 Policy Making and Public Authorities

Relevant public institutions have the potential to influence the way in which individual users (e.g. a farmer) respond to climate forecasts. Such institutions may act via, for instance, proper dissemination of the forecast or by offering incentives, the latter possibly coupled with some form of insurance to spread the risk of responding to probabilistic forecasts.

2.4.10 Retailing Industry

The impact of the climate variability is seen across many areas of the retailing industry (e.g. ice creams, refreshing beverages or air conditioning units, summer or winter clothes). By taking climate forecasts into account, customer demand could be better predicted. In turn, this would mean making the most out of sales and reducing waste through efficient delivery, staffing and stock control. Climate forecasts may influence decisions about provisioning of a particular product, for instance coffee imported from Indonesia rather than Brazil or Central America, or vice versa, although such decisions can be detrimental to the livelihoods of the coffee-growers and their communities.

2.4.11 Transport and Tourism

Climate information is potentially useful for planning in the operation of leisure facilities as well as strategic planning and investment. Predictions of tropical cyclone activity or anomalous climate conditions could, for instance, be used by transport planners and resort owners to prepare for potential impacts such as for storm damage. Equally, these predictions could be used to inform tourists of the likely risks they would incur by travelling to specific regions. Current thinking has it that using climate forecasts for leisure planning may involve risks of litigation (e.g. by resort facility owners or washed-out tourists). It should be noted, however, that the main difference with the sectors mentioned earlier is that in this case information is provided to the general public also. As with the issuing of weather forecasts, experience needs to be built on ways to communicate climate forecasts, along with their uncertainties, to wider audiences.

2.5 The Importance of Climate in Key Development Sectors[7]

Three top priority development sectors are particularly sensitive to climate variability, namely agriculture, water resources and health. The situation is most critical in Africa where the livelihoods of hundreds of millions of people are extremely vulnerable to climate variability. Much improved climate risk management is essential to support more effective development and to help mitigate disasters.

Climate exerts a profound influence on the lives of poor people who depend on agriculture for their livelihoods and sustenance, who are unprotected against climate-related diseases, who lack secure access to water and food, and who are vulnerable to hydro-meteorological hazard. For vulnerable communities, developing flexible, proactive responses to climate variability that enhance resilience is both a crucial step toward achieving the MDGs by 2015, and a foundation for coping with the uncertainties of a changing climate into the future. Furthermore, because climate has a confounding influence on many development outcomes, attention to climate variability is essential for evaluating real progress.

2.5.1 Agriculture in Africa

All current initiatives for development in Africa emphasize the overriding importance of agriculture, both for eliminating hunger, and also as a local and national economic driver. The Millennium Project proposes major scaled-up interventions to enable smallholder agriculture to develop and sustain itself throughout the poorest regions of Africa. These interventions then are designed to be coupled with a ‘safety net’ to protect communities and local economies in disastrous years, so that gains made in better years are not wiped out by unfavourable seasons, as happens so often at present. Such an ambitious programme, designed to bring a

[7] This section is derived largely unaltered, with permission of IRI, from text originally produced for their Position Paper entitled ‘Sustainable Development: Is the Climate Right’, that was published in 2005 and in which a predominantly African perspective was taken; nonetheless this section can be read within a wider geographical context.

hundred million people out of poverty by 2015 in sub-Saharan Africa, explicitly recognises the importance of climate variability in its proposals.

2.5.1.1 Subsistence Farmers

For too many people in Africa, subsistence agriculture is a desperate form of poverty akin to slavery that requires major effort for relatively little return. With reduced fallow periods, smaller farm size, declining soil fertility, lower yields, increasing indebtedness and isolation from markets, such farmers have relatively few choices even before rainfall variability, crop pests and diseases, malaria, AIDS and emigration of young labour make their lives even more onerous.

2.5.1.2 Cash Crop Farmers

Farmers that engage more with local markets and dealers, can also access credit and buy inputs (improved seeds, fertiliser, sprays for pest and disease control) to increase the value of their labours. Such farmers tend to be less risk averse and more proactive in their management choices, and as such, are in more of a position to access and take advantage of weather and climate information, particularly in their choice of seeds and other agricultural inputs.

2.5.1.3 Risk Benefit

Communities who depend on rain-fed farming for sustenance and livelihood in high-risk environments are among those most affected by climate variability, but conversely are also often particularly well poised to benefit from improved management of climatic risk through appropriate use of climate information. It is important to empower rural populations to better manage risk and exploit opportunity by (a) providing relevant, timely information to the target populations; (b) fostering and guiding adaptive management responses; and (c) addressing resource constraints to adaptive responses.

2.5.1.4 Managing the Rural Economy

Without a healthy rural economy, farming communities cannot get the inputs they need to cope better with climate variability and so the cycle of poverty is perpetuated. There are many ways that governments can improve the rural economy (see the Millennium Project proposals for example) in ways that are sensitive to prevailing conditions. For example, modern methods of monitoring crop production from satellite are now routinely used in most regions of Africa. Coupled with seasonal climate prediction, these enable early yield estimation, extend the lead-time of food stock or relief decisions, and facilitate timely implementation of measures to help ensure local food security or cope with harvest surpluses. And knowing in advance the risk of food shortfall/surplus is very important information for central government economic advisers, and local government planners, in order to make contingency arrangements.

2.5.2 Water

Improved water management is recognised as a fundamental requirement for development. In the Africa Water Vision for 2025 the key problems identified are:

1. The multiplicity of transboundary water basins
2. Extreme spatial and temporal variability of climate and rainfall, and climate change
3. Growing water scarcity

It is of prime importance for people in Africa today, and tomorrow, that water from rainfall is managed more effectively. In order to achieve this, the most important step is to ensure that rainfall variability is not simply accepted as an inescapable 'fate from the gods'. Rather, rainfall needs to be regarded as an environmental variable that is influenced by increasingly well-understood physical processes. As such water supplies can and must be managed better by a whole host of decision makers in the diversity of economic and social domains affected by fluctuations in availability.

2.5.2.1 Transboundary River Systems

Much ocean induced climate variability affects large areas of Africa and its effects are particularly noticeable at the scale of transboundary river basins. Attempts are being made through the African Network of Basin Organisations to improve management and decision making in all transboundary river systems, to encourage greater cooperation between stakeholders, and to mitigate flooding and reduce competition and conflict over access to water. To achieve these objectives effectively, it is absolutely essential to incorporate knowledge of seasonal water variability into decision making, and where appropriate, early warning through seasonal forecasting. Capacity building in water authorities to enable people to use these increasingly powerful tools is essential.

2.5.2.2 Reservoir Management

Reservoirs are designed to hold significant amounts of seasonal runoff to mitigate the effects of upriver rainfall variability. The Aswan dam in Egypt provided irrigation through 10 years of drought and sub-normal rainfall over Ethiopia in the 1980s. Very often, however, there are conflicting demands on reservoir managers to provide water for hydropower, irrigation and to manage flood and base flows for the health of lower river communities and ecosystems. Without knowledge of future rainfall, reservoir managers inevitably tend to be conservative. With reliable indicators of future rainfall quantities, reservoir managers are in a better position to make best use of the limited stored water available. Such decisions involve risk, and managing risk is an essential component of making the best of a scarce and highly variable resource such as water.

2.5.2.3 Summary

Climate variability not only affects the design and management of water and sanitation infrastructure, but also plays an important role in the planning and design of water resource systems. It is essential that knowledge of climate variability be incorporated in water management strategies at all timescales, as an integral part of knowledge-driven decision-making: optimal system management is impossible without it.

2.5.3 Health Management

The European heat wave of 2003 had a dramatic impact on mortality causing an excess of about 15,000 deaths in France of which about 1,000 were in Paris alone. The consequences of unusual warm years in Africa pass largely undocumented. We all know from direct personal experience that dry-season illness tends to be different from wet season diseases. But how much does the overall incidence of disease, and hence death rates, depend on climate variability, and hence fluctuate from season to season and year to year? The answer is ‘climate has an enormous impact on health’ and many diseases are recognised by the World Health Organization (WHO) as being climate sensitive. These include: influenza, diarrhoeal disease, cholera, meningitis, dengue fever, chikungunya, avian flu, Rift Valley fever, leishmaniasis and malaria.

2.5.3.1 Malaria

Malaria is widely appreciated as the most important of the climate-sensitive diseases in the world. It is seen as a major impediment to socio-economic development particularly in Africa where 90% of the 1–3 million deaths it causes each year occur. If we are serious about reducing malaria, and associated maternal and child mortality as part of the Millennium Development Goals, then information on the seasonality of climate and its variability must be taken into account when planning and implementing routine health campaigns and epidemic preparedness.

It is estimated that more than 110 million people in Africa live in regions prone to malaria epidemics. The populations affected have little acquired immunity to malaria and are therefore vulnerable to explosive epidemics that can cause high case fatality rates among all age groups. In spite of the severity and the magnitude of the problem, understanding of epidemic malaria is very limited and almost nothing is known of the economic burden of malaria epidemics in sub-Saharan Africa.

For malaria climate is the primary factor in determining at least some epidemics.

- **Temperature** influences development rates of both the malaria parasite and its mosquito host. Higher temperatures, but only up to about 40°C, shorten the parasite extrinsic incubation period and increase the stability of disease transmission
- **Increased rainfall** in semi-arid areas increases availability of breeding sites and therefore augments malaria vector populations if temperature is favourable. It is also associated with increases in air humidity that result in higher adult vector survivorship and therefore greater probability of disease transmission

Epidemics frequently occur when periods of drought (during which people can lose immunity to the disease) are followed by a return to normal or above normal rains in the more arid regions. Combining information on malaria trends and vulnerability with rainfall information can provide warnings for high transmission years prior to the peak malaria season. For example, the case of Botswana has demonstrated a strong impact of December–February rainfall on malaria incidence anomalies, which make it possible to alert the Ministry of Health of increased risk of an epidemic before the peak transmission period of March and April. Seasonal climate forecasts can supply even earlier warning of changes in malaria risk. A seasonal climate forecast in November can provide information about the expected extent of the next malaria transmission period 5 months before the peak of the malaria season and 3–4 months earlier than warnings that are issued based on observed rainfall. Prime interventions include planning integrated vector management; awareness raising campaigns allied to education, as well as timely procurement of drugs.

Part II
The Production of Seasonal Climate Forecasts

Chapter 3
Overview of Seasonal Forecasting

David L.T. Anderson

In many regions of the globe, the largest climate signals after the seasonal cycle are those associated with El Niño and La Niña. These are manifestations of a coupled process in which climate changes occur in both atmosphere and ocean. The origins of El Niño and La Niña (or ENSO, the El Niño/Southern Oscillation) lie in the tropical Pacific, but the effects can be felt to some degree almost globally. It is in the equatorial region that coupling between atmosphere and ocean is strong. In middle latitudes the coupling is much weaker. Models of varying complexity have been developed to study ENSO. So-called intermediate models, where the atmosphere is grossly simplified and only the upper ocean is modeled, have been used to understand the role of equatorial waves in setting the timescales of ENSO as well as being used for seasonal prediction. Seasonal forecasting using complex models of the atmosphere and ocean is relatively recent. The atmospheric component of such models is very similar to what is used in weather forecasting. This is a much more mature science and there is some synergy between weather and seasonal forecasting. Both are initial value problems in the sense that the information on which a forecast is based depends on these starting conditions. The atmospheric initial state for a seasonal forecast is generally provided from the atmospheric state created to initialize a weather forecast. So seasonal forecasting relies on weather forecasting in that sense. There is a much more important reliance on weather forecasting however, which involves the ocean and atmospheric reanalyses. If there were sufficient observations of the ocean, then it would be possible to create ocean initial conditions just from the observations, but this is not the case. Even with today's observing system it is still necessary to augment the ocean observations with knowledge of the ocean gleaned from the past history of the ocean forcing (momentum, heat and freshwater fluxes), as will be discussed in. Some results from complex models are presented to allow an assessment of skill and to contrast the predictability of the tropics with the extratropics. The potential importance of using multi models is introduced which allows some assessment of the importance of model error. A developing field is the *application* of seasonal forecasts. For some applications, only a simple concept is required, but for others quite

---

David L.T. Anderson
European Centre for Medium Range Weather Forecasts

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

complex application models are needed. Developing these modules can be difficult as the functional dependence on weather parameters is not always obvious. One example is given but a full assessment of the difficulties in developing application models is left to later chapters (Chapters 11–13).

3.1 Introduction

Weather forecasting is a discipline familiar to all. Weather forecasts have a limited forecast range on account of the chaotic nature of the atmosphere (see Lorenz 1993); the predictability horizon depends on what variable one seeks to predict and on what scale. It is perhaps just minutes to hours for smaller-scale features such as thunderstorms, but might be as long as a week for large-scale weather systems. If the useful range of weather prediction is only a few days,[1] why then do we discuss forecasts to seasons or even longer? The reason is that we believe the ocean imparts some memory to the atmosphere, at least in some parts of the globe resulting from the fact that the ocean has a much larger heat capacity than the atmosphere and consequently is much more ‘inert’. Anomalies in sea surface temperature last for a few weeks or longer, depending on the spatial scale, whereas the timescale for weather is typically a few days.

The simplest picture of ocean-atmosphere interaction, however, is not of the ocean driving the atmosphere, but of the atmosphere driving the ocean. Hasselmann (1976) postulated that the upper ocean was driven by high frequency (stochastic) variations in atmospheric heat flux. The sea surface temperature integrates the heat flux forcing. That means that the ocean response is the integral of the forcing. If the spectrum of forcing is white (no preferred frequency), then the ocean response will not be white but red, i.e. with more energy at low frequencies. This is quite a good approximation in much of the world.[2] However, in the tropics we believe that changes in the atmosphere influence the ocean and changes in the ocean influence the atmosphere. The atmosphere still has much high frequency (stochastic) variability but in addition there are low frequency variations which result from ocean-atmosphere interaction. Most prominent of these processes are El Niño and La Niña.

---

[1] There are atmospheric phenomena with a longer time scale such as the intraseasonal oscillation (about 40–60 days), sometimes called Madden-Julian Oscillation (MJO), and some aspects of these may have predictability beyond 10 days.

[2] In Chapter 4 there is a fuller discussion of the upper ocean heat budget, including the role of advection in maintaining the local heat budget, and the role of the subtropics in maintaining the heat balance of the upper ocean in the equatorial region.

Figure 3.1a shows the surface pressure difference between the western equatorial Pacific-Indonesian region and the eastern equatorial Pacific. Originally, the index was based on Darwin and Tahiti pressures as these were the stations where there were long data records. The index was called the Southern Oscillation Index (SOI). Figure 3.1a shows the more modern EQSOI which is a better indicator of large-scale swings in mass between the western and eastern sides of the equatorial Pacific. Figure 3.1b shows the time variation of a measure of sea surface temperature (SST) in the central east equatorial Pacific (NIÑO3.4) and shows that in the ocean the dominant timescale is also typically a few years. Importantly, the EQSOI and the NIÑO3.4 records are very highly anti-correlated. What causes these massive readjustments of pressure and changes in SST and what sets the timescales?

The timescales of a few years come mainly from the ocean. Including ocean variability can give rise to enhanced atmospheric predictability if we are dealing with processes that depend on both media interacting. On the other hand, it is quite possible to have some memory in the ocean and some predictability of ocean variability but with little or no associated atmospheric predictability if the ocean is not

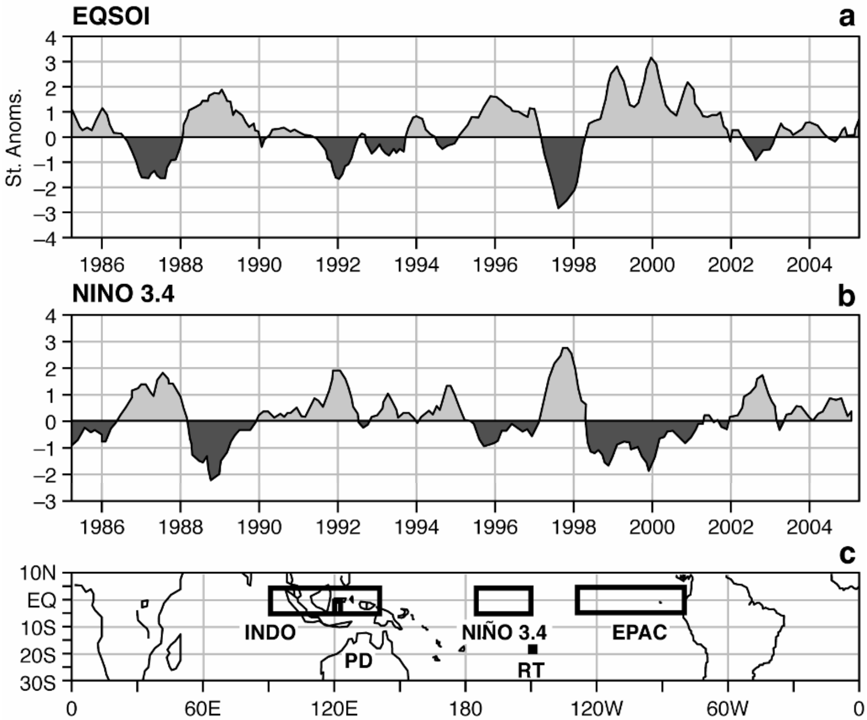

Fig. 3.1 (a) Plot of the EQSOI index as a function of time from 1985 to 2005. (b) Plot of the SST in region NIÑO3.4, (190–240°E/5°S–5°N) as a function of time. The dominant timescales in these two indices are very similar. (c) The locations of the two regions used in the EQSOI. From the Climate Analysis Bulletin, published monthly by NOAA. See: http://www.cpc.ncep.noaa.gov/

driving the atmosphere by the predictable part of its variability (Latif et al. 2002, 2006). So including the ocean in the forecast system does not of necessity lead to enhanced atmospheric predictability. The ocean may have greater predictability than the atmosphere but unfortunately, in general, we are more interested in atmospheric prediction than ocean prediction.

The coupling between the atmosphere and ocean is believed to be quite strong in the equatorial region, giving rise to the best example of climate variability on year to year timescale, viz. that associated with El Niño and La Niña or ENSO as it is frequently now referred to.[3] Although El Niño/La Niña are mainly located in the tropical Pacific, their influence can extend to almost all parts of the globe. Of course in distant regions other processes may also be affecting climate variability and ENSO may not be the dominant process. The predictability coming from ENSO might be quite weak in such a situation.

One conceptual model of weather is that of a series of events which are (for practical purposes) unconnected: the weather next week is essentially independent of the weather this week. An example of such a model is an unbiased coin. If such a coin were tossed several hundred times, one would expect to find short runs of one face or other, say five heads in a row, purely by chance. If the heads are thought of as inactive weather systems, then such a run of heads might correspond to drought conditions. There is no point in seeking a physical cause for such a run. The ‘drought’ is simply the outcome of a series of chance processes and as such it is unpredictable. But weather patterns may not always be purely stochastic. Suppose the coin were slightly biased. Then a sequence of tosses would still throw up heads and tails in pretty much random ways as before, but a more careful analysis might reveal that there were slightly more heads than tails, which in our simple analogue, would correspond to below average rainfall. Individual weather systems may still be chaotic, but the statistics governing them may have been altered by the bias in a deterministic and predictable way. In the earth’s climate system, it is thought to be the slower changes in the ocean sea surface temperature which are most important for imparting a degree of predictability. The amount of predictability is very much a function of position, with the tropics being more predictable than the middle latitudes.

ENSO involves a positive feedback between the SST gradient along the equator and the winds blowing along the equator. It also involves ocean dynamics: whereby information in the west equatorial Pacific can influence events in the east equatorial Pacific months later. To the extent that we know enough about the processes by which this information is propagated eastward via equatorial Kelvin waves and how these come to later influence the atmosphere one has a basis for prediction.

[3] ENSO stands for El Niño Southern Oscillation to reflect the importance of both the ocean and atmosphere. These are now known to relate to the same process though it took several decades to appreciate that this was the case and that they are manifestations of a coupled atmosphere–ocean process.

The tropical Atlantic and Indian Oceans may have zonal modes of this type too but they are less dominant than in the Pacific, less clearly identifiable against a relatively noisier background and of shorter duration. There is some indication that there may also be meridional modes of climate variability but again these have not been clearly identified. See Wang and Picaut (2004), McCreary and Anderson (1991), Neelin et al. (1998), Anderson et al. (1998), Philander (2004), and Chang et al. (2006) for some review articles on ENSO. See also the book edited by Palmer and Hagedorn (2006). Later in this chapter and in Chapters 4 and 6, we will consider how these equatorial waves propagate, theories of how they can give rise to ENSO, and their potential role in ENSO prediction.

In low latitudes, where the SSTs are high, the atmosphere exhibits convection throughout the depth of the troposphere, the location and intensity of which is influenced by the SST. In middle latitudes where the SSTs are cooler, there is less organized deep convection. Consequently SST variability in middle latitudes does not influence the atmosphere as strongly as at the equator. In most of what follows we will discuss primarily tropical processes associated with El Niño.

Not all aspects of climate influenced by El Niño are adverse. Indeed, from a North American perspective El Niño might well have a net beneficial impact (Goddard and Dilley 2004). Winter is warmer and so money could be saved on heating, hurricanes in the Caribbean are less frequent and so hurricane damage might be less.

This latter is a tricky issue. There is no doubt that a major hurricane striking Miami such as Andrew did in 1993 or New Orleans as Katrina did in 2005, causes huge damage – Katrina caused perhaps the biggest financial meteorological damage in history. But how does one show that such a hit occurred because there was a La Niña or avoided because there was an El Niño. This of course cannot be done in a definitive (deterministic) way but one might try by using models to determine the probability of strikes when there is and when there is not El Niño or La Niña. A convincing case on this has not been made as models for seasonal forecasting are not yet of a resolution able to simulate cyclone tracks accurately, though forecasting tropical cyclones in models is now beginning and will likely improve as model resolution increases. Even if the models were able to simulate hurricane genesis and tracking well, the answer to whether, e.g. New Orleans avoided a major hurricane strike because of El Niño can at best be probabilistic. Maybe one can say the probability of New Orleans being struck by a hurricane of category x is reduced by a factor y in an El Niño year.

For models to have realistic hurricane tracks needs quite high horizontal resolution, perhaps a horizontal resolution of about 150 km or higher. It has not yet been evaluated if there is skill in predicting interannual variability in tracks. Models do seem to have some skill in predicting the frequency of occurrence, however, and it has been shown that multi-models do a better job of predicting the interannual variability in frequency than individual models.

There are also processes which can affect the climate which are not generally taken into account. The big El Niño of 1982/83 was preceded by a major volcanic eruption in Mexico (El Chichon) from 28th March until 4th April 1982. A large eruption also occurred in June 1991 in the Philippines (Mt Pinatubo) ahead of the weak 1991/92 El Niño. These eruptions are essentially unpredictable on the seasonal timescale but it is not clear the extent to which they induce or influence ENSO. After they occur, models could be altered to reflect the amount and type of volcanic aerosols ejected, though operational models do not yet generally include this option.

3.2 Modelling the Coupled System

Models of various complexities have been used to represent the atmosphere, the ocean and the coupling in between. The atmosphere is well recognised as a highly complex system. How then can we simplify it? It might come as a surprise but the assumption that on the large scale, the tropical atmosphere is to a considerable degree a slave to the ocean seems to work reasonably well. In particular the assumption is made that if the SST fields are known, then so is the large scale tropical wind. For the timescales of relevance to El Niño or La Niña, only the upper ocean is involved. As we will see later the temperature structure of the ocean is approximately a mixed layer where the temperature is relatively uniform, beneath which is a region where the temperature drops rapidly, called the thermocline. Beneath that the temperature drops slowly to the deep ocean. This means that the simplest approximation to the ocean is that it consists of two layers, one active above the thermocline and one inert beneath the thermocline. A simple model based on these ideas is frequently used in oceanography, called a reduced gravity model or a one mode baroclinic model. A key feature of such a model with simplified vertical structure is that the speed of gravity waves (or Kelvin waves which have the same speed as gravity waves) is approximately $2\ m\ s^{-1}$. This is much reduced compared to external waves or barotropic waves.[4] (See for more description of equatorial waves and vertical modes).

McCreary and Anderson (1984) constructed a coupled model embodying the above simplifications. Figure 3.2 shows the evolution of the thermocline depth in a model when forced with a given wind stress, 1–3 months after the wind is

[4] The speed of propagation of barotropic or external gravity waves is $\sqrt{gH}$, with g the gravity acceleration $9.8\ m\ s^{-2}$ and H the depth of the basin considered, which is over 200 m/s in the deep ocean where the depth is over 4,000 m. Such waves are the agent by which the effects of earthquakes spread, for example the Tsunami of Boxing Day 2004, but are not very important for climate purposes and are frequently filtered from simplified models.

applied.[5] It shows four key processes: first, the movement of the thermocline, up in some places and down in others in response to the wind forcing, second, the

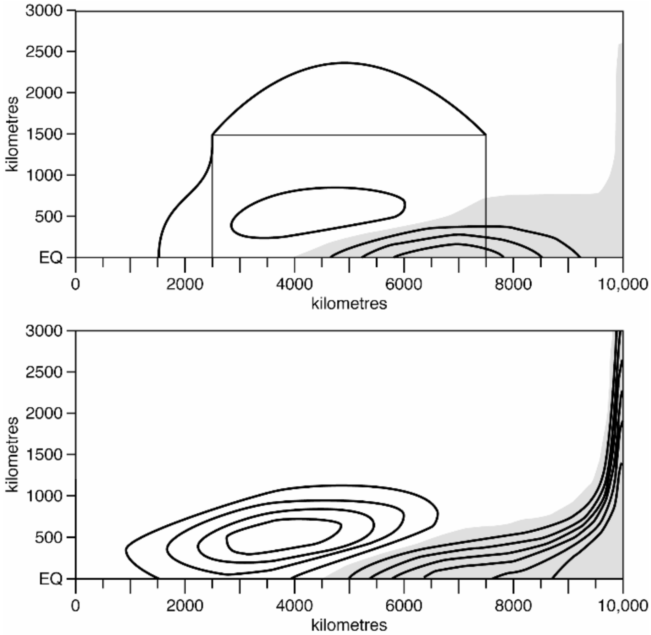

Fig. 3.2 Plot of the depth of the thermocline in a model of the Pacific Ocean. The geometry has been simplified to that of a 'box'. Only the northern hemisphere is plotted. The region where a zonal wind is applied is shown in the upper panel. The wind profile has also been simplified. The curve on the left of the box gives the latitudinal profile: the zonal wind is maximum at the equator dying away by 10N (1,000 km). The wind profile in the zonal direction is given by the curve above the box. The wind is maximum at 5,000 km dropping to zero at 2,500 km and 7,500 km. The model is initially at rest. The upper panel shows the response after 1 month. If the wind is anomalously westerly, then the shading corresponds to a deepening of the thermocline as would happen in the onset of El Niño. The asymmetric response with the equatorial signal propagating to the east is due to the Kelvin wave. After 3 months (lower panel) the signal has reached the eastern boundary and is propagating poleward along this boundary. The eastward asymmetry is very clear. Also evident is the shallowing of the thermocline off the equator. This signal has already started propagating westward as a planetary wave, sometimes loosely called a Rossby wave (From McCreary and Anderson 1984)

[5] Although the ocean is forced by the atmospheric wind, the heat flux and the freshwater flux (the difference between precipitation and evaporation), the wind is the key parameter for processes on the time scales up to a year or two, which is why we look at the ocean response to changing wind.

asymmetry in the equatorial response – the fastest wave is a Kelvin wave propagating only to the east along the equator. When it reaches the eastern boundary the wave splits with energy going polewards in both hemispheres. A third feature is the presence of energy propagating more slowly to the west in the form of a planetary wave, sometimes called a long Rossby wave. The fourth feature of note is that after a few months, a gradient is set up along the equator in the region of the wind forcing. (The key component of the wind is the one blowing along the equator). The balance between the thermocline slope and the zonal wind stress is called Sverdrup balance (see Chapter 4).

Figure 3.2 is a model result but with the development of real-time ocean analysis systems it is possible to detect and monitor these waves quite well. Although they are disturbances of the thermocline, they have a small (a few cms or in a big El Niño, a few tens of centimetres) signal at the surface. These can be detected in ocean analysis systems. With the advent of satellite altimeters which can measure the bending of the top surface of the ocean, it is also possible to detect them from space. Figure 3.3 shows the displacement of the top surface of the ocean from the ECMWF analysis system. There are many similarities between this figure and the very simple model result shown in Fig. 3.2. For example the region of strong pressure gradient along the equator, the poleward extension of the equatorial signal in both hemispheres along the eastern boundary, the region of opposite pressure gradient lying a few degrees off the equator, in both hemispheres though stronger in the northern hemisphere in the case of Fig. 3.3.

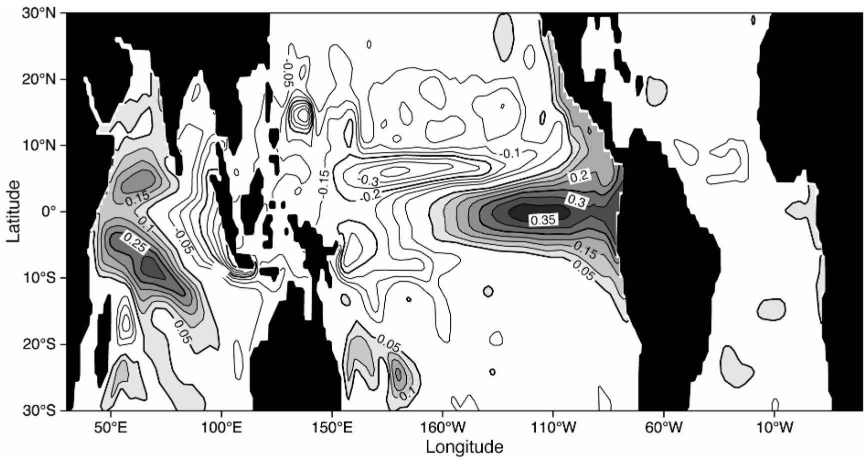

Fig. 3.3 Plot of the sea level from the ECMWF ocean analysis system at the height of the 1997/98 El Niño. The lifting of the top surface (corresponding to a depression of the thermocline) in the east is clearly visible, as is the propagation polewards along the eastern boundary. A depression of the surface near the dateline at around 8ºN is also clearly visible. The similarity to Fig. 3.2 is quite striking, despite the use of simplified wind patterns in the model

Although the model of McCreary and Anderson was rather simple it did capture several features of El Niño. A slightly more complex model was constructed by Zebiak and Cane (1987) and applied not just to simulating El Niños but to predicting them. In particular a retrospective prediction was made of the El Niño of 1982/83 when the model did rather well. Whether this was for the right reason or not can be debated. Regardless, their results greatly boosted the possibility of forecasting El Niño with physically based models.

Since these early days much more complex models have been targeted at El Niño simulation and prediction. Many models are capable of simulating several realistic features of ENSO though one can not be sure that any is truly realistic as even in nature no two El Niños are the same, making it harder to validate the models. The complex models, sometimes called CGCMs – Coupled General Circulation Models are now used routinely to forecast El Niño and changes to climate on a seasonal timescale in general. What is needed to forecast the climate a few months ahead with such models will be discussed below and in greater detail in Chapters 4–6.

It is instructive to look at the different characteristics of the atmosphere and ocean, especially in the tropical region. Figure 3.4a shows a diagram of the zonal wind stress along the equator versus time. This is quite a 'noisy' diagram, reflecting the high frequency nature of the atmospheric wind field. In panel (b) the depth of the 20°C isotherm is shown from the ocean analysis.[6] This is a much smoother field reflecting the fact that the ocean integrates the atmospheric forcing. One can clearly see the eastward propagation of Kelvin waves excited in the west Pacific. This smoother response is not just a feature of thermocline depth. Sea level for example shows very similar though not identical behaviour (panel c). Finally SST is plotted in panel (d). This field is also quite smooth but it does not show the eastward propagation so vividly demonstrated in panels (b) and (c). Some simplified models of ENSO use thermocline depth or depth of the 20°C isotherm as a proxy for SST (see Chapter 4). This plot shows that that is not a good approximation in general.

Figure 3.4 also shows one of the difficulties facing those predicting ENSO. There is no doubt that the MJO or Intraseasonal oscillation can be strong in the west Pacific and can generate ocean Kelvin waves which propagate eastwards. As long as these waves remain beneath the surface displacing the thermocline but not influencing SST they do not affect the atmosphere. Many intraseasonal Kelvin waves are damped in the eastern equatorial Pacific and do not lead to an El Niño or La Niña event. Some, however, do break the surface, generate SST anomalies which then can influence the development or demise of an El Niño or La Niña. At the present time it is not known how to assess the extent to which an MJO in the

[6] Near the equator, the depth of the 20°C isotherm is often used as a proxy for the depth of the thermocline since it lies roughly in the middle of the thermocline.

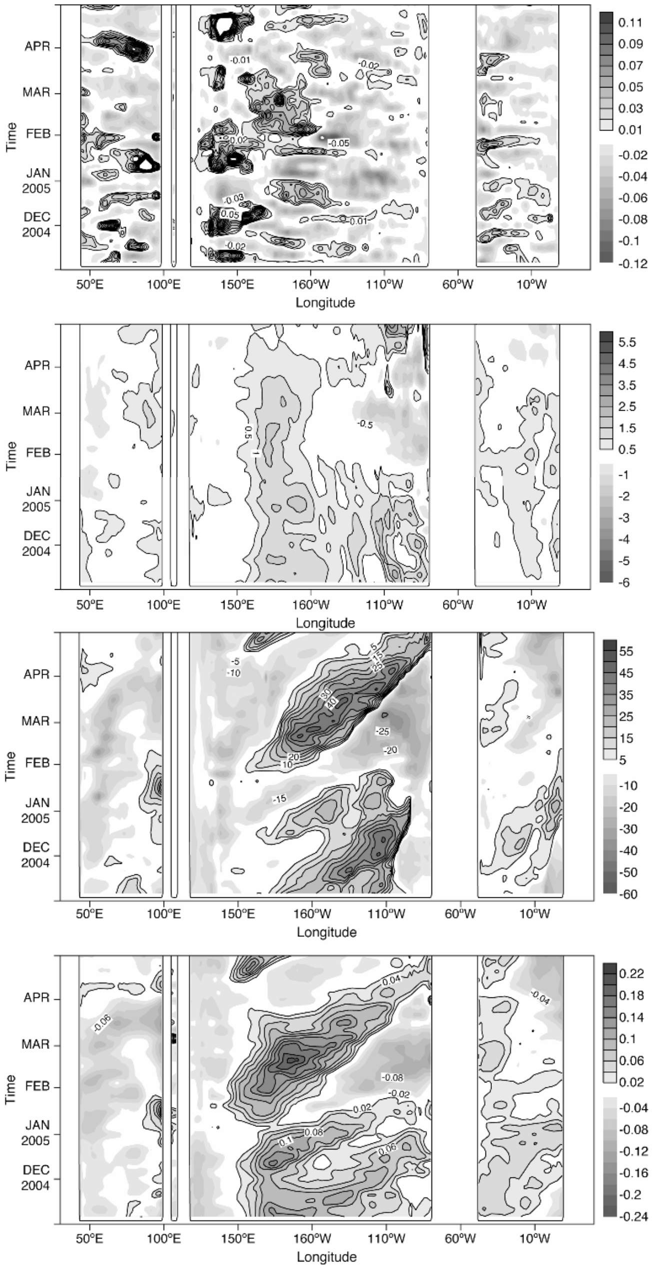

Fig. 3.4 (a) Plot of the zonal wind stress along the equator as a function of time from 1 Nov 2004 until 1 May 2005. The panel on the left corresponds to the Indian Ocean, the middle panel to the Pacific and the panel on the right to the Atlantic. (b) As for (a), but for the depth of the 20°C isotherm. Note the marked eastward propagation. (c) As for (a) but for sea level. The behaviour is similar to (b) though not identical. (d) As for (a) but for SST. Note the absence of eastward propagation

west Pacific will generate Kelvin waves and lead to changes in the SST in the east Pacific. Through ocean analyses one can easily detect the Kelvin waves generated by an MJO but there is no clear concensus on their importance in general. This is a hotly debated topic, especially since there was a strong MJO in Feb 1997 preceding the large El Niño that year. Whether it was key or consequential is still open to debate (van Oldenborgh 2000; Eisenman et al. 2005; Vecchi et al. 2006).

3.3 Ingredients of a Physically Based Forecast System

There are different strategies for trying to predict the climate a few months ahead. The simplest and oldest is to seek correlations between different events at different times. If event X sometimes/often/usually follows event Y by 2 months, then one has a basis for some form of prediction of X. Initially, the predictors (the Ys) were easily obtainable observations. Now they could include parameters which have only recently been available such as temperature in the ocean at say 100 m. Because of reanalyses of the atmosphere and ocean it is possible to get estimates for such quantities even if they were not directly measured. As the historical record gets longer both by new observations, and gets pushed further into the past by better use of past observations and extended atmospheric reanalyses, there is likely to be continued scope for a statistical approach to climate prediction. This will be discussed further in Chapter 7, but in this section we will consider only forecasts based on dynamical GCMs.

To make a forecast from a dynamical model requires knowledge of the current state of the system one is trying to predict. The forecast depends on the state of the atmosphere, ocean and land conditions and therefore initial conditions for the atmosphere, ocean and land are needed. The most common approach is to separately initialise the individual main components, namely the ocean, the atmosphere and the land. The separate initialisations are done mostly for practical reasons as it is easier and less computational demanding to deal with one component at the time. The main drawback of this approach is that the separate initial conditions may not be in balance when forecasts are started and therefore coupling shocks may negatively impact the results of the forecast, from early on in the integration. However, coupling shocks may be considerably alleviated if common boundaries (e.g. the SST seen by both the atmospheric and ocean models separately) are treated in a consistent way (see also Sections 5.1.1 and 6.3.2).

The most important of these initial conditions is likely to be the state of the ocean, and for the seasonal forecast range, the upper 200–300 m is sufficient. It is therefore important to use the available observations in order to produce an analysis of the ocean to be used as initial conditions of the ocean model. This involves a data assimilation system. Because the tropical Pacific is so important to ENSO forecasting, a special observing system for the equatorial Pacific (from 8S to 8N) has been developed. See Chapters 1 and 5 for details on this so-called TAO-TRITON observing system. This observing array has been extended into the tropical Atlantic, spanning a somewhat larger latitudinal range than the Pacific, and is in the process of being extended into the Indian Ocean. The mooring array measures the temperature to a depth of 500 m (some moorings are instrumented to 750 m) at about ten levels. Observations are available every day from this array, relayed via satellites such that the data is ingested into operational analysis systems within hours of being taken. Another interesting development has been the expansion of the ARGO float array (see Chapter 5). These floats measure both temperature and salinity continuously in depth from 1,000 m (some from 2,000 m) to the surface but only every 10 days. The third major component is the XBT array. This has been important in the past though coverage has been declining over the last few years. There are very few measurements of velocity: there are only five moorings taking velocity in the whole of the tropical Pacific. So although velocity is an interesting oceanographic variable, it can only be indirectly obtained from the analysis system. What current measurements there are can be used for validation.

In any forecast system, it is important to give an estimate of the uncertainty of the prediction. One source of uncertainty results from the chaotic nature of the climate system: small uncertainties can grow and give rise to very different sequences of weather events. There are two other broad areas of uncertainty associated with error in the initial conditions and with errors or uncertainties in the models used to make the forecasts. Since we are dealing with an initial value problem, it is necessary to give a measure of uncertainty in the analysis of the ocean state. One way of estimating that uncertainty is through running ensembles – many realisations of the same events but perturbed in some way commensurate with the perceived uncertainty. To estimate the uncertainty in the ocean initial state we can run an ensemble of ocean analyses. This ensemble can be generated by taking into account uncertainty in the atmospheric fields that are used to force the ocean – mainly the wind uncertainty, and uncertainty in the analysis of sea surface temperature (SST). Uncertainty in other ocean measurements should also be taken into account through perturbed measurements but this is not commonly done explicitly at present. It is done indirectly in that when the model first guess is combined with the observations, perceived errors are ascribed to both in order to judge how much weight should be given to each source of information. This will be dealt with further in Chapter 5. Uncertainty in atmospheric initial conditions can also be dealt with to some degree through the use of 'singular vector' perturbations. This is important on the shorter timescales but probably not so important beyond, say, a month. More details of the ocean observing system and analysis procedure and the atmospheric counterpart are given in Chapter 5. In principle, model uncertainties should be included as well, as model errors become important within the seasonal range, but they are normally not accounted for in any individual operational system. They are accounted for, to some degree, by making forecasts with more than one model (the multi-model approach).

A second ingredient of a forecast system is a model of the atmosphere–ocean system. Ideally this should also include sea ice but that is not usually done at present as ice modelling is complicated and probably does not bring a worthwhile increase in predictive skill relative to the effort expended. Land initial conditions are prepared through soil moisture and snow cover and all credible AGCMs include a land package to represent the effects of rainfall, vegetation, run off, etc. Future work will make the parameterisations of land processes progressively more sophisticated. Once the initial conditions have been prepared, the coupled model is run forward for several months.

Coupled models normally capture the various synoptic features such as anticyclones and cyclones quite well. A good model would also reproduce blocking, and the ability to represent droughts, extreme events leading to floods, etc. It should also simulate events such as the MJO. There is an important difference between being able to represent a type of event and being able to forecast a specific event. For these shorter-lived meteorological events one does not try to predict their occurrence beyond a few days but hopes to represent their statistical occurrence correctly. In practice the models are not yet that good since processes such as the MJO, or blocking tend to be under represented.

As mentioned earlier, an ensemble of forecasts is generated as a means of representing uncertainty in the forecasts. Different members of the ensemble will have highs, lows, MJOs, blocks, etc. occurring at different times in the forecasts. For a seasonal forecast one is not trying to say what will happen on a given day 3 months ahead but rather how the average weather over a month or a season might change, i.e. how the lower frequency processes might change. It may be that there will be an increased probability of drought in some place and in that case more than half of the ensemble would show below average rainfall.

The final ingredient in any practical forecasting system is a method for dealing with systematic error, the fact that the models do not represent the climate accurately. One method, adopted by the EUROSIP (see later) project for example is to create a large number of forecasts over past years. Typically this is 15 years but this is really too small, and a longer period of 25 years is being attempted. For any given month for each of these years an ensemble of forecasts is made. This then defines the model climatology for this month. Forecasts are then compared against this climatology, and presented as anomalies (Stockdale et al. 1998). By this means a first-order linear correction for model error can be made. It is a simple approach and does seem to work reasonably well but further, more refined approaches are possible as discussed for example by Stephenson et al. (2005).

3.4 How Accurate are Seasonal Forecasts

Let's start with evaluating predictions of the temperature in the equatorial Pacific. A key region is NIÑO3, an area in the Eastern Pacific Ocean (210–270°E/5°S–5°N). Figure 3.5 shows the growth of error in the forecast as a function of lead-time out to 6 months (solid line). One can see a steady growth over the period. (The rms error between the forecast and the observed value of NIÑO3 temperature is used as a measure of forecast accuracy.) Many forecasts have gone into this figure: all forecasts for all months for all years from 1987 to 2004. Also shown on this plot is the growth of ensemble spread (dotted line). This grows less fast than the error. One can interpret this result in two ways.

The negative interpretation says that the spread is smaller than the error and therefore the forecast system is poorly calibrated: the model forecasts are too confident – it means that the observed SST frequently lies outside of the range spanned by the forecast ensemble. Calibration is discussed further in Chapter 8. An alternative, more optimistic, interpretation is to take the model estimate of spread as a measure of potential predictability by interpreting one ensemble member as truth and measuring the differences of other members from that. This then gives an estimate of the potential limit of predictability in the absence of model error. The system illustrated is far from that limit. So by working harder and reducing model error one should (hopefully) be able to improve the forecasts. Of course the current model might underestimate the limits to potential predictability since the model does not do a good job of reproducing intraseasonal

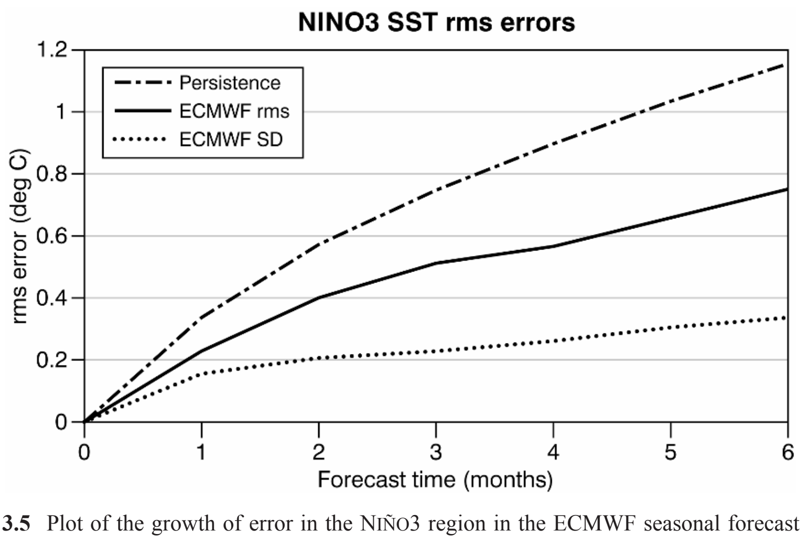

Fig. 3.5 Plot of the growth of error in the NIÑO3 region in the ECMWF seasonal forecast system (solid line). Also shown (dotted) is the ensemble spread. These are average values covering all months for all years from 1987 to 2002. The ensemble spread is obtained by calculating the ensemble mean and the root-mean-square (rms) difference of ensemble members from the ensemble mean. Ideally, in a well balanced system, the spread and the error should be similar. In this example, they are not: the forecasts are too confident, in the sense that the spread is too small, indicating that all of the uncertainty in the system is not being accounted for. Model error, a major cause of forecast error, is not represented since all forecasts are made with the same model. The dot-dashed curve is the skill of persistence, i.e. saying that the anomaly at the start of the forecast will not change with time

variability such as the Madden-Julian Oscillation which, it is thought, might play a role in limiting predictability of ENSO (see also Section 3.2 and Chapter 4). However, even if the optimistic interpretation of the limit of predictability were correct, the reality is that such a level of skill has not yet been achieved as the model error is larger than the spread. One has to work with the practical reality that for now the model is not well calibrated. This limitation is not specific to ECMWF but applies to other models as well. The final curve on this figure (dot-dash) is the growth of error using the simplest of all forecast strategies: that the forecast anomaly for the month ahead, 2 months ahead, etc. is the same as for the current month. It may come as a surprise, but many models fail to beat this measure in the first 1 or 2 months (see for example Fig. 3.6).

One way of improving the forecast reliability is to sample model error in the ensemble probability distribution function (PDF) and one way to do that is to develop a multi-model approach. This has already been done in a non-real-time mode (see Palmer and Hagedorn 2006, Chapter 26). A real-time operational multi-model forecast system, called EUROSIP,[7] has been implemented at ECMWF, which currently consists of forecasts from the Met Office, Météo-France and ECMWF, with planned extension to other models. In order to get some feel for the potential improvement in forecast skill as a result of the multi-model approach, we plot in Fig. 3.6 the rms error for the NIÑO3 region for two models that participate in

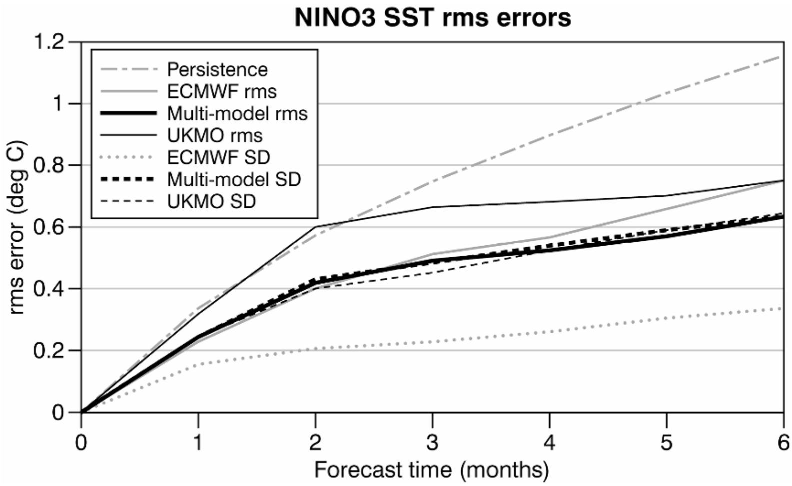

Fig. 3.6 Plot of rms error from a multi-model forecast system using models from ECMWF and the Met Office. The spread and rms error are better matched in the case of the multimodel. This is a necessary though not sufficient condition for a good system. As in Fig. 3.5, the dot-dash curve indicates the skill of persistence

[7] EUROSIP stands for EUROpean multi-model Seasonal to Inter-annual Prediction: http://www.ecmwf.int/products/forecasts/seasonal/forecast/forecast_charts/eurosip_doc.htm.

real-time multi-model predictions at ECMWF. The models shown in Fig. 3.6 are actually from DEMETER (Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction[8]) and as such are earlier versions than are used in real-time operational applications but they should give a fairly good assessment of what to expect. The error growth is shown for NIÑO3 in the Pacific but is evaluated for many regions including the Atlantic and Indian Oceans. The skill in the Atlantic is lower than in the Pacific: actually the error growth is similar but the size of the interannual signal is smaller in the Atlantic, so the error is more serious. Similarly, the anomaly correlation drops more rapidly in the Atlantic than the Pacific, which is probably related to the fact that climate anomalies are shorter lived in the equatorial Atlantic which may in turn be related to the smaller size of the Atlantic basin. Looking at indices gives a concise way of evaluating skill but is not the only way. A more detailed assessment of model skill is given in van Oldenburgh et al. (2005), but further analysis is still needed.

One can also evaluate specific events. Either one can look at a forecast and see where the model (or preferably the multi-model) is predicting a sizeable anomaly and then retrospectively see if this occurred. An alternative is to see where there are or have been major climate anomalies and then to see if the model predicted them. However, the first method of evaluating "real" predictions (i.e. before one knows the outcome of the forecast) is probably more objective than judging forecasts in hindsight.

Further validation is shown in Fig. 3.7. The measure of skill illustrated here is anomaly correlation as this is quite a simple quantity to evaluate and understand. It does not make full use of probability information coming from ensembles of integrations. Such skill measures (e.g. Brier skill score) are discussed in Chapter 10. Figure 3.7 shows the skill for near surface temperature. It is high over the tropical Pacific, related to ENSO. It is generally low over land related to the smaller heat capacity of land compared to sea. There is one intriguing region where the skill over land seems quite high viz over western Europe. This is probably a real signal as it is present in other models too and is likely related to snow cover (Shongwe et al. 2007). It is not present in other seasons.

If one considered rainfall, then the correlations are highest in the tropics. The skill in predicting rainfall is lower than for near surface temperature, even in the tropics. Although the skill is generally low there may yet be applications which can benefit from even modest levels of skill.

[8] A completed EU project concerned with seasonal forecasting. For more information see: http://www.ecmwf.int/research/demeter/.

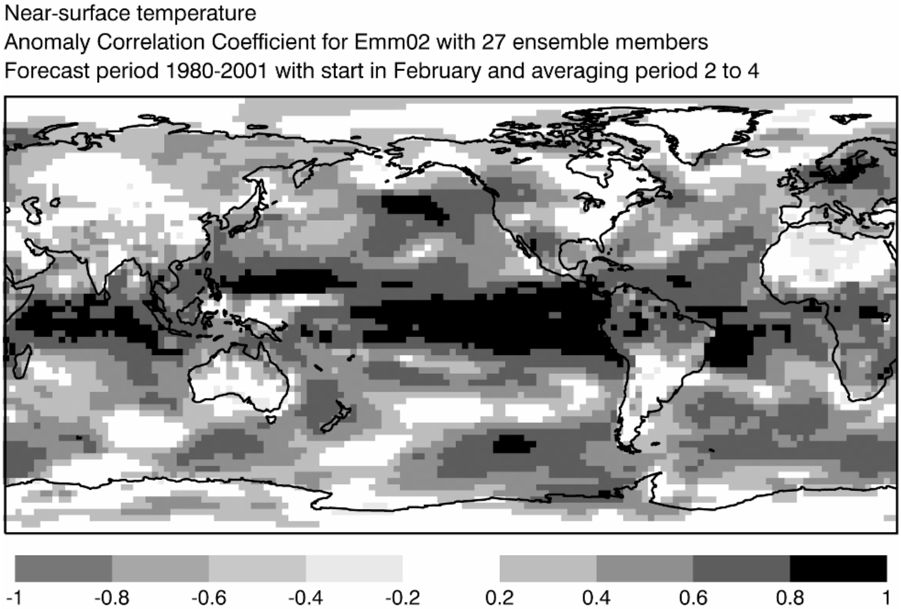

Fig. 3.7 Global anomaly correlation for near surface temperature from the ECMWF model. The results are for predictions starting on 1 Feb for March, April May starting on 1 Feb. Results will vary depending on the season being predicted. In general skill is higher in the tropics than at higher latitudes but the temperature signal over northern Europe is real

3.4.1 Further Verification

For a correct interpretation of seasonal predictions the user needs to complement the forecast products with knowledge of the forecast skill (e.g. by assessing skill measures such as those shown in Fig. 3.7). It is not possible in this book to discuss all the verification methods that have been used but an extensive assessment is available on the ECMWF web site.[9] Estimates of model bias for a wide range of variables, including zonal averages, time series of a set of indices of SST and large-scale patterns of variability such as the Southern Oscillation Index (SOI), the Pacific North American Pattern (PNA) and the North Atlantic Oscillation (NAO) are available. A suite of verification scores for deterministic (e.g. spatial

---

[9] A comprehensive documentation of skill levels, using methods that have been agreed at the international (WMO) level for the evaluation of long-range forecast systems can be found at: http://www.ecmwf.int/products/forecasts/d/charts/seasonal/verification Spatial distributions of the mean errors (biases) are provided at: http://www.ecmwf.int/products/forecasts/d/charts/seasonal/verification/bias/.

anomaly correlation and Mean Square Skill Score Error (MSSE)) and probabilistic forecasts can be viewed for the operational system. The robustness of verification statistics is always a function of the sample size. For the operational seasonal forecast system, the sample size of 15 years is considered barely sufficient.[10] Verification is performed in cross-validation mode (Michaelson 1987) using the whole set of forecast data available, i.e. both hindcasts and real time forecasts. The seasonal forecast skill depends strongly on the season; so forecasts are evaluated separately for different starting months. Issues such as how to evaluate probability forecasts will be discussed in Chapter 9.

3.4.1.1 Applications

The development of seasonal forecasting applications is very much in its infancy; some of the difficulties in developing application models will be covered in Chapter 11. In this section we give just a couple of examples. In the first there is no formal application model. In the second (malaria) there is an application model but the issue of validating the application model is bypassed as the same application model is used for validation as is used for prediction. The difficulties in developing a disease model such as malaria are covered in Chapter 13.

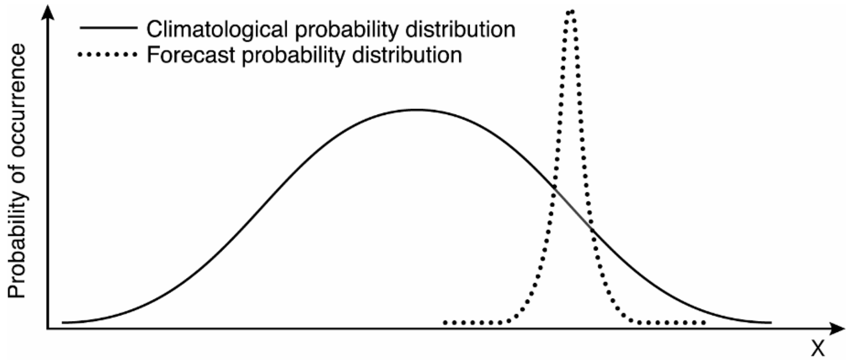

Fig. 3.8 Solid curve: climatological PDF of temperature at Gallipoli (hypothetical). Dotted curve: hypothetical forecast PDF. In this case the two distributions are different in shape as well as being displaced relative to each other. In practice, for such a mid-latitude station, one would not expect to get such a clean separation (Adapted from Palmer 2006)

[10] A longer period of forecasts has been performed in an experimental framework in the EU project DEMETER, but for a limited set of start months. See web for more details.

The climatological PDF for temperature at some place, say Gallipoli in June is shown in Fig. 3.8. There is very small probability of the temperature being extremely high or of it being very low. The most likely temperatures are close to the average values but there is a fairly high chance of the temperature being a few degrees either side. The dotted curve shows the PDF from some hypothetical forecast. If this were an accurate and reliable forecast then there would be much useful information: the most likely temperature is quite a bit higher than average but there is very little chance of it being very hot or being below the climatological average. Readers could use their own 'implicit' application models to decide whether to take a holiday in Gallipoli or not.

Often the distributions are not so clearly separated, especially when dealing with middle latitude situations, but the changes in the PDF of rainfall in the tropics between El Niño and La Niña conditions could be large. If there is a big El Niño then the rainfall patterns are shifted: in the west Pacific rainfall decreases, while in the central Pacific it increases. For many parts of the world the shift might be quite small: like loading the coin only slightly in the analogy in Section 3.1.

For many applications, the dependence on weather might be quite complex – either depending on more than one weather variable or having a non-linear dependence such as a threshold when something only happens if the temperature is above or below some value. All of these dependencies can be easily taken into account and the potential benefit from seasonal forecasts evaluated, provided the transfer function linking the application with the meteorology is well known. Unfortunately, it is often quite complex to relate the weather parameters to an application, and so difficult to develop/define/verify the transfer function. For example, although there is some relationship between say malaria and weather, it might be complicated to evaluate and validate. This will be discussed in Chapter 13 but for now let us assume that some plausible relationship has been found which depends on weather parameters such as temperature and rainfall. One can then use the output from the climate forecast model as input to the 'malaria' model.

The next step is to force the application model with observed quantities from, e.g. ERA-40. The application model is the same model whether driven from the forecasts or from the analysis. Comparing the output from these two then does not verify the application model at all but does give a measure of the forecast skill as applied to this application model. Figure 3.9 shows just such an example for malaria in southern Africa. If the transfer or application model were good, then the potential for malaria prediction in this region is good. Extensive work is needed to develop and validate application models such as the malaria models since the data may be hard to obtain and the application may be influenced by more than weather.

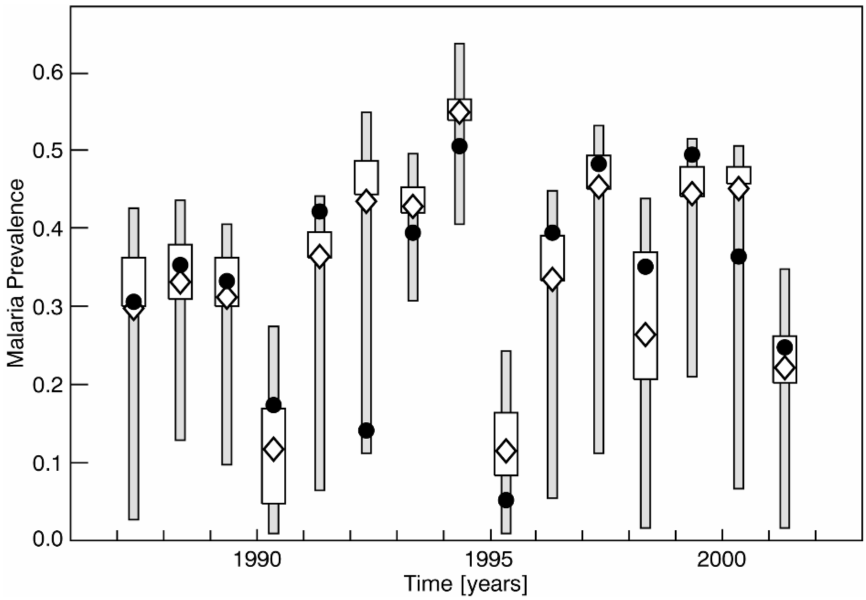

Fig. 3.9 Predicted malaria prevalence in March, April May predicted from 1st February. The spread of the forecasts driven by the ensemble forecasts is depicted by the box-and-whisker representation with the whiskers containing the lower and upper tercile of the ensemble. The diamonds represent the ensemble mean. The reference driven by the ERA-40 data is shown by the black bullets (From Hagedorn et al. 2006)

3.5 Summary

A variety of coupled atmosphere ocean models has been developed and used for understanding and predicting El Niño and La Niña. The most complex of these models are those based around general circulation models of the atmosphere and ocean. Such coupled models are essentially the same as those used for weather forecasting – but with the added complexity of having an interactive ocean module – and they are able to generate weather sequences just like weather forecast models. Change the initial conditions slightly and the model will generate a different sequence of weather. If the initial conditions of the ocean and land and to a smaller degree the atmosphere differ from normal then the predicted climate will differ from the normal climate; anomalies will result. An analogy based on the weather coin was introduced: where the loading of the coin is strong, then the probability of developing climate anomalies is strong. Where the loading is weak, there might be shifts in climate but one might need many realizations to detect them. Thus the tropics in an El Niño event is an example where the loading is strong, middle latitudes even in an El Niño event might be a case where the loading is weak.

To cope with the probabilistic nature of climate prediction, ensembles of forecasts are made. Ideally to detect weak signals, many realizations are needed. In practice only a limited number of ensembles is feasible; typically order 40. Even if models were perfect, forecasts would still be probabilistic: there would still be a need to make ensembles of integrations. But models are not perfect and so strategies are needed to deal with model error. One strategy is to run as many forecasts over as many past events as possible to develop the model climatology. (In practice, the reforecasts are of the order of 5–15 ensemble members spanning 15–25 years.) A forecast PDF is then compared against the model climate PDF and climate anomalies predicted. This approach allows a linear correction for model error for any given error. There may be errors which are not linearly related to the model climate error; these require more sophisticated correction algorithms than those usually applied. Using a single model does not allow one to sample all possible model error. To some extent this can be accounted for by using more than one model; hence the development of operational multi-model activities such as the EUROSIP project at ECMWF. Additional error can arise from uncertainty in ocean or land initial conditions and so the ensemble should include uncertainty in these. This is done to a small degree in the use of an ensemble of ocean conditions within any CGCM system but is not yet done for land initial conditions, though it is in development.

Examples of forecasts using state of the art CGCM forecasts are given, but only using the simplest measure of skill (anomaly correlation). The skill is highest in the tropics, as expected. More sophisticated assessments of skill using probability information from the ensembles are possible but are not discussed in this chapter. Climate forecasts are themselves of interest but greater use is possible if they feed into application models. Such models are rather case specific; many applications models are needed. Developing such models is, however, not a trivial matter. Further discussion of this is postponed to Chapters 12 and 13.

Acknowledgements The author would like to thank all of the Seasonal Forecasting Group at ECMWF but especially Magdalena Balmaseda, Paco Doblas-Reyes and Tim Stockdale for providing figures and Rob Hine for help with graphics.

Chapter 4
Ocean–Atmosphere Basis for Seasonal Climate Forecasting

Brian Hoskins and Paul S. Schopf

There are many phenomena of interest in the atmosphere and ocean, only some of which are relevant for seasonal forecasting. One way of identifying the processes likely to be active is through scale analysis which identifies the important terms in the governing equations and highlights the importance of geostrophic balance. Simple arguments for Rossby waves are given. These waves are important in both atmosphere and ocean as a means of transferring energy over large distances. When the waves are embedded in a westerly flow it is possible for the waves to be stationary, giving rise to the possibility of a coherent remote response. A possible source of stationary atmospheric Rossby waves could be the deep convection over parts of the equatorial oceans where the sea surface temperatures are high. These stationary wave trains may interact with mid-latitude phenomena such as the storm tracks, so changing the occurrence and preferred locations of storms. This is an example of interaction between weather and lower frequency climate changes. Other teleconnections are introduced, such as the link between the Indian summer monsoon and Mediterranean climate. The area of the world where the interaction between the atmosphere and ocean is strongest is in the tropics. It is important to understand how the upper equatorial ocean works and how it is connected to the subtropical thermocline. The connection of the tropics to the subtropics gives a possible mechanism for low frequency variability of ENSO. Various theories of ENSO are introduced in which the importance of equatorial Kelvin and Rossby/planetary waves is highlighted. Simple models illustrate oscillatory behaviour in certain parameter regimes but damped oscillations in others. While these ideas are interesting in generating a framework within which to consider ENSO, the real test comes from the making of forecasts and determining by experience the limits of predictability.

Brian Hoskins
University of Reading

Paul S. Schopf
George Mason University

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

4.1 The Role of the Atmosphere

4.1.1 Scales

The atmosphere and ocean are shallow layers of fluid around the Earth acted upon by gravitational attraction to the almost spherical solid Earth. Using the Earth parameters a, the radius, and Ω, the rotation rate, N the basic buoyancy frequency associated with the stable stratification, and typical scales for the phenomenon of interest for horizontal and vertical length, L and H, respectively, and horizontal velocity, V, we have the following scaling relations:

$$H \ll a \text{ (shallow fluid)} \quad \text{and} \quad H \ll L$$

so that hydrostatic balance is a good approximation in the vertical. Also, important velocity scales and typical numbers for them are:

$$\begin{array}{ccccccccc} a\Omega & \sim & (gH)^{1/2} & > & (NH)^{1/2} & > & V & \\ 465 & & 300 & & 100 & & 20 & \mathrm{m{\cdot}s^{-1}} \end{array}$$

The first number is the speed at which a parcel on the equator moves purely due to the rotation of the Earth. The second, in which the density height scale of the atmosphere (about 10 km) has been used, is the speed of external gravity waves. The next is the speed of internal gravity waves. The comparatively small value of the speed of motion relative to the Earth (V) emphasises the rapid rotation of the planet and the relatively small deviation of the atmosphere (and even more the ocean) in its motion from solid body rotation with the planet. The last inequality emphasises the strong stratification of the atmosphere (and ocean). Behaviour in the local vertical and horizontal directions is therefore very different. In the vertical there is generally stable stratification and a balance between the very large gravitational and pressure gradient forces. In the horizontal, the much smaller Coriolis force associated with the rotation of the Earth can be important. For synoptic scales, and indeed for larger scales away from the equator

$$V/fL < 1,$$

where $f = 2\Omega \sin \phi$, the Coriolis parameter, is twice the local vertical component of rotation of the Earth. This implies that the basic momentum balance in the horizontal is between the Coriolis and pressure gradient forces, and that **v** is approximately geostrophic:

$$\mathbf{v} \approx \mathbf{v_g} = (\rho f)^{-1}\, \mathbf{k} * \mathbf{grad\ p}$$

The potential temperature, θ, is the temperature air would have if it was taken adiabatically (no heat added) to a standard pressure (usually 1,000 hPa). In the absence of heat sources and sinks, it is conserved.

It is very useful to have another quantity that involves the dynamics and is conserved following the fluid in the absence of diabatic and frictional processes but, in these circumstances, the absolute circulation (C) around a closed material line on a constant θ-surface is also conserved. However this is difficult to use directly. If the closed material line shape is used to make a material cylinder between this θ-surface and its neighbour at $\theta + \delta\theta$, then both the mass, m, of the cylinder and $\delta\theta$ are also conserved. Therefore the quantity

$$C \times \delta\theta / m$$

is conserved. Writing the circulation in terms of the absolute vorticity $\boldsymbol{\zeta} = f\,\mathbf{k} + \text{curl}\,\mathbf{v}$, a measure of the local rotation in the fluid, and using derivatives, this conserved quantity may be written:

$$\rho^{-1}\,\boldsymbol{\zeta} \cdot \text{grad}\,\theta.$$

This is called the potential vorticity (PV) and is conserved moving with the fluid in the absence of heat sources and sinks. The PV involves the dynamics as well as the thermodynamics: from the derivation given here it is basically a measure of circulation on a θ-surface divided by mass between isentropic surfaces.

In most large-scale motions of interest there is balanced dynamics involving the Coriolis force. The simplest example is geostrophic motion. In such cases the 3-D distribution of PV, along with suitable boundary conditions, can be inverted to give all the details of the balanced flow. The large stratification is associated with a large vertical component of **grad** θ, and so, on synoptic and larger scales, it is the vertical component of absolute vorticity, $\zeta = f + \mathbf{k} \cdot \text{curl}\,\mathbf{v}$, that is most important for PV and so in the analysis of atmospheric motion.

4.1.2 Atmospheric Phenomena

A vast range of phenomena occur in the atmosphere and it is essential when modelling the system to consider which of these are to be simulated, and what are their characteristics that have to be represented in the model and diagnosed in atmospheric or model data. Probably the most fundamental of these on the larger scale is the Rossby wave. Its nature can be understood by considering the situation shown in Fig. 4.1. The equatorward initial perturbation in contours of the vertical component of absolute vorticity implies a positive vorticity anomaly. Associated with this will be cyclonic motion, as shown. The equatorward wind to the west implies a

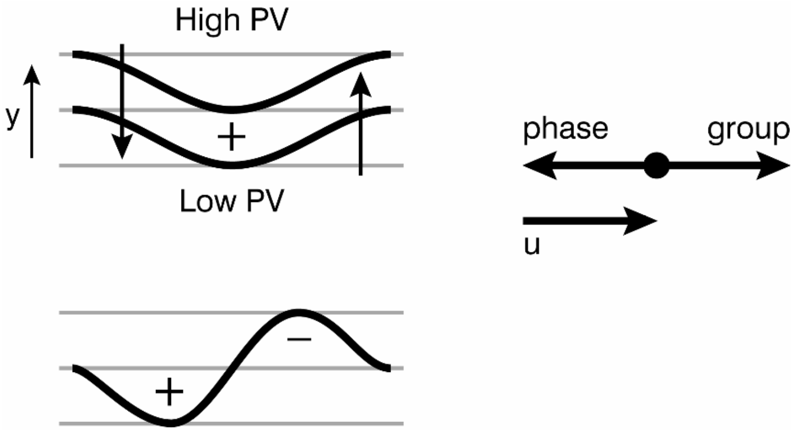

Fig. 4.1 A simple description of Rossby wave dynamics. The basic situation is one with high absolute vorticity (or more generally PV as marked here) poleward and low absolute vorticity equatorward. The starting point is a local equatorward displacement of the absolute vorticity contours, leading to a cyclonic anomaly (represented by a +). This induces north-south flows as shown, and these in turn lead to vorticity anomalies as indicated in the panel below. The result is a westward movement of the cyclonic anomaly (+) and the development of a new anticyclone to the east. These features correspond, respectively, to westward "phase speed" and "eastward group velocity". A basic westerly flow (**u**) will add on to them, giving a reduction of the former and an increase of the latter

tendency to extend the initial positive vorticity anomaly in this direction: a westward "phase speed". The poleward wind to the east of the original positive vorticity anomaly implies a tendency to create a negative vorticity anomaly there. This means that the region of wave activity extends in this direction: an eastwards "group velocity". The propagation of the wave activity is measured by the group velocity which is therefore eastward. If a basic westerly flow is added, then the eastward group velocity becomes larger and the phase speed can become zero, depending on the wavelength. The discussion of Rossby waves given here can be extended to apply to PV on θ-surfaces, to waves with their crests and troughs tilted from the north-south direction, and also to realistic flows on the spherical Earth, in which case propagation tends to be along great circle paths rather than east-west lines.

The existence of such stationary Rossby waves is very important because it means that there can be coherent remote responses to stationary wave sources such as mountains and regions of persistent deep convection such as the western tropical Pacific with its high SSTs. This response can occur on planetary scales in a wide arc on the eastern side of the wave source. Such responses lead to the climatological average waves and also to monthly or seasonal anomalies, which are normally associated with a sequence of height field anomalies of alternating sign. An example for October 2000, which was one with record-breaking rainfall in England and Wales, is given in Fig. 4.2. The low height-field, cyclonic anomaly

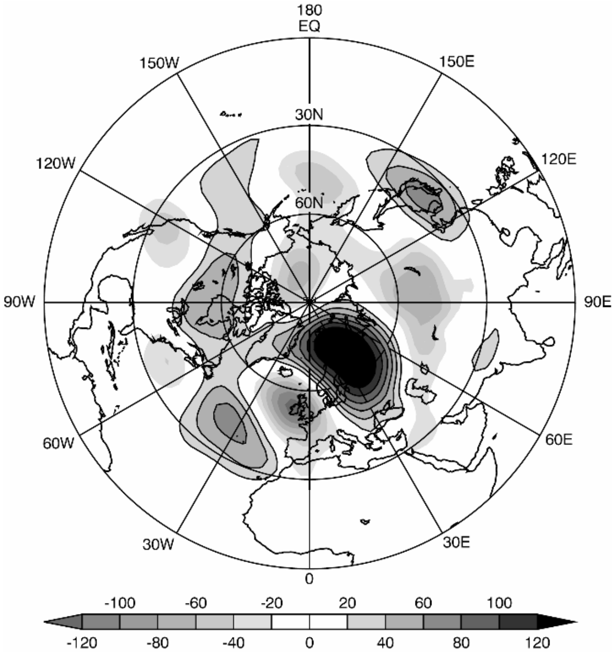

Fig 4.2 The 300 hPa geopotential anomaly from climate for Sep–Nov 2000 (M. Blackburn 2005, personal communication)

over and to the east of the UK is seen as part of a wave pattern. The group velocity arguments suggest that the origin of the anomalous pattern should be sought to the west, in the Caribbean/Americas region.

The picture for October 2000 is probably not as simple as this might suggest. The North Atlantic storm-track extends from the coast of N America towards NW Europe. The weather in Europe is strongly dependent on the position and intensity of the storm-track. The anomalous large-scale flow in October 2000 will have influenced the storm-track. However the storms themselves will have fed back on the larger-scale flow through their vorticity and heat transports, thereby changing it. Fluctuations in the North Atlantic near surface westerly flow and in the storm-track are frequently characterised in terms of the North Atlantic Oscillation (NAO). There is much current interest in possibly predictable monthly to seasonal timescale behaviour of the NAO that may be related to the strength of the lower stratospheric vortex or to sea surface temperature (SST) patterns.

The absence of storms affecting a region can be associated with a phenomenon referred to as blocking, often characterised by a persistent deep positive height field anomaly. It is thought that blocking can occur as an interaction between weather systems and an anticyclonic anomaly, which may itself form part of an anomalous stationary Rossby wave train. Blocking is particularly important for Europe, being associated with anomalously dry or wet weather, depending on location, and warmer or colder weather, depending on the season.

In the tropical region a common occurrence is for frequent deep convection to occur in a large region for many days. This implies large latent heat release in this region. The response of the atmosphere to this heating usually has the general characteristic flow pattern which is shown schematically in Fig. 4.3. The middle tropospheric latent heating is balanced by adiabatic cooling associated with ascent. Off the equator this implies vortex stretching and the generation of cyclonic vorticity below and vortex shrinking and the generation of anticyclonic vorticity above. As in the Rossby waves argument, the lower and upper tropospheric circulations extend and move to the west, the latter process continuing until the parts of the circulations with, respectively, poleward and equatorward moving air are in the heating region. Through considerations of balance, associated with the change in the sense of the circulation with height, the mid-troposphere to the west of the heating must be warm. Such a pattern of circulations can be associated with anomalous heating in any month, perhaps associated with higher or lower SST than usual. It can also act as source for a Rossby wave train propagating into higher latitudes, perhaps like that seen in Fig. 4.2 for October 2000.

A particular example of such heating and associated global anomalies is the tropical Intra-Seasonal (Madden-Julian) Oscillation. Large regions of much intensified or weakened convection move slowly from the western Indian Ocean to the west Pacific and perhaps continue to the dateline on a monthly timescale. Again this offers the possibility of predictive power both in the tropics and in higher latitudes.

When, such as in the Asian Monsoon, the summer tropical heating region extends to high enough latitudes there is an interaction with the extra-tropical westerlies. These westerlies flow down the sloping θ surface on the western side of the circulations with their mid-tropospheric warmth, enhancing the descent there. This descent can be further enhanced and localised by topography. Radiative cooling, in the absence of convective heating, can then produce further enhancement, leading

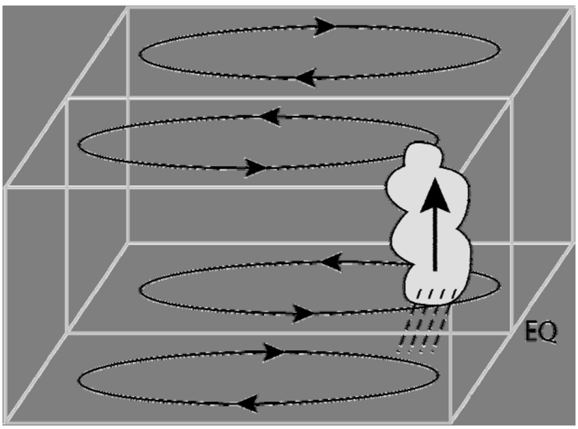

Fig. 4.3 A schematic showing the response to large-scale tropical convective heating. The convection is shown by a cloud and is assumed to span the equator

to very strong local descent. Rodwell and Hoskins (1996) proposed that this is the basic mechanism for the summer climate of the Mediterranean, which is therefore seen as part of the Asian Summer Monsoon. Such remote associations are very important in providing a context for considering seasonal anomalies and their forecasting. Indeed, significant weakening of the Asian summer monsoon may allow North Atlantic weather systems to enter the Mediterranean and then move into southern Europe as in the summer of 2002.

4.2 The Role of the Ocean in the Climate System

The fundamental role of the oceans in the climate system is to (1) act as a buffer for mitigating transients, (2) contribute to the required pole-to-equator heat transport, and (3) provide hidden "memory" in the coupled atmosphere-ocean-land system.

Over much of the planet, the ocean can be considered to be well-represented as a surface mixed layer whose temperature (T) obeys a simple heat conservation law:

$$\rho \cdot c_p \cdot h \cdot \frac{\partial T}{\partial t} = Q_s(t)$$

Here ρ is the density and c_p the heat capacity of seawater. The depth of the mixed layer (h) varies in space and time – it is typically thin during summer months, and thick during the winter. The wintertime depth can reach hundreds of meters or more at high latitudes, while during the summer a depth of 10–20 m might be found. Figure 4.4 shows the climatological mean profile of temperature at 40°N, 170°W in the ocean for February and August. During the winter, the cooling and wind cause the ocean to be well-mixed down to 200 m, while during the summer,

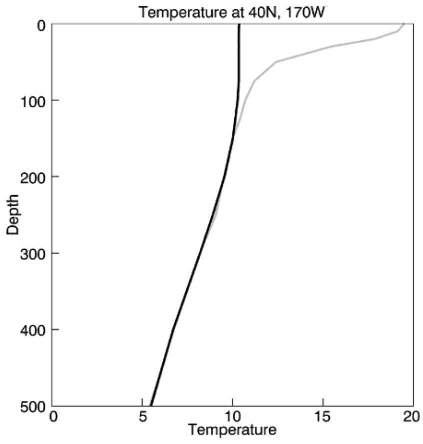

Fig. 4.3 Climatological temperature values at 40°N, 170°W in the North Pacific for February (heavy) and August (light) (Data from NOAA NODC World Ocean Atlas 2005)

the very shallow surface layers warm up considerably. The surface fluxes (Q_s) may include a wide range of frequencies, including diurnal cycles, synoptic atmospheric weather, seasonal cycles and longer term climate changes. Due to thermal inertia, the ocean mixed layer will damp the high frequencies, providing a "reddening" of the spectrum. The deeper the mixed layer, the more pronounced the reddening. For the problem of seasonal climate prediction, modelling this mixed layer behaviour over the open ocean captures most of the essential physics over much of the ocean. Theories and models exist for simulating the 1-dimensional behaviour of turbulent mixed layers under the combined effects of heating and wind.

The above heat balance assumes that there is no heat flux out through the bottom of the mixed layer. For many problems, this treatment is adequate, but such an approximation will not permit any transport of heat from one latitude to another. The circulation in the ocean can carry heat into or out of the mixed layer. In regions of strong currents this extra heating can become important. The mixed layer budget becomes

$$\rho \cdot c_p \cdot h \cdot \frac{\partial T}{\partial t} = Q_s(t) - Q_c$$

In the time mean, a balance must exist between this circulation-induced heating and the surface heat flux, or $Q_s = Q_c$. It is this spatial variation of the circulation-induced heating that enables a net ocean heat transport. Circulation-induced heating can be caused by vertical motion at the bottom of the mixed layer, horizontal currents or turbulent mixing. There are a few distinct regions where the ocean currents strongly affect the surface heat balance: the western boundary currents, such as the Gulf Stream and Kuroshio, the Antarctic circumpolar current, and the tropics, particularly within the equatorial wave guide. The tropics get special attention in the seasonal to interannual climate problem not only because the ocean dynamics plays this strong role, but because the atmosphere responds strongly to the ocean-induced changes. In the mid-latitudes, the oceans carry heat, but the atmospheric response to this heating is not as strong, and does not cause secondary effects that influence the circulation on these timescales.

4.2.1 The Thermocline – Setting the Stage for El Niño

While the ocean is very deep, most of the important dynamics for seasonal to interannual timescales happen within the relatively thin warm region at the top of the ocean known as the thermocline.[1] Oceanographers now understand that the

[1] The term 'thermocline' originally referred to the region of strong thermal gradient, but recently is sometimes associated with the entire upper ocean through the development of the ventilated thermocline theory (see, for instance Pedlosky 1996).

character and shape of the thermocline is described by a dynamical construct known as the ventilated thermocline (Luyten et al. 1983). In a static view, one would expect the thermocline to be deepest at the equator, where the warmest surface waters are found. But instead, the thermocline almost vanishes along the eastern end of the equator in both the Pacific and Atlantic. Cold water from below the main thermocline is exposed to the surface in a feature commonly referred to as the "cold tongue". Figure 4.5 and Figure 4.6 show climatological temperature sections of the top 500 m of the ocean along the dateline and equator, respectively, for February and August. Note how in Fig. 4.5 the 20°C isotherm is deepest at about 20° latitude along the dateline, but that in Fig. 4.6, it comes very close to surface at the eastern end of the equator. We can see that close to the surface, seasonal effects matter, but once deeper than a hundred meters or so, the seasonal effects are smaller.

As we will see later in this section, El Niño models function by predicting perturbations that happen to the thermocline. Its structure affects the sensitivity of the surface temperature to the subsurface ocean variability, which in turn affects the coupling between the ocean and atmosphere. Models for El Niño have shown sensitivity to the sharpness and tilt of the thermocline. A new body of research has emerged on how long-term variations in the thermocline occur and how they might influence the evolution of El Niño.

The equatorial thermocline connects to the subtropical thermocline through a circulation system known as the subtropical cells (STCs) or shallow overturning circulation. Work by McCreary and Lu (1994) has shown that the equatorial "cold tongue" is not simply an accident of having a thermocline and easterlies along the equator, but is an essential property of the STC circulation.

These cells have a three-dimensional circulation structure, with largely poleward surface branches and equatorward sub-surface flow. When the water flows along the surface, it is in constant contact with the atmosphere, and its properties are altered by the surface fluxes of heat and freshwater. nce removed from the surface, turbulent mixing and heating is much smaller, and the water is found to conserve its property over long distances and long times. A common approximation is that the flow is adiabatic, and flows along surfaces of constant density or "isopycnals".

McCreary and Lu showed that at about 15° north and south latitudes in the Pacific, the lower branch of the cell has a net flow toward the equator, when integrated completely across the basin. Where does this water go? Except for a small leakage through the Indonesian throughflow, there is no horizontal outlet. The water coming in at this depth must rise to the surface (upwell) within this tropical band. With the mean easterly Trade winds, the necessary upwelling conditions apply at the eastern end of the equator, along the American coasts, and in a few isolated regions such as the Peru upwelling and the Costa Rican dome.

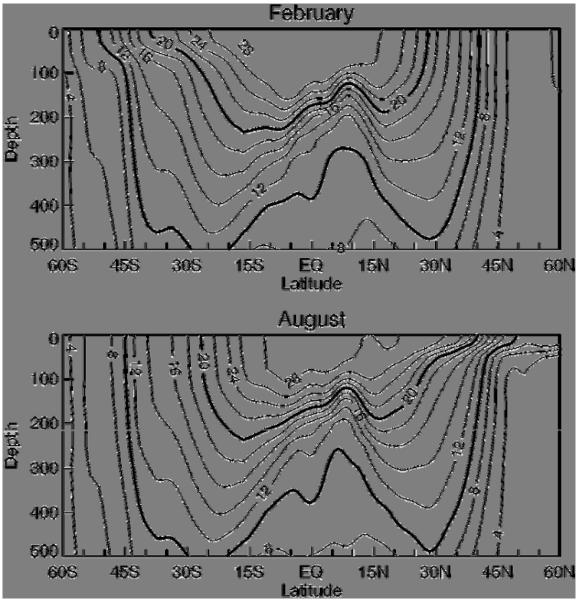

Fig. 4.4 Climatological temperature sections along the dateline for February and August (Data from NOAA NODC World Ocean Atlas 2005)

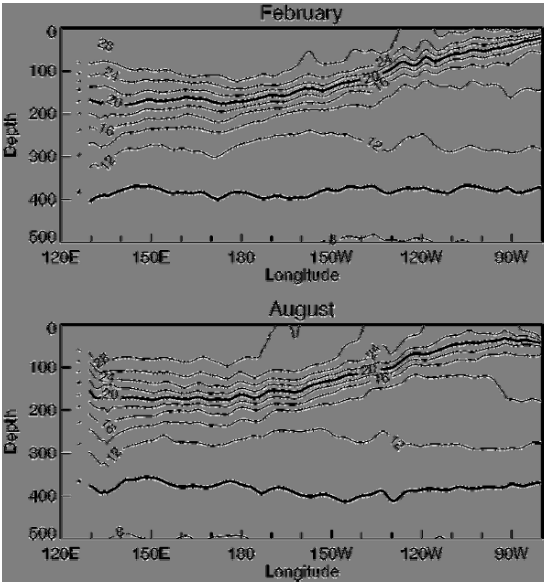

Fig. 4.5 Climatological temperature sections along the equator for February and August (Data from NOAA NODC World Ocean Atlas 2005)

Further studies with more complex ocean circulation models have investigated the source waters of the equatorial under current or equatorial cold tongue,[2] and are largely in agreement that the source waters of the equatorial under current and equatorial cold tongue lie well within the subtropical gyres. Observations support the canonical view of the STCs (Johnson and McPhaden 1999).

In summary, the large-scale circulation of the top several hundred meters of the oceans creates a thermocline with warm water overlying cold, and this thermocline is constrained to be tilted along the equator, with the cold water showing up at the surface at the eastern end. This sets the stage upon which fluctuations act to produce El Niño and La Niña.

4.2.2 *Variations on the Thermocline*

Once we understand that the thermocline should exist, and that it should surface at the equator, it seems natural to ask whether this outcropping is stable, or whether the system can experience an oscillation or be disturbed by local wind and weather effects. The theory of McCreary and Lu basically states that "since on the average, x kg/sec of cool water converges toward the equator from both hemispheres, on the average x kg/sec of cool water must surface". One might think that this sets the temperature of the cold tongue, and that El Niño arises from changes in this process. But this applies only over a suitable averaging interval. We know that this flow can take decades to close the loop. For changes that take only a few months or even a few years, another theory is needed that permits significant variations, and it is these variations that are now known to be the root cause of El Niño and most of tropical climate variability.

For this theory, we turn to the dynamics of internal gravity waves as modified by the special features of planetary rotation – the equatorial Kelvin and Rossby waves that propagate signals east and west in the equatorial waveguide (see Moore and Philander 1977 or Gill 1982 for a synopsis of equatorial wave dynamics).

Because of the special nature of the Coriolis effect near the equator, low frequency planetary waves take on distinct properties, with one wave type propagating eastward (the Kelvin wave) and a set of others with westward propagation (the equatorial Rossby waves). Each wave describes the evolution of the thermocline perturbation (h), the zonal current (u), and the meridional current (v). The Kelvin and Rossby waves are different "modes" of the system. Each mode has a different pattern in the north-south direction, but they tend to keep all the action

[2] Rothstein et al. (1998); Harper (2000); Huang and Liu (1999); Malanotte-Rizzoli et al. (2000); Rodgers et al. (2003); Fukumori et al. (2004).

near the equator. Mode 0 is the Kelvin wave, with a phase speed and group velocity to the east with the same speed as an internal gravity wave. Modes 1, 3, 5, ... are antisymmetric, with h = u = 0 at the equator, and a maximum in meridional velocity at the equator. Modes 2, 4, 6, ... are symmetric with perturbations in height and zonal current that have local maxima on the equator, and no meridional flow across the equator. Note that the gravest symmetric Rossby mode (n = 2) propagates to the west with 1/3 the speed of the Kelvin wave. The higher mode Rossby waves propagate slower still.

Figure 4.7 shows the meridional structure of the Kelvin wave. The Kelvin wave is somewhat special because it has no meridional current (v = 0), and the meridional structure for h and u are the same. Note that the successively higher Rossby modes have amplitude extending further from the equator, and have a more oscillatory behaviour.

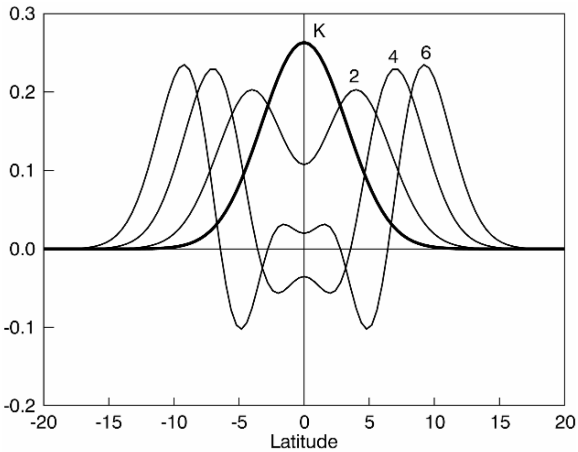

Fig. 4.6 Meridional structure functions for the equatorial waves. These functions depict the relative amplitude of pressure perturbations from the Kelvin (K) wave and the first three symmetric Rossby waves (2, 4 and 6) from linear solutions with an internal gravity wave speed of 3 m s^{-1}. Anti-symmetric Rossby modes also exist, but are not shown

In the ocean, the Kelvin and Rossby waves are largely forced by the wind. Weakening the trade winds in the centre of the Pacific will generate Kelvin waves that cause the thermocline to deepen while at the same time driving Rossby waves that cause the thermocline to shallow. Looking along the equator, one would see the deepening signal head off to the east and a shallowing signal heading west.

Other than the slightly strange meridional structure and modified phase speeds, the Kelvin-Rossby wave set behave like internal gravity waves in a channel. Only one wave can send signals to the east, while all the rest send signals westward. Unlike a bounded channel, however, the reflection properties of the Kelvin and Rossby waves are different. When Kelvin waves reach the eastern end of the equator, they propagate poleward along the coast. These coastal Kelvin waves shed some Rossby wave energy as they go, but to a large extent, much of the energy in the equatorial Kelvin wave is propagated out of the equatorial zone.

At the western boundary, Rossby waves have a somewhat complicated reflection. Cane and Sarachick (1977) demonstrated that Rossby waves reflect into equatorial Kelvin waves. When the zonal mass flux in the Rossby wave is integrated in the meridional direction, there can be no net accumulation of mass. The Kelvin wave that is reflected has sufficient amplitude to balance this mass convergence.

What we see from the wave reflections is therefore a “leaky” system: Rossby waves propagating westward return their energy in a Kelvin wave travelling eastward, but these Kelvin waves leak their energy to higher latitudes when they reach the eastern boundary. This leak means that the system can not simply resonate like a closed channel, with waves forever bouncing back and forth. Instead, any sustained oscillations of this system must involve forcing.

These wave dynamics describe the forced and freely propagating linear response of the upper ocean to imposed wind stresses. They describe the motion of a simplified representation of the thermocline. These motions might be interesting in and of themselves, but until and unless they change the SST, they will not have any influence over how the atmosphere evolves in time. But if the motions do alter the SST, and this change in SST causes changes in the atmosphere that further change the winds, then there can be a feedback loop between anomalous winds, anomalous currents in the ocean, SST perturbations and finally back to the anomalous winds. This loop is known as a closed feedback loop, in which perturbations in the system propagate from one variable to another. Closed feedback loops can lead to instabilities in the system.

The most important modification needed to our theory is to improve the representation of the ocean temperature. Although the upper ocean may be treated as two distinct layers to explain the essential dynamics, the actual ocean thermocline is a region of continuous gradients of temperature and salinity. The forced wave motions that alter the thermocline therefore introduce a continuous change in the surface temperature.

We concentrate on the eastern equatorial Pacific, because this is where the thermocline outcrops and causes the “cold tongue”. Bjerknes (1966) noted that westerly wind anomalies in the central part of the Pacific would drive Kelvin waves to the east that would deepen the thermocline and carry warmer water to the east. Both effects will cause warming of the SST in the east. Theories, models and observations of the atmospheric response to the warmer surface temperature agree that such warmer SSTs will lead to further strengthening of the westerly winds. Thus a positive feedback loop exists that can extract energy from the system. This source of energy is key to overcoming the “leaky” nature of the wave reflection arguments made earlier.

A second important point should be noted about the interaction of the waves with the SST: in the western Pacific, and away from the zone where the thermocline outcrops, changes in the depth of the thermocline do not have a perceptible influence on the surface temperature. Although changes have been introduced into the ocean, and waves are sending signals around, the ocean has "sequestered" the information from the atmosphere. This is a key ingredient of the delayed oscillator mechanism, which we discuss next.

4.2.3 The Delayed Oscillator Theory of Enso

We now have a view of the system where the thermocline tilts up to the east and exposes cold water to the surface. Changes in the position of this thermocline are reflected in changes to SST which perturb the atmosphere. These changes give rise to a positive feedback through the winds to drive an unstable growth. If h_e is the thickness of the thermocline in the eastern Pacific, and $\hat{\tau}$ is the zonal wind stress averaged across the basin, we have

$$\hat{\tau}(t) = Ah_e(t)$$

$$\frac{\partial h_e}{\partial t} = B\hat{\tau}$$

where A and B are proportionality constants. Then,

$$\frac{\partial h_e}{\partial t} = AB\hat{\tau}$$

This is a simple view of the Bjerknes instability.

In his original paper, Bjerknes noted how this feedback can explain the emergence of El Niño events, but he then remarked on the difficulty in finding a reason for the system to turn around and go from warm to cold. (Or, for that matter from cold to warm, as for example at the end of La Niña.) Since we have a rationale for the thermocline to be exposed to the surface in the eastern Pacific over the long term, such perturbations as described by Bjerknes can not take over and control the result forever. This, plus the observed preference for El Niño to occur every 3–7 years led to a search for a mechanism that could explain an oscillation in the equatorial system. One solution can be found in the delayed action oscillator.

As explained above, the winds that perturb the ocean by driving Kelvin waves to the east also drive Rossby waves to the west. They propagate on the deeper thermocline, hidden from the atmosphere. When they reach the western boundary, they reflect into equatorial Kelvin waves, and propagate back to the east. In this case, the equation for $h_e(t)$ must be modified to include the effects of the Rossby waves. The eastward propagation of Kelvin waves along the equator and the poleward propagation along an eastern boundary can be clearly seen in Fig. 3.3. This figure also shows the westward propagation of Rossby waves from the eastern boundary and their generation in mid-ocean and westward propagation to the western boundary.

The important feature of the Rossby waves is that a wind that drives a shallowing Kelvin wave will drive a deepening Rossby wave, and that a deepening Rossby wave reflects into a deepening Kelvin wave. Thus, the evolution of the height field in the east is a combination of the Bjerknes instability and information from some "old" Rossby wave:

$$\frac{\partial h_e(t)}{\partial t} = ABh_e(t) - C\hat{\tau}(t-\Delta t)$$

The factor C includes the effects of how the wind drives the Rossby wave, how efficiently the western boundary reflection works and how the Kelvin wave alters the thermocline thickness. The time $t\text{-}\Delta t$ reflects the fact that the height at present time is influenced by the wind that existed in the past – at the time that the Rossby wave was first generated.

But once again, since $\hat{\tau}$ at any moment is presumed to be proportional to the SST anomaly, which is presumed to be proportional to the thermocline displacement, we can combine all these proportionality factors and arrive at:

$$\frac{\partial h_e(t)}{\partial t} = ABh_e(t) - Dh_e(t-\Delta t)$$

This equation is a differential-difference equation that describes the basic delayed action oscillator. Under certain conditions, this equation can lead to growing oscillations.

In their original proposal Schopf and Suarez (1988) include a cubic damping term which is intended to reflect the fact that SST can not grow without bounds: In our advective model, if the thermocline floods in completely from the west, the surface temperature can not get much above 30°C, because that is the warmest water available. Similarly, because the process works as an uncovering of the thermocline, if too much water is brought up, the surface will see the relatively uniform intermediate water that lies just below the thermocline. Their proposal is therefore

$$\frac{\partial h_e(t)}{\partial t} = ABh_e(t) - Dh_e(t-\Delta t) - \gamma h_e^3(t)$$

This cubic term means that the system will not grow without bound, but will undergo regular oscillations of a fixed amplitude. The modification to a non-linear system is not fundamental to our understanding of the mechanics of the delayed action oscillator.

By rescaling time with the growth rate AB, and the dimensional h_e with $(AB/\gamma)^{1/2}$, we see that the oscillator depends on two parameters:

$$\frac{\partial h_e(t)}{\partial t} = h_e(t) - \alpha h_e(t-\delta) - h_e^3(t)$$

where $\alpha = D / AB$ and $\delta = AB\Delta t$. The system described may undergo self-sustained or damped oscillations, depending on the location of the base system in the parameter space described by α and δ. When oscillations are present, they have a period in excess of twice the delay. They are typically far greater than this. See McCreary and Anderson (1991) for a full coverage of the various types of response as a function of α and δ.

The delayed action oscillator succeeds in describing a mechanism whereby a preferred periodicity for El Niño may exist. As one can see, there are several parameters which are not easy to quantify, and attempts to diagnose whether the system should exhibit self-sustained oscillations or not have been made, but they are inconclusive. Instead, experiments with numerical models have been designed to examine the point, but in the end there is less to be learned from examining the stability question than there is in understanding the elements of the system.

The key elements of the delayed oscillator are:

1. Coupled instability in the east via Bjerknes mechanism
2. Low frequency Rossby wave generation that perturbs the thermocline
3. Reflection of the thermocline displacements into an equatorial Kelvin wave
4. "Coupled reflection" at the east

This last point is an interesting twist on what one would expect if gravity waves bounced back and forth across a closed basin. Instead of a period that is set by the time it takes a wave to go back and forth across the basin, the delayed oscillator operates with a period which is at least twice that time. Recall that the reflection of a deepening Kelvin wave at the eastern boundary causes a weak set of deepening Rossby waves. In the delayed oscillator, the Rossby waves are not generated at the coast, but through the coupling process whereby deepening Kelvin waves give rise to shallowing Rossby waves. This phase reversal is key to the period-doubling inherent in the system.

The delayed action oscillator theory demonstrates that El Niño arises from a coupled instability in the ocean-atmosphere system. Neither an ocean-only nor atmosphere-only model can explain El Niño and its dominant frequency. It implies that the memory in the system lies in the thermocline, off the equator. In Chapters 5 and 6, we discuss the nature of the prediction system, and how model initialization, and particularly ocean data assimilation is essential to successful forecasts of El Niño. One of the main reasons for this lies with the information contained in the ocean thermocline and the dynamics of how that information propagates through the system, only later to show up as changes in the surface temperature.

4.2.4 The Recharge Paradigm

An alternative to the delayed oscillator theory is the recharge paradigm for El Niño. In this view, it is recognized that the Rossby and Kelvin waves cross the basin far more quickly than El Niño changes to La Niña. When fast-moving waves are forced slowly, it is hard to recognize them as waves at all. Instead of describing the changes in the state as due to wave propagation, perhaps we can describe the ocean as in quasi-steady state.

Jin (1997) was able to use this property to develop a simpler system of equations that describe an oscillator. When the waves are fast, the equatorial Kelvin wave can be written in the very simple form:

$$h_e(t) = h_w(t) + a_1\hat{\tau}$$

where h_e is the thermocline height at the eastern end of the equator, and h_w is the height at the west. $\hat{\tau}$ is the average zonal wind stress across the basin, and a_1 is a proportionality factor.

Anderson and Gill (1975) demonstrated how the steady circulation of the ocean (the so-called Sverdrup flow) is established by the net effect of Rossby waves. In the recharge view, the explicit treatment of the Rossby waves of the delayed oscillator are replaced with Sverdrup flow, which causes mass to converge toward the equator, thereby setting up changes in the thermocline in the west. An equation for the thermocline thickness in the west is then

$$\frac{dh_w}{dt} = -rh_w - \beta\hat{\tau}$$

where r is a damping factor, and β is a proportionality factor that builds in the different projection of the winds onto the modes as well as a number of other effects.

The coupling in the recharge oscillator occurs through the SST, as in the delayed action oscillator: the stress is proportional to the temperature in the east. In the delayed oscillator, the relationship between the thermocline depth and the SST is treated as due to several factors:

$$\frac{dT_e}{dt} = -a_2T_e + a_3h_e + a_4\tau_e$$

where T_e is the SST, which is damped by surface fluxes to some equilibrium, a_3h_e reflects the contribution of upwelling, and $a_4\tau_e$ represents an advective feedback due to wind stress local to the east.

These terms are then related to two variables as primary h_w and T_e: the winds (both $\hat{\tau}$ and τ_e) are made proportional to T_e. If time is scaled with the Bjerknes instability growth rate and h_w scaled appropriately, the coupled set of ordinary differential equations can be written

$$\frac{dT_e}{dt} = T_e + h_w$$

$$\frac{dh_w}{dt} = -\hat{r}h_w - bT_e$$

The recharge oscillator and delayed action oscillator share the same ocean dynamics (low frequency forced modes), and depend on two parameters. In both views, one parameter sets how fast the western basin fills in relation to the time-scale of the Bjerknes instability. In the delayed oscillator, it is the Rossby wave propagation time, while in the recharge oscillator, it is the damping parameter $\hat{r}$. Both theories also have a free parameter describing how strongly the conditions in the west influence the SST in the east.

Given their shared view of ocean dynamics and their ultimate dependence on two similar parameters, it is not possible to differentiate the two theories based on observations or model experiments. For most questions, they share similar challenges. For instance, it has been noted that the recharge paradigm depends on the latitude at which one wishes to compute the Sverdrup flow. But the delayed oscillator can consider more than one meridional Rossby mode, with higher modes extending further poleward and travelling at slower speeds. The delayed oscillator is criticized because it may be possible for Rossby waves to propagate through the Indonesian archipelago.[3] But in the quasi-steady Sverdrup flow of the recharge paradigm, the buildup of mass in the west may be returned poleward in western boundary flows or may pass through Indonesia just as the low frequency Rossby waves. In short, there is little to be gained from differentiating these two views.

4.2.5 Conclusion

The ocean is but one part of the climate system. We have discussed how the ocean takes up heat to buffer the high frequency changes induced by the atmosphere. Next we noted that the large scale circulation created by the combined effects of winds and surface heating does not drive the entire ocean uniformly, but leads to a rather shallow circulation that is described by the ventilated thermocline. This thermocline connects to the equator via the shallow tropical cells, and an inevitable consequence of the atmospheric forcing is that this thermocline will emerge at the eastern end of the equator (at least in the Pacific).

Coupling to the atmosphere and a simple deterministic view of the atmosphere led us to discover that this tilted thermocline is perhaps not a stable stationary state, but can possibly have unstable, self-sustained oscillations. These oscillations lie at the heart of El Niño and La Niña. Whether or not dynamics such as the delayed oscillator are strong enough to cause spontaneous changes to the system,

[3] Schopf and Suarez (1990) show that a reflection efficiency as low as 15% is sufficient to permit oscillations.

it is clear that other perturbations to this tilted thermocline are capable of causing significant changes in the equatorial surface temperatures. Storms, sub-seasonal variations, and other unpredictable features of the tropical atmosphere will all leave their imprint on the thermocline. Some may lead to expressions in the SST that will give rise to coupled instability, some will pass through the system as subsurface Kelvin waves with little hope of making a sustained change in the climate system.

The challenge to the problem of seasonal climate prediction via dynamical models is to build models that must capture all of the essential physics – the mixed layers, ventilated thermocline, shallow tropical cells, wave dynamics, and thermodynamics of how the thermocline emerges, and they must be able to be initialized with the important information that contains the dynamics of the evolution.

4.3 The Nature of the Prediction Problem

The problem of seasonal climate prediction is one of attempting to simulate the seasonal average of the weather, not the individual fronts, cold snaps, or storms. These "weather" events have been shown to have no predictability beyond a week or two (see also Chapter 3). If the climate is the sum of weather, but the weather is unpredictable, does this not imply that the climate is unpredictable? In fact, the answer is no, the climate can be predictable considerably longer than the weather. If a forecast system fails to predict a storm 10 days from now, but predicts one 12 days from now, the forecast is wrong, but the average number of storms in the next month will be correct. The prediction problem relies upon the fact that some parts of the system evolve slowly, while others are of short duration. If the short events are unpredictable, but an equation can be written to describe the slow evolution, then the high frequency component can be considered unpredictable "noise", and the challenge is to describe the effect of noise on the solution of the slow equations.

In Chapter 3 and in the previous sections of this chapter we discussed some of the current theories for El Niño/La Niña, which involve the propagation of signals on the ocean thermocline, transformation of these signals to SST anomalies, then coupling to the atmosphere, modification of the winds and driving of the ocean. We derived equations for this slowly evolving part of the system. The oscillator theory is very simplified, however, and much can disturb the process. Each El Niño develops differently, and the magnitude can vary greatly from one event to the next. The examination of this irregularity is fundamental to understanding the prediction problem, because it lies at the core of understanding the "predictability limit".

4.3.1 Predictability Limits

The predictability limit is a concept that describes our recognition that we do not do as well with models as we can, but that even a perfect model and perfect initialization will be unable to forecast the climate forever. If the models are inherently flawed, then the predictability limit may be a gross overestimate of how long we can make a successful forecast, but it is useful to try to approximate this limit. If, for instance, it can be demonstrated that no model/initialization system can forecast for more than 2 months, why bother to build better and better models and more and more expensive observing systems? If, on the other hand, it can be shown that forecasts of up to 3 years can be made, then we had better put a lot more effort into our models, observing systems and initialization methodology.

Unfortunately, there is no absolute way to define a predictability limit. We can study how models behave, using the “perfect model” technique. To study predictability, we want to know how fast a perfect model diverges from nature. Unfortunately, although we can know what nature did over the past, we can not construct a perfect model. Instead, we can examine predictability by replacing nature with a model simulation. For this model simulation, there does exist a perfect model – the model itself. There also exists perfect initial conditions and perfect initialization.

If we run the same computer code on the same computer many times with widely different initial conditions, the solutions will enclose a wide region of phase space that describes the climate and its variability. One should see the seasonal march of temperatures, for instance, but the model simulation for a specific day will vary considerably from one run to the other. This spread in the results of a random collection of model runs is known as the “saturation”.

If we run the same computer code on the same computer many times with exactly identical initial conditions, the model will produce identical results forever, unless a coding error exists. There will be no spread between results. But if we introduce a very tiny error in the initial conditions, the model runs will diverge. At first, if we repeat this experiment over and over with many initial conditions that differ by small amounts, we find that the spread in these results is far smaller than the saturation. But eventually, solutions with even the tiniest of initial errors will reach a spread indistinguishable from saturation. When this occurs, we have reached the predictability limit.

Thus, we can define the predictability limit in the context of a model. But does this model represent nature? Is it close enough? If the experiments are repeated with another model, will the results be the same? If they are, is it because the two models reflect nature, or because they share a systematic bias or systematic error that leads to this behaviour? These are questions which confront the theoretician trying to deduce a predictability limit.

4.3.2 Enso Irregularity and Predictability

Is there a relationship between the fact that El Niño is irregular and its predictability? This irregularity reflects the complexity of the coupled ocean-atmosphere system and hints at the difficulties in predicting ENSO. Is it due to noise in the system, the inability to adequately specify the initial conditions, inherent deficiencies in the models, or to not-yet-understood fundamentals of the physical system.

Theories on the cause of ENSO irregularity can be broadly grouped into three categories that are related to their assumption about the strength and validity of the underlying oscillator and the importance of noise. We have presented the delayed action and recharge oscillators as theories for the dominant periodicity of El Niño, but there are debates as to whether they actually operate. We know that in certain parameter ranges, the equations for these simple systems will describe robust oscillations, while in others, the only solution will be a decaying, damped oscillation. The first is self-sustained, the latter requires some external forcing to keep the system going. The three categories of theory on El Niño irregularity split into a view that the oscillators are self-sustained, that they are damped, or that they are essentially neutral. The role of non-linearity and noise is markedly different in each case, and our view of predictability is different in each.

The first view argues for the importance of non-linearity within the tropical coupled system. The non-linearity arises from strong air–sea feedback that puts the coupled mode in an unstable dynamic region. In this regime, El Niño can not only be described as due to a self-sustained oscillator, but it can interact non-linearly with either the annual cycle or other coupled modes. A common model that is cited in this regime is the Zebiak-Cane coupled model, which can be configured to exhibit strong non-linearities and chaotic behaviour. In this view, the loss of predictability is primarily due to the uncertainty in the initial conditions or in non-linearities in the atmospheric response to the ocean. It relies upon fairly robust ocean wave dynamics that provide the underlying timescales for the problem.

The opposing view to this is the stochastic ENSO theory in which "weather" noise generated by the internal dynamics of the atmosphere plays a fundamental role in not only giving rise to ENSO irregularity, but also in maintaining ENSO variance. In this view, the coupled mode is in a damped regime, and thus the ENSO cycle cannot be self-sustained without external noise forcing. The oscillator describes a tendency for the system to have a preferred period, but does not explain much about the appearance of any single event. It is the exact pattern of the noise and how it forces the weakly coupled modes that determine whether a large or small El Niño or La Niña will next appear. The cyclic nature of the underlying oscillator merely alters the odds a little in favour of one side or the other. In this view, the role of the equatorial Kelvin wave and the equatorial air-sea coupling is important, and the off-equatorial ocean dynamics seems less vital.

In between these two viewpoints is the view that ENSO is very close to the dividing line between self-sustained and damped behaviour. Its behaviour is governed by the temporal characteristics of the single, most dominant coupled mode plus the influence of weather noise. In this scenario of ENSO, predictability comes from the oscillatory nature of the dominant mode, while the loss of predictability is primarily due to noise influence. Different from the stochastic ENSO theory where the noise influences the non-modal growth of the coupled system, the role of the noise in this case is to disrupt the regular oscillation of the dominant mode. In this regime, the ocean wave dynamics and reflection properties must also be sufficient to sustain the oscillation.

An extension of this view is the notion that over decades, the system can wander across the dividing line between self-sustained oscillations and a damped regime – so that predictability may vary from 1 decade to the next (Kirtman and Schopf 1998). This concept of time-varying predictability is an important one to bear in mind when considering the skill of previous forecasts and whether this means that our current forecasts are "better" or "worse" than before.

Pinpointing exactly where in the parameter regime ENSO resides in reality is difficult, if not impossible, given the available observations. Many of the recent studies on this issue are based on relatively simple coupled model simulations and prediction experiments. Some of the evidence supporting stochastic ENSO theory is based on the finding that in the damped regime the coupled model forced by stochastic processes produces the best fit to observed ENSO statistics. But in a non-linear system such as the delayed oscillator with cubic damping, the system will appear as damped, while in fact it will spontaneously generate oscillations. Other evidence comes from the finding that there is a lack of support for a continuous ENSO cycle in the observations. In particular, there is little observational evidence that the initiation of an ENSO event relies on the memory of a previous event, though the termination of an event is generally consistent with the delayed oscillator mechanism. The break in the cycle suggests that the system is in a damped regime and the onset of ENSO relies on external influences. Other studies dispute the stochastic hypothesis by providing evidence that seems to be more consistent with the self-sustained ENSO theory. As demonstrated in Schopf and Suarez (1988) and discussed in Jin (1997), a system with a stable, periodic oscillation in the absence of noise can become irregular with the addition of stochastic forcing, and will present statistics that appear to be more stable. Chen et al. (2004) provide retrospective forecasts of ENSO over a 148-year period and show that all prominent ENSO events can be re-forecasted at lead-times up to 2 years. Such a long predictability is in better agreement with the self-sustained ENSO theory than the stochastic theory. However it remains to be tested in the crucible of an actual forecast.

What has emerged in the consideration of the theory is the conclusion that noise has a profound influence on the system and that ocean wave dynamics are essential to obtaining predictive skill as is the proper description of the air-sea coupling. The non-linear, strongly oscillating view gives the most optimistic view of predictability, the stochastic version gives the most pessimistic.

Finally, the debate over where the system lies may have less importance for the practical forecaster than for the theoretician. If weather noise has an influence on the system, there are two parts to consider: what is the role of noise that occurred in the past, and what can we do about the future weather? The past weather noise has become stamped on the ocean and is propagating in the system. If we had a good observing system and initialization method, one could hope to capture all the influences of the past noise, and march forward to a good simulation. This means that we need more than a single simple metric for the ocean initial state. It is insufficient to look at the depth of the thermocline in the west and make a prediction. It will not work to describe the average amplitude of the gravest westward propagating Rossby wave in the ocean. The past noise is inherent in the very complex and complete ocean state, and extracting as much of this as possible is the key job of the data assimilation systems.

If one might hope to capture the effects of past "noise" or weather with a good observing system, what can we do about the weather events that are going to occur over the upcoming seasons that we are attempting to predict? There is evidence that some features, such as the Madden-Julian oscillation, may be able to be predicted for more than a week, but beyond that time, one has to consider these effects as unknowable. It is ultimately these disturbances that will limit the predictability of seasonal means. Perfect models and perfect initialization will never be able to overcome their effect. Experience with idealized model predictability studies seems to show that the limit of predictability is significantly longer than we currently realize with today's prediction systems. Much work remains to be done, advancing the models and refining the initialization systems.

Chapter 5
Getting the Coupled Model Ready at the Starting Blocks

Joe Tribbia and Alberto Troccoli

The aim of coupled models is to represent the trajectory of the climate system as realistically as possible. Given the chaotic nature of the climate system, it is essential that the starting point of the coupled model trajectories (i.e. the initial conditions) is as close as possible to the observed climate trajectory. In order to ensure that this condition is satisfied, observations are used to modify the coupled model via the *data assimilation* approach. In the context of seasonal forecasting, data assimilation is nothing more than a combination of observations and model data, performed with the aim of achieving the 'best' initial state of the coupled model. However, what constitutes the best initial state is still under debate as it is not obvious whether the best seasonal forecasts are obtained by targeting (a) the most accurate initial state estimate or (b) the most consistent (with the model's own preferred state) coupled state or (c) the coupled state that controls some particular growing modes, or indeed a combination of the three. In this chapter we will present generic data assimilation strategies, including some of their history, which could be adapted to any of these three options. If observations were abundant, one could just use the information given by the observations to obtain the best initial conditions. With the advent of satellites, the last 2 decades of the 20th century have seen an enormous increase in Earth observations. Despite this abundance of observations, large parts of the Earth system still remain unobserved: the interior of the ocean for instance can not be measured remotely. Data assimilation becomes therefore indispensable. At present, data assimilation is applied separately to the individual components of the coupled model. Ideally, the initialisation should be realized using the coupled model directly but this approach, as discussed in this chapter, is still in its infancy and progress is not expected to be fast as many obstacles, not least the presence of serious model errors in both atmospheric and oceanic models, still hinder its way to a full mature phase.

Joe Tribbia
The National Center for Atmospheric Research

Alberto Troccoli
European Centre for Medium-Range Weather Forecasts

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

5.1 Data Assimilation Overview

To some, data assimilation may seem like a daunting mathematical-technical problem. In practice, this might be the case, but not more than in any field that requires rather heavy use of a mixture of mathematics and computer programming languages. The task is made easier, however, if it is clear where one is heading. So, although it is true and unavoidable that the foundations of data assimilation are purely mathematical, what the mathematics are trying to represent is a pretty straightforward concept: data assimilation is a blending of two different representations of the same *system* (e.g. the climate system), also referred to as *truth*. In general, the two representations are constituted by (i) *a model* (normally a dynamical one), which aims at representing the system, and (ii) *some observations*, which actually sample the system. The aim is to improve the description of the system as reproduced by the model by using the information coming from the observations.

5.1.1 Data Assimilation Strategy for Seasonal Forecasts

Three main objectives of data assimilation can be identified, even though in practice they overlap considerably:

1. To obtain a four dimensional picture of the system (*state estimation*), consistent with both measurements and dynamics
2. To provide initial conditions to be used by forecast models, as widely developed for Numerical Weather Predictions (NWPs)
3. To improve the dynamical model in order to get a better physical description of the real system

These objectives are applicable to data assimilation in general – data assimilation is utilised in a variety of disciplines, e.g. satellite orbit determination or dynamic plant models – and these objectives are certainly valid for the atmospheric and oceanic media, with which we are concerned here.

Since the prediction of the climate system on seasonal to interannual timescales is mostly an initial value problem, the role of data assimilation is essentially to provide the best possible initial conditions. As discussed in Chapter 3, the most common approach is to separately initialise the individual main components, namely the ocean, the atmosphere and the land.

In order to prepare the separate initial conditions, an independent analysis to provide a 4-dimensional picture (i.e. in space and time) of the atmosphere is undertaken in which the ocean sea surface temperatures (SSTs) are held fixed. The atmospheric analyses thus produced are then used to provide momentum and heat flux forcings to drive the ocean models into which ocean observations will be assimilated. In so doing, a 4-dimensional estimation of the ocean also becomes available. Similarly to the ocean, the land surface module is driven with surface fluxes of heat and moisture, including precipitation, from the atmospheric model and observations to initialise the land component.[1] The individual models thus initialised are then coupled and integrated forward to produce seasonal predictions (see also Sections 3.3 and 6.3.2).

The success of a data assimilation system is very much dependent on the quality of the model available with which the assimilation of data is performed. For infinitely dense observation networks, i.e. if one could observe all the variables needed by the model, the model used would be irrelevant. In reality, however, the level of detail of a model is generally much higher than that provided by observations and therefore the quality of the model matters a lot. In particular, model deficiencies (also known as model errors) should be taken into account when devising data assimilation techniques. Despite their importance, model errors are generally ignored in standard assimilation techniques: in so doing the mathematics is considerably simplified. More recent approaches have only recently started to take model errors into account hence in this chapter we will mostly focus on standard data assimilation approaches (for more information on model errors in data assimilation see Dee 2005 and references therein).

Data assimilation concepts used by the atmospheric modelling community are dealt with first. Although presented as pertaining to the atmosphere, the approaches discussed are valid for a wider class of fields, oceanic modelling being one of them. After a description of assimilation methods, the types and number of observations are introduced, first for the atmosphere and then for the ocean. The following section presents an investigation of the impact of some assimilation methods on seasonal prediction. An outlook on data assimilation strategies, including a research area which is gaining increasing interest – coupled data assimilation – is subsequently discussed. Before jumping into the formal data assimilation, some examples, with reference to applications in fields other than climate science, are given.

5.1.2 Data Assimilation Beyond Climate Science

Missiles are fired with the intent to hit a predetermined target, be it moving or stationary. Even when their trajectory is accurately computed beforehand using some relatively simple formulae (the *model*), missile progress has to be constantly monitored. It is unlikely that the modelled trajectory describes the trajectory

[1] Although an important component of the climate system on seasonal timescales, land data assimilation will only be referred to in the "Suggested further reading" chapter (p. 465). Suffice to say that methods for land assimilation are often adapted from those discussed here.

actually followed by the missile. Remote adjustments are generally needed in order to bring the actual trajectory closer to the trajectory that the missile has to follow in order to reach the target (the *system* or *truth*). It is in these adjustments that some data assimilation procedure needs to be employed. In fact, the modelled trajectory has to be combined with measurements of the actual trajectory (the *observations*). The result of this operation is an adjustment which is applied to the missile so as to nudge it towards the right track along with an updated modelled trajectory. This procedure is then repeated regularly until the missile reaches the target.

There is no guarantee, however, that any trajectory could be beneficially adjusted. If the missile is fired at an angle such that the actual trajectory lies too far from the modelled one, there may be little possibility of bringing the missile back on track. This might more easily happen in the case of models of more complex systems such as that of climate. The model trajectory might be sufficiently far from the observed one that the adjustment required would just be too large. In such a circumstance, if the adjustment was applied to the model trajectory it would cause the model to crash (or blow up) as the model solution would not be compatible with the data assimilation solution. An alternative approach would be to disregard the adjustment and let the model follow its own trajectory, but then we would not talk of data assimilation anymore. Fortunately climate models are not that bad.

A missile is an easy object to visualise. With a little stretch of the imagination it is actually possible to construct data assimilation examples which are somewhat out of the ordinary. Take for instance the case of the *human mind*. Although this illustrative example may appear slightly controversial, it can nonetheless be thought provoking to some readers (the authors certainly find it intriguing).[2]

So how does the human mind fit in the data assimilation framework? Let's start from the sleeping state. When one sleeps, the mind is free to wander and to access states over which we have little or no control (e.g. dreams). It is this floating state which best describes the analogy between the human mind and the model. We can therefore view the model as a collection (or series) of (apparently?) chaotic states, which ultimately constitute the human mind. Now, when we wake up, a variety of constraints emerge, from the shape of the rooms we live in, to the people we get in contact with, to even the strong sense of direction imparted by the sun. All these constraints act as observations which the human mind (the model) will (try to) assimilate. Clearly, some information is accessible to the human mind even before observing what is around. When one wakes up in the morning she/he knows a lot about the surroundings (in which house and room one is in, whether or not there is a garden, etc.) and this information – in the data assimilation parlance called statistical information or *covariances* – is elaborated along with what is actually observed.

[2] The idea for this example is courtesy of a colleague, Dick Dee.

Unlike for the missile, in the case of the human mind, it is less obvious what the model should describe and what the objectives of the human mind are. Without entering into philosophical debates, it is sufficient to point out that for instance from the perspective of free will, the modelled trajectory is represented by the path towards anyone's personally selected target. Interestingly, even in the case of the human mind as a model for data assimilation, we can draw the analogy of the blowing up of the model: when everyday constraints become too stringent, insanity may arise.

In summary, although we might not realise it, data assimilation is embedded in several instances of our everyday life and the examples above might be useful to help visualise the data assimilation problem. We are now ready to tackle the atmospheric and oceanic data assimilation.

5.2 Data Assimilation for Prediction

5.2.1 Introduction to Atmospheric Assimilation

The primary focus of atmospheric data assimilation has been the production of initial conditions for the purpose of numerical weather prediction. As mentioned above, there can be other rationales for the assimilation of atmospheric data, temporally consistent climate records for example, but the historical development of assimilation in an atmospheric context has benefited from the demands of improving daily operational weather prediction.

The prediction problem is frequently idealized mathematically when it is presented in academic courses and much of the emphasis in textbooks and courses is on the dynamical, mathematical and numerical aspects of computational prediction with little presentation of the specifics of how the initial conditions for the initial value problem are to be obtained. The initial conditions are 'observed' but quite a bit goes on between the observation of meteorological variables, using satellites and rawindsondes, etc., and the first time step of a numerical weather forecast. All of this happens within the domain of atmospheric data assimilation.

5.2.2 Beginnings of Atmospheric Data Assimilation

After L. F. Richardson's heroic effort of computing by hand a numerical weather prediction for a single point during the First World War, interest in computational prediction of the weather began in earnest with the project at the Institute for Advanced Studies at Princeton under the guidance of John von Neumann. In the late 1940s, the first electronic computer (the ENIAC) was developed and von

Neumann chose weather prediction as one of the first applied problems to be attempted with the ENIAC. The famous experiment detailed in the Tellus article by Charney, Fjortoft and von Neuman (1949), computed a 24 hour weather prediction over the continental US using the equivalent barotropic model, forecasting the height of the 500 hPa pressure surface from an analysis of that field at the initial time. The initial analysis was not objectively obtained, however, but was an interpolation of the hand-drawn analysis produced by a synoptic meteorologist.

Clearly, there was a mismatch in the techniques used for producing the initial state (subjective) and the forecast field (objective) which needed to be addressed and so the first incarnation of atmospheric data assimilation, objective analysis, was developed. The first attempt at an objective analysis method was made by Panofsky (1949), who proposed and tested a local polynomial fitting technique. In this method the height field was assumed to be given as a quadratic polynomial in the spatial Cartesian coordinates x and y, i.e.

$$z(x,y) = \sum_{i,j<3} a_{ij} x^i y^j ,$$

with the coefficients, a_{ij}, determined by a least square minimization (see appendix at the end of this chapter) of the difference between the assumed quadratic form and the values of z observed at neighbouring radiosonde observation locations. Panofsky noted a few limitations of this method having to do with the locality of the approximation; (1) the geostrophic vorticity, proportional to the second derivative of z with respect to x and y, was very noisy, (2) if the method was used to analyse the horizontal wind components, the horizontal divergence of the wind could become large if the wind components were independently fitted as quadratic polynomials and (3) since the method used local fitting, the edges/seams of the of the local fits were noticeable.

The next method developed attempted to rectify some of the shortcomings of the Panofsky method and began to utilize the information inherent in the operational prediction of the weather on a daily time schedule. Bergthorsson and Doos in Europe and Cressman in the USA, developed similar schemes for objective analysis that incorporated the 24 hour forecast as information for the analysis and also made use of the notion of the geometric relationship of observations relative to the analysis points in a manner superior to mere function fitting. These were accomplished by (1) using the forecast fields as a ‘first guess’ base field and (2) devising the method so that observations nearer to the point at which the analysed field is required are given more weight than those observations which are farther away. Schematically, the method of successive corrections, as the Bergthorsson–Doos–Cressman scheme is known, can be summarized as follows:

1. Use the forecast field as first guess
2. Fit an increment to the first guess weighting the influence of observations to their distance from the analysis point

3. Successively refine the correction by using fewer remote observations with each iteration
4. Impose geostrophic balance in the iteration scheme

(The mathematical details are specified in the appendix, Section 5.6.)

The methods developed by Panofsky, Cressman and Bergthorsson and Doos were rooted in the idea that a good atmospheric analysis could be obtained through the interpolation of observations to the computational grid of a dynamical model. That such interpolation should be guided by the statistical structure of the field to be interpolated was not a consideration in these methods. The first to put forward such a concept was Arnt Eliassen in 1954 who demonstrated the concept for the surface pressure field. Independently, Lev Gandin in the Soviet Union fully developed this concept in a book which detailed the method of optimal (or optimum) interpolation (OI). Roughly speaking, OI is designed around the fact that the meteorological fields of temperature, wind, humidity and pressure are organized into systems which are of finite scale and so, for example, the temperature anomaly at a point is correlated with the temperature anomaly at nearby points. (If it is colder than normal in Lecce, it is oftentimes also colder than normal in Gallipoli.) The major advances from successive corrections were firstly that the interpolation of corrections to the first guess fields forecasts was free from an ad hoc weighting factor because the weights and spatial dependence were obtained from the statistics of the field itself and secondly the multivariate interpolation was easily handled in the OI scheme since the formalism only depends on field correlations, which could be between differing variables such as the pressure and the wind fields. This second advance made it relatively straightforward to impose well-known constraints on the analysis like geostrophic and hydrostatic balance in a statistically consistent fashion.

As above, the mathematical details of the OI method are spelled out in the appendix. The OI method served the atmospheric prediction community, through the 1970s and well into the 1980s. The limitations of the method which required the development of the current suite of analysis/assimilation methods were associated with the fact that OI is ideally suited for the analysis of conventional observations of model state variables coming from radiosondes, for example, and not directly adaptable to satellite observations which measured radiances. Thus, up to the late 1980s satellite radiances were first 'converted' into inferred temperatures before being assimilated into an analysis. A second issue was the lack of temporal consistency in the analysed fields since the only memory in the analysis cycle comes in through the (first guess) forecast, which represents only one particular realisation, namely that at analysis time.

5.2.3 Four-Dimensional Data Assimilation

The constraint of temporal consistency and continuity has been one of the main tools of the subjective analyses produced by synoptic meteorologists in order to maximize the information used in the analyses. These subjective analyses were digitized and used to initialise the earliest numerical weather predictions. Temporal continuity and consistency is an obvious constraint on the atmosphere and so even in the relatively early days of NWP Thompson (1968) devised a variational method to analyse the atmosphere in both space and time. This was the first 4-dimensional analysis method proposed but it (like Richardson's numerical weather prediction) was ahead of the computational means necessary for practical utility. However, as noted above, a first step towards temporal consistency was effected by using the short-range forecast as a first guess in the operational forecast-analysis-forecast cycle, in the successive corrections first and with the OI later. In this way, the information from the previous assimilation was retained, although not optimally. A first step to optimal interpolation in time and space required that both the first guess (forecast) and the observations to be assimilated be treated as random variables with quantifiable error characteristics. An illustrative example which can be considered the essence of modern techniques of assimilation is the following. Suppose one is given two distinct estimates of the temperature at a locale, T_1 and T_2, with errors associated with each. (For example, one temperature could be a forecast temperature from a model and the other temperature could be a measured temperature from a thermometer.) Suppose also that the expected errors associated with each estimate as measured by its standard deviation is σ_1 and σ_2. Then, as shown in the appendix, the linear combination of the temperatures with the least error on average is given by:

$$T_{optimal} = \alpha T_1 + \beta T_2 \text{, with } \alpha = \sigma_2^2 \Big/ \left(\sigma_1^2 + \sigma_2^2\right) \quad \text{and } \beta = \sigma_1^2 \Big/ \left(\sigma_1^2 + \sigma_2^2\right).$$

That is the weight given to each temperature is inversely proportional to the relative accuracy of the temperature estimate.

Thinking of one of the temperatures as coming from a forecast to consistently incorporate a memory of past analyses and optimally use the forecast, one needs to weight the forecast and the observations according to the accuracy in each. The missing piece needed to move OI toward four-dimensional assimilation is to use an estimate of forecast accuracy in weighting the first-guess field. This 'quasi-four-dimensional' OI method was used until the 1990s by the global operational weather centres.

In the early 1990s the operational centres began preparing for truly four-dimensional data assimilation by formulating the assimilation in a purely variational context. This seemingly orthogonal perspective actually coincides with the statistical formulation of the assimilation problem, under certain conditions. To see this we consider once again the combination of two independent estimates of temperature but his time utilize a variational formulation of the problem of obtaining an optimal estimate of temperature. A reasonable way of formulating this problem is to hypothesize that the analysed temperature should be close in some sense to both T_1 and T_2. This being so, a penalty function $J = W_1(T - T_1)^2 + W_2(T - T_2)^2$ is defined. By minimizing this penalty function one is assured of obtaining a temperature close to both estimates. Minimizing J gives:

$$T_{opt} = W_1 T_1 + W_2 T_2$$

which agrees with the statistically optimal estimate given above if

$$W_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad \text{and} \quad W_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

These weights can be alternatively expressed as

$$W_1 = \frac{\sigma_1^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}} \quad \text{and} \quad W_2 = \frac{\sigma_2^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}},$$

showing that the weights are inversely proportional to the uncertainty in each individual estimate.

This makes perfect sense; we wish to weight each estimate proportional to the amount of confidence we have in the estimate. If one estimate is very uncertain (unreliable) compared to the other estimate, it should be given a much smaller weight than the reliable estimate. This is all that any data assimilation method is trying to accomplish.

5.2.4 The Current Practice in Operational Atmospheric Data Assimilation

Several operational centres, ECMWF, Météo-France, the UK Met Office, Canada and Japan have adopted a variational implementation of four dimensional data assimilation, called 4D-Var. Some of these, and others, are also experimenting with a statistical approach to four dimensional assimilation based on the Kalman filter. These approaches, advanced variational assimilation and Kalman filtering, have been developed because of some of the limitations to the standard OI approach of the 1980s.

Foremost amongst the limitations is the awkward manner in which non-traditional data, like satellite radiances must be incorporated into data assimilation in OI. All observations must be first ‘converted’ into meteorological field variables before being interpolated and merged in the OI scheme. This meant that a separate independent step was necessary which transformed satellite radiances into vertical profiles of temperature and humidity mimicking radiosonde observations. This independent step mitigated the optimality of ‘Optimal Interpolation’. A second, less serious drawback to OI is that despite the conceptual simplicity of the method, in practice, optimality was difficult to preserve because the computational cost of an optimal interpolation was prohibitively high. Computational shortcuts like approximating covariance matrices with spatially compact and vertically separable forms were necessary which also lessened the accuracy in the statistical sense, i.e. the optimality, of the analysis product.

In 1990 both ECMWF and NCEP (began implementing a three dimensional variational method of assimilation which, in principle, was merely a reformulation of OI in terms of a variational penalty function. However, this reformulation removed the limitations alluded to above. With regard to nonstandard data, the variational problem posed worked directly with the observed variables, e.g. satellite radiances, by structuring the cost function to be minimized in terms of an observation increment. If y_i is the i-th observation and $\vec{x}$ is the state vector of meteorological fields of winds and temperatures, etc. at the analysis points, then the cost function is formulated in terms of the mismatch between the observation and the value of the observation determined from the state vector, i.e. the state vector interpolated onto observation space. Thus, $y_i - H_i(\vec{x})$ is a variable in the cost function, where $H_i(\vec{x})$, is the function which relates the meteorological state variables to the observed quantity. In the case of radiances from satellite measurements, this would correspond to the physical laws of radiative transfer relating the radiation at a certain wavelength to the absorption and re-emission of radiation in the atmosphere which is dependent on the vertical structure of temperature and moisture. The cost function used in three dimensional variational assimilation (3D-Var) is:

$$\begin{aligned} J &= \tfrac{1}{2}(\vec{x}-\vec{x}_b)^T \mathbf{B}^{-1}(\vec{x}-\vec{x}_b) \\ &\quad + \tfrac{1}{2}(\vec{y}-\vec{H}(\vec{x}))^T \mathbf{R}^{-1}(\vec{y}-\vec{H}(\vec{x})) \end{aligned}$$

where **B** and **R** are the error covariance matrices for the background (first guess) field and the observations respectively. Because the observations are used directly in the cost function, the minimization of J leads to a consistently optimal use of the observations whether standard or nonstandard. A side benefit of posing the assimilation problem in a variational manner is that a great deal of computational science has been developed in the past 30 years with the specific goal of producing computational algorithms which can economically solve variational problems in high dimensional spaces like the 3D-Var problem. Using conjugate-gradient methods and their extensions, no compromises in the fidelity of the optimization are necessary for computational reasons, in contrast to OI. Restructuring the assimilation problem as a variational problem essentially eliminated the major drawbacks of OI. The gory details of the solution to the 3D-Var problem are given in the appendix.

As computational power increased in the 1990s, the implementation of temporal consistency/optimality became feasible using the computational economy of the variational formulation and ECMWF led the move to 4D-Var. Formally, 4D-Var appears to be a small extension to 3D-Var in which the cost function is treated as both a function of the state variables and time:

$$J(\vec{x}(t_0)) = \tfrac{1}{2}(\vec{x}(t_0) - \vec{x}_b)^T \mathbf{B}^{-1}(\vec{x}(t_0) - \vec{x}_b) \\ + \tfrac{1}{2}\sum_k ((\vec{y}_k - \vec{H}(\vec{x}_k))^T \mathbf{R_k}^{-1}(\vec{y}_k - \vec{H}(\vec{x}_k))$$

where the summation over k is over all the discrete times in which observations are made and the explicit dependence of the cost function on the state vector at the beginning of the interval, $\vec{x}(t_0)$, is meant to denote that J is minimized with respect to the initial state vector. Although the modifications in the cost function are seemingly minor, an enormous amount of complexity has been added to the variational problem in going from 3D-Var to 4D-Var. All of the complications have been hidden in the need to update $\vec{H}(\vec{x}_k)$. This requires the time evolution of the state variables, a forecast, and the minimization of J using an iterative algorithm like conjugate-gradient requires a linearisation of the forecast dependence on the initial state. Both of the above are obtained using linear approximations to the forecast model and the transpose, or adjoint, of the linearised model as explained further in the appendix.

The additional complexity necessitated by a linearised version of the forecast model places an enormous burden on the modelling effort since every component of the forecast system must be both linearised and transposed. While the linearisation is straightforward for the dynamical aspects of an atmospheric forecast model, difficulties often arise in developing suitable linear approximations for parameterized physical processes especially those associated with precipitation. The nonlinearity in these parameterizations is typically of high order with a critical threshold triggering different physical behaviours. As an example, convective precipitation is frequently tied to critical values of moist stability leading to a non-differentiable functional relationship in the convective parameterization. A great deal of testing is needed to ensure that linear approximations to this threshold behaviour are accurate enough to be of benefit in the assimilation cycle. Because the infrastructure development is so demanding and the computational cost so prohibitive in 4D-Var, researchers have begun experimenting with Kalman filtering methods. These were first discussed in the meteorological analysis context by Peterson (1968) and brought to prominence there by Ghil and colleagues in the 1980s (Ghil et al. 1981).

In the original development of Kalman (1960) and the interjection into the meteorological literature by Peterson and Ghil, the Kalman filter (KF) is applied to a linear system of prognostic equations for the state vector $\vec{x}$. The linear system is assumed to only approximate the true evolution of the state vector and so a representation of model deficiencies is included in the system in the form of an additive noise term which is uncorrelated in time (so-called white noise). Note that no such allowance for model imperfection exists in the standard 4D-Var assimilation at the major NWP centres. The KF requires not only a linear prognostic equation for the state vector but also requires that an explicit prediction be made of the expected covariance of the error in the forecast using the same linear dynamics. The KF then utilizes the standard statistical perspective of determining the best linear unbiased estimate (BLUE) given the forecast error covariance of the prediction and the known error covariance of the observation at any given time. This is essentially a vector version of the problem which was discussed above, combining two estimates of the scalar temperature at a single point. The power of the KF approach is that, if all the assumptions of linear dynamics and of model errors being uncorrelated in time hold, then Kalman showed that the BLUE is determined sequentially with no need to include any history of the past observations. The history is included by optimally using the information in the dynamics. The mathematical detail of the standard KF is given in the appendix, but even without the detail several of the drawbacks of the method can be easily seen. First, the meteorological equations are not linear and so, as in 4D-Var, a linearisation of the equations must be undertaken. Second, prediction errors using even the full non-linear prediction equations are not uncorrelated from one time step to the next in a prediction model and so white noise is probably not a good approximation to such errors when a linear forecast system is used.

The first two drawbacks reflect limitations in the Kalman mathematical framework for the problem of assimilation for weather prediction and these will compromise the optimality of the assimilated state vector. The third and most important drawback of the KF is that it requires a covariance prediction. Since a typical weather prediction model has a state vector with order $n \approx 10^6$ elements, the error covariance of the state vector is a matrix with order $n^2 \approx 10^{12}$. This is computationally impractical and so to consider the KF algorithm at all in an atmospheric context, a strategy must be devised to limit the dimensionality of the covariance prediction. After several years of experimentation, one of the most promising approaches to limiting the dimensionality is the use of ensemble methods to estimate error covariance, as first suggested and developed by Evensen (1994, 1997) and elaborated upon by Tippett et al. (2003).

In the ensemble Kalman filter (EnKF), the approach is similar to ensemble prediction in which a multitude of forecasts are made with slightly deviating initial conditions in order to give a distribution of probable states of forecasted weather. For ensemble weather prediction, as practiced at ECMWF, NCEP and in Canada, research has focused on the art of ensemble construction and been principally concerned with the method of specification of the initial ensemble realizations. The singular vector technique (Palmer et al. 1994) and the bred vector method (Toth and Kalnay 1997) have been the strategies devised to limit the dimensionality (and thus the number of realizations needed) in ensemble prediction at ECMWF and

NCEP respectively. The success of ensemble prediction with a moderate numbers of realizations (order 50–100) gives hope that a similar strategy can be made to work for data assimilation. However, as opposed to the circumstance in ensemble prediction, more than the forecast uncertainty, the forecast error covariance is needed for the EnKF. The advantage of the EnKF is that 100 realizations of short range forecasts is far less expensive to compute than a forecast for n *x* n elements of a full covariance matrix. There are two other advantages that come along with the EnKF: first, the issue of initialisation of the ensemble realizations using bred or singular vectors diminishes since the post-observation ensemble is required to estimate, as accurately as it can, the uncertainty in the forecast system at that time. The assimilation ensemble is an appropriate set of realizations to initiate an ensemble prediction, although one might wish to augment this ensemble with realizations spanning additional directions, e.g. selected singular vector directions. Second, since the model used to advance the state variable in each of the ensemble realizations is the full non-linear prediction model, no linearisation approximation is made and so EnKF is an efficient algorithm for extending the Kalman filter to the non-linear dynamics domain (the so-called extended Kalman filter).

The EnKF is sequential so a description of its basic cycle begins with an ensemble of forecast states just after observations have been incorporated into the system. All members of the ensemble are updated using the forecast model to the time at which a new observation is available. At this point the forecast error covariance is estimated using the ensemble of realizations. The observation and the estimate of the observed quantity are optimally combined using the forecast error covariance and the observational error variance to determine the BLUE for the quantity. Using the forecast covariance all the state variables which are correlated with the observation are updated in each realization of the ensemble in a manner which updates the mean and covariance of the ensemble so that it is consistent with the BLUE (i.e. linear Kalman filter) expected reduction in error. This description is for a single observation but in fact the algorithm can be shown to be independent of order if multiple simultaneous observations are to be assimilated as is the case for the radiosonde network. Each individual observation can be assimilated sequentially as if their arrival were at differing times without changing the assimilated state. At any point in the cycle an ensemble of realizations reflecting the uncertainty of the state vector is obtained and can be used for ensemble predictions (details of the EnKF formulation are given in the appendix).

As one can note from the description of the EnKF, there remain at least two weaknesses in the EnKF as currently formulated. The first is the standard concern in using ensembles: that with a small number of realizations in a high dimensional system, sampling errors can be substantial. The optimistic perspective regarding sampling errors is that while an atmospheric model has several million degrees of freedom, the dynamical structure of the atmosphere is such that at least locally the relevant number of degrees of freedom for the atmosphere is small and that the prediction model evolution does an excellent job of singling out the most relevant and important degrees of freedom for error growth. This is the rationale for bred and singular vector ensemble sampling strategies and it is hard-wired into the EnKF. The second weakness is related to the mix between linearity and non-linearity in the current version of the EnKF. As described above, the sole extension to non-linearity is the use of the fully non-linear equations to advance the state vector and error covariance. All other aspects of the algorithm are identical to the linear KF. This restriction is not absolutely necessary but it is convenient and can be considered to be equivalent to fitting a Gaussian distribution to the ensemble of realizations for the purpose of determining the weighted compromise between the observations and the forecast realization. This is similar to what is done in 4D-Var where a linearisation of the forecast model is used. Thus, the EnKF as currently used will suffer in the same highly non-linear threshold situations (e.g. rain/no rain) that are problematic for 4D-Var. This limitation can be removed using a Bayesian formulation, but doing so is beyond the current state of the art in EnKF.

5.2.5 Atmospheric Initialisation

As developed earlier in this section and in operational practice, all observations are compared with the 'virtual' observations that would have been obtained from the background guess field at the time of observation. This requires a so-called forward operator, denoted $H(\vec{x})$ earlier, to convert from the model state vector to the particular observation. In the case of direct observations of state variables like temperature or winds this is usually only a spatial interpolation operator. In contrast, in the case of satellite radiances, extensive radiative transfer modeling is needed to connect the observed radiances with the state vector at a given time.

The need for forward models is just one of the challenges associated with the incorporation of satellite data into the assimilation cycle. In addition to this is the challenging nature of the spatio-temporal structure of the data itself (see Section 5.3.1). Despite the challenges associated with satellite data, one of the great successes in moving to a variational framework for assimilation has been the improvements in forecasts directly attributable to the improved extraction of information in satellite radiances. This is primarily attributable to the use of forward operators instead of independently derived equivalent vertical soundings in the assimilation. The most convincing demonstration of this is shown in Fig. 5.1 which depicts the improvements in Southern Hemisphere forecast skill at the ECMWF. Since ground based observations are much more sparse in the Southern Hemisphere, the duplication of the Northern Hemisphere skill by forecasts for the Southern Hemisphere is due almost entirely to the improvements made in incorporating satellite data into the assimilation.

Another aspect of data assimilation that was ignored in the discussion of methodology is the need for imposing a so-called balance constraint in the process of assimilating observations. The need for a balance constraint comes from the

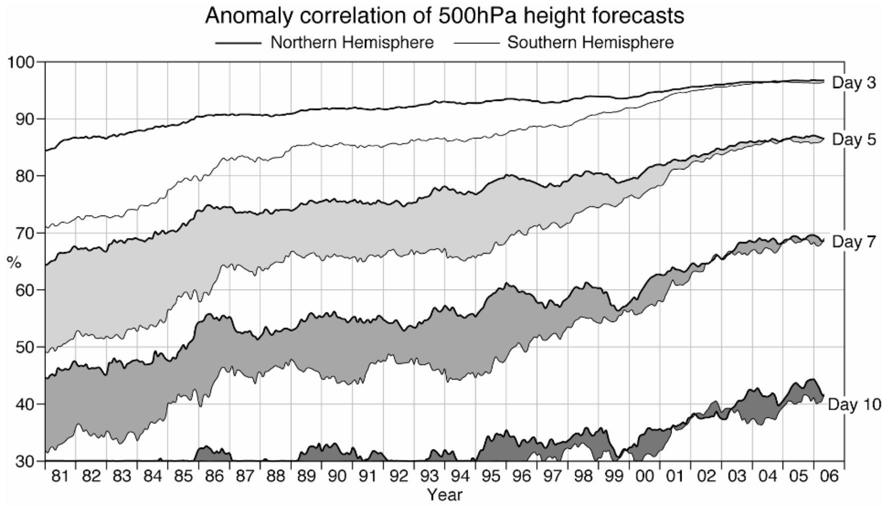

Fig. 5.1 Improvement in the anomaly correlation skill score in ECMWF forecast system over the past 25 years. Improvement is shown for various forecast lead-times up to a 10 day lead. The upper curve for each lead-time shows skill in the Northern Hemisphere while the lower curve shows skill in the Southern Hemisphere

necessity of dimensional reduction that was discussed with regard to the Kalman filter. In particular, the use of localization in the ensemble covariances, necessary because of severe statistical under-sampling problems, can generate unbalanced motions in a balanced ensemble. A balance constraint on the assimilation provides information that restricts the degrees of freedom allowed to respond to data. The basic constraint used in this restriction is that the atmosphere (and thus a prediction model of the atmosphere) should not evolve too fast. Specifically, the atmosphere and atmospheric models support motions on a wide span of space and timescales, from sound waves to quasi-stationary variations due to slow forcing of the atmosphere by the ocean. Weather variations evolve and progress on a timescale of days to weeks for synoptic scales and hours to days for convection and mesoscale disturbances. The fact that weather does not propagate as fast as acoustic and high frequency buoyancy oscillations is used in variational assimilation methods to maximize the information associated with an observation. The process of filtering high frequency oscillations from the assimilated state requires a compatibility condition (i.e. a balance of the terms responsible for the oscillation) resulting in a balance condition. This is described mathematically in the appendix. The dispersion diagram for waves in an equatorial beta plane model of the tropical atmosphere is shown in Fig. 5.2. For the external mode of the atmosphere and the equatorial beta plane model, the phase speed of the Kelvin wave shown in this figure is the same as a sound wave, propagating at 300 m/s. This is far faster than the speed of Rossby waves which propagate at the advective speed of ~10–20 m/s.

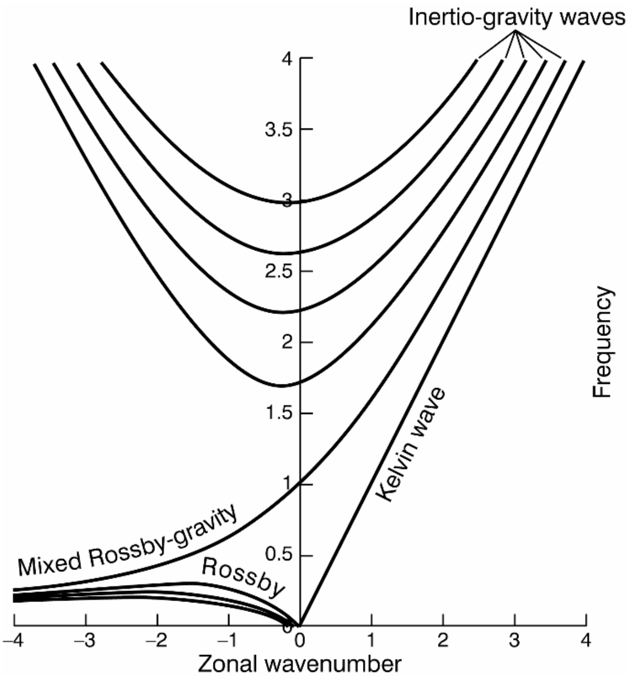

Fig. 5.2 The dispersion diagram showing the scaled frequency vs wavenumber for waves on an equatorial beta-plane. Eastward propagating waves have positive wavenumber while westward propagating waves are denoted with a negative wavenumber

With 4D-Var the temporal consistency of the dynamics is such that the need for balanced initialisations independent of the assimilation cycle has diminished and in practice it is no longer carried out. In the atmosphere it is more convenient to enforce balance as a weak constraint, i.e. penalizing imbalance in the penalty function. Similar balance constraint arguments apply in oceanic data assimilation too.

5.2.6 Introduction to Oceanic Assimilation

For climate prediction on timescales longer than a few months the main source of predictability comes from the ocean component (e.g. Palmer and Anderson 1994). Therefore particular attention has to be devoted to the way in which the ocean is initialised. On the other hand, climate prediction is a field considerably younger than weather prediction and so data assimilation in the ocean has lagged behind its sister discipline in the atmosphere. In fact, most techniques used in ocean data assimilation were first developed for atmospheric data assimilation (one exception is the EnKF, which was initially tested in an oceanic context), and these were described in the previous section. Climate forecasting is not the only driver for ocean data assimilation, but it has provided the main impetus. Given that data assimilation methods have already been introduced, only some practical aspects of the methods used for the ocean are described in this section.

5.2.7 Methods used by Prediction Centres for Preparing Ocean Initial Conditions

As mentioned in the introduction (Chapter 1), several research and/or operational centres, such as ECMWF, the UK Meteorological Office and the Australian Bureau of Meteorology Research Centre (BMRC), routinely produce seasonal forecasts. The data assimilation systems used to initialise the ocean component are normally based on either optimal interpolation or 3D-Var. Although the basic formulation of the background errors in the OI or 3D-Var frameworks is prescribed, and therefore fixed in time, it is possible to introduce more complex flow-dependent features. So, for instance, even when only temperature observations are available at a certain location and they are directly assimilated, salinity corrections can be applied too by exploiting ocean physical and dynamical features (e.g. the preservation of the temperature-salinity relationship, as in Troccoli and Haines 1999). This procedure introduces a flow-dependent feature to the background errors of the OI system. It is therefore analogous to including a time-varying dimension to the OI. In addition, the salinity corrections are designed in such a way that imbalances in the density field, normally present when temperature is modified independently of salinity, are markedly reduced (Troccoli et al. 2002).

5.3 Observing Systems

Up to now, only the methodology of data assimilation for the purpose of prediction has been described without any discussion of the observational data that goes into the process. Since observations are the essential ingredient of a data assimilation system, it is crucial to know what the main components of the atmospheric and ocean observing systems are. A comparison between these two systems is also explored.

5.3.1 The Atmospheric Observing System

Atmospheric observations can be segregated into two types: in situ measurements of variables which require sensors to be collocated with the measurement and remotely sensed measurements which rely on inferring physical variables from afar through the inversion of a radiated signal. The radiosonde measurements and satellite temperature retrievals are prototypical examples of in situ and remotely sensed data streams used in atmospheric assimilation. Figure 5.3a shows the temporal evolution of satellite sources as used in the ECMWF data assimilation system (i.e. most of the available satellite sources). In Fig. 5.3b it can be seen how the increasing number of satellite data sources have reflected on the growth in the number of data used in the ECMWF analysis: in excess of 5 million data are used each day.

The spatial data coverage of atmospheric measurements is shown Fig. 5.4. The different coverage for the in situ observations (top two panels) and that for satellite observations (lower two panels) it is noticeable. Quantities normally measured by in situ instruments are temperature, wind velocity, pressure, humidity and precipitation. Currently, there are about 600,000 in situ observations available per day on average (see Fig. 5.3b). Typical measurements from instruments on board satellites are radiances (which depend on temperature and humidity), wind speed and cloud products. The number of satellite observations is much larger than the number of in situ observations: as shown in Fig. 5.3b they are about ten times larger and they are predicted to grow considerably over the next decade. Table 5.1 below details the data sources and the variables they measure.

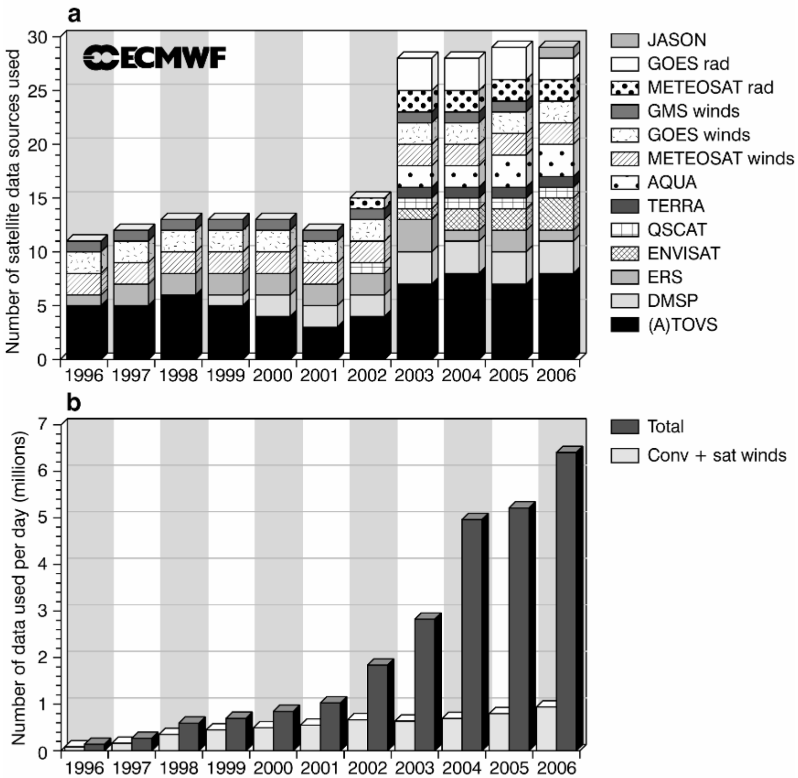

Fig. 5.3 (a) The growth and mix of satellite observations used in ECMWF operational analyses. (b) The total number of observations used each day in the operational ECMWF analysis as a function of time

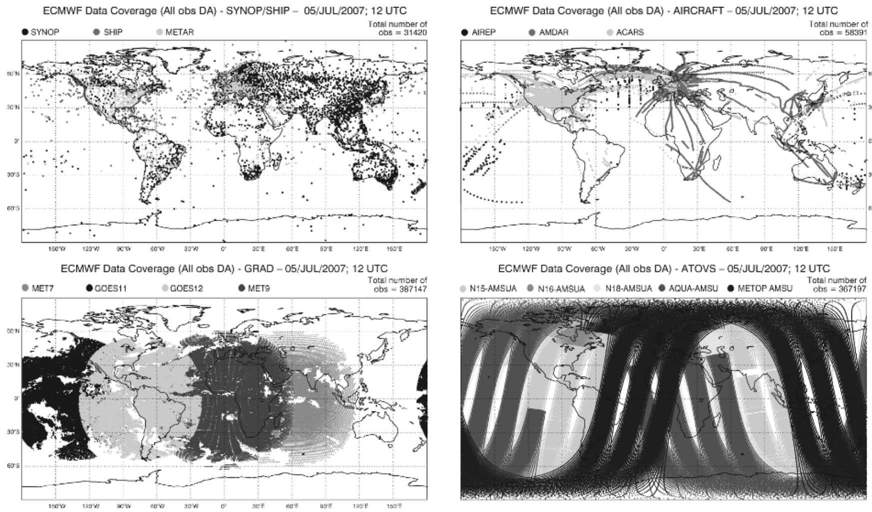

Fig. 5.4 Typical data coverage over a 6-hour period for different data sources. (a) Synop/ship instrument which measures surface temperature, wind velocity, pressure and precipitation; (b) Aircraft for temperature and wind; (c) GRAD for geostationary radiances; (d) ATOVS for polar orbiting radiances. (a)–(b) are in situ measurements whereas (c)–(d) are satellite measurements (as can noticed from their different coverage)

Table 5.1 A summary of the main observation sources (in parenthesis is their technical name) and the variables they measure. The variables are denoted by u and v for wind components, z for geopotential height, p for surface pressure, T for temperature, rh for relative humidity and q for specific humidity, respectively

Observation source	Measured quantities
IN SITU	
Synoptic surface observations (SYNOP/ SHIP)	u, v, p (or z), rh
Aircraft reports (AIRCRAFT)	u, v, T, q
Drifting buoy reports (BUOY)	u, v, p
Radiosonde soundings (TEMP)	u, v, T, q
Wind soundings (PILOT/PROFILER)	u, v
SATELLITE	
Satellite cloud track winds (SATOB)	u, v
Geostationary radiances (GRAD)	Radiances
Polar orbiting radiances (ATOVS)	Radiances
Wind scatterometer (SCAT)	Wind speed

5.3.2 The Oceanic Observing System

Despite the volume of the ocean being only about three times smaller than that of the atmosphere (ca. $4.2{\cdot}10^{18}$ m^3 for the atmosphere and ca. $1.4{\cdot}10^{18}$ m^3 for the ocean),[3] the number of observations in the ocean is considerably smaller than those in the atmosphere. This disparity has contributed to the slower progress in oceanic data assimilation.

In the 1980s the number of oceanographic observations available was several orders of magnitude smaller than its meteorological counterpart. However, with the advent of satellite oceanography, starting in the 1980s, things have changed considerably. Still, there is only so much satellites can measure over the ocean: most notably, the sea surface height (SSH), the SST, the sea surface salinity (SSS) and ocean colour. The oceanic medium is, in fact, opaque to most electromagnetic waves, the most transmittable frequencies being in the visible range (corresponding wavelengths are from ca. 1 μm to ca. 100 nm), for which the penetration has an e-folding scale of about 20m at most (Apel 1987). In practice, therefore, remote measurements are only used to observe the surface of the ocean. To learn about the ocean subsurface, direct measurements are necessary, via so-called in situ instruments.

In terms of surface measurements, the SST is one of the most relevant quantities. Instruments which remotely measure SST are on board several satellite missions. With the objective of developing a new generation of global, multi-sensor, high-resolution (~6 hours and 10 km) SST products, an international project, GHRSST[4] (The GODAE High Resolution SST), has recently started.

Another remote measurement is SSH, which has been measured from space since the mid-1980s. The first mission was Geosat launched in 1985, followed by ERS-1, ERS-2, Geosat follow-on, TOPEX/Poseidon and currently Jason-1. Sea surface salinity is also a quantity that can in principle be measured from space, although the accuracy of remote SSS measurements is not as good as that of SST. Despite the fact that both salinity and temperature affect the density of seawater, salinity is still poorly observed. Plans are underway, however, for satellite missions (SMOS – Soil Moisture and Ocean Salinity – due for launch in 2008 and Aquarius due for launch in 2009) to measure SSS.

Up to the beginning of the 21st century, the majority of in situ instruments only measured temperature profiles. These included instruments like XBTs (eXpend-

---

[3] The volumetric ratio is actually much larger if we consider that only the upper ocean is relevant for seasonal predictions. This implies a more dramatic effect on the disparity of number of observations between atmosphere and ocean. However, it should be born in mind that volume consideration is only part of the story because the time and space scales of processes in the two media are different too. The bottom line is that roughly speaking the ocean is less well observed than the atmosphere.

[4] http://www.ghrsst-pp.org/

able BathyThermographs), mostly along shipping lanes, and the TAO-TRITON array (Tropical Atmosphere Ocean/Triangle Trans Ocean Buoy Network), which consists of approximately 70 moorings in the Tropical Pacific Ocean.[5] The Tropical Ocean Global Atmosphere (TOGA) programme provided the framework into which these moorings, as well as some XBTs and other in situ measurements like tide gauges were developed from the mid-1980s to the end of the 20th century. In addition, campaigns or fixed term projects such as WOCE (World Ocean Circulation Experiment) also provided in situ observations. Since the year 2000, a new observation system called Argo has been introduced. This system has largely modified the way in which the ocean subsurface is observed. Before Argo, observations were mostly taken at the same location (e.g. TAO array) or along tracks concentrated along shipping routes (e.g. XBT profiles) or within limited regions (e.g. Conductivity-Temperature-Depth, CTD) during research campaigns. With Argo, which consists of free-drifting profiling floats that measure the temperature and salinity of the upper 2,000 m of the ocean, most of the ocean can in principle be covered. A large number of Argo floats have been deployed so far: in mid-2006 there were about 2,500, and this should reach about 3,000 in 2007–2008; their measurements are available in near real time.[6]

To maintain the collected data also requires considerable and concerted efforts. Under the CLIVAR[7] (Climate Variability and Predictability) and GODAE umbrellas, several regional projects have taken the challenge to contribute to the development of continuous, automatic, and permanent ocean observation networks. For instance, in the USA the USGODAE[8] project is the reference point, while in Europe, Coriolis[9] has taken the lead. In addition, researchers such as those at the UK Met Office, produce and maintain a range of gridded datasets of meteorological variables for use in climate monitoring and climate modelling.[10]

5.3.3 Comparison Between Atmospheric and Oceanic Systems

Figure 5.5 shows the temporal evolution of in situ observations in the ocean, for temperature and salinity separately as well as for their sum. Temperature observations have consistently outnumbered salinity observations. There are two important

[5] http://www.pmel.noaa.gov/tao/

[6] http://www.argo.ucsd.edu

[7] http://www.clivar.org/

[8] http://www.usgodae.org/

[9] http://www.coriolis.eu.org/

[10] http://www.hadobs.org/

reasons for this: salinity plays a secondary role in the sea water density and salinity is much more expensive to measure. However, with the advent of Argo floats this discrepancy has been considerably reduced (see Fig. 5.5 after the year 2000). More generally, with Argo the oceanic observing system has drastically improved, both in terms of coverage and in terms of number of observations. Given the immensity of the oceans, large gaps still exist and more observations are needed in order to reduce the uncertainties in the oceanic circulation, as also concluded by the recently completed EU project ENACT (ENhanced ocean data Assimilation and ClimaTe Prediction, see the ENACT web site for more info[11]).

It is apparent then that by comparison with the ocean, the atmosphere is much better observed with an average of about 600,000 in situ observations per day coming from instruments like Synop, Aircraft, Pilot/Profilers, Buoys and Temp. Since the in situ observations in the ocean are only about 15 thousand per day (Fig. 5.5), the observing systems in the two media differ by a considerable factor of forty.[12]

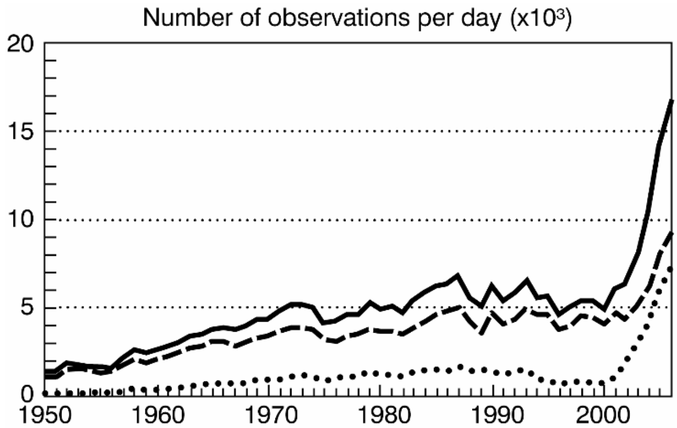

Fig. 5.5 Global number of in-situ oceanic observations on typical model levels as a function of time (dotted line: salinity, dashed line: temperature, solid line: sum of the two). Temperature observations have historically been more abundant than salinity ones. The noticeable downturn in observations in the 1990s was due to the reduction in XBT (eXpendable BathyThermographs) profiles. Since 2000, however, with the advent of the ARGO floats (see text), salinity – as well as temperature – observations have considerably increased.

[11] http://www.ecmwf.int/research/EU_projects/ENACT/index.html

[12] However, this comparison does not take into consideration differences in the physical characteristics of the two media, such as the different radius of deformation (smaller in the ocean) and the different timescales of variability (longer in the ocean). Even taking such differences into consideration, it is likely that the ocean remains under-observed compared to the atmosphere.

5.4 Impact of Oceanic Data Assimilation on Seasonal Forecasts

One of the purposes of data assimilation in the ocean is to provide improved initial conditions which should lead to improved seasonal forecasts. In the absence of oceanic data assimilation facilities, a fallback solution is to use the oceanic initial conditions created by a forced-only integration, i.e. an integration with exactly the same settings as those used in a data assimilation experiment but without assimilating any data.

Despite seasonal forecast performance being germane to the testing of ocean data assimilation systems, there is a surprisingly modest amount of literature which addresses this issue. Some of the reasons for this paucity will become clearer later in this section. Alves et al. (2004), one of the few such references, showed that the use of an OI data assimilation system improved considerably the performance of their coupled model seasonal integrations. More recent results, outcome of the ENACT project, are presented here (Davey 2006). In Fig. 5.6, the assessment of the impact of a 3D-Var assimilation system of seasonal hindcasts (or retrospective forecasts or even re-forecasts) is made in terms of SST anomaly correlations for four different start dates (1st Feb, 1st May, 1st Aug and 1st Nov) so chosen to resolve the annual cycle. These correlations are compared to another set of hindcasts, started from a

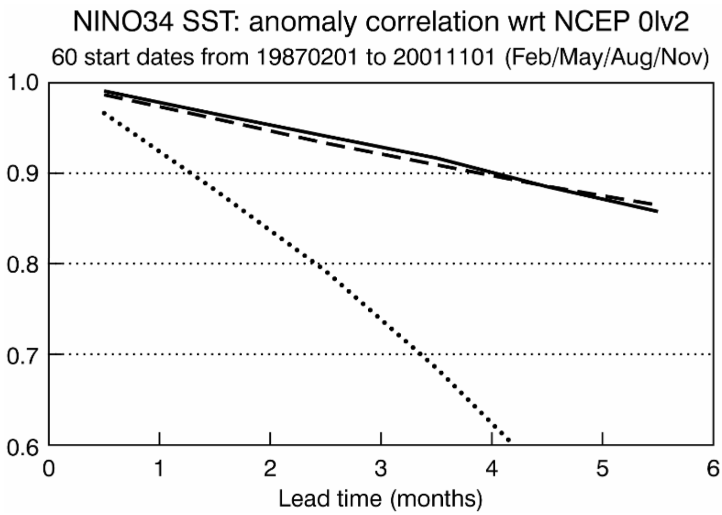

Fig. 5.6 Anomaly correlation for SST as a function of the hindcast lead-time for 3D-Var (solid) and forced-only (dashed). The dotted curve uses persistence as the predictor, i.e. using the value of the month zero for all subsequent months. The assessment period is 1987–2001 and the region is NIÑO3.4 [170–120°W, 5°S–5°N]. Correlation is very high for both hindcast sets and it is slightly superior for the 3D-Var case

forced-only ocean run (dashed line), and to predictions made using persistence, i.e. using the value of the month zero for all subsequent months.[13]

In the central equatorial Pacific (the NIÑO3.4 region), correlation values for both sets of hindcasts (forced-only and 3D-Var initial conditions) are very high – in excess of 0.85 at 6 month lead-time – and markedly better than persistence – less than 0.6 at 5 month lead (Fig. 5.6). The high correlation in these hindcasts is a manifestation of the ability of the coupled model to capture well the SST variability on seasonal timescales in this region. This ability seems to be independent of the two methods used to provide the initial conditions, however, as the correlation for 3D-Var is only marginally better than that for forced-only. This similarity is an indication that, in this region, both the surface forcings used to produce the ocean initial conditions and the characteristics of the coupled model (and its errors) play a prominent role. However, these two factors, the surface forcings and the coupled model errors, play different roles. Whereas improvement in the surface forcings, as have happened in recent years, bring the forced-only and the data assimilation hindcasts closer to each other, improvements in the coupled model, though much more difficult to achieve, would increase their separation as the effect of better initial conditions via the use of data assimilation should reflect on the dynamical evolution of the coupled system on the seasonal timescale. These results thus indicate that, currently, the error in the coupled hindcasts is dominant as it reaches about 0.5°C in root-mean-square error (RMSE) in NIÑO3.4 at lead-month 6 (Fig. 5.7), which is large considering that the interannual variability is only about twice that (not shown). This large RMSE is believed to be largely due to errors in the coupled forecast model that manifest themselves both at the beginning, via the so-called coupling shock, and during the integration of the coupled models; these model errors are one reason why assimilation of ocean data makes only a modest impact on the seasonal hindcast skill.

It is instructive to assess forecast performance in regions other than the equatorial Pacific, hitherto undoubtedly the primary target of seasonal forecasts given the prominence of the ENSO signal. Correlations in the North Atlantic (Fig. 5.8) are smaller than in NIÑO3.4 as expected from the map in Fig. 3.7, although their values are large in absolute terms. Unexpectedly, however, data assimilation does not seem to have a positive impact on correlation. Closer inspection indicates that correlations in the assimilation hindcasts are approximately constant over the 6 month range and do not suffer from the sharp drop at month 6, present in both the forced-only run and in persistence. This behaviour seems to suggest that subsurface information contained in the 3D-Var initial conditions might have a positive impact on the hindcast performance, possibly via the emergence of subsurface signal. It is also worth noting that oceanic regions other than the tropical Pacific

[13] Persistence is the cheapest way to make predictions.

have received less attention in the context of data assimilation (e.g. balances in density and velocity fields used near the tropics may not hold at higher latitudes) and hence the potential for improvement may be substantial.

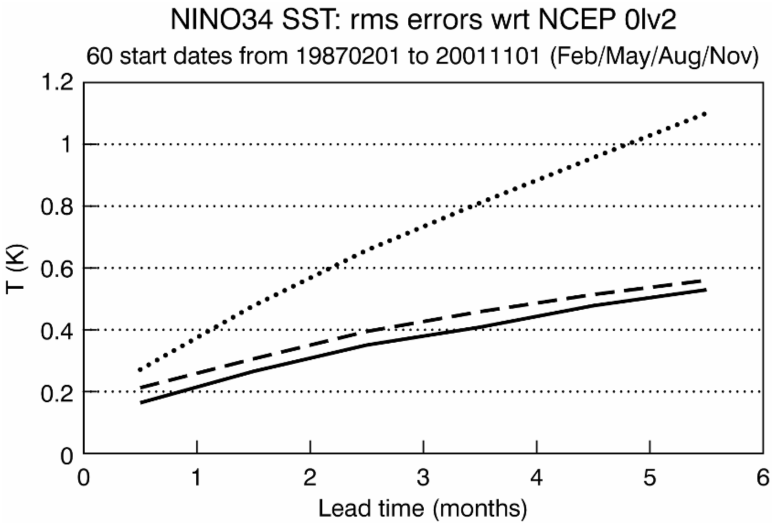

Fig. 5.7 As in Fig. 5.6 but for the root-mean-square error (RMSE). As for the correlation in Fig. 5.6, the RMSE in both hindcast sets is considerably reduced with respect to persistence. Moreover, the initial conditions provided by the 3D-Var assimilation yield improved hindcasts RMSEs for all lead-times compared to the hindcasts started from forced-only oceanic initial conditions

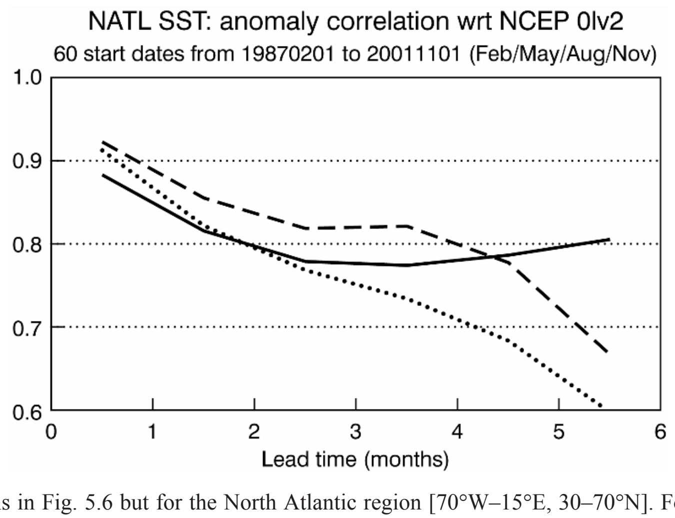

Fig. 5.8 As in Fig. 5.6 but for the North Atlantic region [70°W–15°E, 30–70°N]. For this mid-latitude region, the two hindcast sets do not perform as well as in Niño3.4 and the 3D-Var initial conditions do not seem to yield improved correlations in the first few months of the hindcast. Ocean data assimilation in mid-latitude has thus far received less attention than in tropical regions. Persistence, on the other hand, is a better predictor than in Niño3.4

In the case of the ocean, it is arguable whether "less sophisticated" methods such as OI or 3D-Var, together with improvements as discussed in Section 5.2.7, might be as effective as the "more advanced" methods when seasonal forecast performance is taken as the measure of success. As part of the ENACT project, two of the more advanced methods, 4D-Var and EnKF, were also used to provide initial conditions for seasonal forecasts. Albeit preliminary, results from these two methods showed no significant improvement in forecast performance. While it is true that further development is needed for methods such as 4D-Var and EnKF for the ocean, the apparent lack of impact is likely related to the size of coupled model errors, as discussed above.

5.5 Data Assimilation: An Outlook

5.5.1 What Assimilation Methods are Going to be used in the Medium Term for Initialising the Ocean?

In order to address this question, it is useful to recall the main features in a data assimilation system used to create ocean initial conditions. The principal quality of an oceanic initial condition is the optimal use of as many available observations as possible accompanied by the attainment of a well balanced oceanic state. Such an initial condition would provide an accurate representation of the ocean state and at the same time should help reduce errors due to coupling shocks. In principle, the two best candidate methods to achieve this twofold objective are 4D-Var and EnKF. As presented in Section 5.2, the former has the advantage of providing a dynamically consistent representation of the system over the relatively short assimilation window, whereas the main advantage of the latter is the representation of flow-dependent features which have evolved over time, derived from an ensemble of realisations. The disadvantage of both systems is that they are computationally demanding and this limits the amount and speed of experimentation. Moreover, systematic errors in the coupled model are a significant cause of forecast error. Neither 4D-Var nor the EnKF, in their normal forms as described here, are designed to cope well with systematic errors. This is why, at present, systems such as improved OI and 3D-Var are the preferred choice of most, if not all, seasonal forecasting systems with ocean data assimilation capabilities.

In summary, although 4D-Var and EnKF are considered as the best options for the medium term future of ocean data assimilation, further developments are needed before firm conclusions can be drawn in terms of their impact on seasonal forecasts. It should also be kept in mind that model errors arising during the coupled integration might hinder improvements in the initialisation procedure and therefore the seasonal forecast metric may not be the best to assess the quality of a data assimilation system.

5.5.2 Coupled Data Assimilation: Arguments for and Against

Up to now, the discussion has been concerned only with data assimilation performed separately in the model components, namely in the atmosphere and ocean. As seen in Section 5.1.1, there is only a very weak coupling, via the commonly used SST, in the assimilation across component models and so there is no attempt to reach an optimal coupled state of the system for prediction purposes. Since seasonal prediction depends critically on coupling the atmosphere to the ocean and land surface, the question of assimilation in coupled climate models should be addressed.

The idea of performing data assimilation in the coupled system, rather than carrying it out in each component separately as is currently done, has appealed to scientists for quite some time, as illustrated by Miyakoda as far back as 1986. Although it is considered as a fascinating area of research, coupled data assimilation has not yet properly taken off.

There are three main reasons why this is so. First, as described earlier, the most advanced methods of assimilation are extremely computationally expensive and coupling an atmospheric model to either an ocean or land surface model for joint assimilation only exacerbates an already severe resource limitation. The dimensionality of the problem increases by the dimensionality of the additional ocean model's degrees of freedom and the consequent need for covariance information of co-varying component model variables like SST (ocean) and cloudiness (atmosphere). Second, there is a large mismatch between the natural timescales of the atmosphere, the ocean and the land surface; the ocean and land can be treated as stationary boundary conditions for the atmosphere over an assimilation cycle without much error. Because of this, the decoupled assimilation method sketched in Section 0 works well. Lastly, there are very large systematic biases in coupled models which can overwhelm the benefits of consistency and information sharing that are the raison d'etre for coupled assimilation. An example of the speed with which such biases can impact coupled forecasts is shown in Fig. 5.9, depicting the systematic 2 m temperature biases at month zero (top) and month 1 (bottom) lead-times in an ensemble of ENSO predictions using the ECMWF coupled model.

In spite of these technical and conceptual difficulties, the interest in the scientific community to tackle the issue of coupled assimilation is rising and works such as that by Galanti et al. (2003) might signal the beginning of this new phase in data assimilation.

5.6 Mathematical Appendix

Because so much of data assimilation requires mathematical expression to describe the methods used, many of the mathematical details of the methods presented in the core of the contribution have been relegated to this appendix. It is hoped it can

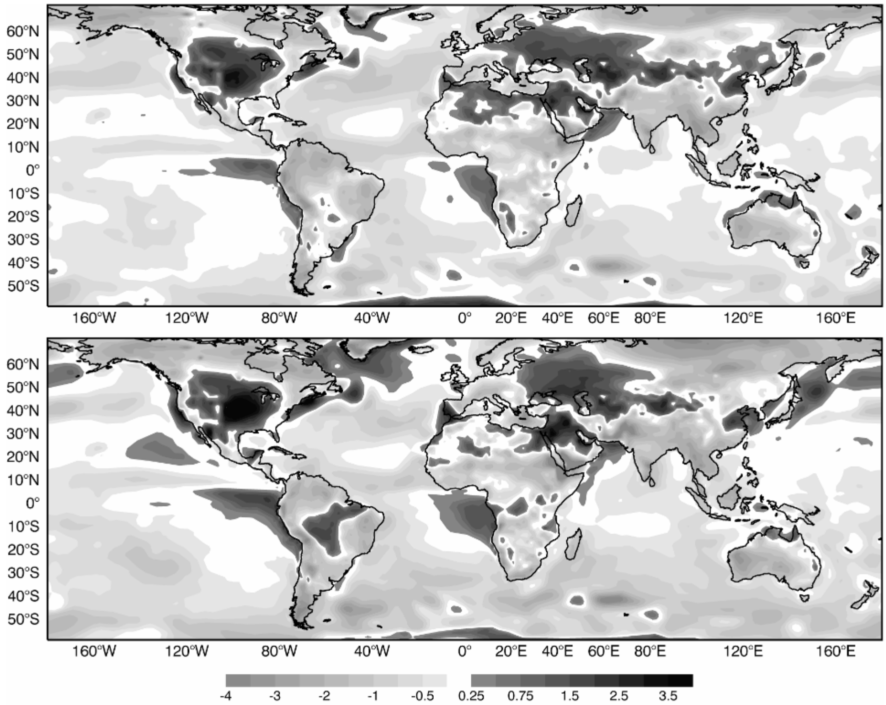

Fig. 5.9 The evolution of the 2 m temperature systematic error (or bias) in °C at month zero (top) and month 1 (bottom) lead-times in an ensemble of seasonal forecasts using the ECMWF coupled model (System 3). Note how the bias generally increases with increasing lead-time. The ensemble of cases is taken from asset of forecasts initiated at the beginning of each July from 1981 to the year 2000 (hence top panel shows July mean and bottom panel August mean)

be used as a concise reference for those actually attempting to practice the art of assimilation. A note of caution has to be spent: practical applications of any assimilation method are always distinctly different from their theoretical formulation. Therefore, if you intend to implement any of the following methods, it would be prudent to consult the given references.

5.6.1 Least Squares Minimization

The method of Panofsky used a least squares polynomial fit to the height field, i.e.

$$z(x, y) \approx \sum_{i+j \leq 3} a_{i,j} x^i y^j .$$

The coefficients $a_{i,j}$ are determined by minimizing the misfit at the observation points. Thus

$$I \equiv \sum_{obs} (z - z_{obs})^2$$

is minimized by differentiating I with respect to each coefficient and setting the derivative equal to zero,

$$\frac{\partial I}{\partial a_{i,j}} = 0 .$$

This results in a linear algebraic system of equations to be solved for the $a_{i,j}$'s.

5.6.2 Cressman Scheme

The scheme of successive corrections developed by Cressman and Bergthorsson and Doos assimilates deviations from a background or first guess. So for any state variable, say the temperature T at point j, its increment is given by:

$$T_{anal}(j) - T_b(j) = \frac{\sum_{i=1}^{N} w(i,j)(T_{obs}(i) - T_b(i))}{\sum_{i=1}^{N} w(i,j)},$$

where T_b is the first guess temperature and the sum is over all i = 1,2,…N temperature observations present at analysis time. The weight function given by Cressman is

$$w(i,j) = \max(0, \frac{r^2 - d(i,j)^2}{r^2 + d(i,j)^2}),$$

where $d(i,j)$ is the distance between the points i and j. The analysed temperature is thus a weighted average of the background and the increments to the background, or innovations, from the observed temperature.

5.6.3 Optimal Interpolation (OI)

This method, pioneered by Gandin (1965), uses the statistical covariance between fields at different points in space to determine the interpolation weights. As in the Cressman scheme it uses the anomaly of both the observations and the analysed field from a background field supplied from climatology or a forecast field. This is done as follows. The field is broken into the guess and the anomaly,

$$f_{grid} \approx f_{guess} + f'_{grid}$$

where the guess field, assumed to be bias-free, is climatology in Gandin's original formulation and the anomaly is assumed to be a weighted average of the anomalies in the observations, i.e.

$$f'_{grid} = \sum_k w_k f'_{k,obs}$$

The weights are then selected using the minimum mean square error (least squares) criterion, with

$$\overline{E^2} = \overline{(f'_{grid} - f'_{true})^2}$$

being the mean square error and at each grid point then

$$\overline{E_i^2} = \overline{(f_i - f_{guess} - \sum_k w_k f'_k)^2}$$

where the over bar denotes the statistical average and *i* and *k* represent the analysis point and the observation points respectively. Minimizing, results in:

$$\frac{\partial \overline{E_i^2}}{\partial w_k} = 0 \Rightarrow \overline{(f'_i f'_k)} = \sum_j w_j \overline{(f'_j f'_k)}$$

which is a linear algebraic equation for the weights in terms of the two-point covariances of the field *f*.

OI is usually implemented by solving these simultaneous equations in matrix form (they turn out to be the same equations as for 3D-Var below, see Lorenc 1986). There are as many equations as observations, so to reduce computational costs only a local selection of observations is used in each location.

5.6.4 Variational Assimilation

Both 3D-Var and 4D-Var have a similar structure with only minor modifications in the final form of the assimilated state. Of course, the computational cost and actual assimilated state are generally very different. 3D-Var begins with the cost function defined as:

$$J = \tfrac{1}{2}(\vec{x}-\vec{x}_b)^T \mathbf{B}^{-1}(\vec{x}-\vec{x}_b) + \tfrac{1}{2}(\vec{y}-\vec{H}(\vec{x}))^T \mathbf{R}^{-1}(\vec{y}-\vec{H}(\vec{x}))$$

where the various terms have been defined in Section 5.2.4. Differentiating this gives an equation for its gradient with respect to x:

$$\nabla_x J = \mathbf{B}^{-1}(\vec{x}-\vec{x}_b) - \mathbf{H}^T\mathbf{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_b)) + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}(\vec{x}-\vec{x}_b)$$

where J has been expanded about the background state, $\vec{x}_b$. Setting the gradient to zero to find the minimum gives:

$$\vec{x}_a - \vec{x}_b = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\vec{y}-\vec{H}(\vec{x}_b))\,.$$

The distinguishing feature of 3D-Var and 4D-Var algorithms is that they only need software to calculate the scalar J, and its gradient grad_x J; the latter is the size of a model state and so no more complicated to manipulate. With these, and an appropriate descent algorithm, the x which minimises J can be found to any desired accuracy, without ever needing to explicitly represent or manipulate the matrices B and R – which would be impossible for the size of models and numbers of observations used.

The equivalence between this approach and OI can be seen by deriving an explicit equation for the $\vec{x}_a$ which minimises the penalty function. At the minimum $\nabla J = 0$, giving:

$$\vec{x}_a = \vec{x}_b + \mathbf{K}\left(\vec{y}-\vec{H}(\vec{x}_b)\right).$$

The Kalman gain matrix **K**, so-called because it appears also in the Kalman filter equations below, is given by:

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\,.$$

But this form is not the most convenient because of the many matrix inverse operations. With some matrix manipulations an equivalent form,

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\,,$$

can be developed which is computationally less demanding and is of the form where the uncertainties in the background and observations are more easily seen as contributing to the weight of each. Neither of these forms for **K** can even be stored, let alone calculated, for the huge models and many observations in our problems. The second form, made cheaper by considering at one time only a limited local area and data selection, is implemented in OI.

It is common to use an incremental form of 4D-Var, in which one assumes that the problem can be well approximated in a linear fashion in terms of $\delta\vec{x} \equiv \vec{x}(t) - \vec{x}_b(t)$. Linearity enters through the use of a linear model to update $\delta\vec{x}(t_i) = \mathbf{M}_{i,0}\delta\vec{x}(t_0)$, where $\mathbf{M}_{i,0}$ is the linear operator (matrix) propagating the initial value $\delta\vec{x}(t_0)$ to the ith time interval. The 4D-Var cost function is then given by:

$$
\begin{aligned}
&J(\delta\vec{x}(t_0)) = (\delta\vec{x}(t_0))^T \mathbf{B}^{-1}(\delta\vec{x}(t_0)) \\
&+\sum_k ((\delta\vec{y}_k - \vec{H}(\mathbf{M}_{k,0}\delta\vec{x}(t_0)))^T \mathbf{R}_k^{-1}(\delta\vec{y}_k - \vec{H}(\mathbf{M}_{k,0}\delta\vec{x}(t_0)))
\end{aligned}
$$

where $\delta\vec{y}_k \equiv \vec{y}(t_k) - \vec{H}(\vec{x}_b(t_k))$) and the linearised propagator $\mathbf{M}$ has been used to update the deviation state vector $\delta\vec{x}$. Note that the summation over the index *k* corresponds now to a summation over all the observation within the time interval over which *J* is defined. As above, the analysis increment to the background can be written in terms of a gain matrix of the form

$$\mathbf{K} = \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1} \text{ or } \mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{RH})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

so that the analysis increment is:

$$\delta\vec{x}(t_0) = (\mathbf{B}^{-1} + \sum_i (\mathbf{HM}_i)^T \mathbf{R}(\mathbf{HM_i})^{-1} \sum_i (\mathbf{HM}_i)^T \mathbf{R}^{-1}(\mathbf{HM}_i)\delta\vec{y}_i$$

An additional step normally taken is to further re-arrange the terms so as to avoid directly defining the matrix $\mathbf{B}$. This is done by defining $\mathbf{B}$ as $\mathbf{B} = \mathbf{UU}^T$ After a little math: $J_b = \delta\vec{u}^T\delta\vec{u}$.

Useful references for atmospheric applications of 4D-Var are Klinker et al. (2000), Fisher (2003), Lorenc and Rawlins (2005) and for the oceanic applications Vialard et al. (2003), Weaver et al. (2003, 2005).

5.6.5 The Kalman Filter

The main advance in the Kalman Filter is the use of a forecast of the error covariance. Consistent with linear dynamics

$$\delta\vec{x}(t_i) = \mathbf{M}_{i,0}\delta\vec{x}(t_0),$$

the forecast error covariance at time step $i + 1$ is given by

$$\mathbf{P}_f(i+1) = \mathbf{M}_{i+1,i}\mathbf{P}_a(i)\mathbf{M}_{i+1,i}^T + \mathbf{Q}$$

where $\mathbf{Q}$ is a model error covariance taken to be white noise in time and $\mathbf{P}_a(i)$ is the analysis error covariance at time step i. The gain matrix is:

$$\mathbf{K}(i) = \mathbf{P}_f(i)\mathbf{H}^T(i)(\mathbf{H}(i)\mathbf{P}_f(i)\mathbf{H}^T(i) + \mathbf{R}(i))^{-1}$$

so that the analysis increment is given by

$$\vec{x}_a(i) - \vec{x}_f(i) = \mathbf{K}(i)(\vec{y}(i) - \mathbf{H}(i)\vec{x}_f(i))$$

where the increment is the difference between the analysis and forecast state vector. To complete the algorithm the analysis error covariance must be specified; for minimum mean square error it is:

$$\mathbf{P}_a(i) = (\mathbf{I} - \mathbf{K}(i)\mathbf{H}(i))\mathbf{P}_f(i).$$

If the model M is linear, and if we either set Q = 0 here or allow for non-zero Q in 4D-Var, then these equations have an identical solution to 4D-Var. The difference is in the algorithm – 4D-Var iterates over a time window so that it can avoid explicit representation of the error covariances, whereas the Kalman filter does not iterate and can be integrated forward indefinitely, at the additional cost of explicit representation of the covariance matrix.

5.6.6 The Ensemble Kalman Filter

In the Ensemble Kalman filter (EnKF) the idea is to assume the ensemble covariance is a good estimate of the forecast error covariance:

$$\mathbf{P}^f \approx \mathbf{P}_e^f = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)^T}$$

(In practice the ensemble covariance is often localized in space to minimize the sampling errors inherent in small magnitude, geographically distant correlations.)

The analysis on each member of the ensemble:

$$\mathbf{x}_j^a = \mathbf{x}_j^f + \mathbf{P}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}\right)^{-1} \left(\mathbf{y}^o - \mathbf{H}(\mathbf{x}_j^f)\right)$$

The same relation holds true for the ensemble mean:

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{P}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}\right)^{-1} \left(\mathbf{y}^o - \mathbf{H}(\overline{\mathbf{x}^f})\right)$$

Useful references for oceanic applications are: Keppenne et al. (2005) and Leeuwenburgh (2005).

5.6.7 Balance

The atmosphere supports free variations on multiple timescales and some high frequency variations, like sound waves, typically have only very small amplitude. This fact can be used to constrain assimilation. A model for the atmosphere can be written in the following fashion

$$\frac{d\vec{x}}{dt} + L\vec{x} = \vec{N}(\vec{x})$$

where L represents linear terms in the equation and N represents the non-linear terms. The fact that an atmospheric model can support high frequency variations is evident in the structure of the characteristic frequencies associated with the eigenvalues of the matrix L. The development of a balance relationship begins with the separation of the state vector $\vec{x}$ into low frequency components $\vec{y}$ and high frequency components $\vec{z}$ using the eigenvectors of L to effect this decomposition. The high-frequency equation can be written:

$$\frac{d\vec{z}}{dt} + L_z\vec{z} = \vec{N}_z(\vec{y}, \vec{z})$$

Balance conditions are derived by noting that high frequency variations can be suppressed by minimizing the time derivative term in the equations in some fashion. A simple way to minimize the time tendency is to set $\vec{z} = 0$, corresponding to geostrophic balance in a primitive equation model. More refined balance conditions posit a slave relationship between the high-frequency part of the state vector and the low frequency components; i.e. $\vec{z} = \vec{M}(\vec{y})$. This ensures that the dynamics, at least for a short period of time, resides on a slow manifold of possible solutions and the constraint of balance can be used either diagnostically in assimilation by requiring the covariance structures obey the balance constraint or directly as part of the variational assimilation by imposing balance as a weak constraint.

Although (normal mode) initialisation was used for many years at weather centres such as ECMWF, it is no longer used now that 4D-Var is operational having been replaced by a weak constraint. It is not generally used in ocean analysis.

Chapter 6
Modelling the Atmospheric, Oceanic and Coupled System

Brian Hoskins, Paul Schopf, and Antonio Navarra

A model of the coupled system consists of component models of the atmosphere and of the ocean as well as software to link the two. The atmosphere must see the slowly evolving ocean sea surface temperature (SST) while the ocean must see the rapidly changing weather, in the form of the surface exchange of momentum, heat and freshwater. Although one should consider the coupled system as a whole, much progress has been made by considering aspects of the atmosphere and ocean modules separately. Computer restrictions mean that both the atmospheric and oceanic components have to be simplified. Many processes in both media take place at scales smaller than can be resolved by the component models and so must be parameterised as they cannot be explicitly resolved. Methods to deal with limitations to parameterisation are discussed. A variety of ways of validating testing and improving atmospheric general circulation models is considered. This can be done by making long runs of atmospheric models with observed SSTs or with simplified earth systems, such as aqua-planet models, to determine the importance of the process withheld. Often several different models are used in order to intercompare results. Examination of imbalances early in the model forecasts can also give clues as to model deficiencies. This latter use is interesting as it brings together to some degree weather and climate forecasting. Some other promising options are to run models at very high resolution for limited periods and to use the model results to guide and test the development of parameterisations appropriate for lower resolution. Early models of ENSO used ocean models of intermediate complexity in which the 'essential' physics was included but many processes were either excluded or heavily parameterised. However, there are now many coupled models using ocean general circulation modules. These are, like

Brian Hoskins
University of Reading

Paul Schopf
George Mason University

Antonio Navarra
Centro EuroMediterraneo per i Cambiamenti Climatici

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

their atmospheric counterparts less prescriptive, with many more degrees of freedom. Parameterisation of subgrid processes is important particularly mixing along and across density surfaces. How to deal with salt, poor representation of the flux of freshwater across the surface, and the run-off from rivers and melting ice pose particular challenges. Examples of model error in some key atmospheric fields such as precipitation when the atmosphere is forced by the observed SSTs and when the atmosphere is coupled to the ocean are shown. The errors may well get bigger in the coupled case as errors in either the atmosphere or the ocean may cause errors in the other medium and a positive feedback may result. This results in climate drift, a major problem facing all modellers of climate, be it on a short or a long climate timescale.

6.1 Atmospheric Models

Atmospheric models have been developed for the weather forecasting problem. They form the central component of predictions on longer timescales, including projections of climate change due to human activity and of predictive Earth system models with the inclusion of extra ingredients such as interactive vegetation, atmospheric composition and ice sheets.

The basis for the models is the set of equations for momentum, mass, thermodynamics, and water vapour content. To put them on to a computer, these equations for a continuous fluid have to be turned into a description of the system using a finite set of numbers. This is usually performed using the values at a discrete number of levels in the vertical, in the horizontal by using discrete points alone or in combination with the coefficients of a representation in terms of functions (spherical harmonics), and using discrete intervals in time. There is a wide range of choices to be made over how to represent derivatives in space and time, but there is considerable experience in atmospheric modelling and other computational fluid dynamics applications to guide this choice.

As will be discussed later, processes should either be represented explicitly or parameterised, i.e. their effect on the model variables is represented in terms of the model variables themselves. Considerable effort has been put into developing many different parameterisations of radiative processes, clouds, convection of various kinds, large-scale latent heat release, surface exchanges and boundary layer turbulence, the drag associated with gravity waves triggered by mountains and perhaps by rapid events in the atmosphere, and interior mixing processes. These have been developed in the context of observational studies including special targeted observational programmes and, sometimes, detailed modelling of the phenomena under consideration. For one version of the ECMWF model, the temperature tendencies over the first time-step given by the radiation and convection parameterisations zonally and over 1 year are given in Fig. 6.1a and b, respectively, and the total tendency given by all the parameterisations in Fig. 6.1c.

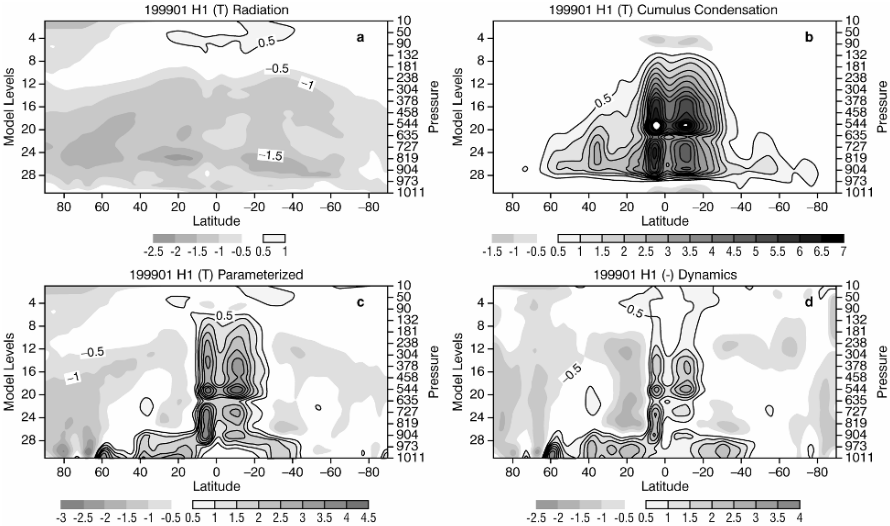

Fig. 6.1 Parameterised temperature tendencies computed from single time-steps in the ECMWF model and averaged zonally and over 1 year: (a) Radiation, (b) Convection, and (c) Total. (d) shows the residual between the total parameterised tendency (given in (c)) and the tendency due to the dynamics

The resolution of the model and the detailed nature of the individual parameterisations used depend on the timescale for the prediction to be made and the computational power available. For example weather forecasting for 1 day may require only crude representation of some radiative processes, whereas climate change simulations will need to reflect accurately the detailed effect of trace constituents on the residual between short and long wave radiation. Currently, the horizontal grid of the model may be 30–50 km for a global weather prediction model and 100–250 km for a climate model. The vertical resolutions used in the various contexts tend to be more similar, typically 0.5–1 km, respectively in the two contexts.

6.1.1 Model Performance Evaluation

6.1.1.1 Introductory Comments

The fundamental evaluation of model performance is through comparison with the real system. There are a number of difficulties in this evaluation. Firstly, there is uncertainty over the state of the real system at any instant and consequently in any statistics derived over a period. Observations are limited in space and time and in the components observed. Estimates of, for example, precipitation can be made from rain-gauge data together with inferences from satellite radiance measurements. Alternatively, use can be made of the data assimilation performed at operational forecast centres where all the available observations and the previous short range forecast are combined together in the context of the model. However the product of such data analysis will reflect any errors in the model used. Routine operational analyses over a period suffer from the fact that the observational system changes in time, e.g. there are few satellite data before 1979, and also that the analysis system itself undergoes changes. To overcome the latter problem various centres have performed reanalyses for many decades of observational data using modern analysis systems. They also try to use data that may have not been available for the analyses performed at the time. Details about reanalyses may be found at the web address below.[1]

One particular reanalysis is that at ECMWF for the period 1957–2002 (Uppala et al. 2005, see also Chapter 3). An atlas compiled from the so-called ERA-40 data is available on the web[2] as well as a special quick access web version.[3] However, whether in the routine operational process or in the special reanalyses there remain considerable discrepancies in some fields, such as precipitation, between the products from different analysis centres.

The second uncertainty in evaluation of atmospheric model performance is associated with the extent of the system concerned. For example in the seasonal context if the atmospheric model is coupled to an ocean, any errors may, or may not be due to defects in the ocean model. If the sea surface temperatures (SSTs) and sea ice are held constant throughout the forecast then this will lead to errors. If in hindcast mode the SSTs and sea ice are specified from observations, errors could still arise due to the lack of two-way interaction between the atmosphere and the underlying ocean.

6.1.1.2 Model Intercomparisons

Models may also be compared with each other as well as with the real system. There has been a lot of such activity, focussed mainly on the World Climate Research Programme/Working Group on Numerical Experimentation Atmospheric Model Intercomparison Project (AMIP). A large number of comparisons of particular aspects in many models run with specified SSTs and sea ice for a 17-year period have been performed. One aim is to learn about the skill that may be possible in predictions on monthly or seasonal timescales using current atmospheric models. AMIP and other model intercomparison studies have proved very successful in this regard. A second aim is to determine whether particular model abilities or defects can be related to particular model ingredients. However, such

---

[1] See: http://dss.ucar.edu/pub/reanalyses.html

[2] See: http://www.ecmwf.int/publications/library/do/references/list/192

[3] See: http://www.ecmwf.int/research/era/ERA-40_Atlas/index.html

associations have in general proved elusive because of the large number of non-linear interactions and the large variety of processes represented in models.[4]

One approach to obtaining information on how fundamental aspects of model results are related to its ingredients involves the making of drastic simplifications. To learn about the representation of the basic equations, the so-called dynamical core experiments (Held and Suarez 1994) reduce the physical parameterisation package to a relaxation to a specified thermodynamic structure on a timescale of about 1 week plus a simple boundary layer drag. To investigate the interaction of the physical parameterisations with the dynamics, full models have been retained but with the underlying planet simplified to be water covered everywhere with a specified zonally symmetric SST plus possibly simple anomalies. A model inter-comparison has been performed using such aqua-planet models.[5] It has been found for example that under certain circumstances some models give a single convective maximum on the equator, whereas others give two maxima either side of the equator, a double Inter Tropical Convergence Zone (ITCZ).

6.1.1.3 Systematic Errors

For an operational weather forecast model, errors can be examined at various timescales. Using the technique based on summing initial time-step tendencies pioneered by Klinker and Sardeshmukh (1992), referring again to Fig. 6.1, if the forecast model were consistent with the initial data given to it, then on average the thermal tendency given by the parameterisations (Fig. 6.1c) would be exactly cancelled by that associated with the resolved dynamics. Figure 6.1d shows that the actual cancellation is far from exact. This could be due to errors in the observational data or its assimilation, in the equation representation in the model or in the physical process parameterisation. In the latter cases it could be associated with a systematic error or with a short timescale spin-up/adjustment process. The systematic error grows and evolves in time. However, as discussed in Jung (2005) the growth saturates and the seasonal mean error for the same model run in hindcast mode with specified SSTs (Fig. 6.2a) has many features in common with the 10-day error (Fig. 6.2b), and the long-term anticyclonic error over the western and central North Pacific is in fact present even at day 3 (not shown).

6.1.1.4 Extratropical Systems

Two important aspects of the extratropical weather and climate are the storm-tracks and blocking highs. The storm-tracks are the regions in which mid-latitude

---

[4] See: http://www-pcmdi.llnl.gov/projects/amip/ for details of AMIP.

[5] See: http://www.met.reading.ac.uk/~mike/APE/ for details.

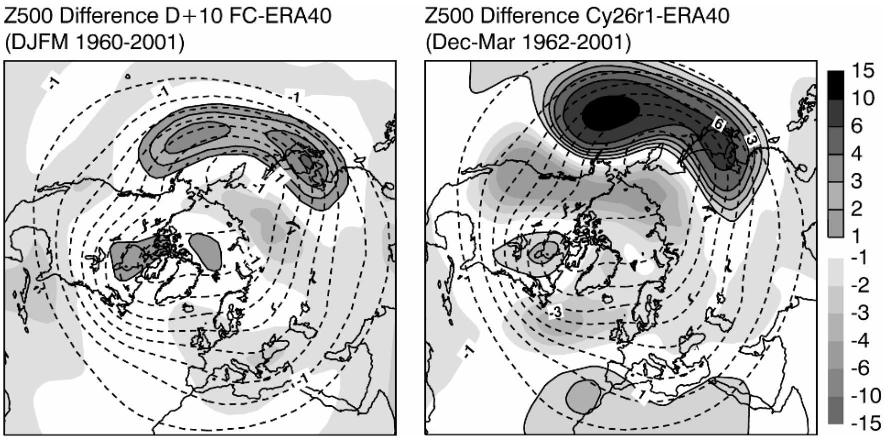

Fig. 6.2 Systematic errors for Dec–Mar in the ECMWF model in the period of about 1960–2000 for 10 days (a) and seasonal forecasts (b)

weather systems characteristically grow move and decay. Any changes in the Northern Hemisphere storm-tracks in any particular season are important, in particular for North America and Europe that are situated at the end of the two main tracks. Figure 6.3 shows the average winter storm-tracks determined from ERA-40 (as in Hoskins and Hodges 2002) and from three AMIP models (which will have been updated since this figure was made). It is clear that the models capture the general storm-track structure but have errors in the detail that vary from model to model.

The frequency, intensity and positioning of blocking highs are crucial for any season in many mid-latitude regions, and in particular for Europe. Blocking is associated with a reversal of the usual westerlies on the equatorial flank of the high, and strong meridional flows upstream and downstream and these features tend to be fairly stationary. All these aspects lead to anomalous weather. On seasonal and longer timescales it has been found that most models capture the regions in which blocking tends to occur, but they under-estimate its frequency and intensity.

6.1.1.5 Tropical Behaviour

Models have general success in simulating tropical phenomena such as monsoons on average though their variability is difficult to capture. On the seasonal timescale there is great interest in ENSO which is a coupled ocean-atmosphere phenomenon for which there is some predictive skill. However the ENSO onset may be strongly influenced by westerly wind bursts associated with the progression of large regions of organised convection from the Indian Ocean to the West Pacific on a timescale of weeks. This "Intraseasonal or Madden-Julian Oscillation", and also convectively coupled equatorial waves and even the diurnal cycle all still provide big challenges for atmospheric models. See also Chapters 3 and 4.

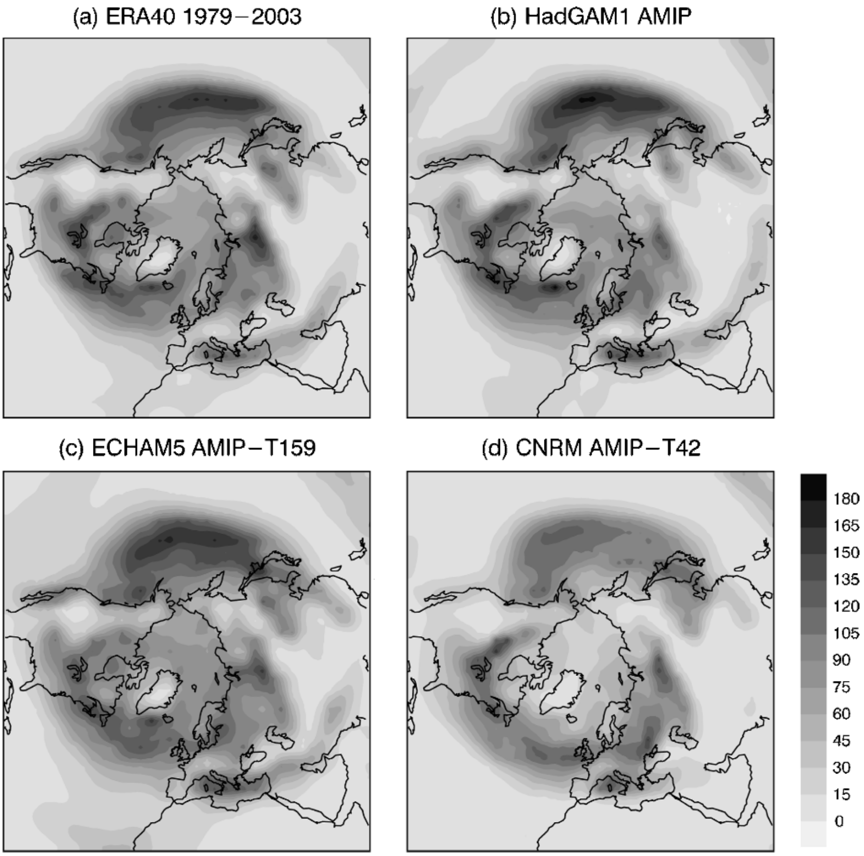

Fig. 6.3 Northern hemisphere winter storm-tracks in ERA data and in three AMIP (Atmospheric Model Intercomparison Project) models. The field shown is the track density of cyclonic features at 850 hPa (K. Hodges 2005, personal communication)

6.1.2 Prospects for Improving Models

A number of different approaches for improving models will be discussed.

6.1.2.1 Comparison with New Observations

The advent of a new observational instrument gives an opportunity for a fresh view of model performance through a comparison with its observational data. For example, in the SINERGEE project, the observations of outgoing long wave radiation (OLR) and albedo made by the GERB satellite have been compared with

Hadley Centre model data.[6] One conclusion was that in the model the Sahara desert did not reflect sufficient solar radiation. Alternatively the new observations may come from one of the special limited-time, focussed observational programmes.

6.1.2.2 Diagnoses of Observational and Model Data

The development and application of new diagnostics based on theoretical understanding may give new insights into model behaviour. The analysis of equatorial waves with embedded deep convection provides one example. A diagnosis of convection in the tropics, or equivalently of high, cold cloud shows that the models do not correctly capture the peaks in wave number and frequency that are present in the observations and the phase speeds are wrong. Going further than this, using a new technique, based on equatorial wave theory, it is possible to isolate particular wave structures in the observations and the models. Initial indications from one such study are that models are seriously in error in some aspects of the wave structures. The challenge will be to understand why and produce model changes that improve the representation, without introducing other deleterious effects.

As discussed above, initial tendencies or single time-steps in models initiated from good data analyses can indicate spurious imbalances in the model terms. The data analysis system itself can also yield useful information through study of the increments added to the model first guess in the data assimilation procedure. For example, if the observations are always trying to moisten the model tropical atmosphere, it suggests that the model parameterisations are producing too much drying in the region.

6.1.2.3 Performance on Weather-Seasons-Climate

Looking at their performance on timescales other than seasonal can provide interesting insight into which aspects of models that are to be used for seasonal forecasting need attention. On the short timescale, the ability of the models to represent individual synoptic systems or regimes of flow can be assessed by using them for re-forecasting the weather. On longer timescales, running the models for one or more years will give information on the tendency of the model to drift to a different climate. Such a tendency may be damaging for seasonal and shorter timescale predictions but may be less clear as to its nature.

The models can be run either with specified SST or in coupled mode. Seasonal hindcasts themselves can also be run in coupled or uncoupled mode. Controlled

[6] http://www.nerc-essc.ac.uk/Research/Atmospheric/Atmospheric.htm#sinergee

experimentation can elucidate the relationship between errors in the model performance; for example, by warming the SSTs to artificially improve the convection in the tropical west Pacific, the impact of this error on the model results elsewhere can be assessed (Turner et al. 2005).

6.1.2.4 Results for Simplified Problems

As discussed above, running a full atmospheric model with a simplified lower boundary, for example an aqua-planet, can yield useful information on the representation of the dynamics and the parameterisations, and the interaction between them. Going even further, by drastically simplifying the parameterisations, dynamical core experiments can provide useful comparative information on numerical representations of the basic equations.

6.1.2.5 Very High Resolution Runs

Operational forecast models are run at a spatial resolution (particularly in the horizontal) that is determined by the availability of computer power. However in case-study mode, global models can be, and are being run with an order of magnitude finer resolution, 10 km or less. Such studies give information on benefits that may be had through, and problems that may be solved by, going to finer resolution. They may also yield ideas on how the parameterisations at lower resolutions could mimic the higher resolution behaviour and therefore perform better. Another version of this investigation is to run a limited area version of the model in which even higher resolution will be possible. The negative aspect in this case is that there is now dependence on the imposed boundary conditions.

6.1.2.6 Different Computational Approaches

In atmospheric modelling and in the wider area of computational fluid dynamics there are new ideas on numerical methods that may improve seasonal forecast models. Different grid meshes to cover the sphere, and different ways of determining derivatives in space and of integrating in time are all being raised and tested. Options for the vertical coordinate, such as potential temperature (θ) are being assessed. This coordinate has the advantage that air would move along the coordinate surface unless heat was added to it. However the disadvantages associated with θ-surfaces intersecting the Earth's surface have currently limited its application. Given the important role played by it (see Chapter 4), another theoretically very attractive option is to use potential vorticity (PV) as one of the model variables.

6.1.2.7 Development of New Parameterisations

Apart from Section 6.1.2.6, the approaches discussed above are generally aimed at giving information on possible defects in the parameterisations. The hope is that this will suggest how they may be modified or new parameterisations may be developed. They also provide means for testing models that include these modified or new parameterisations.

Another way of developing a new convective parameterisation, for example, is to use experimentation with a cloud resolving model, based on a grid of perhaps 100 m or less in a box the size of a few grid cells in the global model, and look for the equilibrium tendencies associated with the simulated convection for steady boundary conditions. This can be seen as setting the target for the convective parameterisation. This technique is now gaining wide popularity. However, help in capturing the crucial interaction of convection and dynamics in tropical waves may require resolutions of a few hundred metres in more realistic limited area models of large regions in the tropics.

In future, parameterisations may be introduced for processes that have not previously been considered to need any scheme. For example, most models exhibit regions in the upper troposphere in the tropics where the absolute vorticity is of opposite sign to the Coriolis parameter. As expected from theoretical considerations, a linear version of the ECMWF model suggests that these regions are unstable ("inertial" instability). However any such perturbations have very limited growth in the full non-linear model and are presumably removed by other processes in it. It is not clear whether the net result of this removal is as observed in

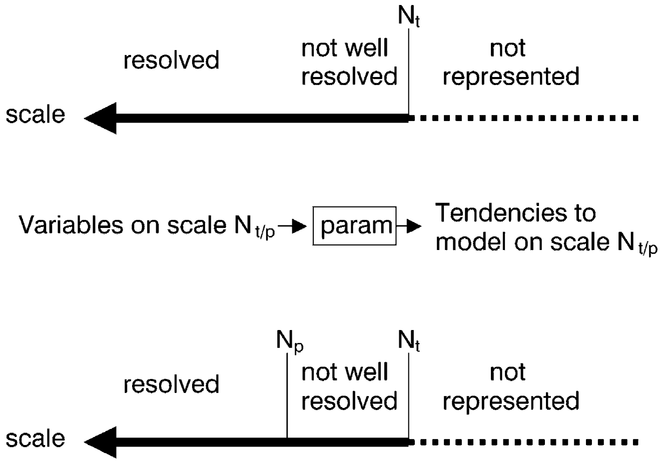

Fig. 6.4 Frameworks for parameterisation, as discussed in the text. The upper panel shows the current method and the lower panel a method that may be preferable. N_t is the truncation scale of the model and is the scale at which parameterisation is currently performed. N_p the scale at which the model fields are "believable", and it is arguable that this is the scale at which parameterisations should take their input fields and give their tendencies

the real atmosphere. When the latter is understood better it may be decided that an explicit parameterisation is needed in the model.

It is possible that the whole approach to parameterisation should be rethought. In current schemes, following Lander and Hoskins (1997), the framework for parameterisation can be described as in the top part of Fig. 6.4. The truncation wavenumber (N_t) or scale divides the range of scales into those not represented and those represented. However the scales slightly larger than N_t (arguably up to about four times this scale) are poorly resolved. Information is taken from the model at the grid-scale, where it is not to be trusted. It is fed into a parameterisation scheme and the resultant tendencies are fed back onto the grid-scale at which the model is not able to treat them correctly. This approach was necessary when the range of scales included in a model was not very large. However, given that this is no longer the case, the alternative scheme shown in the lower part of Fig. 6.4 may be preferable. Here information given to the parameterisation is taken from the model at the scale that is believed (N_p) and tendencies are fed back on the same scale, a scale at which the model is able to handle them properly. There are questions about which parameterisations should be handled in this manner and what to do about singularities such as coastlines. It is probable that parameterisations that are essentially dissipative should be applied on the scale N_t, but that those that can be viewed as forcing should be applied on the larger scale N_p. This approach is now being tested.

Until recently, parameterisations have been deterministic in the sense that a particular set of grid-scale values of model variables will always yield the same tendency from, for example, the convective parameterisation. However it is now increasingly recognised that it is unrealistic to think that this is the case, and representations of the random element are being included in "stochastic" parameterisations (e.g. Palmer 2001). Also, organised behaviour on sub-grid scales may feedback energy onto the retained scales: again representations of such "backscatter" are being tested (e.g. Shutts 2005). Finally there is also now research on representing convection in particular by embedding a specialised prognostic model that interfaces with the full model near the grid-scale. One version of this ("super-parameterisation", see Grabowski 2001) is a cloud resolving model simplified for computational reasons to use only one horizontal dimension. Another is to have a model in which convective clouds are represented by so-called cellular automata (Wolfram 1994) that obey simple rules which determine their growth, interaction and decay.

6.2 Ocean Modelling

Ocean models are a critical component of seasonal forecast systems. As outlined in Chapters 3 and 4, there are a few essential roles for the ocean in the climate system, and it is important that numerical models for the ocean capture these features and represent the physics included in them.

There are two opposing goals of modelling – to elucidate the essential dynamics or to simulate and forecast. Often, elucidation means removing processes believed to be non-essential, paring away process after process until only the simplest possible model remains which is able to capture the phenomenon. Each process successfully removed can then be identified as not essential to the physics, and the "most elegant" model can be constructed. Once the simplest model can be built, second- and third-order approximations can be constructed to make more and more faithful renditions of nature.

This strategy of isolating the essential physics, then incrementally adding complexity can lead to relatively successful prediction systems, as exemplified by the Zebiak and Cane (1987) coupled model for El Niño. They derive their success from two sources: first, they contain most of the relevant physics, second, they avoid systematic drifts and biases which can quickly corrupt more complex models. (Every time complexity is added to a model, additional unknown parameters are added to the system as well as additional ways to introduce substantial biases.)

We have outlined a view of the processes in the ocean as one-dimensional mixed layer physics, ventilated thermocline mechanics, shallow tropical cells, surfacing of the thermocline and details of the interaction of the thermocline with the sea surface.

6.2.1 Models for El Niño

Over time a hierarchy of models has been used to simulate El Niño. They fall into the following categories:

6.2.1.1 Single Layer Reduced Gravity Models

These models capture the equatorial wave dynamics of the upper ocean. They represent the ocean as a relatively shallow layer of warm water sitting atop an infinitely deep, dense abyss. The displacement of the interface between the warm and cold water introduces horizontal pressure gradients. Combining the effect of these pressure forces with the Coriolis force and surface stress yields the momentum balance describing the horizontal flow. This flow is then used in the mass equation to compute the changes in interface displacement.

$$\frac{\partial \mathbf{v}}{\partial t} + f \cdot \mathbf{k} \times \mathbf{v} = g'\nabla h + \frac{\boldsymbol{\tau}}{H}$$

$$\frac{\partial h}{\partial t} + H\nabla \cdot \mathbf{v} = 0$$

where $g' = g{\cdot}\Delta\rho/\rho_o$, and $\Delta\rho$ is the density difference between the surface layer and ρ_o is the abyssal density. H is the mean layer thickness.

By choosing the mean thickness of the layer and the density difference between the surface and the abyss, the model can be made to have a wave propagation speed ($\sqrt{g'H}$) consistent with the various baroclinic modes of a continuously stratified ocean – usually the gravest mode is chosen. Such models provide a useful solution to simple wave propagation questions in the presence of complex boundaries, including islands, or in the presence of complex wind forcing. Examples of such models include the studies of Busalacchi and O'Brien (1981) and Cane and Patton (1984).

The most obvious limitation of the single layer reduced gravity model is that it fails to simulate the SST. A parameterisation of the SST is often made by relating the anomalous layer thickness with an anomalous subsurface temperature. A thin surface layer is presumed to indicate that there is less warm water near the surface and the abyssal water is brought up closer to the surface. Imagining that some form of mixing acts to modify the temperature, the shallow thermocline is then mapped to a cooler surface temperature. An alternative derivation can be made in a statistical sense, relating the observed record of thermocline displacement to temperature anomalies. This latter method can be used to make arguments that localize the SST response, because in some regions the change in thermocline is not as strongly related to the change in SST as in others (see Fig. 3.4).

6.2.1.2 Two Layer Reduced Gravity Models

To overcome the objection that SST is not predicted in a single layer model, a number of models were built which included two reduced-gravity layers, where the top layer had a representation of a shallow surface mixed layer, and the second layer described the remainder of the water above the thermocline. The deeper ocean was still represented as an infinitely deep abyss.

These models differ largely in their treatment of the second layer temperature and how that temperature interacts with the mixed layer during upwelling. The Schopf and Cane (1983) model carried an explicit equation for both the temperature of the layer and its vertical gradient. The surface layer also carries a full thermodynamic equation for the SST, including non-linear advection, surface heating, diffusion, and vertical advection across the base of the mixed layer. The model in Zebiak and Cane (1987) treats the second layer density as fixed, and then parameterises the temperature profile based on the thickness of the layer or position of the thermocline. This model has served as the ocean component of one of the first successful ENSO prediction systems, which combined this ocean with a simplified non-linear atmosphere model. Its success supports the notion that theories based on ocean wave dynamics (the delayed oscillator and recharge paradigms) represent the dynamics of ENSO, while by no means proving the point, since it also failed to predict major ENSO events such as that in 1997/98.

6.2.1.3 Multi-layer Reduced Gravity Models

While the two layer models add an equation for the surface temperature, they depend heavily on the parameterisation of the sub-surface thermal structure, and the representation of its dynamics through a single vertical mode. Models with several layers in the vertical were then developed as a natural extension to these systems. Such models can represent the ventilated thermocline very well (see Chapter 4). They share the reduced gravity approach to the theory, allow for several outcroppings of isopycnal surfaces, and are fully non-linear. When coded with care, they exhibit very good conservation of potential vorticity – a requirement for the ventilated thermocline – and can be run in nearly global mode for thousands of years, reaching an equilibrium state very quickly.

These models connect the mid-latitude thermocline with the equatorial zone, and therefore provide a means for completing the shallow tropical cells, the equatorial Kelvin and Rossby waves, and the mixed layer dynamics. A model of this complexity is in use for routine seasonal prediction at the NASA Goddard Modeling and Assimilation Office (Schopf and Loughe 1995). A variant on this approach uses sigma-coordinates to treat the thermocline water, rather than isopycnals (Gent and Cane 1989), and has also been used for many coupled ocean-atmosphere studies.

6.2.1.4 Ocean General Circulation Models

Ocean general circulation models (OGCMs) have become the model of choice for coupled climate prediction systems. The reduced gravity models take advantage of the very small changes found in the background state of the deep ocean over the seasonal to interannual timescales. The simplest single layer models have no problem maintaining the proper wave speeds, since they are built in to the equations as external parameters, not predicted by the model. The two layer reduced gravity models share this constraint. The multi-layer reduced gravity models are the first in our hierarchy which have to maintain their own climatology. If their only drawback were the inability to represent the slow thermohaline circulation, it would likely have little impact on the simulation of seasonal to interannual climate. But one serious deficiency affects reduced gravity models: their inability to represent the effects of topography, in particular the effects of the shallow passages and straits through the Indonesian archipelago.

An ocean circulation model does not require or impose the "background conditions", and instead proceeds to solve the primitive equations of motion and thermodynamics in a direct fashion. They must create the background state, the abyssal water, thermocline and near surface mixed layers. Since it is hoped that they mimic the dynamics of the real ocean, some features, such as the very slow evolution of the deep water, should be reflected in their solutions. It is common practice to start these models from some climatology, and so it is hoped that the deep ocean, which would take millennia to equilibrate, can be very nearly in equilibrium at the start of an integration. If the ventilated thermocline theory is meaningful, it implies that the pressure gradients in the deep ocean are very small, and this slow evolution of conditions in the deep is consistent with our assumption. The ventilated thermocline dynamics must be reflected in solutions to the primitive equations, and an accurate integration of these equations should give similar behaviour to the multi-layer reduced gravity model, and hopefully, to the real ocean.

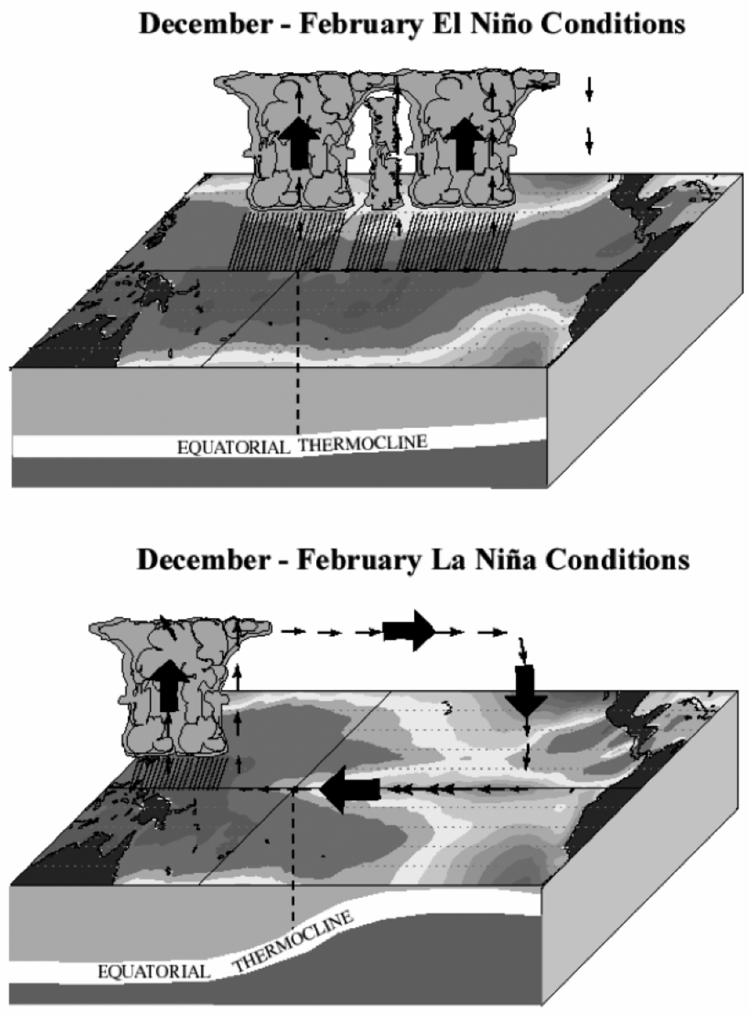

Fig. 6.5 Schematic of El Niño and La Niña

But the challenge to the OGCMs is great. Figure 6.5 shows a typical cartoon of El Niño vs. La Niña: the thermocline changes dramatically, theconvection shifts over large regions, and the ocean is depicted as a single layer. Figure 6.6 shows observed sections of temperature along the equator (Johnson et al. 2002). The top two panels show the climatological tempera tures for 6 months around the annual cycle. The bottom two panels show the conditions during a canonical El Niño and La Niña. What should be apparent is that the El Niño is not a large shift of the thermal structure of the ocean, but rather one that is about the size of the annual cycle itself. It is also apparent that an approximation of a two-layered system seems plausible when looking at the equator. If one looks at meridional sections across the equator, however, the complexity of the thermocline becomes more apparent, becoming more diffuse toward the west and more diffuse at higher latitudes. Figure 6.7 shows these sections from Johnson et al. (2002).

While the two layer models exploit the sharp thermocline, the OGCM must produce it. Small errors in the numerics or physics can easily lead to substantial changes in the simulated mean state. Inspection of Figs. 6.6 and 6.7 reveals that a 10 m vertical displacement of an isotherm could lead to a 1°C or 2°C temperature anomaly. Experience with several ocean models used in El Niño prediction reveals that these models tend to share a number of problems, not the least of which is an overly diffuse thermocline.

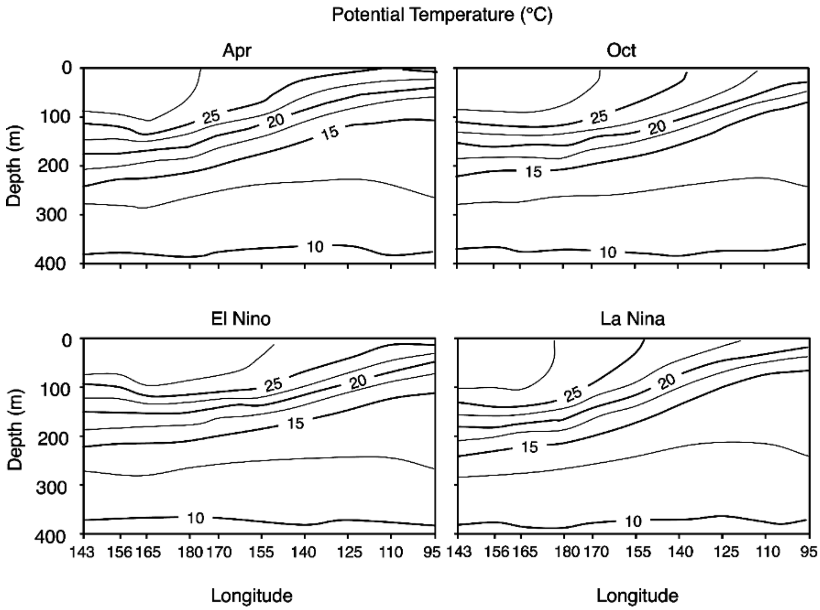

Fig. 6.6 Observed temperature profiles along the equator for April and October (top panels) and for El Niño and La Niña (bottom) (after Johnson et al. 2002)

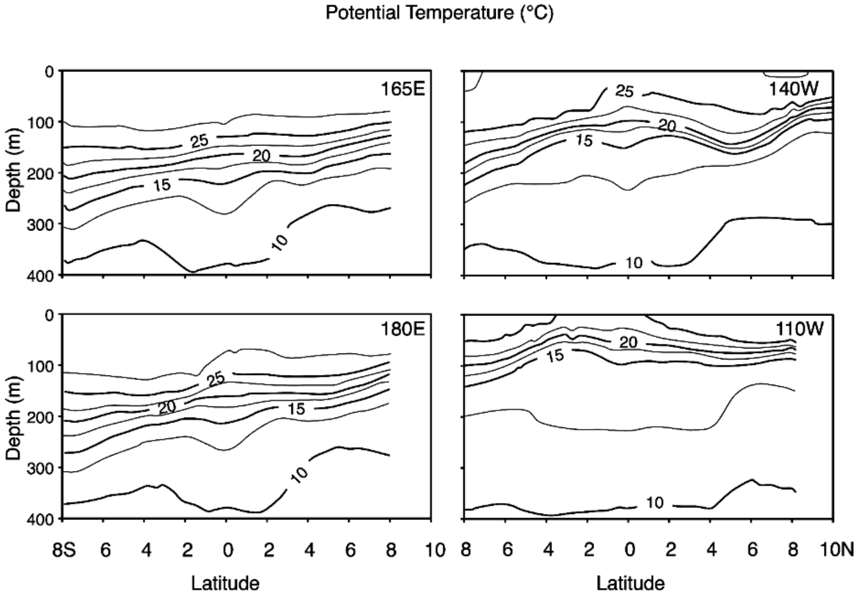

Fig. 6.7 Sections of mean potential temperature across the Pacific at longitudes of 165°E, 180°E, 140°W and 110°W (after Johnson et al. 2002)

6.2.2 Challenges and Improvements to Ocean Models

There are two essential components of modern ocean models, usually referred to as "dynamics" and "physics". Dynamics integrates the equations of motion on scales down to the resolution of the computational mesh, while physics includes the parameterisations of effects of finer scale motions, which usually take the form of "mixing". While research continues on improving the hydrodynamic codes, it is clear that modern ocean models suffer more from errors in the treatment of small-scale motions than from errors in the simulation of well-resolved hydrodynamic processes.

Mixing occurs in several regions in the ocean, and can be very vigorous and important in establishing the character of the flow. In other regions, we believe that mixing is very weak, especially mixing across surfaces of constant potential density (isopycnals). This type of mixing is known as diapycnal mixing. When diapycnal mixing occurs, there is a significant increase in the potential energy of the water column. There must be an equivalent source of turbulent energy in order to accomplish this mixing. Mixing of tracers along isopycnal surfaces does not involve a change in potential energy, and such mixing can happen much more effectively than diapycnal mixing. The rate of isopycnal mixing appears to be about 10^7 or 10^8 times greater than diapycnal. Although diapycnal mixing is far,

far smaller than that along such surfaces, diapycnal mixing is essential for the effective transport of heat in the climate system. Its accurate representation in climate system models is therefore important, if for no other reason than obtaining the proper distribution of heat. On the other hand, the ventilated thermocline theory depends on the diapycnal mixing being low. The multi-layer reduced gravity models discussed above are able to run with essentially no diapycnal mixing below the surface turbulent layer. Many other consequences arise, however, if we consider the transport of salt and other significant materials.

The first numerical models of the ocean circulation were constructed in the relatively straightforward geometrically based coordinate systems on which the hydrodynamic codes could be efficiently and easily represented. Numerical schemes for the equations of motion are either highly dissipative in their nature, or need additional damping of the finest scale motions. For many years, models were so coarse in resolution and needed so much "numerical glue" that the parameterisation of diapycnal and other mixing processes had no need to be physically based. While the numerical schemes still need suppression of the finest scale motions through mixing, modern modelling practice has chosen to add the damping in a way which aligns the effects along isopycnal surfaces (either by rotating the mixing tensor or by using isopycnal coordinate systems). With the combined advent of higher resolution and an appreciation for improved numerical techniques, we have now arrived at a point where diapycnal mixing can be set by physically based parameterisations.

Although mixing along isopycnal surfaces is far larger than diapycnal, the problem of adequately representing such mixing is far from solved. The highly energetic eddies of the ocean circulation cause intense motions on scales that are still not well resolved except by the most ambitious computing efforts. For the seasonal prediction problem, there is advantage that the size of these eddies increases near the equator, and it is now fairly common to resolve them within 10° of the equator. But returning to the premise that the equatorial thermocline depends on the shallow tropical cells, the effect of unresolved eddies has been shown to be very important in the subduction process which creates the source waters of these cells, as well as causing a mixing of potential vorticity in the cells themselves. For the class of models used for climate simulations and forecasts, it is likely that it will be many years before eddy resolving ocean models are used in these integrations. Further, the class of eddy-resolving models that can be foreseen in the next 20 years will still need sub-grid scale parameterisations of still smaller motions in order to close the problem. This leaves the problem of parameterisation of eddy fluxes as another key issue to improving coupled model simulations of climate.

For the seasonal to interannual prediction problem, the most important mixing processes are those of:

- Surface boundary layer processes
- Shear driven upper ocean mixing

- Internal wave breaking
- Interaction of eddies with mixed layers
- Mixing within the thermocline at lateral boundaries
- Double diffusion and salt fingering

The first three directly affect the simulation of the tropical and equatorial profiles of temperature, salinity and currents. The last three influence the establishment of the shallow tropical cells, and therefore have an indirect influence on the equatorial thermocline. The importance of these latter three arises if we are attempting to simulate the coupled climate in a drift-free model. Modern data assimilation methods can probably obviate the need for exact solution of the subtropical thermocline, but these effects have been shown to have a direct bearing on models' abilities to simulate El Niño. Meehl et al. (2001) showed that significant changes in El Niño amplitude in coupled models seem to be related to the background diffusivity used in the ocean models (see Fig. 6.8). In numerical experiments with a simplified coupled model, much of this sensitivity comes from the diffusive effects outside the equatorial belt, further emphasizing the importance of simulating the STCs in establishing the equatorial thermocline.

In equatorial regions the thermocline is often relatively strong and shallow, which allows cold water to be maintained near the surface, where it can be made available for cooling the surface. Variations in the strength of such cooling are primary factors in the El Niño cycle of equatorial SST. Since the diapycnal mixing of heat in the thermocline continually works to destroy this thermocline it is important to understand the processes involved, and to represent them accurately in ocean models. One such process is thought to be mixing due to breaking internal

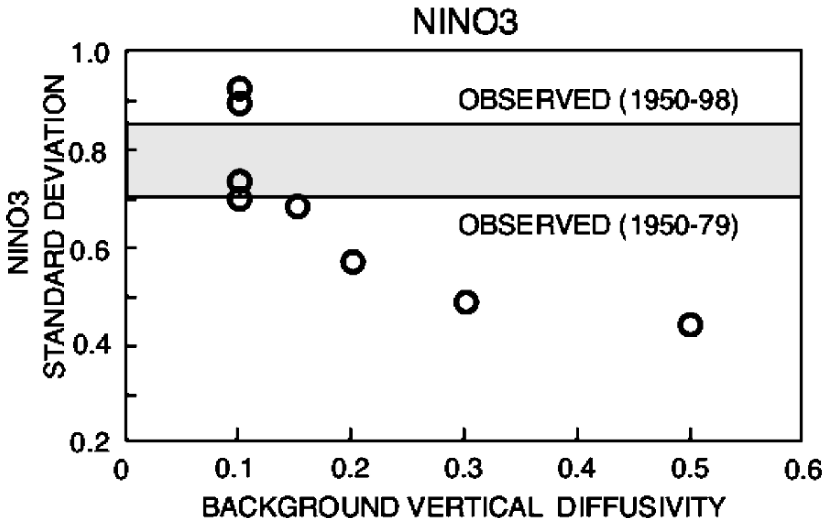

Fig. 6.8 Effect of background vertical diffusivity on simulated El Niño amplitude from a variety of coupled models at varying resolutions (after Meehl et al. 2001)

waves. It is commonly represented in ocean models as Fickian diffusion with a constant coefficient (1–2 $cm^2\ s^{-1}$) given by extratropical tracer release experiments (Ledwell et al. 1993). Despite the lack of observational support, models also need the corresponding viscosity and typically use a constant Prandtl number of about ten. However, there would be significant consequences should these values prove to depend on the Coriolis parameter and hence for there to be smaller diffusivity at the equator.

Another mixing process derives its energy from the shear between the surface and the equatorial under current (EUC). Microstructure measurements have provided

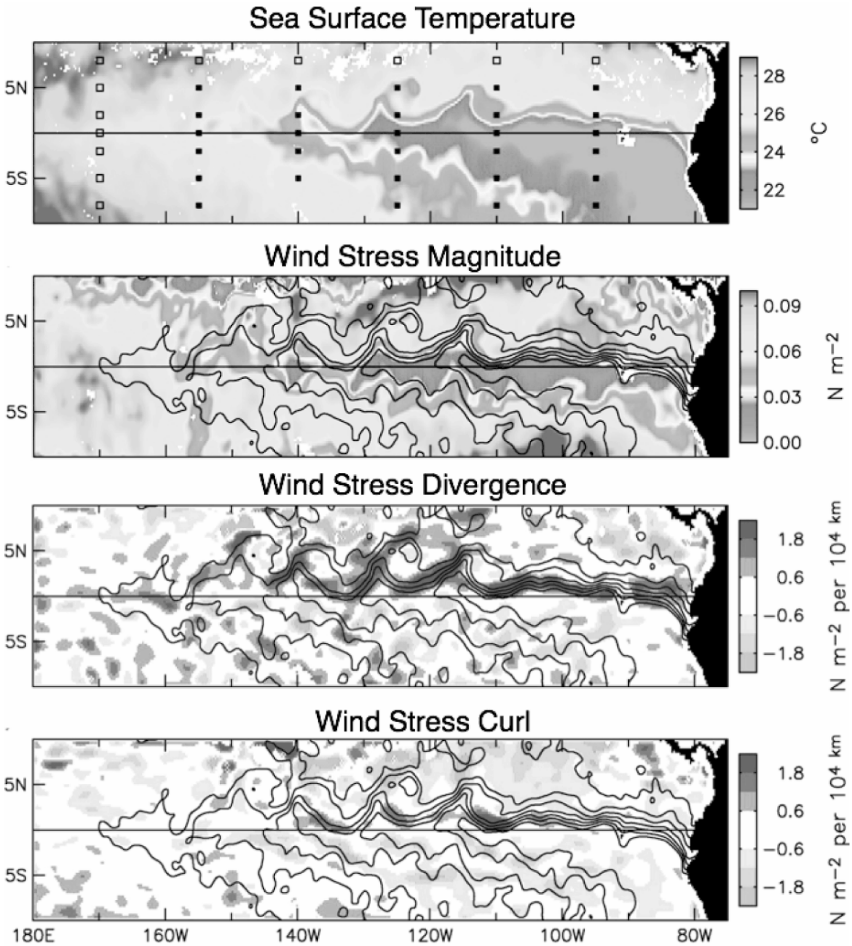

Fig. 6.9 September 1999 Sea Surface temperature (SST), wind stress magnitude, wind stress divergence and wind stress curl observed from satellites (after Chelton et al. 2001). The contours in the lower 3 panels are the SST from the top panel

many estimates of its diffusivity, which has been formulated as a function of local Richardson number (Peters et al. 1988). Usually ocean model implementations follow Pacanowski and Philander (1981). The sensitivity of the equatorial structure to these parameterisations was demonstrated in Yu and Schopf (1997).

A third process, which may be a combination of the first two, is known as deep-cycle turbulence. It is observed on the equator in the central Pacific, where the mean shear of the EUC and a daily cycle of daytime solar heating and night time surface convection combine to produce strong turbulence extending tens of meters into the stratified water below the surface layer.

Above, we asserted that "physics" and unresolved motions give larger problems in ocean modelling than the solution of the hydrodynamic equations, but even the solution of the hydrodynamics becomes problematical when the features are barely resolved by the computational mesh. For the problem of El Niño, the most important issue here are the tropical instability waves – a series of ocean eddies on the scale of 700 km that appear seasonally just north of the equator in the eastern Pacific, and across the Atlantic. The challenge is to accurately model the fine scale structure of the ocean eddy field and its relation to the surface fluxes – particularly the surface stress. Chelton et al. (2001) demonstrated the strong relationship which appears to exist between the SST and the surface stress, even downto the scale of the tropical instability waves. Figure 6.9 shows the SST and surface stress properties for a week at the start of September 1999. The strong relationship between the SST and the stress shows the influence of coupled ocean-atmosphere effects down to a very small scale. Of particular importance are the divergence and curl of the wind stress, which show up as very clearly linked to the SST. These scales can now be represented in ocean circulation models, but are only beginning to be resolved in the most ambitious atmospheric GCMs.

6.2.2.1 Summary

Ocean general circulation models are the basis for most coupled forecast systems currently in use at national forecast centres. They have the strong advantage of being able to represent almost all the important physics for El Niño and La Niña. Today, when coupled to atmospheric models, they produce climate simulations with significant biases that lead to rather rapid degradation of the forecasts (see Figs. 5.9 and 6.16 later in this chapter). Data assimilation and forecast assimilation techniques can be used to improve the prediction and to correct for the systematic biases that arise in the model, but it is clear that the models themselves can be substantially improved by the inclusion of better treatments of the small scale physics that are so important in controlling the overall, long term simulation. The interests of the ocean modelling community for seasonal to interannual prediction are not so very far removed from those studying the role of the ocean in climate change.

General circulation models are designed to *include* dynamics and physics. The early two-layer models succeeded in large part by excluding physical processes. By including physics, the models enable better representation of nature, but by including physics badly, the models can rapidly go astray. It is clear that ocean models will go to ever increasing resolution; whether they can include more and more accurate representation of the physics remains to be seen.

6.3 Coupled Modelling of the Atmosphere-Ocean System

6.3.1 Teleconnections

As the equatorial Pacific is engaged in a fancy dance of waves, the rest of the world can hardly miss it. Just as a couple of top dancers quickly draw the attention of the whole ballroom, so the dance of the Pacific is felt around the world. The large SST anomalies in the equatorial Pacific activate large areas of deep convection, enormous amounts of heat are released into the atmosphere and anomalous circulations quickly set in. The anomalous heating displaces the normal distribution of east-west (Walker-like) circulation cells in the global equatorial zone. The normal distribution of vertical velocity is modified and subsidence appears in unusual and distant places, carrying drought, or exceptional rains where instead upward motion is moved. All along the equator, in Brasil, East Africa, the Indian subcontinent, the consequences of the Pacific dance are sorely felt. These *teleconnections* are a dramatic consequence of the coupled mechanisms in the equatorial region; they transfer the impact of ocean-atmosphere processes to distant regions and, exploiting the slow timescales of the ocean, can influence the atmosphere for a long time.

Teleconnections are not limited to the equatorial area. The equatorial atomsphere is very sensitive to anomalous heating that quickly generates high level vorticity areas that become the sources for the generation of atmospheric planetary Rossby waves, propagating into the mid-latitudes in both hemispheres. The signal of the Pacific is now carried away tens of thousands of kilometres, stretching like beads of a necklace across the Pacific to North America, and sometime to Europe. Similar chains of anomalies extend into the Southern Hemisphere, affecting South America, Australia and South Africa.

Figure 6.10 shows some of the climate impacts of the teleconnections stemming from El Niño. One can see regions as far away as North America and southern Africa which are affected by El Niño. These connections are statistically derived: i.e. they often occur when there is an El Niño, but there is no requirement that any or all of these patterns will occur in any given El Niño.

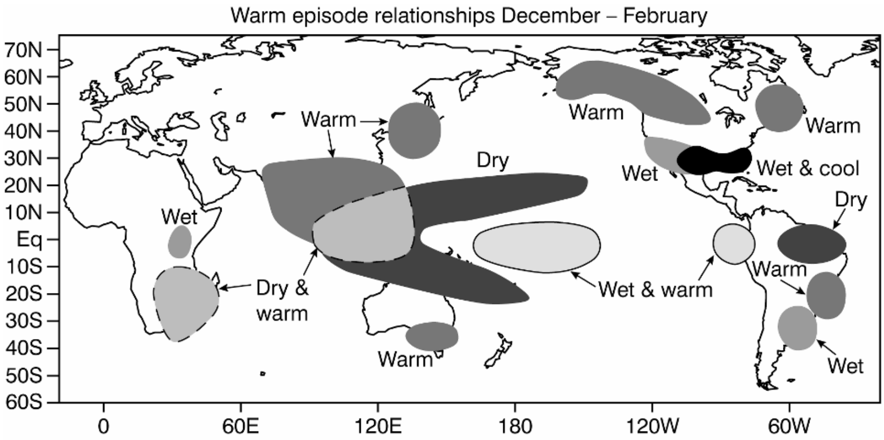

Fig. 6.10 Plot of the frequently observed climate anomalies in temperature and precipitation associated with El Niño. This plot is for Dec–Feb, often the peak phase of El Niño. Other seasons will have other climate anomalies (teleconnections). A given El Niño will not necessarily show all of these climate anomalies

6.3.2 Developments in Coupled Modelling

The progress of science has been gradually formalized into a rather well-accepted pattern. Experiments provide results that require explanations, theorists try to come up with a theoretical framework that might explain existing experimental results and possibly make some predictions that can be tested by further experiments. Eventually a *crucial experiment*, i.e. an experimental set under controlled conditions, will be designed, allowing the selection between competing theories and providing for the time being, the best explanation.

This process is clearly not possible in the case of the dynamics of the ocean and atmosphere. The climate system can be observed and measured, processes can be identified, diagnosed and budgets of conserved quantities can be calculated, but we cannot make "experiments" in the sense used in many other fields of science. We cannot change the system artificially to verify ideas and theories, monitor in detail the evolution of processes, under conditions different from what we see on our Earth. Unfortunately, this interplay between theory and experiments is the main driving force behind the development of science. The scientific consideration of climate has been therefore seriously jeopardized by the difficulties encountered trying to apply the general paradigm.

The transformation of climate science into a quantitative science has been made possible by the development of numerical models. The equations that regulate the evolution of the atmosphere and oceans were already known at the end of the

19th century, but their mathematical complexities placed them beyond the mathematical-solving capabilities of the time. The spectacular advances in computational techniques and computational capabilities in the late 20th century have allowed the realization of numerical models of the climate system. Models have become very advanced constructions, from the early pioneering work (Manabe and Bryan 1969; Manabe et al. 1975, 1979; Bryan et al. 1975; Schlesinger 1979; Washington et al. 1980). They are now capable of simulating the mechanisms and the evolution of the climate system, even if we are far from a totally satisfactory simulation of mean climate and its variability. The climate system is composed of many subsystems, but the main elements remain the atmosphere, the oceans and to some extent the sea ice. We have seen in the previous pages how the atmospheric and oceanic modules are realized and we will describe how they can interact with each other. The interactive model is simply called a coupled model.

The basic scheme for a coupled model is described in Fig. 6.11. The scheme describes how as the atmospheric model and the ocean model evolve they can exchange the information that is necessary for the interaction.

The ocean model requires from the atmospheric component several physical fields necessary to complete the momentum and energy balance at the ocean surface. The momentum input is provided by the wind stress fields, whereas the energy budget is obtained by providing the short wave radiation flux (solar), the net long wave radiation flux and the net sensible heat flux. The mass flux is obtained by the net freshwater flux, precipitation minus evaporation. Up to the late 20th century, many ocean models used in coupled modelling used the rigid lid approximation which implies that the net freshwater flux does not change the total

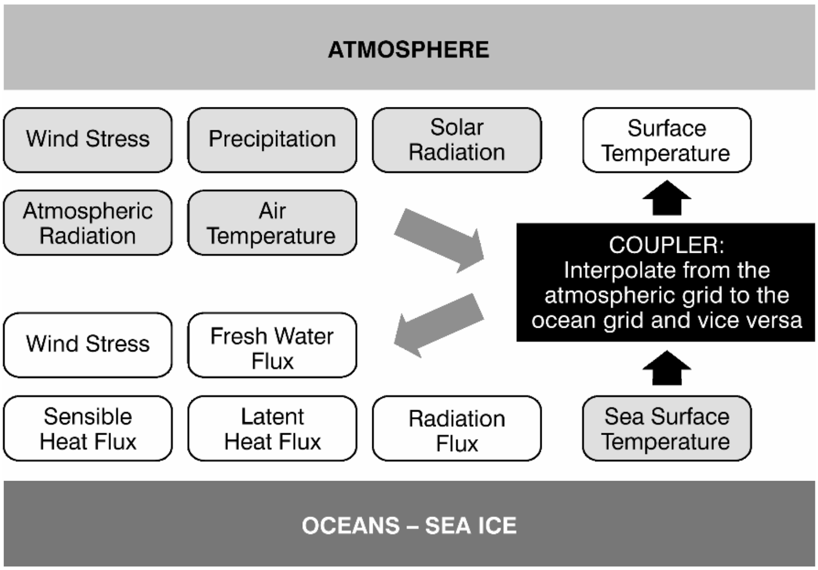

Fig. 6.11 A schematic representation of an atmosphere-ocean coupled model

mass, but rather modifies the salt concentration in the salinity equation. More modern models relax this constraint and allow the free surface to vary (the so-called free surface). The net effect is to allow surface gravity waves, which, however, impose constraints on the time step.[7]

The impact of the ocean on the atmosphere is exerted mainly through the distribution of SST. The SST distribution is a major forcing factor especially in the equatorial area by its effect in regulating convection. Deep convection in the tropics can start easily over warm ocean waters. As a rule of thumb the value of 28°C is often mentioned as an empirical rule. The tropical convective towers release massive amount of latent heat in the condensation/precipitation processes and they are therefore the main energy source for the deep atmosphere.

We must introduce a third component to manage the delicate dance between the atmosphere and the ocean, the coupler. The coupler is the software component that keeps track of the time, directing the right traffic between the models and making sure that the right data is available at the right time. The role of the coupler, like a conductor in an orchestra, is to harmonize all the communications, manage the software messaging and interpolate the data from the grid in the ocean model to the atmospheric model. Some implementation of coupled models can expand the role of the coupler by including in the coupler the calculation of the fluxes themselves, but in general they include only the interpolation. Several couplers have been developed in the past few years. OASIS, the coupler developed at CERFACS (Valcke et al. 2004) has become popular in Europe; in the USA several couplers are being consolidated under the Earth System Modeling Framework.[8]

The basic physics of the coupling may look deceptively simple. If the models exchange fields at the proper time the right interactions are properly represented, with a beneficial result on the accuracy of the simulation. However, the devil is in the details. The fields to be exchanged are rarely on the same numerical grid. The most commonly used models for the atmosphere use spectral techniques, based on spherical harmonics, that require a latitude-longitude grid that is regular in the longitudinal direction, but uses special latitudes based on the zeros of the Legendre polynomials. Such a grid is known as a “Gaussian grid”. Modern ocean models use orthogonal curvilinear grids which locate the poles in unusual locations. Regular grids have been discontinued in the past because they generate the *pole problem*,

---

[7] The reason why the rigid lid approximation was introduced is that it filtered out external gravity waves, but had the undesirable effect of altering the propagation of long external Rossby waves. Without the rigid lid approximation the time step has to be very short (order of 10 minutes) or some other approximations have to be made. In both cases, a substantial computational effort is needed to solve for the external mode which has only a minor effect on climate. Very few ocean models now use the rigid lid approximation. Preference is given to dealing with the barotropic mode separately and in many cases stepping it forward with short time steps of a few minutes. Killworth et al. 1991.

[8] Collins et al. (2006) and Delworth et al. (2006). See also: http://www.esmf.ucar.edu

namely the fact that the grid points get closer and closer in the polar regions because of the convergence of the meridians.

The ever decreasing size of the spatial cells at the poles decreases the numerical stability of the calculation, by increasing the chance that a basic numerical stability threshold (the Courant-Levy-Friedrich or CFL limit) is crossed. Atmospheric models had to resort to complex schemes of filtering at the poles to control the problem, before the introduction of spectral models that finessed the entire issue, since spectral models by construction do not suffer from this ailment. Because of continental boundaries, global spectral models are not viable for the ocean; the pole problem for the ocean can be treated by using almost-global models, i.e. cutting the domain short of the extreme North polar areas. Limiting the domain to non-polar regions, however, is not a satisfactory solution. It is not realistic and it prevents the possibility of simulating the evolution of the Arctic Ocean that has important climate effects, especially on the North Atlantic. A better solution has been to apply a conformal transformation to the regular latitude-longitude grid of the ocean. Conformal transformations rotate and stretch the grid, moving the poles to other positions. In a general case moving around the North pole on the sphere will not be very helpful: it will merely be shifted to another position. In the case of the ocean we can place the new poles over land, effectively eliminating them from the numerical solution of the ocean.

Conformal grids are a very elegant solution to the pole problem, but there are no free lunches. The backdrop is that the grids become highly distorted on the sphere, assuming strange orientations and there is almost no chance that the ocean grid boxes will fit the atmospheric grid boxes of the Gaussian grid. The interpolation problem between the atmosphere and ocean grids becomes a delicate affair. A badly designed interpolation can introduce systematic errors in the energy budget at the surface, resulting in fictitious sources or sinks of heat that prevent the achievement of a closed energy budget. No interpolation is perfect, of course, but a serious effort must be made to keep the imbalances at a minimum. The developers of OASIS, for instance, have introduced an energy-conserving interpolation that successfully minimizes the errors.

The definition of the complete coupled problem requires also that we determine how often the coupling is realized, i.e. how often the atmosphere, the ocean and the other components that we may include have to exchange their data. In reality, the interaction is continuous: the ocean is affecting the atmosphere via the temperature at the same instant in which the wind is affecting the ocean. In the discontinuous world of the models the atmosphere and the ocean are marching at finite steps. The fast atmospheric processes require short time steps, whereas the relatively slow ocean evolution can be described with longer time steps. One may have for instance a time step of 2 hours for the ocean and a time step of 30 minutes for the atmosphere. In principle, the best strategy would be to couple, i.e. exchange data, every time step, but it would be a waste to use the minimum denominator and evolve the ocean with the fast atmospheric time step. Rather, the longer time step is usually chosen as the coupling time step. In some cases, the coupling has been performed only every 24 hours using average fields, but this is becoming less common as it does not allow an accurate resolution of the coupling on the time-scale of the daily cycle.

As can be seen in Fig. 6.12, the time march of the coupled model is rather regular and the coupling is occurring at frequent intervals, even to the point of being able to resolve the daily cycle. This kind of strategy for coupling is known as synchronous coupling, since there is essentially no time shift or lag between the fields that are exchanged. It is obviously computationally demanding since it requires the calculation of two models at the same time, discarding the opportunity offered by the different timescales for the ocean and the atmosphere.

It is possible to use a different strategy, particularly in the case of very long simulations, that is also somewhat computationally cheaper. The price to be paid in this case is that it is going to be less accurate. In this approach the coupling is not realized at all times, but only in selected periods. The large scale thermal structure of the ocean requires a long time to be generated from atmospheric forc ing. If in the particular application the details of the coupling are not needed, but there is only interest in forcing the ocean with a statistically realistic atmosphere, then the approach has some merit. In practice, the atmosphere and the ocean are allowed to interact for a limited time, then the ocean model is evolved persisting the atmospheric condition until the next coupling period and so on. This approach

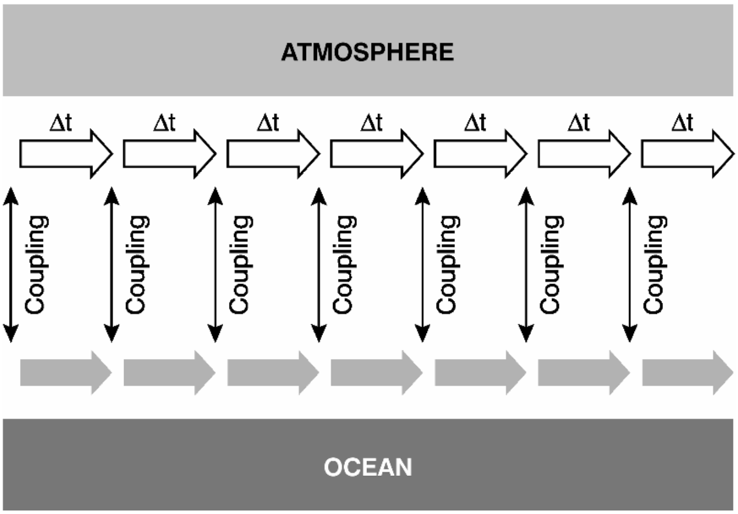

Fig. 6.12 A possible strategy for the simulation of a coupled model: synchronous coupling. In this case the models are executed at the same time and the data are exchanged in principle every time step. In practice, data are exchanged at some congruent number of time steps depending on the size of the time step in the atmosphere and the ocean. Typically the atmosphere has a timestep considerably shorter than that of the ocean. So in practice, the coupling might be done every ocean timestep, but not every atmospheric timestep (for a better representation, the ocean timestep is only twice that of the atmosphere in this schematic)

is called asynchronous and it is sometimes used to create statistically balanced initial conditions, when hundreds of years of simulation are required to reach a statistically stable steady state. It is of limited value in the case of the forecasts at seasonal range, since in this case the details of the interactions can not be neglected.

The initialisation problem has also some peculiar issues that are unique to coupled modelling. In general, the objective of initialisation is to provide a smooth start to the simulation, trying to avoid initial shocks and long adjustment processes. It is desirable to minimize the adjustment process in short term seasonal forecasts in order to avoid the distortion of the solution in the initial phases of the forecasts. In the case of long climate coupled simulation, a poorly adjusted initial state may show up in lengthy adjustment that may take tens of years, as the model slowly searches for its own equilibrium.

Weather forecasting has developed a sophisticated set of procedures to produce an initial condition that is a faithful representation of the state of the atmosphere at a given time and provide a physically consistent description of the wind, temperature and in general all the other fields that are necessary for the evolution of the model. This procedure is known as data assimilation (see Chapter 5). It basically consists of a space time interpolation of the available data, constrained by the set of physical relations that we know must be valid in the real world. The embodiment of the known consistency relations is realized in our numerical models that represent in a way our best estimate of the laws that regulate the evolution of the atmosphere and ocean. Data assimilation methods therefore basically mix observations and model evolution, yielding our best estimate of the instantaneous state of the atmosphere or ocean. These methods are discussed in Chapter 5 and there is no need to discuss them here in detail, but it is worth while discussing the specific issues for the coupled system case.

Data assimilation methods have been extensively developed for the atmosphere and, in recent years, also for the ocean. As discussed in Chapter 3, the coupled evolution is dominated by the ocean and balancing the ocean component of the initial condition is essential for the forecast. At the present time, true coupled data assimilation systems, capable of assimilating observations both in the atmosphere and in the ocean at the same time, are still in a developing stage (see Section 5.5). The existing system tries to exploit the short memory of the atmosphere and they concentrate all effort on the balancing of the ocean side. Figure 6.13 shows two possible arrangements for the initialisation of a coupled model.

In the first approach, called *Robust Diagnostic*, the system is evolved in time as a coupled model, exchanging data usually using a synchronous approach, but in the exchange process one or more of the exchanged fields at the ocean-atmosphere interface is substituted with the real observed one(s). In principle, both SST and surface salinity can be constrained, in practice a typical choice is to use some sort of constraining of just SST, using usually strong relaxation constants. In this way it is possible to prevent the SST field deviating too strongly from reality and adversely affecting the evolution of the atmospheric model. Another possibility is to

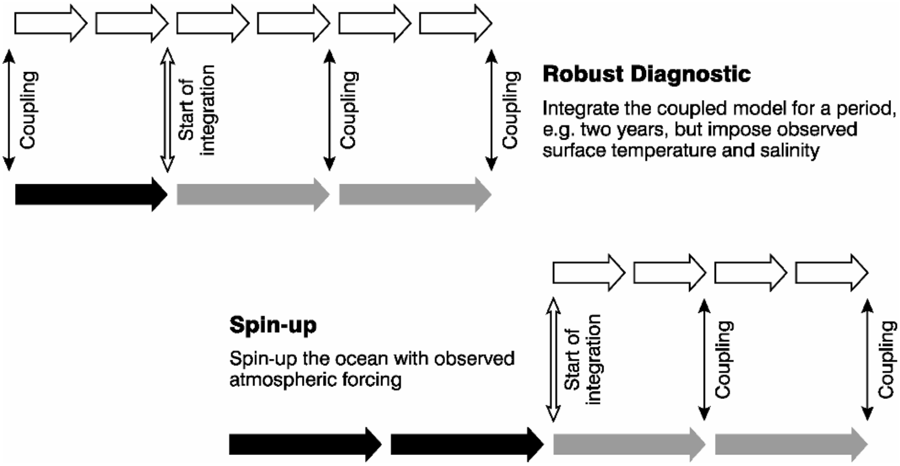

Fig. 6.13 Possible strategies for the initialisation of a coupled model: Robust Diagnostic (top) and Spin-up (bottom). The red arrows indicate the initialisation periods. The yellow arrow is the start of the free coupled integration

do the same on the surface winds, in this case it is the winds from the atmospheric models that are discarded and the real observed surface winds or stresses are provided to the ocean model.

The capability of the ocean to be strongly determined by the atmospheric forcing that is partially exploited by the *Robust Diagnostic* technique can be more strongly employed in another strategy of initialisation. In this case the ocean model is time marched alone, but forced by observed atmospheric fields. This approach, called simply *Spin-up*, has the advantage of being very simple to implement and is also relatively cheap since it involves only the integration of the ocean model (see also Sections 3.3 and 5.1.1). The disadvantage is that it makes no use of subsurface data. It must rely entirely on the ocean model parameterisation and numerics to carry the observed atmospheric signal into the ocean depth and to generate a realistic current pattern. There are no rules for the length of the period to be used for the spin-up, but values of 10–20 years are common for producing initial condition for seasonal forecasts. Spin-up for coupled simulations used in climate scenario experiments, may extend to hundreds of years due to the presence of sea-ice that introduce a timescale that is considerably longer.

The quality of the interaction between ocean and atmosphere is such that the qualitative behaviour of a coupled model may be quite different from that of its constituent components. It would be tempting to use a modelling strategy that would call for an accurate preparation, for instance, of the atmospheric model separately from the ocean model, maybe exploiting the expertise of a specialized modelling group. Model design is based on basic physical and in general, scientific principles, but there are still large numbers of parameters that are insufficiently known. The modellers have some liberty to assign these values. Physical processes are represented through parameterisations that represent in bulk, processes that are too fast or too small-scale to be explicitly calculated by the time-marching equations. Details and numerical settings in the parameterisation are designed to get the best overall result in reproducing phenomenologically the atmosphere and the oceans.

The models are trying to obtain the best global results: in practice the net effect of the overall exercise is a delicate balance of errors. Figure 6.14 shows an example of this kind; the precipitation distribution for boreal summer for the atmospheric model forced by observed, prescribed SST is shown on the left column for three different resolutions with the 'observed' precipitation shown in the bottom panel for comparison. The atmospheric model is only partially successful in reproducing the observed precipitation patterns, even when the correct, i.e. observed, SSTs are provided. For example, the monsoonal precipitation in the Bay of Bengal is largely underestimated and there is too much precipitation in the equatorial Indian Ocean. In the Philippines and East of the Philippines the precipitation is positioned too much along the lines of the parallels, rather than gently moving toward the equator. A gap appears at the equator between the precipitation in the Philippines and the South Pacific Convergence Zone (SPCZ) that is absent in the observations. Increasing the horizontal resolution from T30 (top) to T42 (middle) to T106 (bottom) has some beneficial effects. The artificial patterns visible in the low resolution simulation that modulates the precipitation disappear, the SPCZ is better defined and the precipitation over the Bay of Bengal and the Western Ghats also appears to improve, but the major deficiencies are still there. A large precipitation deficit appears in the Central Pacific, separating the SPCZ from the areas north of the equator and cutting the ITCZ north of the equator into two separate pieces.

When the same atmospheric model is coupled to a fairly advanced ocean model, without other changes to the formulation or to the optional parameters in the parameterisations we obtain the column on the right (Cherchi and Navarra 2006). One can see that coupling has not been able to eliminate some of the large inconsistencies of the left column and appears to have introduced new error areas. The coupled model has its own typical signature of error pattern, it has its own systematic error. The precipitation seems to be much more zonal than in the previous case, though some improvements seem to be confirmed in the Indian subcontinent and in the equatorial Indian Ocean. The results of the coupled model do not seem particularly better than the atmospheric only model, in some cases they are worse.

It is difficult to separate the preparation and development of a coupled model into its separate components. The developments must be done on the model that is the final target of the development itself. If the goal is a coupled model, then all the development must be done with the coupled model, possibly at the same resolution that will be used in the end for the scientific application or forecasts. In practice, there is no guarantee that, once coupled, the model will retain the nice properties identified in the original component models. This phenomenon is due to the known large sensitivity of the climate models to small perturbations. Small

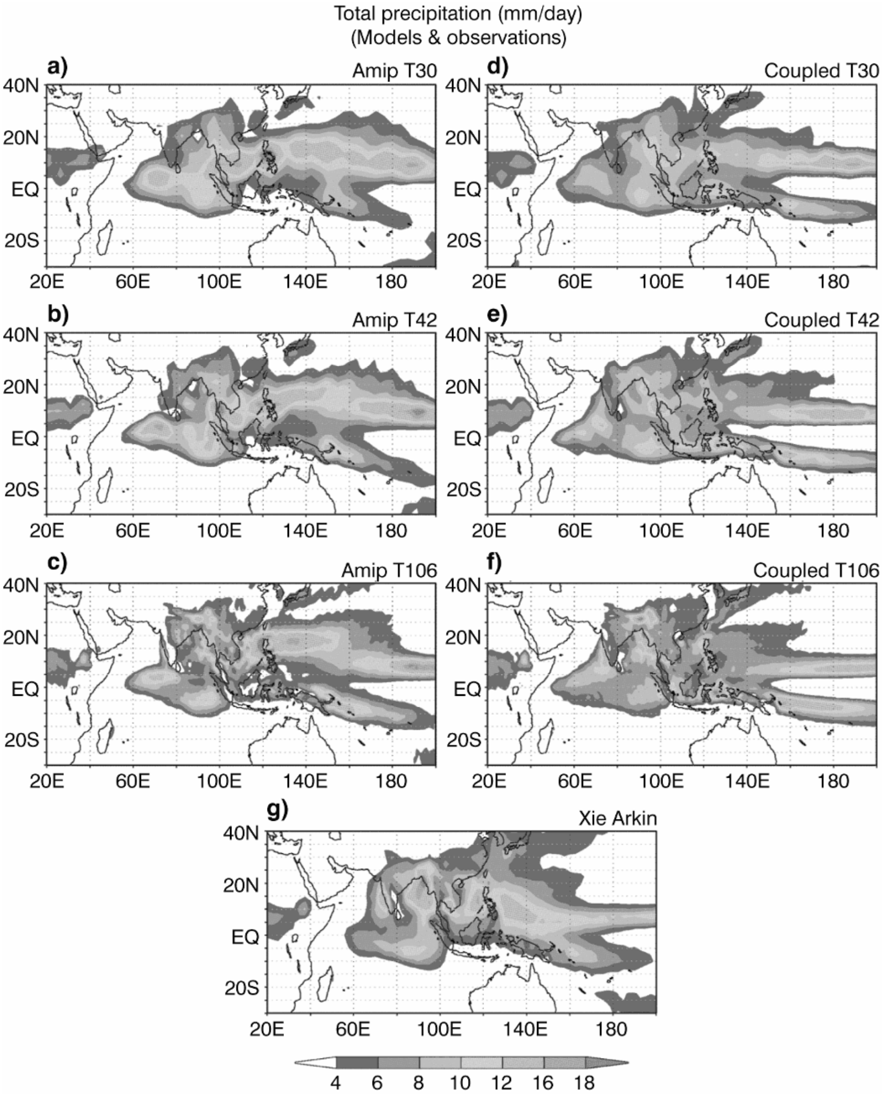

Fig. 6.14 The rain distribution for Summer (July-August-September) obtained in atmosphere only models (left column) and coupled models (right column) compared to observations (bottom panel). The experimental set-up is such that the atmospheric model used in the prescribed SST experiments (left) is the same used in the atmospheric component of the coupled model (right). Nevertheless, it is possible to see how the distribution simulated by the coupled model is qualitatively different from the atmospheric mode. These experiments show that the structure of the SST that determine the distribution of precipitation in the tropics, even with the same model. All values are in mm/day

imbalances in the model can amplify more readily in the coupled model than in atmosphere or ocean only models. The reason is that coupled models are less constrained than the individual models and so small errors can grow in more ways than before, leading to new characteristic patterns for the systematic error.

This discussion certainly applies to the tropical areas, where SST and convection are strongly linked, leading to a very close relation between SST and precipitation, but the effect of the coupling is also felt outside of the areas where coupling is strongest. The strong coupling in the equatorial Pacific modulates the position and intensity of the large convective areas in the region that are crucial for the maintenance of the extra tropical general circulation. The variability in the mid-latitudes is therefore also remotely affected by the complex atmosphere-ocean interaction that is taking place.

The coupled processes in the tropics can conceptually affect the mid-latitudes in two ways. They can directly force stationary Rossby wavetrains that show up as chains of anomalies at monthly and seasonal timescales, arching from the Pacific toward North America and occasionally deep into the North Atlantic. They can also affect the statistics of the mid-latitude internal variability, making slightly more probable some flow configurations than in normal conditions and changing the frequency and pattern of the dominant modes of variation. It is conceivable that these two processes are both active at the same time, sometimes dominating each other or contributing more equally to the variability.

It is not surprising than the modes of variability in the mid-latitudes are affected by coupling. The extra degrees of freedom offered by coupling make the simulation more challenging and difficult. The model is much more sensitive and errors have more ways to interact with others and amplify. An example is given in Fig. 6.15, showing the first EOF mode of the Boreal Winter (January-February-March – JFM) from the ERA reanalysis, and simulations. The simulations are chosen as a 40-year simulation using an atmosphere-only model forced by observed SST, also sometimes called AMIP-like integrations, and two 200-year simulations with coupled models with the same atmospheric model. The horizontal resolution of the atmospheric model used is T30.

The observations show a familiar picture. The leading mode of variability has active centres over the North Pacific, extending over North America and the Atlantic. The dynamical interpretations of the alternating anomalies can be elusive, since many processes have been proposed to explain the peculiar shape of the anomalies, such as forced Rossby wavetrains and internal non-linear interactions, though a final explanation is probably still missing. The shape of the anomalies is however less controversial. The anomalies have a dominant spatial scale in the zonal direction, but there is definitely variability in the zonal direction. It is possible to see how the positive anomalies are interrupted by negative anomalies around the latitude circle at 40°N. These features are captured by the prescribed SST model, with some noticeable differences. The main centre of action over the Pacific is weaker than in the reanalysis and the break in the positive anomalies

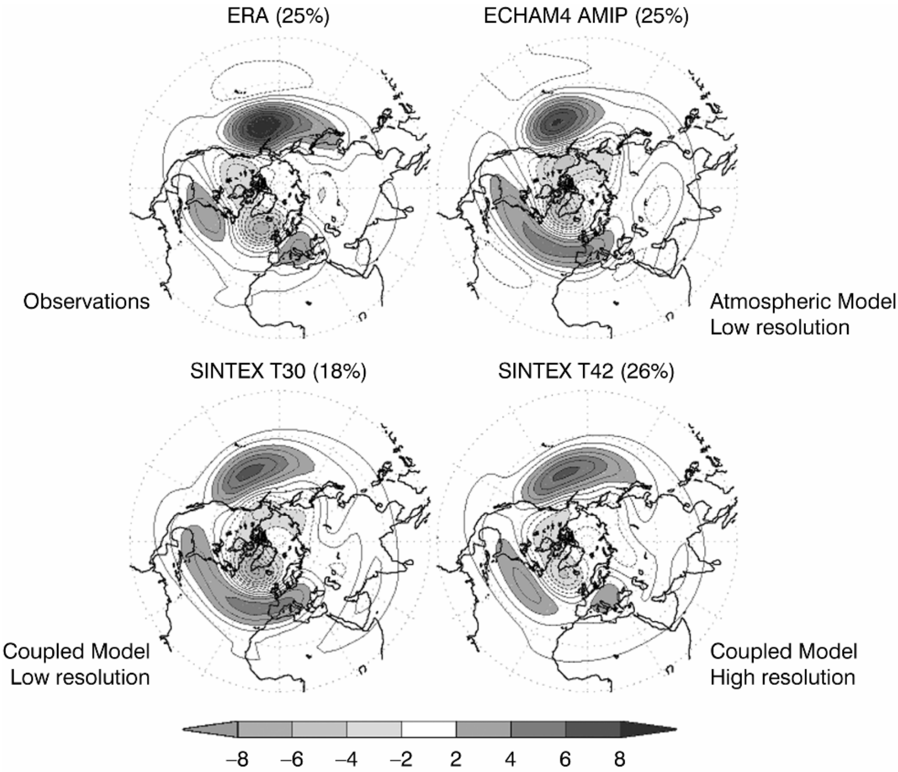

Fig. 6.15 The leading modes of variability shown as the first mode of an Empirical Orthogonal Function analysis of the Winter (January-February-March) 500 mb geopotential heights, for the Observations (ERA reanalysis, top, left), an atmosphere model forced by prescribed observed SST (top, right) and two coupled models with different atmospheric resolution (bottom right and left). The main centres of action are in the North Pacific and in the North Atlantic and they are captured by all simulations, but it is possible to note how the systematic errors are qualitatively different even if the atmospheric model used in the simulation is the same

over the Atlantic sector is missing. The dominance of the zonal spatial scale is enhanced and the centres of action are stretched in the latitudinal direction. These are the typical errors produced by a model under the chosen conditions. They are peculiar to a particular model, but they have also some characteristics that are common to many models.

The coupled models yield different results. The zonalization is further enhanced and the main centre over the Pacific is larger, slightly tilted in its orientation and shifted to the north. The variations in the zonal directions are severely damped and along the 40°N parallel there is almost no zero crossing line for the anomalies.

It is not surprising that these results and those shown in the previous picture indicate that overall the coupled models give worse simulations of the climate variability than the atmosphere-only model. Coupled models have a much more difficult task, since they have to reproduce with fewer constraints the climate variability. Even with the help of realistic initial conditions, as is the case in seasonal forecasting, coupled models have a difficult time to reproduce anomalies. The extra degrees of freedom generated by the release of the ocean SST constraints provide other paths through which errors can grow and distort the evolution. Initial errors in the ocean initial conditions can now feedback by altering the surface winds and in turn modifying the SST again, but also errors in the model numerical formulation and unrealistic features of the parameterisations can feedback between the ocean and the atmosphere, amplifying to finite magnitude in new ways.

This is particularly evident in long simulations. The long term drift in the SINTEX coupled model (Gualdi et al. 2003) shows as a slow, secular increase of the zonal average of the surface temperatures, which is more pronounced in the high latitudes, especially in the North Atlantic. Different models will have different patterns of drift. An example from a seasonal forecast system is given in Fig. 6.16 which shows the average error in the Met Office coupled model after 4–6 months of integration. In this model at this time of year the ocean is biased warm in the upwelling regions of the Pacific and Atlantic oceans. The picture over land is more complex: India and parts of the central United States are also warm but many other land areas have a cold bias. At different times of year the bias will be different (cf. Fig. 5.9). Although the size of the drift is not small compared with the signal one is trying to predict, the presence of a drift by itself is not necessarily a damning feature for a model and several techniques to (partially) eliminate its

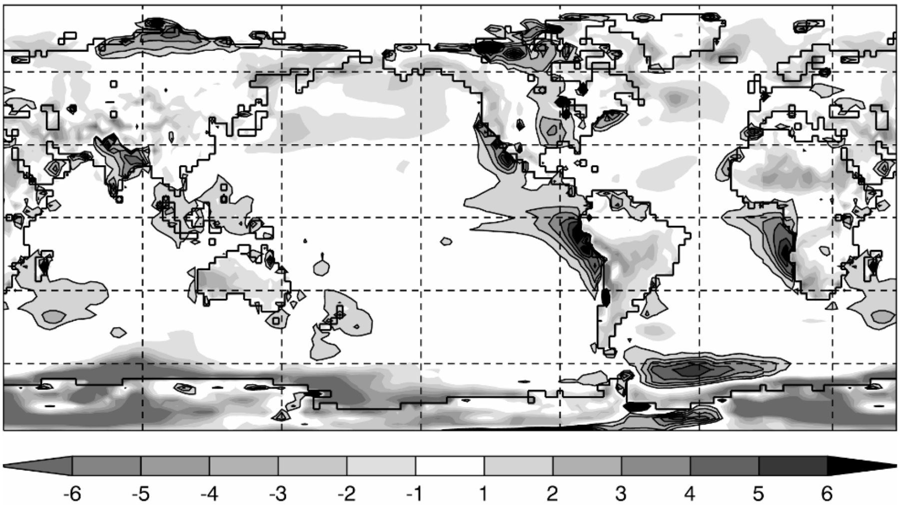

Fig. 6.16 Plot of the bias in the predicted near surface temperature from the UK Met Office seasonal forecast model. This plot shows the bias in May, June July season from 1959 to 2001 for forecasts started in February. Other start dates or models will have different drift patterns (see: http://www.ecmwf.int/research/demeter/d/charts/verification/bias/). In general, the size of the drift is not small compared to the size of the signal one is trying to predict

effects on the data analysis are available, but it is an indication of some imbalance in the coupled model and it would be desirable to keep it to a minimum (see for example Stockdale et al. 1998).

The evaluation of coupled models is usually performed along similar lines as stand-alone atmosphere or ocean models, i.e. comparison with observations for the mean state and variability, budget analysis to verify internal consistency and idealized experiments to verify the over all physical consistency. In the case of seasonal forecast, the skill score of the forecast is in itself a form of verification, but it is less stringent than in the case of weather forecasts because of the probabilistic nature of the seasonal forecast itself. Seasonal forecasts have the advantage that it is possible to generate ensembles with relatively small perturbations in the initial conditions and perform the related analysis. Ensembles for longer climate simulations are much more difficult to produce and less effective in sampling the phase space. New techniques that take into account the special features of the coupled system have to be developed, like the Coupled Manifold (Navarra and Tribbia 2005).

The future development of coupled models will probably involve better dynamical cores to minimize the inconsistencies between the numerical formulations of the ocean and atmosphere. The grand challenge is to eliminate, or at least attenuate, the systematic errors in the basic representation of climate that are still persistent after more than 20 years of research, of which probably the most important is the double ITCZ that models present in the tropical Pacific.

Chapter 7
Statistical Modelling

Simon J. Mason and Omar Baddour

Statistical models provide an alternative approach to using dynamical models in seasonal climate forecasting. In statistical models relationships between one set of data, the predictors, and a second set, the predictands, are sought. Common predictands include seasonal mean temperatures and accumulated precipitation, and are typically predicted using antecedent sea surface temperatures primarily within the tropical oceans. Predictions are made on the assumption that historically observed relationships are expected to apply in the future. There are many conditions for such an assumption to be valid, including the need for high-quality datasets to ensure that the historical relationships are robustly measured, and the need for relationships to have a sound theoretical basis. Because of the possibility of identifying spurious relationships between the predictors and the predictands, the statistical model should be tested carefully on independent data. Most statistical models are based on linear regression, which provides a "best guess" forecast under the assumption that a given change in the value of a predictor results in a constant change in the expected value of the predictand regardless of the value of the predictor. Modifications to the linear model can be made or alternative statistical procedures used when there is good reason to expect a relationship to be non-linear. However, other weaknesses of linear regression may also require these alternatives to be considered seriously. The primary problems with linear regression are multiplicity, multicolinearity, and non-normality of the predictands. Multiplicity refers to the effects of having a large number of candidate predictors: the danger of finding a spurious relationship increases. Multicolinearity arises when more than one predictor is used in the model and there are strong relationships between the predictors which can result in large errors in calculating the parameters of the model. Finally, a linear regression model may not be adequately constructed if the data being predicted have a strongly skewed or

Simon J. Mason
International Research Institute for Climate and Society

Omar Baddour
Moroccan Meteorological Service

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

otherwise non-Gaussian distribution; seasonally accumulated precipitation often exhibits such problems. Alternative forms of linear and non-linear statistical models can be applied to address such distributional problems.

7.1 Introduction

Whereas seasonal climate prediction using general circulation models is based upon successful modelling of the physics of the interactions between the atmosphere and the earth's surface (primarily the sea surface) and of the dynamics of these components of the climate system (Chapters 3–6), the earliest scientific efforts at forecasting seasonal climate anomalies were based on empirical observations of the atmosphere alone. In the late-19th and early-20th centuries, Gilbert Walker, working on the problem of predicting the Indian monsoon, discovered that seasonal anomalies in different parts of the tropics were connected. For example, droughts in India and Australia would often occur in the same year. In such cases where there is a lag between the observed climate of one region and that of another, prediction may be possible. The most important pattern of connected climate anomalies identified by Walker was the Southern Oscillation, which describes opposite changes in sea-level pressure between the western and eastern Pacific Ocean, and involves major disruptions to the trade winds across the southern Pacific. Such relationships between climate anomalies in different areas are known as "teleconnections", and constituted the basis for early empirical methods of seasonal climate forecasting.

Teleconnections are suggestive of some large-scale forcing of the atmosphere, but it has only been since about the mid-1960s that forcing mechanisms have been identified and understood. The Southern Oscillation, for example, is closely related to the state of the sea surface temperatures (SSTs) in the equatorial Pacific Ocean: occasional large-scale warming and cooling of the equatorial Pacific Ocean, known as El Niño and La Niña respectively, simultaneously require and cause prolonged changes in the trade winds over the Pacific Ocean. These changes are associated with large-scale shifts in the location of areas of heavy rainfall, and, in turn, can affect climate conditions in other parts of the globe. Anomalous SSTs outside of the equatorial Pacific also can affect regional climate (for example, changes in the meridional SST gradient of the tropical Atlantic Ocean have important implications for rainfall over north-eastern Brazil and over much of West Africa). Most of the statistical prediction models used currently in operational forecasting attempt to model such relationships between observed climate and anomalous SSTs.

In this chapter, the basic principles of statistical modelling for seasonal climate prediction are introduced in Section 7.2. Section 7.3 discusses in some detail the mathematics of linear regression, which is the most commonly used statistical prediction method used in practice. Linear regression forms the basic framework for a range of more sophisticated statistical techniques, and these, and other statistical techniques, are introduced in Section 7.4, after a discussion of some of the limitations of linear regression.

7.2 Statistical Modelling for Climate Prediction

Although some statistical seasonal climate prediction systems are built upon observed atmospheric teleconnections, the most common approach is to model historical relationships between the climate anomalies to be predicted and the underlying forcing mechanisms – specifically, observed SST anomalies. Statistical methods have been used by centres such as the Met Office (United Kingdom), the Bureau of Meteorology (Australia), and the National Centers for Environmental Prediction (USA) for a number of decades, and supplement the dynamically based models that these centres also use. In the late 1990s, facilitated by extensive capacity building programs and an increasing availability of computing power, statistical methods of seasonal forecasting have been adopted by many national meteorological services throughout the world. These statistical models are constructed primarily to generate forecasts of seasonal precipitation totals, but air temperature forecasts are made also.

7.2.1 Requirements for Applying Statistical Methods in Climate Prediction

Statistical methods aim to identify relationships between two sets of variables through statistical analyses performed on the historical records of the data known as time series. The two sets of variables are:

- A set of variables to be predicted (often denoted Y), and called predictands or response/dependent variables, such as seasonal total rainfall, and monthly average maximum and minimum temperatures
- A set of variables used to make the predictions (often denoted X), and called predictors or explanatory/independent variables, such as SSTs or atmospheric indices (e.g. Southern Oscillation Index – SOI)

The intention is to identify within the historical records a “significantly” consistent relationship between observed values of the predictors and of the predictands. A “significantly” consistent relationship is one that is strong enough to be unlikely to have occurred by chance, and so provides a reasonable level of confidence with which to make a prediction. Of course, for a prediction to be made, a lag between the observations on the predictors and on the predictands is implicit. The lag defines the lead-time of the forecast: by convention, the lead-time is defined as the time period between the end of the recording time of the predictors and the beginning of the target period. For example, if the average SSTs for June are used to predict the total rainfall for the 3-month period August–October, the lag is 1-month (the last observation of the SSTs is made on 30 June, and the target period starts on 01 August). For any significant (lagged) relationship between the predictors and the predictands to be identified, there are some basic data requirements that must be met. These requirements are described in the following sections.

7.2.1.1 Data Quality Issues

If relationships between predictors and predictands are to be modelled reliably, both sets of data need to be of high quality. The quality of a dataset is determined by the accuracy of the recorded values, the spatial and temporal resolution of the data, and the length of available records.

Apart from the problems of human and instrumental errors in recording climate variables, inaccuracies in historical records can arise from changes in instrumentation, relocation of recording sites, and/or changes in the recording environment. For example, the relocation of a thermometer even just a short way down slope could introduce an artificial jump in recorded temperatures because of adiabatic effects and changes in exposure. Any such changes in the recorded climate that are not a reflection of real changes are known as "inhomogeneities". Statistical models are designed to "explain" the observed variability in the predictand data by reference to the observed variability in the predictor data. If part of the variability in the predictand dataset is a result of inhomogeneities, the statistical model will try to "explain" this component as if it were real. Similarly, if part of the variability in the predictor dataset is a result of inhomogeneities, the statistical model will try to use this component of the variability to "explain" the variability in the predictands. Correction for inhomogeneities is therefore an important component of the statistical model-building procedure. There are a variety of checks for data inhomogeneities, the most reliable of which make use of metadata. Metadata are information about the data themselves, and include, for example, information about any changes in instrumentation or changes in the location of the recording site.

Inhomogeneities in data can also be introduced by changes in the temporal resolution of the recordings. For example, the introduction of continuous temperature recordings has allowed a more accurate calculation of the daily mean temperature than was previously possible using only the average of the maximum and minimum temperatures. The average of the maximum and the minimum tends to be higher than the integrated average, and so a change in the way the daily average is calculated could introduce an artificial change in the computed temperature. The temporal resolution of the data can also affect the quality of the information that can be communicated as part of a seasonal climate forecast. For example, although seasonal precipitation forecasts are usually communicated as some form of information about the total rainfall to be expected over a 3-month period, if higher resolution data are available it may be possible to provide some information about the statistics of weather within the season. There are strong relationships between seasonal rainfall totals and rain-day frequencies and heavy rain-day frequencies in many parts of the world, and so a forecast of above-normal seasonal rainfall could be translated into statements about the numbers of days of rain (or heavy rain) that might be expected. However, these additional details are possible only if precipitation measurements are available at the daily timescale.

In addition to the temporal resolution, the spatial resolution of the data is of direct relevance to data quality issues. Station-based data, for example, are site specific, and forecasts that have been derived from models using station data may not be applicable to neighbouring areas. For precipitation, the applicability of a forecast for a nearby site can decline much more rapidly over short distances compared to that for temperature because of the highly localised nature of precipitation, especially in areas of convective rainfall. For precipitation forecasts, therefore, a relatively high density of stations would be advantageous. Sometimes forecasts are made for area-averaged precipitation or temperature. The area-averaging generally improves the forecast performance because the locally specific and unpredictable component of variability is reduced by the averaging. A downside, however, is that the forecast loses its specificity for individual locations, and so some form of translation is required to make the forecast relevant for specific locations. This translation is known as "downscaling" (see Chapter 8).

Other aspects of data quality, such as the presence of missing values and outliers, relate directly to sampling issues, and are discussed separately in the following section.

7.2.1.2 Sampling Issues

The extent to which a modelled relationship between predictors and predictands accurately represents the true relationship depends in part upon the number of records available. Inevitably there will be some errors in estimating the form and strength of this relationship because of the limited number of years for which climate observations are available, and such errors will contribute to inaccurate predictions. These errors typically are larger for short records than for long records. For most statistical models used in seasonal climate forecasting it is recommended that at least 30 years of data be available for constructing a model in order to reduce the effects of sampling errors to an acceptable level.

There are three kinds of sampling errors that can occur when constructing a statistical model:

- The wrong predictors are selected
- The wrong forms of the individual relationships between each predictor and the predictands are selected

- The strength of the individual relationships between each predictor and the predictands is estimated incorrectly

In practice, as the complexity of the model is increased each of the three forms of sampling error become more severe, and sample sizes need to be increased to compensate. To guard against the first two forms of error, statistical significance tests are performed as an attempt to estimate the probability that the error in question has occurred (i.e. that a spurious relationship has been identified). Because these tests are not foolproof, and are subject to problems (Section 7.4.1), they should always be supplemented by theoretical considerations; a sound physical explanation should accompany any relationship that is implied by a statistical model. The theoretical basis can be supplied by research using GCMs, and/or by more detailed statistical analyses, perhaps using other climate datasets to investigate moisture fluxes, for example.

The poor availability of sufficient historical data to construct a robust statistical model is compounded by the presence of missing values. The simplest option is to omit the cases in which there are missing values from the analysis, but this approach easily can leave few or no cases with which to construct a model. Instead, attempts can be made to estimate the missing values. These procedures typically rely on relationships between various climatological variables. For example, if SSTs are to be used as predictors missing SST records could be estimated either from records for nearby locations and the spatial correlation structure of the temperatures, and/or from records immediately prior to and subsequent to the missing values and the temporal correlation structure for that location. Alternatively, if rainfall data are to be used as predictands, missing rainfall values could be estimated from the observed values for neighbouring stations, and/or from station values for variables that are not missing, such as temperature and humidity.

An additional aspect of sampling problems that should be addressed is the presence of outliers. Outliers are values either that are extreme in their own right, lying well outside of the range of the majority of the other data records, or are values that are inconsistent with relationships with other variables. In either case, it has to be decided whether the outliers accurately represent what really happened because if they are retained they will have a large effect on most statistical models. If the outliers are considered accurate, it may still be desirable to reduce their impact on the model so that the data assumptions implicit in constructing the model are not violated (see further discussion in Sections 7.3.3 and 7.4.1). For example, seasonal precipitation data for many parts of the globe are positively skewed[1]; the largest seasonal totals therefore can have an undue influence on many statistical models, and this influence can be reduced by applying the model

[1] Positive skewness occurs fairly commonly in meteorological data, and is evident in seasonal precipitation totals for many parts of the globe, most notably in arid and semi-arid areas. Maximum air temperatures in continental interiors can be weakly negatively skewed.

to the logarithms of the precipitation totals. The logarithmic transformation is often effective in reducing the positive skewness of data.

7.2.1.3 Trends

Before attempting to build a statistical prediction model, it is common practice to remove any long-term trends in both the predictors and predictands. The argument for removing the trends is that if trends are present in the predictand(s) and any of the predictors the probability of identifying a spurious empirical relationship is increased. Effectively, the assumption of independent model errors is violated (Section 7.3.3) unless the trends are removed. However, there are two situations under which it would be unadvisable to remove the trends: if there are prior reasons for expecting trends in the predictands to be caused by trends in any of the predictors; if trends are present in any of the predictands or of the predictors, but not in both. In the latter case, if there is a trend in a predictor, but not the predictand, it seems unreasonable to expect the higher frequency variability of the predictor to provide predictive skill, but for the long-term trend to be unrelated to the predictand; if there is a trend in the predictand, then a good statistical model would seek a predictor for this trend.

7.3 Building a Statistical Prediction Model

In this section the primary steps in constructing a statistical model for climate prediction are detailed. The focus is on using SSTs as predictors and seasonal rainfall totals as predictands, although the procedure is similar for other variables. Linear regression modelling is used as a statistical model, while alternative statistical procedures are considered in Section 7.4.2.

7.3.1 Definition of Predictands

Assuming that the necessary data quality control has been conducted, the first step in constructing a statistical model for seasonal climate prediction is to define the predictand. Seasonal rainfall totals are by far the most commonly used predictand, although increasing attention is being given to prediction of the intra-seasonal statistics of seasonal rainfall, such as the number of rain-days. Only one seasonal total per year is used in the model; other seasons are modelled separately because of the seasonally varying nature and influence of the forcing mechanisms that make seasonal climate prediction possible. The standard procedure is to define a season as a 3-month total or average, but care should be taken to ensure that the season is defined appropriately; specifically, within a season the predictand should have a consistent response to the underlying forcing mechanisms. For example, in much of southern Africa, rainfall in November is positively associated with warm ENSO events, but the relationship in December and January is negative. It would therefore be inappropriate to forecast a November–January season.

If forecasts are to be made for regional averages rather than individual stations, the regions need to be delimited. The regions should be defined on the basis of similar relationships with the forcing mechanisms (for example, similar correlations with SSTs). There are numerous ways of defining the regions, and no single method has been identified as universally preferable. The most commonly used techniques include grouping stations with highest loadings on the same principal component (see Section 7.4.2 for further discussion of principal components), and cluster analysis. Once stations have been allocated to a region, a regional rainfall index, r_k^*, is then calculated for each year, k, typically using the following equation:

$$r_k^* = \sum_{i=1}^{m} w_i \frac{r_{k,i} - \overline{r_i}}{s_i}, \tag{7.1}$$

where w_i is a weight applied to the ith of m stations, $r_{k,i}$ is the rainfall at this ith station during year k, and $\overline{r_i}$ and s_i are the average and standard deviation of the station's rainfall, preferably calculated over a common reference period. The weights are defined to sum to unity, and can be set to avoid favouring unduly the contributions of clusters of stations to the regional index. In practice, if the station network is reasonably even, for the sake of simplicity the weights often are set equal for each station. The subtraction of the mean and division by the standard deviation standardises the data at each station and is designed to avoid giving stations with large mean and variance excessive weight (See Chapter 8, Section 8.3.3, for further discussion about standardisation, including some of its limitations).

7.3.2 Definition of Candidate Predictors

The most commonly used predictors in statistical models for seasonal climate prediction are SSTs. There are a number of global SST datasets available with varying spatial resolution (from 10° × 10° to 1° × 1°), and some extend as far back as the mid-19th century (although data quality is considerably improved from about the 1950s). Whichever dataset is used, there are a large number of grids from which to choose, and some kind of pre-selection of grids and area-averaging of SSTs should be performed. Some area-averages have been predefined, such as the NIÑO3 index (5°S–5°N, 150–90°W), but similar averages may be required for other areas if SSTs here are thought to have an important effect on rainfall variability in the region of interest. These area-averages should be defined based on theoretical considerations and extensive supporting statistical research. Simple correlations between the rainfall index and global SSTs followed by delimitation of areas with high correlation should be avoided because of problems with fishing (section 7.4.1) and subsequent problems of potential overestimation of the performance of the statistical model.

The temporal resolution of the predictors is not necessarily the same as that of the predictands. Because SSTs change much more slowly than the atmosphere, a 1-month average is less noisy than a 1-month average of some atmospheric variable, and more faithfully highlights recent trends in temperatures compared to a 3-month average. As a result, statistical models are frequently constructed using SSTs for the latest month available. Of course, for an operational forecast to be made, the predictor data must be available before the beginning of the target period. The lag between the availability of the predictor data and the beginning of the target period defines the lead-time of the forecast (section 7.2.1).

7.3.3 Statistical Model Construction

7.3.3.1 Model Formulation – Simple Linear Regression

The simplest statistical model consists of a single predictand and a single predictor. In this case a regression model assumes a linear relationship between the predictor, x, and the predictand, y:

$$y = \beta_0 + \beta_1 x + \varepsilon\,, \tag{7.2}$$

where β_0 and β_1 are parameters to be estimated, and ε is an "error" term representing the unpredictable component of the predictand. The parameter β_0 is often called the "regression constant" or the "intercept", while β_1 is referred to as the "regression coefficient" or the "slope". The predictable component, $\hat{y}$, is given by:

$$\hat{y} = \beta_0 + \beta_1 x\,. \tag{7.3}$$

The objective in fitting a regression model is to estimate the parameters β_0 and β_1 so that the differences, or "residuals", between the estimated[2] values of the predictands, $\hat{y}$, and the observed values, y, are minimised. From Eqs. (7.2) and (7.3):

[2] In this chapter $\hat{y}$ is referred to as "estimates" or "fitted values" when applied to cases within the training period (i.e. to cases used to estimate the regression parameters), and to "predictions" only when new values of x are applied. See Sections 7.3.3.3 and 7.3.3.4 for a definition and discussion of the training period.

$$\begin{aligned}\varepsilon &= y - \hat{y} \\ &= y - \left(\beta_0 + \beta_1 x\right) \cdot \end{aligned} \tag{7.4}$$

For a set of n years of data, the sum of the squares of these errors, SS_E, is minimised,[3] i.e.:

$$\begin{aligned}\min SS_E &= \min \sum_{k=1}^{n} \varepsilon_k^2 \\ &= \min \sum_{k=1}^{n} \left[y_k - \left(\beta_0 + \beta_1 x_k\right)\right]^2 .\end{aligned} \tag{7.5}$$

Equation (7.5) is minimised by setting its first partial derivatives to zero:

$$\begin{aligned}\frac{\partial SS_E}{\partial \beta_0} &= \frac{\partial \sum_{k=1}^{n} \left[y_k - \left(\beta_0 + \beta_1 x_k\right)\right]^2}{\partial \beta_0} = 0 \\ &= -2 \sum_{k=1}^{n} \left(y_k - \beta_0 - \beta_1 x_k \right) = 0 \quad ,\end{aligned} \tag{7.6a}$$

and:

$$\begin{aligned}\frac{\partial SS_E}{\partial \beta_1} &= \frac{\partial \sum_{k=1}^{n} \left[y_k - \left(\beta_0 + \beta_1 x_k\right)\right]^2}{\partial \beta_1} = 0 \\ &= -2 \sum_{k=1}^{n} x_k \left(y_k - \beta_0 - \beta_1 x_k \right) = 0 .\end{aligned} \tag{7.6b}$$

From Eq. (7.6), the two regression parameters can be obtained as:

$$b_1 = \frac{\sum_{k=1}^{n} \left[(x_k - \overline{x})(y_k - \overline{y}) \right]}{\sum_{k=1}^{n} (x_k - \overline{x})^2} \tag{7.7a}$$

---

[3] The minimisation of the sum of the squared errors is by far the most commonly used form of estimation in seasonal climate prediction. The only other minimization criterion that has been used to any notable degree is that of minimising the sum of the absolute errors, and is known as "least absolute deviation" (LAD) regression. See Section 7.4.2.2 for further discussion of LAD regression.

and

$$b_0 = \overline{y} - b_1\overline{x} \ , \tag{7.7b}$$

where b_0 and b_1 are estimates of the parameters β_0 and β_1, respectively.

The regression coefficient is closely related to Pearson's product moment correlation coefficient,[4] r:

$$r = b_1 s_x s_y^{-1} \ , \tag{7.8}$$

where s_x and s_y are the standard deviations of x and y, respectively. The correlation coefficient is a widely used measure of the strength of linear association between the predictor and the predictand. Although it can be estimated using Eq. (7.8), it is more commonly calculated using:

$$r = \frac{\sum_{k=1}^{n} (x_k - \overline{x})(y_k - \overline{y})}{s_x s_y} \tag{7.9}$$

The numerator in Eq. (7.9) is related to the covariance by a factor of n, and will be positive if positive anomalies in both the predictor and the predictand tend to occur in corresponding cases, and will be negative if opposite anomalies tend to occur. Equation (7.9) defines the correlation as the standardised covariance. Frequently the correlation is squared, and it can then be interpreted as the proportion of the variance of the predictand that can be 'explained' using the predictor.

As an example, December–February 1961/62–2000/01 rainfall over Lusaka, Zambia, is shown as the y variable in Fig. 7.1, and is regressed against the October value of the NIÑO3.4 index. Lusaka is located in part of southern Africa where El Niño (La Niña) conditions are frequently associated with below-normal (above-normal) rainfall. The correlation is –0.49, and is statistically significant at a 1% significance level, indicating that there is a strong statistical basis for making a prediction. The figure shows that rainfall tends to decrease over Lusaka as the equatorial Pacific becomes warmer. The relationship with October values of the NIÑO3.4 index implies that a prediction can be made with a lead-time of 1 month using the formula:

$$\widehat{\text{rainfall}} = 607 - 81 \times \text{October NIÑO3.4} \ . \tag{7.10}$$

[4] There are other correlation coefficients, but Pearson's is by far the most widely used, and unless specified otherwise, the term "correlation" refers to Pearson's correlation.

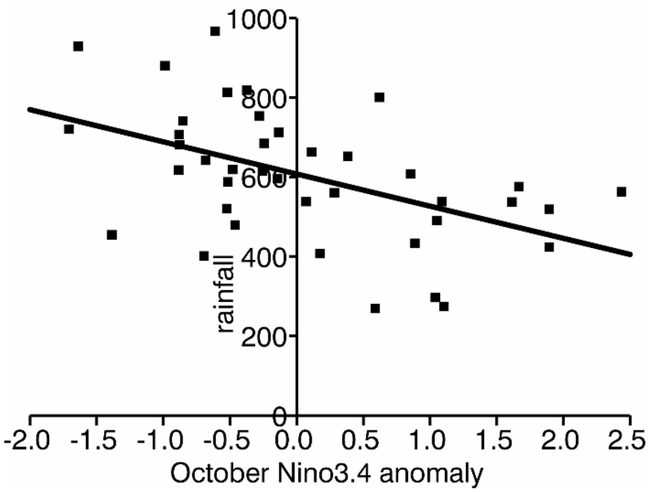

Fig. 7.1 Example of a linear regression model in which October values of the Niño3.4 index are used to predict December–February 1961/62–2000/01 rainfall totals for Lusaka, Zambia. The solid line represents the regression model

The negative regression coefficient in Eq. (7.9) means that the expected seasonal rainfall decreases by more than 80 mm for every 1°C increase in temperature in the central equatorial Pacific.

7.3.3.2 Model Formulation – Multiple Linear Regression

When more than one predictor is used, a multiple regression model assumes the following form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \varepsilon \; . \tag{7.11}$$

Given m predictors and n cases (years of data), the regression model becomes:

$$\hat{y}_k = \beta_0 + \beta_1 x_{k,1} + \ldots + \beta_m x_{k,m} \, . \tag{7.12}$$

Equation (7.12) has $p = m + 1$ parameters, and can be simplified in matrix notation to:

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} \, , \tag{7.13}$$

where **X** is a $n \times p$ array in which the rows represent each year of data, and the columns represent each predictor, with the first column containing unity,[5] and the $i + 1$th column containing the ith predictor.

As with simple linear regression, the objective is to estimate the parameters $\boldsymbol{\beta}$ so that the sum of squares of errors is minimised:

$$\begin{aligned} \min SS_E &= \min \sum_{k=1}^{n} \varepsilon_k^2 \\ &= \min \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ &= \min \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^T \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) \\ &= \min \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \end{aligned} \quad . \tag{7.14}$$

Similarly, Eq. (7.14) can be minimised by taking the first derivatives:

$$\frac{\partial SS_E}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{0} , \tag{7.15}$$

which can be rearranged to give:

$$\boldsymbol{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{y} \quad . \tag{7.16}$$

In practice, the inverse in Eq. (7.16) is difficult to calculate and can be prone to rounding errors if the predictors are inter-correlated, and so most statistical packages use alternative formulations and advanced linear algebra techniques, such as the singular value decomposition, to obtain the parameter estimates.

7.3.3.3 Predictor Selection

Unless the predictors to be used are predefined, the candidate predictors would normally be tested for inclusion in the final model that is to be used to make predictions. The standard approach is to include only those predictors in the final regression equation that contribute to a significant reduction in the size of the errors. Since the addition of *any* additional predictor into the model will always reduce the size of the errors, this reduction needs to be significantly large, i.e. the estimates need to improve sufficiently for us to be confident that the inclusion of the added predictor will effect an improvement in real-time predictions.

---

[5] This extra column is used for the regression constant, which is given as the first element of $\boldsymbol{\beta}$.

A commonly used procedure for selecting predictors is stepwise regression. There are three main forms of stepwise regression:

- *Forward selection*: Predictors are added one-by-one, with the remaining candidate predictor that reduces the size of the errors the most being added next, and continuing until the errors cannot be significantly reduced.
- *Backward elimination*: All candidate predictors are initially included, and then predictors are removed one-by-one, with the predictor that increases the size of the errors the least being removed next, and continuing until the errors can only be increased significantly.;
- *Stepwise selection*: Predictors are added one-by-one in the same way as for forward selection, but at each stage the included predictors are retested so that if the removal of any of these predictors results in an insignificant increase in the size of the errors they are removed.

All of these stepwise procedures require a criterion for deciding whether the change in the size of the errors is significantly large. The approach generally used is based upon the F-statistic, and involves a decomposition of the total sum of squares about the mean of the predictand, SS_T:

$$\begin{aligned} SS_T &= \sum_{i=1}^{n} y_i^2 \\ &= \mathbf{y}^T \mathbf{y} \quad , \end{aligned} \tag{7.17}$$

where y has been centred around zero by subtraction of the mean. The SS_T is decomposed into two components: the explained component as modelled by the regression model, SS_R, and the unexplained component or sum of the squares of the errors, SS_E, as defined in Eq. (7.14). The regression sum of squares is calculated as:

$$\begin{aligned} SS_R &= \sum_{i=1}^{n} \hat{y}_i^2 \\ &= \hat{\mathbf{y}}^T \hat{\mathbf{y}} \\ &= (\mathbf{Xb})^T \mathbf{Xb} \\ &= \mathbf{b}^T \mathbf{X}^T \mathbf{Xb} \quad , \end{aligned} \tag{7.18}$$

so that $SS_T = SS_R + SS_E$. The F-statistic tests whether the change in SS_R for the predictor under consideration is significantly large compared to the mean of the squared errors, MS_E, after including the predictor. The MS_E is the SS_E divided by $n-p-1$. Under the assumption that the predictor is unrelated to the predictand, the F-statistic is drawn from an F distribution with one and $n-p-1$ degrees of freedom. A predefined value of this statistic can be defined for a given level of significance (typically 0.05), and if the calculated F-statistic exceeds this value the predictor results in a significant improvement in the estimates of y. The procedure, however, is problematic, partly because of the sensitivity of the F-statistic to distributional assumptions (Section 7.3.3), and because of problems related to multiplicity (Section 7.4.1), which invalidate the significance tests.

Nevertheless, given the definitions of SS_T and SS_R in Eqs. (7.17) and (7.18), the ratio SS_R/SS_T provides an indication of the proportion of the total variability in the predictor that can be explained by the regression model. This proportion, denoted R^2, is known as the coefficient of determination, and is the multivariate equivalent of the squared correlation coefficient (Section 7.3.3). An adjusted R^2 is sometimes calculated to correct for the number of parameters in the model.[6] The procedure described above based on the F-statistic is equivalent to selecting which of the two models (the one with and the one without the predictor under question) has the larger adjusted R^2.

None of the stepwise procedures guarantees that the best possible set of predictors (i.e. the one that minimises the errors) is selected, and so one option is to search through all possible combinations to find that subset that reduces the size of the errors most significantly. Since this search can be computationally prohibitively expensive if the number of candidate predictors is large,[7] an alternative is to modify the simpler forward selection and backward elimination procedures described above by swapping out at each step any predictors that can effect an improvement in the model. The predictors are swapped one-by-one with the predictor that improves the model the most being introduced as replacement. The swapping continues until no further improvement is possible.

A somewhat different approach is to identify a model that makes a good set of independent predictions, as opposed to one that minimises the errors in estimating the data used to construct the model. The problem with minimising Eqs. (7.5) and (7.14) is that the model is optimised only to describe the relationship between the predictors and predictand over a set period, known as the training or calibration period (the period of the data used to construct the model), but there is no guarantee that this model will make good predictions when it is applied over a different period. Some procedures search for the set of predictors that make the best set of independent predictions by using only part of the data to construct the model and then examining the predictions for the data that was withheld. Techniques for performing this independent assessment are discussed in further detail in Section 7.3.3.

[6] Note that the adjusted R^2 cannot be interpreted as the proportion of variability explained.

[7] Given k candidate predictors the number of possible combinations is $2^k - 1$.

7.3.3.4 Model Assumptions

Before assessing how well the regression model can predict the response variable, it is important to assess the validity of the model. If various assumptions about the data used in constructing the model cannot be upheld, the model parameters may be estimated incorrectly, and the predictions made in real-time will then be less accurate than expected. These assumptions are enumerated below. Alternative procedures for when these assumptions are invalid are discussed in Section 7.4.

- Errors are identically and independently distributed (iid)

The forecast errors (Eq. 7.4) should show no tendency to increase or decrease in size either in the long-term or for identifiable sub-periods of the data. Similarly, the variance of the errors should not be related to values of the predictors ("homoscedasticity"). This latter restriction is often a problem when constructing statistical models to predict precipitation because forecast errors typically increase as the forecasted precipitation increases simply because there is a lower bound to precipitation.

In addition, the errors are assumed to be independent of each other. This assumption means that the model should show no tendency to underestimate or overestimate the observed values over a string of years. In combination with the assumption of a zero mean-error, the independence of errors means that each time a new prediction is made, the probability of overestimating (or underestimating) the observed value is 0.5 in all cases.[8] The Durbin-Watson test is recommended for testing independence of the errors, and works by identifying whether there is any autocorrelation in the errors (i.e. is it possible to "predict" the errors from previous errors?).

- Predictand is normally distributed

Although strictly it is only the model errors that need to be normally distributed, in practice, this distributional assumption about the errors is more often met when the predictand itself is normally distributed. In addition, if the predictand is not normally distributed, the regression parameters can be heavily influenced by the more extreme values. Since seasonal rainfall totals for many areas have a positively skewed distribution (see, for example, Fig. 7.1), it is often advisable to transform the data so that the transformed data are normally distributed. Commonly used transformation functions include the logarithm, and the square root and other power transformations.

[8] More generally, because of the assumption of fixed variance, the probability that the error will exceed any pre-defined value is a constant.

- Linear relationship

If the relationship between (any of) the predictor(s) and the predictand is non-linear, Eqs. (7.2) and (7.11) are of the wrong "form". The true form of the relationship(s) may be unknown, but more complex relationships can be examined using alternative regression models (Section 7.4.2). Apart from testing for improvements in the predictions if a more complex model is used, it can be useful to reorder the predictions so that they are sorted by the value of (one of) the predictor(s) rather than chronologically, and then re-conducting the test for independence. If the true form of the relationship is quadratic, for example, but is assumed to be linear, the residuals will be of a similar sign at the beginning and end of the re-ordered series, and of the opposite sign in the middle.

- Uncorrelated predictors

For multiple regression, the model parameters can be estimated inaccurately when there are strong correlations between the predictors. The presence of strong correlations between predictors is known as multicolinearity, and is discussed in further detail in Section 7.4.1.

7.3.3.5 Model Evaluation

Since measures of the errors in estimating the y values ("goodness of fit" measures), are as much a function of the number of parameters included in the model as they are of the quality of the model's ability to describe the variability in the predictand, they are not necessarily very informative. In order to estimate how well the model can predict new values, a separate set of data that was not used to construct the model is required. Two approaches are used, and in both cases the data is divided into a "training" or "calibration" period, and an "independent" or "verification" period:

- *Cross-validation*: One year is withheld (together, optionally, with additional years immediately preceding and succeeding; this omitted period is known as the cross-validation window), and the remaining years are used to train the model. A prediction is made for the omitted year or the year in the middle of a window larger than one, and the procedure is repeated until a prediction has been made for each year (Fig. 7.2a and b).
- *Retroactive validation*: The model is trained using only the first few years of the data, and a prediction is made for the year immediately after the end of the training period. The model is then updated, adding the year just predicted to the training period, and a prediction for the following year is made (Fig. 7.2c). This procedure is continued until a prediction for the last year has been made. (Sometimes the subsequent k years are predicted, where $k > 1$, and the model is only updated every k years).

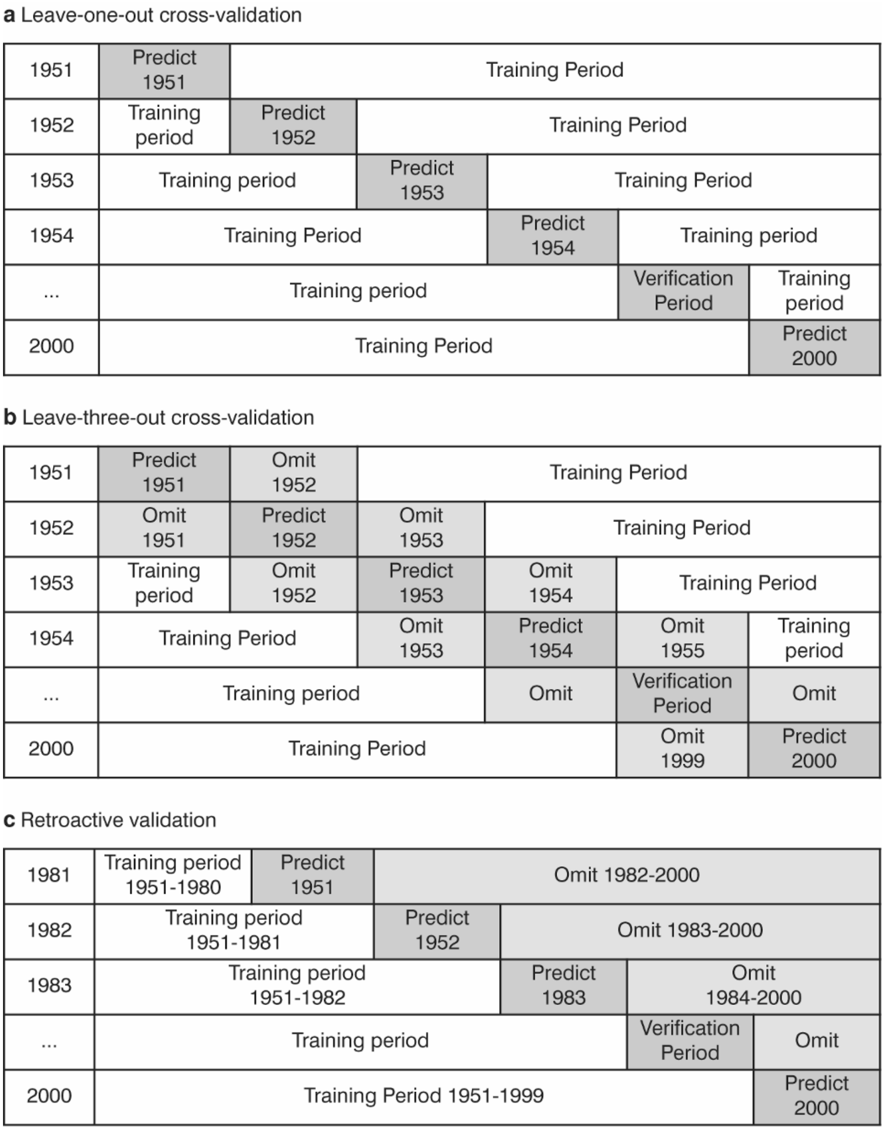

Fig. 7.2 Schematic diagrams illustrating the procedure for (a) leave-one-out cross-validation, (b) leave-three-out cross-validation, and (c) retroactive validation

In each case, the objective is to generate a set of "out-of-sample" predictions. These predictions need to be independent of the data used in the training set, but assuring complete independence is exceptionally difficult, particularly with cross-validation. One of the main ways in which "leakage" of information from the training to the verification sample is allowed to occur is through a failure to re-select the predictors adequately at each step. It is important that the predictors are allowed to be reselected rather than only allowing the model's parameters to be recalculated.[9] Ideally each training period should be independent of each other, but since that is impractical because of limited sample sizes, some effort to ensure that at least some of the training periods differ should be made. In cross-validation this independence can only be achieved by using a fairly large window.

Retroactive validation closely mimics the operational generation of predictions, and so should give a realistic estimate of how well the model would have performed if it had been operational since the first year of the independent predictions (although selection of candidate predictors by using all the data can bias the results). The downside of retroactive validation is that predictions are made only for a subset of the data, and so the small sample size will contribute to large errors in the estimates of the quality of the predictions.

In cases where the predictor(s) is (are) specified and the distributional assumptions described in the previous section do not hold, bootstrapping of the model parameters should be conducted. Bootstrapping involves randomly re-sampling pairs of predictor and predictand values, and then recalculating the regression using the resample. There are many ways of designing a bootstrap procedure, but the standard approach is to generate a sample that has the same number of cases as the original sample. The cases are drawn with replacement, for otherwise the bootstrap sample would be identical to the original sample. A large number of bootstrap samples are generated, and regression models constructed for each one. The distribution of the regression parameters provides an indication of the uncertainty in estimating the "correct" parameters.[10]

7.3.3.6 Scoring Metrics

Given a set of independent predictions, the most commonly used metric to calculate how well these predictions match the observed outcomes is the correlation coefficient. The correlation coefficient was introduced in Section 7.3.3, where it was used to measure the strength of the linear association between the predictor(s) and predictand. To use the correlation for forecast verification, simply replace x

[9] By reselecting the predictors at each step it is quite possible that the actual set of predictors that are used to make an operational forecast are not actually selected in some or even any of the cross-validation steps. This failure to test using the operational predictors may seem problematic, but an essential part of the cross-validation procedure is to test the predictor selection process.

[10] Although not widely performed, one way of estimating the uncertainty in a prediction would be to make a suite of predictions using models constructed using the bootstrap samples. More widely used methods are discussed in the following section.

with $\hat{y}$ in Eq. (7.9). Note, however, that the correlation is not a measure of forecast accuracy for two reasons: the subtraction of the means of x and y in the numerator eliminates any bias in the forecasts, and the division by the respective standard deviations eliminates any variance bias. (See Section 7.3.3 and Chapter 8, for definitions of accuracy, bias, and variance bias.) As a result, predictions of rainfall, for example, that are consistently too wet or too dry, and vary too much or too little can still achieve a perfect verification score. In the context of statistical models, such problems are not usually very severe because the predictions should be reasonably well calibrated over the training period. As a result, the mean bias should be fairly small, although in most cases the variance will be underestimated, simply because in an imperfect model predictions err towards the climatological mean.

The squared correlation coefficient is often quoted as the percentage of variance of the observed values that is successfully predicted. While technically correct, this percentage is often misinterpreted as some measure of how frequently the forecasts are "correct". In the context of the deterministic predictions from regression models, "accuracy" is a more appropriate quality of the forecasts than correctness because the predicted and observed values will always differ if only by a very small amount, and so the predictions are never "correct" in a strict sense. Accuracy generally is indicated using an average of some measure of the errors. The mean squared error, introduced in Section 7.3.3, is a natural choice because it is a quantity that has been minimised when the model was constructed, but is not particularly intuitive otherwise. The root mean squared error resolves the conceptual problem of interpreting squared errors, but the mean absolute error is the simplest to understand: it indicates by how much, on average, the predictions differ from the observed outcomes. A still more informative approach would be to indicate in a contingency table or histogram how frequently errors of different magnitude occur.

Other widely used metrics are based on the contingency table: it has become popular to assign the observed values to one of three equiprobable categories, labelled "below-normal", "normal", and "above-normal", with "below-normal referring to the driest/coldest third of cases, and the other categories defined accordingly.[11] The deterministic forecasts can be classified into one of these three categories, and a table comparing the forecast and observed categories can then be constructed. An example is shown in Table 7.1a for 30 years of cross-validated predictions of December–February Lusaka rainfall using only the NIÑO3.4 index as predictor. The "correct" predictions are shown in the diagonal cells from top left to bottom right.

[11] Categories do not have to be equiprobable, and more (or less) than three categories can be defined. The principles of verification remain the same, however.

Table 7.1 (a) Contingency table and (b) variance-adjusted contingency table of cross-validated predictions of December–February 1971/72–2000/01 rainfall totals for Lusaka, Zambia, using the October NIÑO3.4 index as sole predictor. The categories are equiprobable, and are marked B for below-normal, N for normal, and A for above-normal

(a)

		PREDICTIONS			
		A	N	B	TOTAL
OBSERVATIONS	A	3	7	0	10
	N	0	7	3	10
	B	1	5	4	10
	TOTAL	4	19	7	30

(b)

		PREDICTIONS			
		A	N	B	TOTAL
OBSERVATIONS	A	5	4	1	10
	N	2	4	4	10
	B	3	2	5	10
	TOTAL	10	10	10	30

There is a wide range of summary measures of such contingency tables, but they are not discussed here because the loss of information as a result of the categorization of the observations and predictions, and deterministic nature of the predictions mean that such an interpretation of the climate prediction information is undesirable. The interested reader is referred to Jolliffe and Stephenson (2003) and Wilks (2005) for details.

The number of predictions of the normal category is higher than for the other categories because of the lower variance of the predictions compared to the observations. As a result, the variance of the forecasts is sometimes increased artificially so that the number of predictions in each category is equal. The resulting contingency table is shown in Table 7.1b. Such variance adjustment is problematic because the squared errors are no longer minimised, and it can be seen from Table 7.1b that there is no improvement in the total number of correct predictions (5 + 4 + 5 compared with 3 + 7 + 4), while there is an increase in the number of two-category misses (i.e. predictions of above-normal when below-normal occurred, or vice versa). Variance-adjustment should therefore be discouraged.

Ideally, if the forecasts are categorised they should be expressed as probabilities. Methods for generating probabilistic forecasts from the deterministic predictions of regression models are discussed in the following section. The verification of probabilistic forecasts is a complex issue, and is discussed in detail in Chapter 10.

7.3.3.7 Generating Probabilistic Forecasts

Once the regression model has been constructed, predictions can be made using Eqs. (7.3) and (7.13) given new values of the predictor(s). However, these equations give only a “best-guess” of the outcome, and no indication of the uncertainty is provided. There are a number of ways in which this best-guess forecast can be converted to a probabilistic forecast, but the most reliable procedure is to use information about the variance of the errors in estimating previous known values. The error variance is widely used to define a prediction interval on the forecast, although it is possible to obtain probabilities for predefined categories as well. If the errors in the forecasts are assumed to be Gaussian, these probabilities can be calculated by integration of the t-distribution using the best-guess as the mean and the error variance as the variance. (See Chapter 8, Section 8.5.1 for discussions on different ways of communicating forecast uncertainty.) The error variance is normally calculated from the fitted values, although the errors in the cross-validated forecasts could be used instead, and may be more reliable.

Alternative approaches include using contingency tables that compare the category of the forecast with the observed category for a set of forecasts. Then if 60% of the times that the forecast has indicated below-normal rainfall the observation was also below-normal, for example, the forecast would specify a 60% probability of below-normal rainfall the next time the forecast indicates below-normal. There are two problems with this approach: very large samples are required to estimate the probabilities reliably, and; no distinction is made between the probabilities issued when the forecast indicates well below-normal rainfall, and when it indicates marginally below-normal. The large differences in the amount of rainfall that can be classified as “below-normal”, for example, could be offset by increasing the number of categories, but only at the cost of requiring still larger samples. Given these problems, the use of contingency tables to obtain forecast probabilities is not recommended. Instead there is a suite of statistical procedures that can be used to obtain these probabilities directly rather than estimating a best-guess and then trying to account for the uncertainty subsequently. These procedures are discussed in Section 7.4.2.

7.4 Alternative Statistical Methods to Linear Regression

Linear regression forms the basis for a number of more sophisticated statistical techniques that have been used in seasonal climate prediction. Some of these techniques are discussed in Section 7.4.2, all of which have in common an attempt to estimate a “best-guess” forecast. Some alternative statistical techniques that estimate forecast probabilities without providing a best-guess are considered in Section 7.4.2. However, to understand the motivation for using any of these methods, it is first helpful to consider some of the limitations and potential pitfalls of linear regression, and these issues are outlined in Section 7.4.1.

7.4.1 Problems with Linear Regression

The problems and potential pitfalls listed in this section are not exclusive to linear regression, but are listed to provide a context for understanding the more sophisticated techniques described in Sections 7.4.2 and 7.4.3. In many cases the alternative techniques attempt to address only a subset of the problems listed below.

7.4.1.1 Multiplicity

One of the primary difficulties in using linear regression for seasonal climate forecasting is identifying the predictors to use in the model. Most frequently, predictors used are measurements of SSTs, but land-surface characteristics and atmospheric indices are also used for forecasting in countries such as India where the use of such variables has been supported by extensive research on seasonal predictability. Whether or not SSTs are used exclusively, the pool of candidate predictors is vast, and the problem arises of which subset of these predictors should be included in the regression model. The temptation is to choose the predictors that are best correlated with the predictands, but the probability of identifying highly, but spuriously, correlated predictors increases[12] as the pool of candidate predictors is expanded. This problem is known as "multiplicity", and the search for predictors by repeated testing of the strength of statistical relationships is known as "fishing", and almost invariably results in the creation of a statistical model that performs worse than anticipated when used operationally.

One reason why "fishing" results in models that perform poorly in operations is that standard tests of statistical significance used in constructing a statistical model assume that the predictors to be used in the regression model have already been selected, and these tests become invalid when only the models that give the best results are selected. If a number of regression models are tested with the aim of identifying those that work well, then problems of multiplicity arise. Standard significance tests require adjustment for multiplicity, otherwise there is an increased danger of accepting predictors that should not be included in the model, and/or of overestimating the strength of the model's predictive capability. This selection of spurious, or of too many, predictors is known as "over-fitting".

[12] I.e. the probability of making a type-I error increases.

Cross-validation (Section 7.3.3) is used to test for over-fitting. Leave-one-out cross-validation appears to be the standard approach in the atmospheric sciences (leave-*k*-out is used if the data are autocorrelated, but *k* typically is set only to a maximum of twice the decorrelation time). However, it is not widely recognised in the atmospheric sciences literature that a substantial proportion of the data needs to be omitted to obtain unbiased results. How much data should be omitted remains a question for further research, but there have been suggestions that it should be as much as 40–60% (Xu and Liang 2001). Frequently, therefore, the problems of multiplicity are not adequately addressed.

An aspect of multiplicity is evident not just when constructing a model with a large pool of candidate predictors, but also when constructing a number of models, perhaps for different stations and/or seasons. If numerous models are constructed, the probability of finding at least one that gives spuriously "good" predictions increases, and so the statistical significance of the overall set of results needs to be assessed. Tests for "field significance" have been designed to address this question. Multiplicity problems can apply to GCM forecasts as well since forecasts are made for a large number of locations, variables, lags, and target periods.

7.4.1.2 Multicolinearity

Multicolinearity is a problem that sometimes arises when more than one predictor is used in a regression model. If the predictors used are themselves highly correlated with each other, errors in estimating the model parameters can become substantial. The errors in these parameter estimates can give poor predictions when new values of the predictors are applied to the model, and can also create problems in interpreting the regression coefficients. Whereas multiplicity results in bad forecasts because of the inclusion of incorrect predictors in the model, multicolinearity can cause bad forecasts even when the correct predictors are included simply because the regression parameters may be poorly estimated.

To illustrate the difficulty of interpreting regression parameters when predictors are correlated, consider a simple multiple regression model to predict the March values of the NIÑO3.4 index from the January and February values. Using data for 1971–2000, the regression coefficients for January and February, respectively are −0.395 and 1.216, which seems to imply that the March value is negatively correlated with the January value, whereas one would expect a slightly weaker positive correlation than for February. However, when the January and February values are used in separate models as the only predictors, the coefficients change to 0.628 and 0.761, respectively, showing that the values in both months are positively correlated with that in March.

7.4.1.3 Non-linearity

Linear regression assumes a linear relationship between the predictor(s) and the predictand. This assumption means that for a given change in the value of a predictor, (e.g. a 1°C increase in SST in a specified area), the expected change in the predictand (e.g. an increase in seasonal rainfall of 100°mm) is the same regardless of the actual sea temperature, and regardless of the values of the other predictors. Given the non-linear nature of the atmosphere the linearity assumption seems inherently unreasonable, and the flexibility to model non-linear relationships statistically may be desirable. In practice, however, the linearity assumption is often a reasonable approximation, and even where it is not, the degrees of freedom required to identify the correct form of the relationship are likely to be lacking.

7.4.1.4 Assumptions About Data Distribution

In addition to the linearity assumption, linear regression assumes that the predictand (but not necessarily the predictors) is normally distributed. While this assumption may be quite reasonable for variables such as geopotential heights, for other variables the data may not be normally distributed, and fitting a linear regression then becomes problematic. Although the distribution of surface air temperatures is skewed, this can generally be ignored because the skewness is not usually severe. However, with precipitation, skewness can be marked (see examples in Chapter 8, Section 8.3.1), and there is the related problem that precipitation has a lower limit of zero. It makes no sense for a regression line on precipitation to extend below zero since negative precipitation is meaningless, but a linear regression model is unaware of such a constraint. The lower limit on precipitation also means that even if a regression model is fitted, the errors are usually larger for larger precipitation rates than for rates close to zero. This increase in the variance of the errors in estimating precipitation for larger precipitation amounts violates the homoscedasticity assumption of multiple linear regression. Although these problems could be addressed by using certain forms of generalised linear models (see Section 7.4.3), they are frequently ignored, or assumed not to be problematic.

7.4.2 Regression-based Statistical Prediction Techniques

7.4.2.1 Power and Non-linear Regression

Even when the relationship between the predictor and predictand is non-linear, a transformation of the values of the predictors and/or predictand may make it possible to treat the problem as linear. The most commonly used transformations are power transformations (e.g. using the square or the square-root of the predictors), and adding these to the pool of predictors. The resulting models, known as power regression models,[13] have been used extensively in statistical predictions of the Indian monsoon, for example. However, caution has to be taken since the problem of multiplicity is exacerbated by expanding the number of candidate predictors, and theoretical justifications for the power transformations should be supplied. In addition, multicolinearity is introduced with most power transformations. Power regression is sometimes used in seasonal forecasts of climate impacts, where non-linear relationships between climate variables and the application data in question have a theoretical basis (e.g. Chapters 12 and 13). Other examples of non-linear regression include exponential models, which are used more frequently in forecasting impacts than in forecasts of seasonal climate per se.

Compared to power regression models, neural networks constitute a considerable increase in the complexity with which non-linear relationship can be modelled. Neural networks are a recent development in seasonal climate prediction, but have been applied successfully, and have been implemented as the statistical atmospheric component in hybrid coupled models. The neural networks are constructed by optimizing sets of weights applied to the predictors, which are then transformed using a non-linear function (usually the hyperbolic tangent), and then further weighting functions are applied to provide estimates of the predictand values. The weights are optimised so that the squared errors in the estimates are minimised, as with linear regression. Because of the large numbers of model parameters involved, care has to be taken to avoid over-fitting.

7.4.2.2 Regression Models for Non-normally Distributed Data

Although linear regression assumes that the data being analysed are normally distributed, the procedure can be generalised to allow for predictands with alternative distributions. These generalised linear models (GLMs) are discussed in more detail in Section 7.4.3, where versions of GLMs for estimating probabilities are considered. However, there are forms of GLMs for data with a Poisson distribution that are suitable for modelling data that are recorded as counts, and these have been applied in seasonal forecasting of tropical cyclones. Versions are also available for data with a gamma distribution that would be suitable for forecasts of rainfall, but these have not been widely used.

A primary reason why linear regression becomes problematic when the predictands are not normally distributed is that the more extreme observations (for example the very wet years) have an undue influence on the regression parameters. While GLMs address this problem by making it possible to assume distributions

[13] Polynomial regression models are special cases of power regression, allowing only integer powers to be used.

that are more representative of the data, another alternative is to use regression models that are less sensitive to extreme values. There are two ways in which this sensitivity can be reduced. In robust regression, one option is to reset all errors (i.e. squared differences between the observed and the fitted value) exceeding a maximum value to this threshold. The procedure is not widely used. The second approach is to redefine how the errors are calculated: specifically, instead of squaring the errors, which tends to exaggerate the magnitude of large errors, the absolute errors can be used. This procedure is known as least absolute deviation (LAD) regression, and has been used in tropical cyclone forecasting, for example.

7.4.2.3 Ridge Regression

Ridge regression constitutes an attempt to address the problem of multicolinearity by placing constraints on the model parameters. In effect the procedure artificially inflates the variances of the predictors relative to their covariances, and thus underplays the effects of the inter-correlations when estimating the model regression coefficients. Ridging is used in the constructed analogue procedure, in which a least squares estimate of the spatial pattern of the most recently observed values of the predictands is obtained by weighting the patterns for all years in the historical data. As an example of a simple constructed analogue model, consider the problem of forecasting the December Niño3.4 index from the June value. Assume that the June and December values of the index are known for 1971–2000, and that the June 2001 value is available to make a forecast for December 2001. Weights would be assigned to the June values for 1971–2000 to estimate the June value for 2001. These same weights would then be applied to the December 1971–2000 values to construct a forecast for December 2001. Given that the number of weights to be calculated (30) is larger than the number of values being estimated (1), there is no unique solution to the weights, but the ridging helps to provide a stable solution.

7.4.2.4 Principal Components Regression

Principal components regression (PCR) improves on ridge regression by addressing some of the problems arising from both multicolinearity and multiplicity. The only difference between multiple regression and PCR is that in PCR the principal components of the predictors are used in the model instead of the original predictors themselves. Principal components are optimal summaries of large sets of data, obtained by defining sets of weights, or "loadings", that are applied to obtain a linear combination of the original data. They are ideally suited to the problem at hand, since they will reduce a large candidate pool of predictors to a much smaller number, while retaining much of the information in the original data. In addition, each of the principal components is uncorrelated with all the others, and so problems of multicolinearity are avoided. More complex versions of principal component analysis can be used in PCR that represent, for example, modes of variability that have an evolutionary component, and are discussed further in the next section.

In theory it is possible to expand a PCR equation into an equivalent multiple regression equation given the PCR coefficients and the loadings used to define the principal components. The coefficients of this expanded multiple regression have smaller error variance than if the coefficients had been estimated directly, because the negative effects of multicolinearity are usually associated with the higher order principal components that would generally be omitted from the analysis. However, the coefficients are biased, and so problems of interpretation remain. Despite these issues, and problems in determining the number of principal components to retain in the model, principal components regression is an attractive alternative to multiple linear regression.

7.4.2.5 Maximum Covariance Analysis, Canonical Correlation Analysis, and Redundancy Analysis

When making predictions for a number of different stations or gridpoints, principal components regression can be an inefficient procedure since separate models have to be constructed and tested for each location. In addition, if the predictands are inter-correlated, it is possible for predictions at one or more of the locations to be somewhat inconsistent with those at others because of different sampling errors in the estimated regression coefficients, or even in the selection of predictors, for models at neighbouring sites. There are various techniques that can be used to make predictions at a set of locations. These techniques include canonical correlation analysis (CCA), redundancy analysis, and maximum covariance analysis[14] (MCA). These techniques are widely used in spatial downscaling problems (Chapter 8).

The basic principle behind all of these techniques involves forecasting modes or spatial patterns of variability spanning across the region of interest rather than making forecasts for individual locations. In this context, a mode is akin to a weighted average[15] of the individual locations. More than one mode can be predicted, and the predictions for these modes are then superimposed to construct

[14] Maximum covariance analysis is frequently referred to as singular value decomposition (SVD) or SVD analysis. This nomenclature, however, is confusing because SVD is often used to perform other analyses, including multiple regression, principal components analysis, and CCA. Von Storch and Zwiers (1999) propose calling the technique maximum covariance analysis.

[15] More strictly, because the sum of the squares of the weights rather than the sum of the weights per se, is required to be unity, the modes are a "weighted sum" or a "linear combination".

forecasts for all locations. The modes are predicted using a second set of modes obtained from the predictors so that spatial patterns of variability in the predictors are used to predict spatial patterns in the predictands. If $\mathbf{U_X}$ and $\mathbf{U_Y}$ are the weights for the predictors and predictands, respectively, the modes, or new variables, are:

$$\mathbf{Z_X} = \mathbf{X}\mathbf{U_X}, \tag{7.19a}$$

$$\mathbf{Z_Y} = \mathbf{Y}\mathbf{U_Y}. \tag{7.19b}$$

Maps of the weights are frequently plotted to indicate the coupled spatial patterns. As an example, the first coupled mode (obtained using CCA) of September SSTs for the Indian Ocean and October–December precipitation over part of East Africa is shown in Fig. 7.3. The mode suggests that warming in the western tropical Indian Ocean with cooling in the eastern tropical Indian Ocean and far western Pacific (Fig. 7.3a) can be used to predict anomalously wet conditions over the bulk of Tanzania and Kenya (Fig. 7.3b). The opposite precipitation pattern would be predicted given a reversal of the anomalous zonal temperature gradient in the tropical Indian Ocean. The temporal variability of these modes is shown in Fig. 7.3c; the correlation between the modes is 0.706.

The differences between MCA, CCA, and redundancy analysis are in the properties of the weights that define the modes:

- In MCA each pair of modes has maximum covariance
- In CCA each pair of modes has maximum correlation
- In redundancy analysis the explained variance in the predictand modes is maximised

(Compare principal component analysis, in which the aim is to define a set of weights for either the predictors or the predictands that generate new variables with maximum variance.) For MCA, the covariance between the modes is:

$$\mathbf{C} = \mathbf{Z}_\mathbf{X}^T\mathbf{Z_Y}. \tag{7.20}$$

The covariance matrix $\mathbf{C}$ is a diagonal matrix with the diagonal elements defining the covariances of the coupled modes of predictors and predictands. Equation (7.20) can be written in terms of $\mathbf{X}$ and $\mathbf{Y}$ by substituting from Eq. (7.19):

$$\begin{aligned}\mathbf{C} &= \left(\mathbf{X}\mathbf{U_X}\right)^T \mathbf{Y}\mathbf{U_Y} \\ &= \mathbf{U}_\mathbf{X}^T\mathbf{X}^T\mathbf{Y}\mathbf{U_Y}\end{aligned} \quad . \tag{7.21}$$

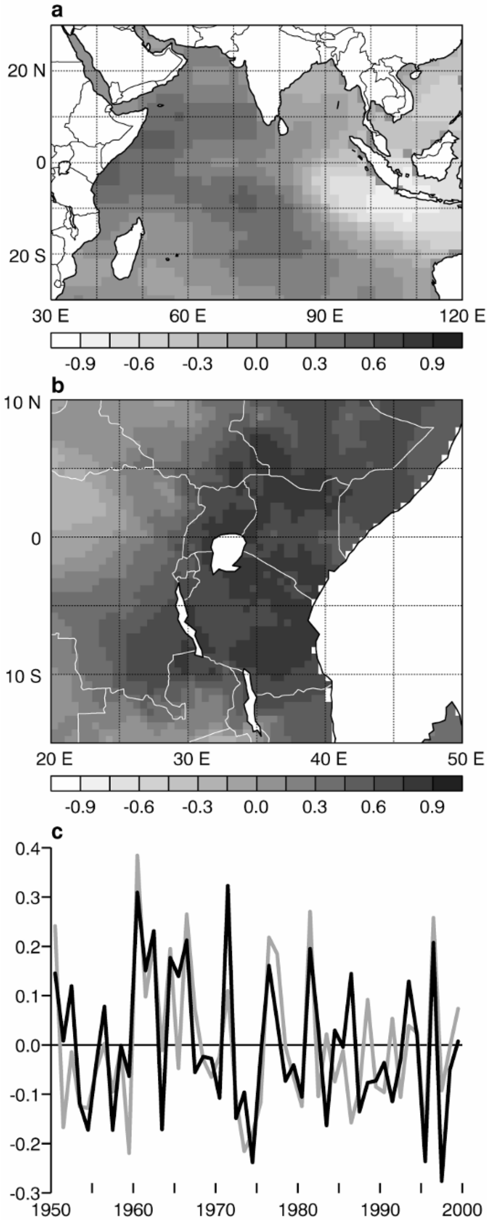

Fig. 7.3 Example of the first coupled mode of (a) September 1951–2000 sea surface temperatures for part of the Indian Ocean used to predict (b) October–December 1951–2000 precipitation over East Africa. Both datasets were pre-filtered by using only the first few principal components. The maps show the correlations between the original gridded data and the respective temporal scores (c) for the predictor (black) and predictand (grey) components of the first canonical coupled mode

$\mathbf{X}^T\mathbf{Y}$ is the covariance of $\mathbf{X}$ and $\mathbf{Y}$ (i.e. the covariance matrix of the original predictors and predictands) and so Eq. (7.21) can be rearranged to express this covariance matrix, $\mathbf{C_{XY}}$, in terms of the diagonal matrix $\mathbf{C}$, and two orthogonal matrices:

$$\mathbf{C}_{\mathbf{XY}} = \mathbf{U}_{\mathbf{X}}^{T}\mathbf{C}\mathbf{U}_{\mathbf{Y}}. \tag{7.22}$$

In other words, the weights $\mathbf{U_X}$ and $\mathbf{U_Y}$ that maximise the covariances between the spatial modes of predictors and predictands can be obtained from a singular value decomposition of the covariance matrix of the original predictors and predictands. Then, given a new set of predictors, $\mathbf{x}$, forecasts, $\hat{\mathbf{y}}$, can be generated:

$$\hat{\mathbf{y}} = \mathbf{x}\mathbf{U}_{\mathbf{X}}\mathbf{\Sigma}_{\mathbf{X}}^{-1}\mathbf{C}\mathbf{U}_{\mathbf{Y}}^{T}, \tag{7.23}$$

where $\mathbf{\Sigma_X}$ is a diagonal matrix containing the variances of the $\mathbf{Z_X}$. Only those coupled modes that explain a large proportion of the total variance are used in the prediction, and so typically only the first few coupled modes are retained. Effectively, the smaller diagonal elements of the matrix $\mathbf{C}$ effectively are set to zero.

However, Eq. (7.23) does not provide least-squares estimates of the predictands, and so MCA is not regularly used in seasonal climate forecasting. Instead MCA is more useful in identifying coupled modes of, for example, SST fields and rainfall that may provide a basis for seasonal forecasting. A much more commonly used variant of MCA in prediction problems is CCA, which aims to identify alternative sets of weights, $\mathbf{V_X}$ and $\mathbf{V_Y}$,[16] that maximise the correlations rather than the covariances between the modes of variability. In CCA the modes defined in Eq. (7.19) are first standardised, replacing $\mathbf{U_X}$ and $\mathbf{U_Y}$ by $\mathbf{V_X}$ and $\mathbf{V_Y}$, respectively, so that $\mathbf{C}$ in Eq. (7.20) becomes a squared correlation matrix, $\mathbf{R}$. Predictions, given a new set of predictors, are then given by:

$$\hat{\mathbf{y}} = \mathbf{x}\mathbf{V}_{\mathbf{X}}\mathbf{R}\mathbf{V}_{\mathbf{Y}}^{-1}. \tag{7.24}$$

In practical terms, CCA identifies linear combinations of predictors that can successfully predict linear combinations of the predictands, regardless of how much of the total variance either linear combination explains. Consequently, there is a danger of identifying well-correlated modes of variability that do not explain much of the total variability. Although the objective in MCA of maximising the covariances rather than the correlations between the modes may seem more pertinent, MCA is also problematic in that the covariances are maximised in part by the variances of the modes for the predictors, and so it is possible that the total

[16] Note that $\mathbf{V_X}$ and $\mathbf{V_Y}$ are not orthogonal matrices, whereas $\mathbf{U_X}$ and $\mathbf{U_Y}$ are.

explained variance of the predictands is low. Further, both methods are subject to interpretation problems, and neither approach is likely to identify robust and easily interpretable modes of variability. Redundancy analysis is a third option that deserves further attention. Redundancy analysis replaces $\mathbf{Z_X}$ in Eq. (7.19a) with the standardised values, and thus seeks to maximise the explained variance in the predictands without necessarily using the largest modes of the variability in the predictors. Redundancy analysis can thus be seen as intermediate between CCA and MCA. In practice, differences in the results of the various techniques are usually minimal.

In most applications of MCA and CCA in the climate literature, the observations and forecasts are pre-filtered by using a subset of the principal components of the data. While the pre-filtering simplifies the solution of the CCA or MCA, the computational gain is lost through having to calculate the principal components. Instead, the main advantage of the pre-filtering is that the noise levels in both the forecasts and the observations are reduced, and so the chances of finding spurious relationships are decreased. This advantage is likely to be greater for CCA than for MCA because the former does not require the coupled modes to represent large proportions of the total variance of the original data.

7.4.2.6 Other Principal Component Analysis-related Techniques

As mentioned in Section 7.4.2.4, principal components can be useful as predictors. There is a hierarchy of sophisticated ways in which these components can be defined. In the simplest formulation, the principal components are defined using a set of predictor variables all of which represent measurements synchronous with each other. Prediction using principal components of SSTs at various locations, but all measured at the same time, would be an example. This form of principal components regression is discussed in Section 7.4.2.4.

If the predictors are measured at a number of different lags, the principal components become "extended" empirical orthogonal functions (EOFs),[17] whose computation is equivalent to that of multi-channel singular spectrum analysis. For example, SSTs for a set of locations measured at a number of different times of the year are sometimes used to predict future SSTs. If a single predictor is used in this context so that the principal components are calculated only from the auto-correlation (or auto-covariance) of this series, the technique is known as singular spectrum analysis (SSA). Although SSA has not been used widely in seasonal

[17] Empirical orthogonal functions are the loadings that define the principal components. Although some authors have drawn a distinction between principal component analysis and empirical orthogonal function analysis based on the normalization of the eigenvectors (Richman 1986), this distinction is not widely adhered to and the two are in most cases synonymous (von Storch and Zwiers 1999; Joliffe 2007).

climate forecasting, it has been used in an attempt to identify the predictable component of the Indian monsoon variability. Similarly, complex EOFs have been used in predictability studies, but have not been widely applied in seasonal climate forecasting. Complex EOF analysis, sometimes called Hilbert singular decomposition, involves advancing all oscillatory components of any wavelength in the data by 90°, and including these as imaginary components in a principal component analysis. The procedure allows lags to be identified in modes of variability.

Principal oscillation pattern (POP) analysis is fundamentally different to the techniques described above. It performs an eigenvalue decomposition of the matrix of first order autoregressive (AR-1) coefficients, and hence identifies optimal multivariate AR-1 models that can be used for prediction purposes. POP analysis has similar objectives to complex EOF analysis in seeking to identify evolutionary modes of variability, but has been more widely used than the latter in seasonal prediction. Linear inverse modelling is a version of POP analysis.

7.4.2.7 Autoregressive Models and Optimal Climate Normals

Linear inverse modelling and POP analysis are sophisticated versions of simpler models known as autoregressive models. Autoregressive models are mathematically the same as linear regression models except that the predictors are the same variable as the predictand, only measured at different lags. So, for example, if the NIÑO3.4 index is forecasted with a regression model using only earlier values of the index, then this model would be autoregressive. The best known example of such a model is the CLIPER (CLImatology and PERsistence) model that has been used to forecast the ENSO phase using lagged and autoregressive relationships. The basic principle involved is that some variables, such as SSTs, change slowly, and so recent evolution can be used as a guide to future values. The name CLIPER implies that future values are predicted using a combination of: the seasonal mean value (climatology) towards which the value of the predictand is expected to drift at increasingly long lead-times, and; the most recently observed anomalies, that are expected to decay[18] only slowly (persist).

A special case of using persistence and climatology as a forecast is that of optimal climate normals (OCNs). In most cases of seasonal climate forecasting, a forecast is made by projecting the most recently observed climate state into the future, i.e. from the previous day (or month or perhaps season) into a coming season. However, with OCN a forecast is made under the assumption that a good guide to the climate conditions for the target season are the conditions that have been observed for the same season over the last few years. The forecast for the

[18] It is possible, such as when forecasting ENSO anomalies at certain times of the year, for anomalies to grow in a CLIPER model (Knaff and Landsea 1997), but such cases are unusual.

coming season is then simply the average of the last few years, and the objective is to identify the number of years to average to give the best forecast. The idea is that the 30-year standard climatological period can be improved upon in some cases when there is low-frequency variability (e.g. inter-decadal variability or trend) in the climate. Using OCNs is sometimes a useful option in areas with inherently low seasonal predictability.

7.4.3 Probabilistic Statistical Prediction Techniques

Rather than trying to estimate a best-guess forecast value and then accounting for the uncertainty in this forecast, there are a number of statistical techniques that can be used to estimate forecast probabilities directly. Some of these methods are alternative versions of the regression models mentioned in Section 7.4.2, and are described in further detail in Section 7.4.3.1, while others are based on classification problems, and are discussed in Section 7.4.3.2. In Section 7.4.3.1 statistical procedures that are similar to ensemble forecasting are described.

7.4.3.1 Generalised Linear Models

Although multiple regression can be used to estimate probabilities as the dependent variable, this is not generally advised because there is no constraint that the estimated probability is between zero and one, and because the distributional assumptions of the procedure are violated (Wilks 2005). Instead a variety of models that are ultimately based on linear regression are available. Although these generalised linear models are closely related to linear regression they are discussed separately in this section.

Generalised linear models are based on the standard linear regression equation:

$$\eta = \boldsymbol{\beta}^T \mathbf{x}\,, \tag{7.25}$$

where $\boldsymbol{\beta}$ is the set of regression parameters, and $\mathbf{x}$ is the set of predictors. The linear predictor η is related to the predictand, which in this case is a Bernoulli variable with mean $\hat{p}$. via a link function. The three most commonly used link functions for Bernoulli variables are:

$$\eta = \log\left[\frac{\hat{p}}{1-\hat{p}}\right], \tag{7.26a}$$

$$\eta = \Phi^{-1}\left[\hat{p}\right], \tag{7.26b}$$

$$\eta = \log\left[-\log\left[1-\hat{p}\right]\right], \tag{7.26c}$$

where Φ^{-1} is the inverse normal distribution function. These link functions are known as the logit, probit, and complementary log-log functions, respectively. In practice, the differences between the three are minimal, but the logistic link is the most widely used, and easiest to compute.

Instead of training the model using observed rainfall or temperatures, for example, the predictand has to be categorised into one of two groups. For example, in Fig. 7.1a December 1950–2000 values of the NIÑO3.4 index are shown as anomalies and plotted against the June values. The regression line and the scatter of values imply a reasonably strong relationship between the phase of ENSO in June and that 6 months later. In Fig. 7.4b, all the values of the December NIÑO3.4 index that exceed the upper quartile are converted to a value of 1, and all the values less than the upper quartile to a value of 0. The values on the x-axis (the June NIÑO3.4 index) are left unchanged. Rather than trying to fit a straight line to the data points, an S-shaped curve is used. Eqs. (7.26a–c) are different ways of converting a straight line to an S-shaped curve that ranges between 0 as a minimum, and 1 as a maximum.

In this example of a generalised linear model, observations are listed either as 0s and 1s, and the fitted curve is interpreted as providing an estimate of the probability that future values will exceed the threshold used to define the categories (i.e. the probability that the December NIÑO3.4 index will exceed the upper

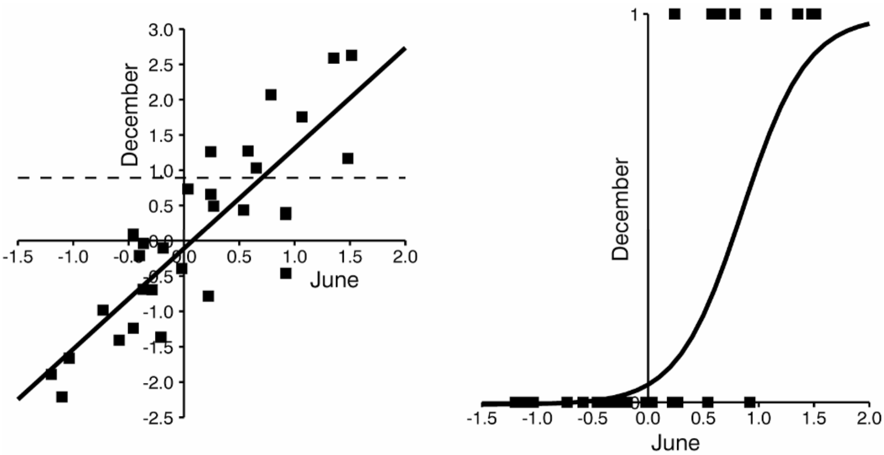

Fig. 7.4 Example of (a) a linear regression model and (b) a generalised linear regression model. June values of the Niño3.4 index are used to predict December 1971–2000 values. The dashed horizontal line represents the upper quartile of December values of the index

quartile). The limitation to only two categories can be too restrictive, but it is possible to further divide the categories either by nesting models, or by simultaneous fitting of parallel models.

The forms of generalised linear models described above, resolve issues related to data distribution assumptions, of indicating forecast uncertainty, and, to some extent, that of linearity, but do not address the problems of multiplicity and multicolinearity. The latter two problems can be addressed in similar ways to that for linear regression, e.g. by using principal components as predictors.

7.4.3.2 Classification Procedures

Classification procedures have been used in seasonal climate forecasting more extensively than generalised linear models. As with generalised linear models, the observations are assigned to one of two or more categories, and then probabilities are calculated that a new observation will be within each of the categories given new values of the predictors. An important distinction, however, is that categories are nominal in classification procedures, so that if there are three or more, the procedures do not know, for example, that they are ordered as below-normal, near-normal, and above-normal. In most cases of seasonal climate forecasting the fact that the categories are nominal in classification procedures is likely to be a disadvantage because relationships between predictors and predictands are most often likely to be monotonic.

Discriminant analysis is the most widely used classification procedure in seasonal climate forecasting. The values of the predictand are assigned to one of the categories, and the mean values of the predictors are then calculated for each category separately. If the predictors have good discriminatory power then the differences in the means of the predictors between the various categories will be large. For example, if seasonal rainfall is strongly influenced by the ENSO phenomenon, then the difference in the average value of the NIÑO3.4 index when rainfall is above-normal compared to when rainfall is below-normal will be large. Given the covariances of the predictors in each category the probability that a new observation will be in each category can be calculated from the new values of the predictors, and from knowledge about the prior probabilities of each category. Mathematically, it is simpler to assume that the covariances are the same for each category, and a linear classification can be defined to identify the most likely category. If this assumption of equal covariance is dropped, the classification function becomes quadratic. The quadratic function only performs noticeably better than the linear function when the differences in covariance are marked.

Canonical variate analysis has had limited application in seasonal climate forecasting, but it has been used in predicting the phase of the ENSO phenomenon. The technique is similar to discriminant analysis, but has some similarities to canonical correlation analysis as well. Just as canonical correlation analysis identifies optimal linear combinations of the predictors to maximise correlations with linear combinations of the predictands, canonical variate analysis seeks optimal linear combinations of the predictors, but in this case to maximise the discrimination between the categories. The discrimination is defined by the ratio of between-group to total variance.

An example is provided in Fig. 7.5, where canonical variates are computed using monthly NIÑO3.4 indices from January–November to predict the ENSO phase for the following December. Three phases are defined based on the outer quartiles of the December value of the index, and are represented by the different symbols: the open circles represent years in which the December NIÑO3.4 index was below the lower quartile (i.e. La Niña events), the open triangles years in which the index was above the upper quartile (i.e. El Niño events), and the open squares years in which the index was within the inter-quartile range (i.e. neutral events). The *x*-axis represents the first canonical variate (a linear combination of the NIÑO3.4 indices for January–November), which maximises the distances between the mean values of canonical variate scores for the three categories, as represented by the solid symbols. This canonical variate therefore maximises the distances along the *x*-axis between the three solid symbols. The first canonical

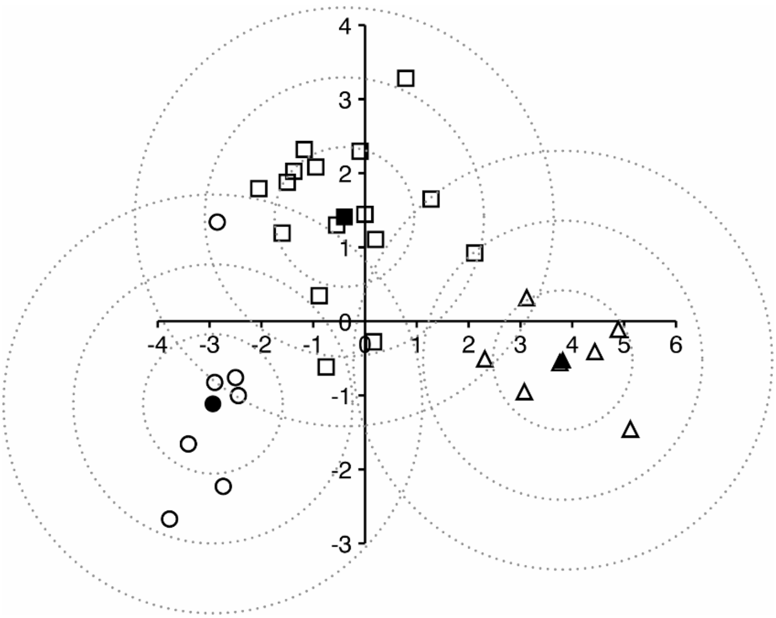

Fig. 7.5 Example of a canonical variate analysis model. The x-axis represents the first canonical variate of monthly NIÑO3.4 indices from January–November, and the y-axis the second. The hollow symbols represent observed scores on the canonical variates for 1971–2000, and the solid symbols the corresponding mean values. The circles represent years in which the December NIÑO3.4 index was below the lower quartile, the triangles years in which the index was above the upper quartile, and squares years in which the index was within the inter-quartile range. The large dashed circles represent distances of one standard deviation

variate successfully distinguishes the three categories, but is most effective in identifying the El Niño events (represented by the triangles). The second canonical variate maximises the distances between the categories along the *y*-axis, and helps to distinguish the La Niña events (circles) from the neutral events (squares). The dashed circles indicate distances in multiples of one standard deviation from the category means, (assuming that the variances in all three categories are equal), and can be used to visualise in which category a new observation is most likely to occur.

Classification procedures address a number of the problems listed in Section 7.4.1. Because the predictands are categorised in both discriminant analysis and canonical variate analysis, no assumptions are made about their distribution. However, it is assumed that the predictors are normally distributed, and linear discriminant analysis is sensitive to violations of this assumption. Quadratic analysis is more robust, except when the data are highly skewed. As with the forms of generalised linear models discussed in Section 7.4.3.1, multiplicity and multi-colinearity remain as problems, but can be addressed in similar ways to that for linear regression, e.g. by using principal components as predictors.

7.4.3.3 Analogue Procedures

Analogue procedures have some similarity to classification procedures, but are listed separately because of a number of important differences from discriminant analysis and canonical variate analysis, and because of a wide flexibility in how the analogues can be used to make a prediction. The essential step is to identify years from the historical records in which the states of the predictors were similar to the states for the current forecast. Some index of similarity (or of dissimilarity) is used to calculate how closely current conditions resemble previously observed conditions. A frequently used measure of similarity is the Mahalanobis distance, which is similar to the squared distance, but which compensates for correlations between the predictors.

The distinction between this step of identifying similar years and classification is that the similarity of individual years, rather than of the mean of a predefined category of years, is investigated. However, in some of the simpler analogue procedures, often, but not exclusively, used when there is only one predictor, the predictor(s) is (are) classified into one of a set of predefined classes, and other years within this category are treated as analogues. A widely used example of this classification step in an analogue procedure is the Southern Oscillation phase system, in which the current state and recent evolution of the Southern Oscillation Index are classified into one of the five categories rapidly falling, rapidly rising, consistently positive, consistently negative, and consistently near-zero.

Once analogue years have been identified, a forecast is constructed using the observed values for these selected years. The forecast can be constructed in a number of ways, the simplest of which is to use the mean value, although normally the variability within the analogue years would also be considered to provide some indication of the uncertainty in the forecast. If the forecast sample is sufficiently large, the probability that the predictand will exceed a threshold value could be obtained by counting the proportion of times it was exceeded in the analogue sample (although errors in calculating this proportion are likely to be large). A more reliable approach would be to fit an appropriate distribution to the analogue and to derive a forecast from this fitted distribution. The problem is essentially identical to that of constructing a forecast from an ensemble of GCM predictions. Each analogue year can be treated as an ensemble member. Procedures for obtaining a forecast from an ensemble are discussed in Chapter 8 (Section 8.5.2).

A special case of an analogue procedure is the constructed analogue, which combines *all* previous cases. The procedure is a form of ridge regression, which is discussed in further detail in Section 7.4.2.

Part III
Calibration and Assessment of Seasonal Climate Forecasts

Chapter 8
From Dynamical Model Predictions to Seasonal Climate Forecasts

Simon J. Mason

Producing a seasonal climate forecast from a dynamical model involves a great deal more than simply running the model and viewing the results. The first problem is to decide which dynamical model(s) should be run given the practical constraints of computing resources. In this chapter the pros and cons of using the more computationally intensive fully coupled models compared to atmosphere-only models are discussed. After running a dynamical model, regardless of its complexity, corrections need to be made for systematic errors because the model's climatology and that of the observed climate are invariably different. Some simple procedures for correcting these systematic errors are assessed, but more sophisticated methods are advisable to adjust for spatial displacements of the model climate. Since the model predictions represent large spatial averages, and generally are presented as seasonal averages, downscaling may be required to make the forecast relevant for specific locations, and to provide more detailed information about the statistics of weather within the season. Commonly used spatial and temporal downscaling procedures are described. Some procedures for describing the uncertainty in the forecast are discussed (further details are provided in Chapter 9). Evidence is presented that forecasts can be improved by combining outputs from different models. Finally, the reliability of the forecast needs to be determined by verification of a historical set of forecasts. Verification procedures are discussed in Chapter 10.

8.1 Introduction

In Chapter 7, the procedures for constructing a statistical model for generating seasonal climate forecasts were described. Although dynamical models have been described in detail in Chapters 3 and 6, the procedures for using such models to

Simon J. Mason
International Research Institute for Climate and Society

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

produce forecasts are far from straightforward. These procedures are described in this chapter, beginning with a discussion of alternative methods of running the dynamical models (Section 8.2), followed by explanations of how to correct for systematic errors in the outputs of the models (Section 8.3) and to tune the predictions so that they become valid for specific locations (Section 8.4). Procedures for obtaining a probabilistic forecast from an ensemble of model predictions are then discussed in Section 8.5, and finally methods for combining predictions from different models are outlined (Section 8.6).

8.2 One-Tiered and Two-Tiered Forecasting

As discussed in Chapters 3 and 4, seasonal climate forecasting is premised upon feedbacks between the atmosphere and boundary conditions at and near the earth's surface. When producing seasonal climate forecasts using general circulation models (GCMs), there are a number of fundamentally different ways of modelling these interactions between the atmosphere and the lower boundary. These approaches range in complexity: in the simplest case, only the atmosphere is forecast using dynamical models while the boundary conditions are specified by persisting the most recently observed values; in the most complex case the atmosphere together with all the various components of the lower boundary thought to be of importance to atmospheric variability at seasonal timescales are modelled as fully interacting. These two extremes, as well as some intermediate options, are discussed in further detail below (Section 8.2.1), and arguments for and against the various levels of complexity in the modelling are considered in Section 8.2.2.

8.2.1 One- and Two-Tiered Forecasting Designs

The simplest method of dynamically modelling the climate system at seasonal timescales is to model only the atmosphere while specifying values for the various parameters of interest in the lower boundary. If forecasts of the atmosphere are to be made, future values for the boundary conditions have to be specified, and so these values have to be forecast prior to integrating the atmospheric model. A "two-tiered" forecast is thus required: forecasts of the boundary conditions are made first, followed by forecasts of the atmosphere with the forecast boundary conditions prescribed (Bengtsson et al. 1993).

Two-tiered forecasting systems invariably involve a system in which sea surface temperatures (SSTs) are forecast first, while procedures for forecasting the other components of the atmospheric boundary are not explicitly mentioned. Forecasts of land-surface conditions, for example, generally are produced by coupling a land-surface model to an atmospheric model, even in two-tiered systems in which SSTs are prescribed. Forecasts of SSTs have involved methods from as simple as persistence of the latest observed conditions, through statistical forecasts and partial-ocean hybrid model forecasts, to basin forecasts from fully coupled models, or some combination of the above. Forecasts of land-surface conditions, including of the biosphere, remain relatively primitive compared to forecasts of the sea surface, primarily because of a paucity of observational data, and there are even substantial problems using the best estimates of the latest observed conditions (Anderson and Ploshay 2000).

Two-tiered approaches allow the boundary to influence the atmospheric variability over the period of model integration, but do not permit the atmosphere to feedback to the boundary. Rather than specifying the boundary conditions at the ocean surface and allowing no feedback from the atmosphere, highly simplified models of the oceans can be coupled to the atmospheric model. Although fully non-linear ocean models coupled to simplified atmospheric models, known as hybrid models (Barnett et al. 1993), have been popular, their counterpart models have not been widely used in seasonal forecasting of the atmosphere. Such slab ocean models would allow two-way heat fluxes between the atmosphere and ocean, but do not involve ocean circulation. This restricted feedback of the atmosphere to the ocean may have advantages over the standard two-tiered approaches, and such models deserve further attention.

The most complex method of modelling the climate system at seasonal timescales is to model all components of the climate system thought to be relevant at seasonal timescales. Operational examples of such models involve separate models for the atmosphere and ocean that are run synchronously and interactively. Such “fully-coupled” models generate forecasts of the atmosphere and of the boundary conditions simultaneously, and so sometimes are referred to as “one-tiered” forecasting systems.

8.2.2 Advantages of One- and Two-Tiered Forecasting Designs

One-tiered forecasting systems, or fully coupled models, are widely acknowledged to represent the state-of-the-art in seasonal climate forecasting. However, comprehensive comparisons of one and two-tiered systems are lacking (see Graham et al. 2005 and Guérémy et al. 2005 for some preliminary results), and regardless of relative performances, there are advantages to two-tiered approaches that are likely to contribute to their continued use for the next several years at least. Some of these advantages are outlined in Sections 8.2.2.2–8.2.2.4 after a brief summary of the advantages of one-tiered systems (Section 8.2.2.1).

8.2.2.1 Advantages of One-Tiered Forecasting Designs

One-tiered forecasting systems represent the most comprehensive attempt to incorporate all the components of the climate system thought to be relevant for understanding atmospheric variability at seasonal to interannual timescales. Because they allow for feedbacks between the atmosphere and the other components of the climate system, coupled models should, theoretically, provide the most realistic representation of how the real climate system operates, and hence should be able to generate better forecasts than their two-tiered counterparts. An implicit assumption in two-tiered systems is that the atmosphere responds to SST forcing, but does not in turn affect the oceans. As indicated in Chapters 4 and 6, strong feedback between the ocean and the atmosphere occurs within the equatorial Pacific Ocean, for example, while atmospheric influence on tropical Indian Ocean variability appears to be stronger than the influence of the ocean on the atmosphere. Similarly, in the extra-tropics, pioneering research on ocean-atmosphere interaction over the North Pacific indicated that the ocean variability is more a response to atmospheric variability than vice versa.

In a two-tiered system, where the atmosphere is uncoupled from the ocean, unrealistic forcing of the model atmosphere can occur. For example, Indian monsoon rainfall in most uncoupled models is positively correlated to tropical Indian Ocean SSTs because of higher moisture fluxes, but in coupled models, and in the real world, negative correlations are evident because the ocean surface heats in response to changes in the trade winds (Wu and Kirtman 2005). The imposed forcing in two-tiered systems can therefore result in incorrect simulations, whereas the coupling permitted in one-tiered designs should result in a more realistic representation of observed climate variability. Although coupled models do not currently perform much better because of moisture flux problems (Wu et al. 2006), improvements in the model physics should result in more realistic simulations, whereas improvements in the physics of an uncoupled atmospheric model will not necessarily resolve the problem.

8.2.2.2 Computational Advantages of Two-Tiered Forecasting Designs

Fully coupled models require huge computational resources, and so currently are used for operational forecasting only at some of the so-called Global-Producing Centres (GPCs). Because of the computational costs, forecasts are compromised, either in the resolution of the model atmosphere and/or ocean, the ensemble size, the lead-time, the frequency of forecast production, and/or the generation of retrospective forecasts used for assessing forecast performance and calibrating for model errors. For example, of the seven models that constituted part of the DEMETER experiment (Palmer et al. 2004), only three have hindcasts extending back more than 40 years, and these for only four initialization dates during the year and for a sufficient number of ensemble members to estimate the models' respective mean responses only in the tropics. Alternative savings involve coupling a global atmospheric model to a single-basin ocean-model, and prescribing sea temperatures elsewhere (Ineson and Davey 1997).

The computational advantages of two-tiered forecasting systems could permit the integration of the atmospheric model at higher resolutions than are possible when the same model is run in one-tiered mode, or the generation of a larger ensemble. In countries where only moderately powerful computing resources are available, the computational advantages enable two-tiered dynamical seasonal forecasts to be generated locally. These computational advantages are enhanced by a relatively weak improvement in forecast quality with increased spatial resolution in two-tiered systems compared to their one-tiered counterparts. Apparently the coupling of the ocean and atmosphere is modelled most effectively at high resolutions, whereas if the atmospheric model is uncoupled many of the benefits of improved resolution are lost.

Additional computational advantages can be achieved if no attempt is made to assimilate observed data into the atmospheric model. While there is some resultant loss of predictability from initial conditions in the first few weeks of the forecast, the loss of skill at longer lead-times is considered minimal, and is partly offset by avoiding problems associated with model drift (Chapter 6). The computational costs involved in data assimilation are substantial, and are an essential component of ocean forecasting (see Chapter 5), and so assimilation is dispensable only if no ocean model is to be run.

8.2.2.3 Atmospheric Predictions from Improved Sea Surface Temperature Predictions in Two-Tiered Forecasting Designs

The quality of seasonal climate forecasts of the atmosphere is intricately related to the quality of the forecasts of the lower boundary forcing, particularly of SSTs. If coupled model forecasts of the lower boundary can be improved by using other forecasting methods, it may be possible to improve on the atmospheric forecasts by using these superior boundary forcings in a two-tiered scheme. For example, since forecasts of persisted SST anomalies are difficult to outperform at lead-times of less than about 3 months, prescribing SST anomalies at short lead-times may provide improved skill in two-tiered atmospheric predictions. While fully coupled models can outperform two-tiered systems in which SSTs are prescribed from simple statistical models, the two-tiered systems may perform at least equally as well as fully coupled systems when more skilful SST forecasts are used. More detailed research on the comparative performances of one- and two-tiered systems is required.

8.2.2.4 Research Value of Integrations with Controlled Boundary Conditions

Apart from the benefits of two-tiered forecasts in an operational setting, atmospheric GCM integrations uncoupled to ocean models can be of considerable research value. Some of the more valuable examples of such research are discussed in this section.

Atmospheric GCMs forced with observed SSTs have been analysed extensively. Such experiments attempt to provide estimates of the potential predictability (an indication of the upper limit of predictive skill) of the climate at seasonal and longer timescales. Typically estimates of potential predictability involve comparing the variability in the simulated atmospheric responses across different ensemble members (intra-ensemble variability) with the inter-annual variability of the ensemble mean to obtain an estimate of the contribution of the SST forcing to the total variability: if the intra-ensemble variability is small compared to the interannual, then the SSTs are evidently constraining the (model's) atmospheric variability, implying that there is predictability. Alternatively, if ensemble size is small, a more reliable approach may be to compare the interannual variability of the simulated atmosphere when forced with observed as against climatological SSTs. Other strategies include, for example, comparing the forecast distributions to the climatological distribution, or examining the distribution of the proportion of ensemble members exceeding the climatological median. However, all strategies are based on estimating how much of the atmospheric variability is forced, and how much is free internal variability. Detailed investigations of the potential predictability of the atmosphere were conducted as part of the PRediction Of climate Variations On Seasonal to inter-annual Time-scales (PROVOST; Branković and Palmer 2000; Palmer et al. 2000), and Dynamical Seasonal Prediction (DSP; Shukla et al. 2000) projects.

Differences in the skill of simulating observed atmospheric variability when a model is forced using persisted instead of observed SST anomalies can be used to diagnose the loss of predictability that results from having imperfect SST forecasts. In the Sahel, for example, where rainfall variability is strongly affected by SSTs in the tropical Atlantic Ocean, the weak persistence of SSTs from 1 month to the next effects poor forecast skill of seasonal rainfall over the region, but skill increases markedly with decreasing lead-time (Ward 1998).

Alternative experiments have considered the effects of prescribing SSTs in only one (or occasionally two) of the three main ocean basins, or in specific areas thought to have important influences on atmospheric variability. Such experiments are valuable in diagnosing model systematic errors and also forecast biases that may result from using incomplete forecasts of SSTs in operational settings. However, because of their artificial nature, coupled with the fact that the total oceanic impact on the atmosphere may not be a simple linear combination of the individual oceanic impacts, they cannot adequately provide answers to questions concerned with the influence of SSTs in specific areas on the global (or regional) atmosphere. Other problems with these kinds of experiments result from the creation of artificial SST gradients at the edges of the domain of perturbed temperatures, even when the temperatures are reduced to climatology smoothly.

8.3 Systematic Model Error Correction

Regardless of how seasonal climate forecasts are made using atmospheric GCMs, substantial differences between the observed and model climates invariably are evident, and need to be corrected in order to provide reasonable forecasts. Definitions of various types of systematic error are provided in Section 8.3.1. Statistical tests for identifying errors in model output are detailed in Section 8.3.2, and are followed by a critique of commonly used methods for correcting for these errors (Section 8.3.3). Discussion on the correction of spatial errors in model output is provided in Section 8.3.4.

8.3.1 Systematic Model Errors

Systematic errors refer to any difference between the observed and the model climatology (implied definitions in the literature vary). The simplest form of systematic error is the mean bias: more generally, the central tendency of the model climatology differs from that for the observations. An example is shown in Fig. 8.1a, which compares observed[1] with simulated June–August precipitation rates for the 50-year period 1951–2000 averaged over a large area of eastern Africa (10°N–10°S, 3050°E). The precipitation was simulated using the ECHAM 4.5 model (Roeckner et al. 1996) at a resolution of about 2.8° and forced with observed SSTs, and the statistics were obtained using 24 ensemble members. The graph shows the frequencies of average precipitation rates over the 3-month period, and clearly indicates a bias in the model: simulated rates are consistently too high. This bias in the mean precipitation rate is known as an unconditional bias because the model rate is too high regardless of the actual simulated (or forecast) rate.

As well as indicating a mean bias, Fig. 8.1a indicates that the variance of the simulated precipitation rates is larger than the observed variance. Variance biases can occur even when the mean bias is minimal, as shown in Fig. 8.1b, which shows precipitation rates for March–May instead of June–August. Variance biases

[1] The New et al. (2000) gridded rainfall data were used. These data are based on station observations interpolated to a grid.

are also known as conditional biases because the model anomalies are consistently too strong (weak) when the model variance is larger (smaller) than the observed variance. Systematic errors in reproducing the shape of the climatological distribution can also occur: in Fig. 8.1c, the model's mean and variance are too high, while the skewness is too low. This example is for June–August precipitation averaged over part of southern Africa (20–30°S, 15°–25°E).

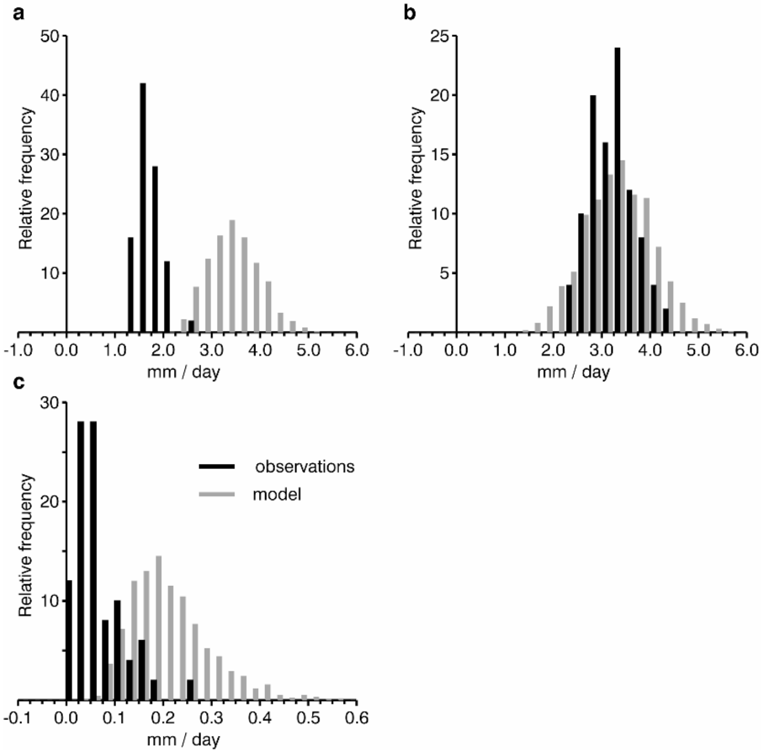

Fig. 8.1 Example of model systematic errors: (a) area-averaged June–August 1951–2000 observed (black) and simulated (grey) daily precipitation intensities over eastern Africa showing a mean bias; (b) area-averaged March–May 1951–2000 observed (black) and simulated (grey) daily precipitation intensities over eastern Africa showing a variance bias; and (c) area-averaged June–August 1951–2000 observed (black) and simulated (grey) daily precipitation intensities over southern Africa showing mean, variance, and shape biases

Any differences between the observed and model climatologies are symptomatic of differences in behaviour of the real and model atmospheres. However, these differences should be distinguished from predictive errors, which relate to differences in the observed and simulated/forecast climate for specific cases. Predictive errors relate to the skill of the model forecasts, and are not necessarily symptomatic of systematic errors. In the absence of any inherent predictability the individual forecasts will not generally correspond well with the observations, but the model climatology may be realistic.

8.3.2 Detecting Systematic Model Errors

Standard error statistics, such as the (root) mean-squared error and mean absolute error, measure differences between paired observations and simulations/forecasts. As a result, these metrics do not distinguish systematic model errors from predictive errors. Ideally these two forms of error should be distinguished. In the following sections, selected tests for systematic errors are described. A summary of the tests is presented in Table 8.1, where a few additional tests that are not discussed in the following text are mentioned. [See Sheskin (2007) for further details.] A selection of tests for predictive skill is provided in Chapter 10.

Table 8.1 Statistical tests, and respective distributional assumptions, for identifying systematic errors. All tests assume independence of the samples. Additional details of these tests can be obtained from Conover (2001) and Sheskin (2007)

Systematic Error	Test	Additional assumptions
All	Kolmogorov-Smirnov	
	Fisz-Cramér-von Mises	
	Relative entropy	
Central-tendency		
Mean	Student's *t*	Equal variance; normality
Median	Mann-Whitney *U*	Equal variance; similar shape
	Median	
Spread	*F*	Normality
	Siegel Tukey, David's, Mood's	Equal central-tendency; symmetry
	Moses	Similar shape

The standard test for systematic model errors is the two-sample Kolmogorov-Smirnov test, which compares the cumulative distributions derived from the model and the observed climatologies.[2] The test compares the maximum vertical difference between these two empirical distributions, D, against a null distribution for the statistic; if the maximum vertical distance is large, the two distributions are likely to be different, and so the model climatology does not match that for the observations. The null distribution for D, and for all the other statistics discussed in this section, depends upon the number of cases used to construct the empirical cumulative distributions, and so depends upon the number of years and the ensemble size. Systematic errors can be identified more robustly given large numbers of cases.

The two-sample Kolmogorov-Smirnov test does not distinguish between different forms of systematic error. Separate tests are available for identifying mean- and variance biases, while biases in skewness and higher order moments (collectively referred to as errors in the shape of the distribution) are not widely

[2] Alternatives include the Fisz-Cramér-von Mises test (the integral of the squared differences between the two cumulative distributions) and relative entropy (Elmore 2005).

used. Mean biases are commonly identified using Student's t-test, which compares the differences in the climatological means. The test is highly sensitive to distributional assumptions (the observed and model climatologies should both be Gaussian), and so alternative tests are required that are not sensitive to these assumptions. The alternative tests compare differences in medians rather than means, since the median is not strongly influenced by the presence of a few extreme values, and so they test for a bias in the central tendency rather than strictly testing for a mean bias. The Mann-Whitney U-test is the most frequently used alternative to the t-test. The U-test effectively calculates the probability that a randomly sampled observation is larger (or smaller) than a randomly sampled forecast. This probability should be 0.5 if there is no bias in the central tendency of the model climatology. Strictly, the U-test should not be used if there is a variance bias, in which case the median test is preferable. The median test calculates the proportion of observations (or simulations) above the pooled median, and is free of any assumptions about other forms of systematic error. Again the proportion should be 0.5 if there is no mean bias.

Tests for variance bias (or, more generally, dispersion bias) are numerous. The most commonly used is the F-test, which compares the ratio of the variances of the observations and simulations to Fisher's F distribution. The ratio should be 1.0 if there is no variance bias, but the test is highly sensitive to distributional assumptions, and should probably be used infrequently. Unfortunately, there is no obvious alternative test to use; there are of the order of 100 candidate tests, but virtually all of them carry some distributional assumptions. A Moses-type test, which is designed to compare the frequencies of extreme values in two samples, can be recommended if the assumption that there are no errors in the shape of the distribution is reasonable. There are a number of variations on this test, but the core idea is to compare the central tendencies of measures of dispersion of random sub-samples of the observations and simulations (Kössler 1999). If there is no dispersion bias, the central tendencies will be similar.

Although these tests are used widely when considering climate change simulations, in seasonal climate forecasting systematic errors are usually removed using a simple statistical correction (Section 8.3.3) and are then ignored, and so the tests are rarely applied. As long as there is some predictive skill, forecast accuracy need not be adversely affected by such errors. If the model's atmosphere is responding to anomalous boundary forcing in the correct direction (for example, the model indicates unusually dry conditions when unusually dry conditions occur) then this variability is believable regardless of any conditional and unconditional biases.

8.3.3 Correcting Systematic Model Errors

Although the terms are often used in different ways in the climate literate, a distinction is sometimes drawn between "calibrated" model output, which has been corrected for systematic errors, and "recalibrated" model output, which has been corrected for model skill in addition to systematic errors. The procedures described in this section perform model calibration. Some model recalibration schemes are discussed later (see also Chapter 9, Section 9.3).

Removal of systematic errors usually involves application of the generalized formula:

$$\hat{z}_o = g_o[g_m[z_m]]^{-1}, \tag{8.1}$$

were z_m is the modelled value of the parameter of interest, $\hat{z}_o$ is the calibrated value, g_m is a function that transforms the modelled values onto a new distribution, and g_o is a function that transforms the observed values onto a distribution that is assumed to be the same as that for g_m.

In the simplest case, g_m and g_o are functions that centre the data to have a mean of zero (i.e. $g[z] = z - \overline{z}$, where g is a transformation function applied either to the model or the observed values, z is a model or observed value, and $\overline{z}$ is the corresponding climatological mean). In this case Eq. (8.1) simply subtracts the difference in the sample means between the model and the observations from the model climatology, thus removing the mean bias. An alternative option, which is suitable when correcting for variables with a zero bound (such as precipitation), scales by the ratio of the observed and simulated means (i.e. $g[z] = z/\overline{z}$). This scaling affects the variance (but not the shape) of the bias-corrected model climatology, unlike the centring procedure.

Scaling assumes that any errors in the variance are simply a function of the mean bias (i.e. that the coefficients of variation for the model and observed climatologies are identical). Since this assumption is frequently invalid, corrections for both mean and variance are generally made by standardizing the data (i.e. $g[z] = \left(z - \overline{z}\right)/s$, where s is the climatological standard deviation). Standardization is a widely used procedure that successfully removes mean and variance biases, but can be problematic when used on data with a zero bound, and/or when there are systematic errors in the shape of the model's climatological distribution. These problems occur because it is generally implicit that application of Eq. (8.1) implies application of the formula

$$\hat{z}_o = F_o[F_m[z_m]]^{-1}, \tag{8.2}$$

where F_m is a cumulative distribution function for the model data and F_o is a cumulative distribution function for the observations. Specifically, when data are standardized, F_m represents the normal distribution function fitted to the model data, and returns the quantile associated with the corresponding standard normal deviate; the corresponding quantile from the normal distribution fitted to the observed data is then used to obtain the transformed value.[3] This procedure works only to the extent that the normal distribution provides a good fit to both sets of data, otherwise errors in estimating the quantiles of the two distributions can result in unreasonable transformations. Consider the effects of standardizing the June–August precipitation data for southern Africa, described above: the model and observed data are plotted as empirical distribution functions in Fig. 8.2, and the fitted normal distribution functions are superimposed. For model precipitation rates of less than about 0.1 mm/day (the driest 5–10% of cases), the transformed precipitation is negative, as illustrated by the corresponding vertical legs of the dotted line.

Unless both the model and the observed data are normally distributed, standardization should not be performed. Instead more appropriate distribution functions should be applied in Eq. (8.2). While the empirical distribution functions could be used, the function for the model data isknown better than for the observations because of the larger sample size provided by the multiple ensemble members. The relatively poor representation of the empirical distribution function for the observations can create problems particularly when transforming extreme values. The alternative is to use a fitted distribution other than the normal distribution. The two-parameter gamma distribution is an attractive option for data that are positively skewed and zero-bound, and its parameters are easy to estimate

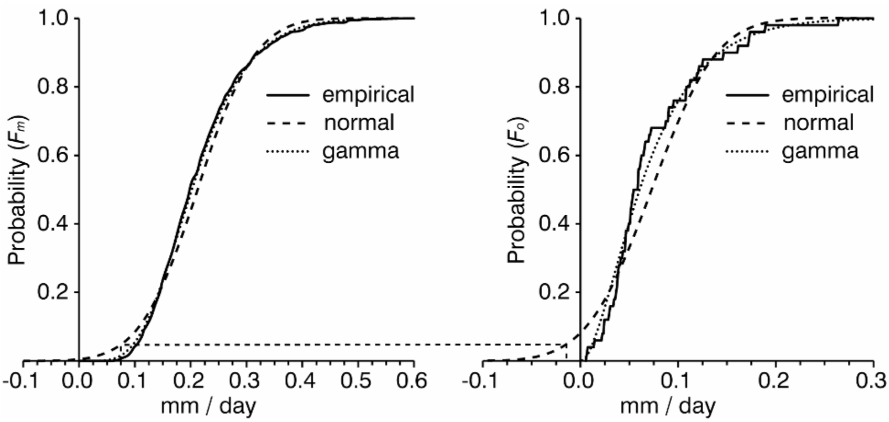

Fig. 8.2 Examples of empirical and fitted distribution curves for area-averaged June–August 1951–2000 simulated (left) and observed (right) daily precipitation intensities over southern Africa. The dotted line represents a transformation of model precipitation by standardization to remove mean and variance biases

[3] The conversion from a deviate on the standard normal distribution is redundant, but the application of the cumulative normal distribution function in Eq. 8.2 is implicit, as evident when standardization is viewed graphically, as in Fig. 1.2.

when the skewness is not too marked (Wilks 2005). Fitted gamma distributions for the model and observed data in Fig. 8.2 are shown, and the improvement in the estimation of the quantiles over the fitted normal distributions is evident not just in the tails of the distribution.

The procedure of fitting appropriate distribution functions[4] and applying Eq. (8.2) requires methods for estimating the distribution parameters. In most cases, the simplest procedure is to use the method of moments: for a given distribution the mean and variance of the distribution can be calculated analytically in terms of the distribution parameters, and so these parameters can be set to give a distribution with the same mean and variance as the sample data. These parameter estimates can be sensitive to outliers, and so a more robust procedure, known as L-moments, has been developed based on order-statistics (Hosking 1990). A more popular approach, however, is to use maximum likelihood estimation: the parameter values that maximise the likelihood of yielding the sample data are obtained. In a few cases, such as with the normal distribution, these values can be derived easily, but for most distributions they have to be obtained using iterative procedures.

8.3.4 Correcting Spatial Errors in Model Output

One aspect of systematic error that has not been addressed in Section 8.3.3 is the problem of spatial errors in model output; climate features in the model are often displaced, as shown by example in Fig. 8.3. The figure compares the first principal components of ensemble-mean forecasts of October–December precipitation for eastern Africa from the ECMWF model (Palmer et al. 2004) and of observed rainfall for the same period (New et al. 2000). While the model successfully forecasts rainfall variability over much of the region to the east of about 30°E, the main mode of variability, which involves region-wide anomalously wet or dry conditions, is displaced to the west by about 15°. Such displacements can result in poor predictions if they are not corrected.

If climate features in the model are displaced relative to the observations, even by only short distances, comparing the model output at any grid with the corresponding observations using the types of methods described in Section 8.3.2 is inappropriate a priori. Instead, the spatial structure of the model output requires correction prior to correcting any systematic errors in the climatological distributions for individual gridpoints. Standard methods for correcting such spatial errors involve multivariate statistical techniques that typically address mean and dispersion

[4] Different distributional forms could be used for the model and the observed data if their distributions have different shapes.

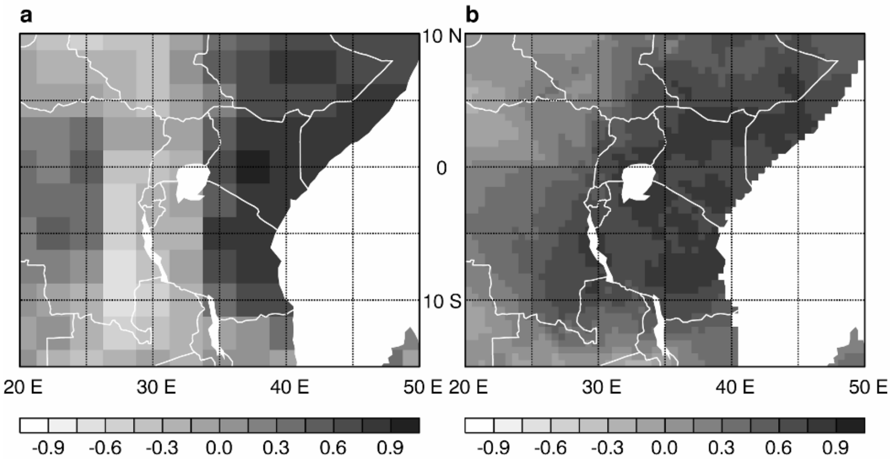

Fig. 8.3 Correlations with the respective first principal components of (a) forecast precipitation and (b) observed precipitation over eastern Africa for October–December 1961–2000. The precipitation forecasts are from the August runs of the ECMWF model generated as part of the DEMETER project (Palmer et al. 2004)

biases[5] at the same time. In practice, most techniques used to correct for spatial errors address forecast skill, and so perform model recalibration rather than simply model calibration, as distinguished in the previous section. A sample of spatial correction techniques is presented in this section. [See von Storch and Zwiers (1999) and Wilks (2005) for further details.]

The two most widely used statistical techniques for correcting spatial systematic errors are extensions to multiple linear regression, namely maximum covariance analysis (MCA) and canonical correlation analysis (CCA). The procedures are essentially identical to those described in Section 7.4.2.5 of Chapter 7, and so are discussed only briefly here. The idea is to use the model predictions as the predictors in a statistical prediction model. In both MCA and CCA spatial patterns of precipitation variability, for example, in the model are identified that have similar temporal variability to spatial patterns in the observations. Since the similarities are defined only in terms of the temporal variability there is no explicit attempt to match the spatial patterns. Consequently, in practice, MCA and CCA may be able to identify a feature of the climate such as the PNA pattern whose temporal variability may be predicted well because of a realistic modelled response to El Niño conditions, but which may be displaced in the model (as in Fig. 8.3). Both procedures will effectively transform the model's imperfect PNA prediction to a more realistic prediction of PNA variability.

[5] Non-linear statistical downscaling techniques such as neural networks could theoretically correct for shape biases in addition to mean and variance biases.

Whichever approach is used for correcting systematic spatial errors, the size of the domain(s) used requires consideration. If the objective is simply to correct for the displacement of climate features in the model, then forecasts only from nearby areas should be considered. However, multiple CCA or MCA corrections would then be necessary, and these would have to be blended somehow. Using larger domains helps to avoid artificial spatial noise in the corrected fields, and is computationally more efficient, but the statistical correction procedures are likely to identify teleconnection patterns, and so are no longer conducting purely spatial correction. Whether or not the identification of teleconnection patterns is undesirable is an open question, and the general question of domain selection requires further research.

8.4 Statistical Downscaling

A typical gridpoint in a GCM used to make seasonal predictions represents an area of about 50,000–100,000 km^2, which is invariably much coarser than the spatial scales at which opportunities to apply seasonal climate forecasts exist. The GCM output therefore needs to be "downscaled" to resolutions and/or locations commensurate with user-requirements. Downscaling involves the translation of a forecast to a spatial (and/or temporal) resolution that is finer than that at which the forecasts are produced. Reasons for performing downscaling are discussed in more detail in Section 8.4.1, and some examples of spatial downscaling using statistical models are provided. An introduction to some statistical techniques to downscale seasonal forecasts to finer temporal resolutions is given in Section 8.4.2. Dynamical methods of downscaling using limited area models are not discussed.

8.4.1 Spatial Downscaling

Since GCMs are designed to represent planetary scale processes, those processes that operate at spatial scales smaller than the model resolution have to be parameterized. Computational constraints make it impractical to operate GCMs at resolutions that would permit more realistic reproductions of regional climate, and even if computational resources were available, careful re-parameterization of the models would be required (parameterizations are tuned to work at specific model resolutions). Apart from the inevitable errors that arise from the imperfect representation of the real world because of the discretization of space (and time) within GCMs, downscaling is required even for a model that reproduces the observed climate perfectly because of the detailed spatial variability of climate. Such issues are discussed in further detail in the following paragraphs.

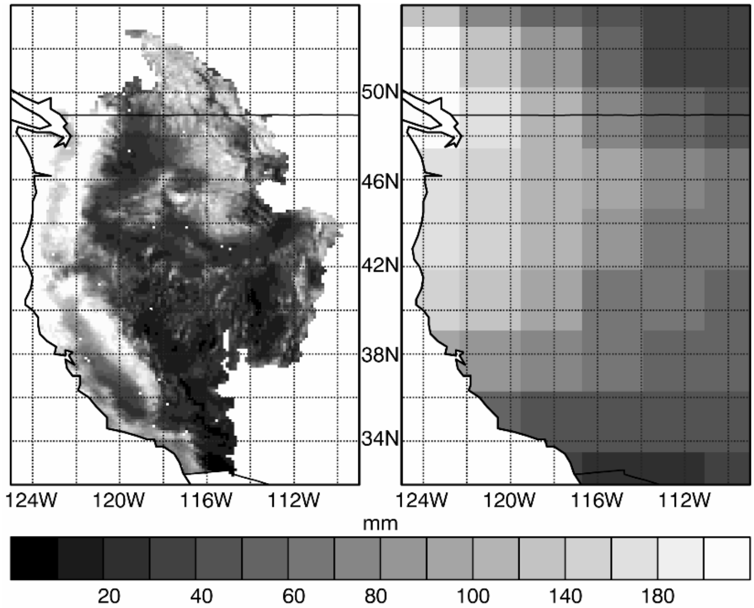

Fig. 8.4 Observed (a) and simulated (b) January–March mean precipitation for 1950–1999

Gridded model output represents an average of an essentially arbitrary area, and so, even if the model reproduces the area-averaged climatology realistically, there may be substantial systematic "errors" when these forecasts are interpreted as representative of specific locations. For example, Fig. 8.4 illustrates averaged January–March precipitation totals for 1950–1999 over part of North America (Fig. 8.4a) together with simulated precipitation totals using the ECHAM 4.5 model averaged over the same period (Fig. 8.4b). The observed data were obtained from the Surface Water Modeling Group at the University of Washington (Maurer et al. 2001, 2002). This dataset is derived from station data spatially interpolated to a grid resolution of 0.125° latitude × longitude over land, which should be compared with the approximately 2.8° resolution of the ECHAM model data. Apart from any errors in the reproduction of the broad-scale climate features, the variability of climate within any of the GCM grids is obvious, and so, at a minimum, GCM grid averages would have to be rescaled to become representative for any specific location.

Detailed spatial variability of mean climate not only affects the systematic "errors" for specific locations, but also translates into detailed variability in the predictability of climate. As a simple illustration, the correlations between the NIÑO3.4 index and observed January–March precipitation over part of North America are shown in Fig. 8.5. Within short distances large differences in the correlation are evident, and imply that GCM output could give highly misleading forecasts for sub-grid areas even after correcting for systematic errors. In addition, because seasonal

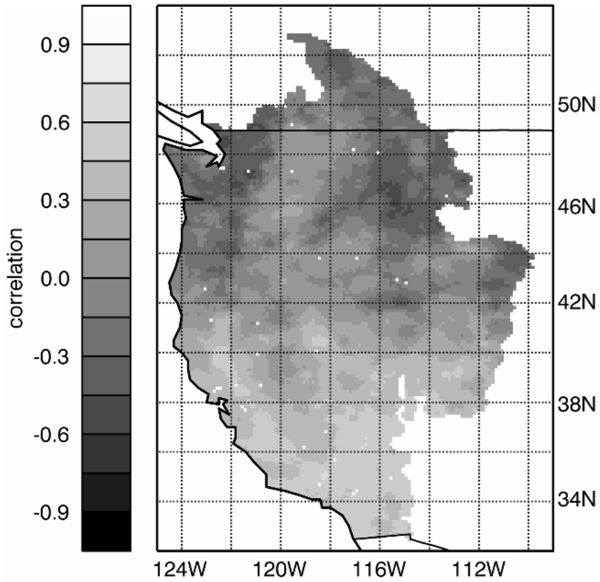

Fig. 8.5 Spearman's correlations between observed January–March seasonal precipitation for 1950–1999 and simultaneous values of the seasonally averaged NIÑO3.4 index

predictability of climate generally is greater for large compared to small area-averages (Gong et al. 2003), performance measures comparing GCM output with commensurate observational data do not necessarily give reliable indications of the accuracy of the models at the spatial scales at which seasonal climate forecasts are to be used. Downscaling is thus required to assess locally specific systematic as well as predictive errors.

If high resolution observational data or data for specific locations are available, detailed spatial corrections can be made to provide forecasts at resolutions that the GCM itself is unable to resolve. To illustrate, the precipitation data for the 50-year period January–March 1950–1999 were used to downscale simulations of precipitation from the ECHAM 4.5 model. A canonical correlation analysis(Chapter 7, Section 7.4.2) was used to downscale the GCM data. Results are shown in Fig. 8.6, which compares the skill of the downscaled predictions with the skill achievable by linearly interpolating the output for surrounding GCM gridpoints. The skill score used (Spearman's correlation) considers only the predictive errors, not any remaining systematic errors.

In Fig. 8.6, results are shown for downscaling the GCM precipitation fields directly, but there have been a number of successful attempts to downscale to station precipitation using other outputs from the GCM. For example, the model's geopotential heights are used frequently, sometimes with more than one level being considered simultaneously. Potential vorticity fields have also been used successfully. However, little attention has so far been given to downscaling multiple fields; if downscaled predictions of precipitation and of temperature are required,

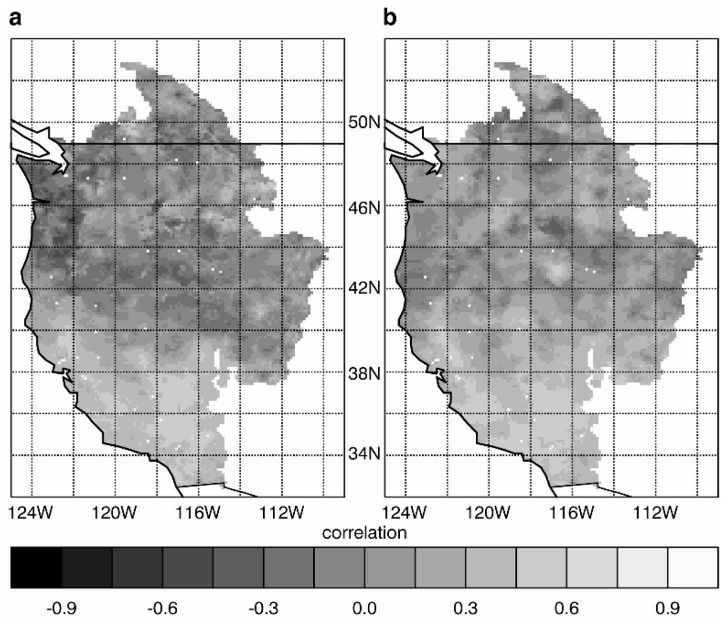

Fig. 8.6 Spearman's correlations between observed and simulated January–March 1950–1999 precipitation. Results are shown for (a) ECHAM 4.5 output linearly interpolated to the 0.125° resolution of the observational data, and (b) ECHAM 4.5 output spatially corrected using canonical correlation analysis. All results are cross-validated using a leave-five-out cross-validation window

for example, these are generally conducted independently, which could result in locally inconsistent results. In contrast, greater attention to correlations between different weather variables has been given in methods of temporal downscaling, and these methods are discussed in the following section.

8.4.2 Temporal Downscaling

Apart from the incompatibility between the spatial resolution of the forecast and that of the observations, other problems with GCM output preclude their application without additional downscaling. An important constraint to the use of GCM output and, more generally, of seasonal climate forecasts, is the temporal resolution of the predictions. As discussed in Chapter 3, seasonal climate is predictable only when the forecast is considered as an aggregate of weather over a period of typically about 3 months; it is not possible to provide accurate predictions of the weather on any given day within the season. However, for many application models, including hydrology and crop models, it is necessary to have forecasts for each day of the season. While the sensitivity of the predictions from such application models to the precise weather on specific days may be low as long as the seasonal weather statistics are accurate, some means of obtaining atmospheric forecasts at the required temporal resolution is required. In this section, various means of obtaining seasonal forecasts at high temporal resolution are discussed.

Since GCMs are generally run at a temporal resolution of about 20 minutes to generate seasonal forecasts, the simplest solution to the need to obtain weather statistics over the period of the seasonal forecasts would be to use the GCM output. However, there are some severe biases in GCM weather data, which are perhaps best illustrated by considering the frequency distribution of daily precipitation intensities. An example is shown in Fig. 8.7, which compares the frequencies of simulated and observed daily precipitation amounts for San Diego for the 50-year period 01 January 1950–31 December 1999. The model clearly underestimates the frequency of dry days (note that the *y*-axis is logarithmic) and of precipitation intensities exceeding about 4 mm/day. In other words, the model generates too much drizzle, a problem that is characteristic of GCM-based forecasts for all timescales.

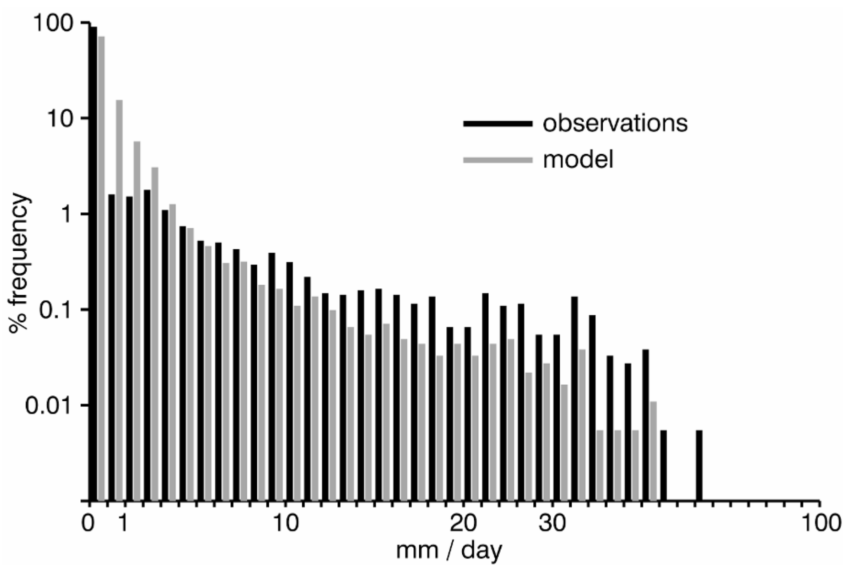

Fig. 8.7 Relative frequencies of observed and ECHAM 4.5-simulated daily precipitation intensities for San Diego for the 50-year period 1951–2000. The simulated precipitation is for the gridpoint nearest to San Diego. Note the logarithmic y-axis, and the uneven intervals on the x-axis

An alternative to using the GCM daily output is to disaggregate the seasonal forecast using statistical methods. Disaggregation involves computing sub-seasonal weather statistics that are consistent with the seasonal forecast. One commonly used statistical procedure is the analogue method, which uses the observed sub-seasonal statistics of seasons that are similar to the forecast for the target season. (See Chapter 7 for more details on statistical forecasting techniques.) Such procedures can be limited severely by sample size, and so are most commonly used in places such as Australia where datasets are relatively long, and so the number of analogue years large, compared with those for many other countries. An alternative approach is to use simple statistical relationships between seasonal climate and sub-seasonal weather statistics. For example, simple relationships between seasonal rainfall totals and the frequencies of raindays or of heavy raindays can be regressed. Such relationships could then be used to estimate weather statistics contingent upon the forecast for the seasonal aggregate.

Based on observed relationships between seasonal climate and sub-seasonal weather, fairly sophisticated statistical techniques for simulating weather over a season have been developed. These procedures are based on "weather generators", of which there are a wide range of different designs (Wilks and Wilby 1999). Most weather generators have been constructed to generate series of daily precipitation, and invariably consider the question of precipitation occurrence separately from precipitation amount. Precipitation occurrence is modelled in one of two ways: either as a chain-dependent process or by spell-lengths. As a chain-dependent process, the probability of precipitation is calculated contingent upon the occurrence of precipitation on the previous (day), which is equivalent to modelling precipitation occurrence as a Markov process. For example, the seasonal cycle of probability of precipitation in San Diego given that the previous day was wet is compared for that given that the previous day was dry in Fig. 8.8a and b, respectively. Throughout the year the probability of a wet day is considerably higher given that the previous day was wet compared to when the previous day was dry. These differences in precipitation probability are indicative of the persistence of weather in San Diego, indicating that spells of weather tend to last a few days, rather than weather changing randomly from day to day. Weather generators based on Markov models simulate a series of precipitation occurrence by randomly generating wet and dry days by considering the weather generated on the previous (few) day(s), and should thus generate weather spells with realistic duration. The second approach to modelling precipitation occurrence is to generate a string of

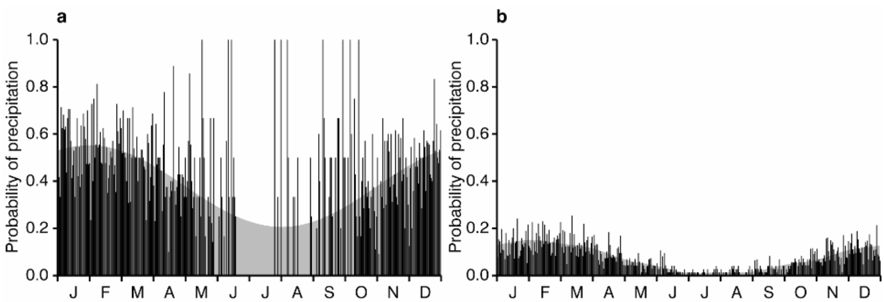

Fig. 8.8 Annual cycle of the probability of precipitation occurrence for San Diego for the 50-year period 1951–2000, given that (a) the previous day was wet, and (b) the previous day was dry. The black vertical bars show the probabilities calculated for each day, while the grey shading indicates smoothed probabilities using the first few harmonics of the annual cycle

alternating wet and dry spell-lengths. The frequency distributions of observed wet and dry spell-lengths are usually modelled using a negative binomial distribution. The spell-length generator operates by randomly drawing alternating random spell-lengths drawn from the corresponding negative binomial distributions.

Whichever way precipitation occurrence is modelled, the generated occurrences of precipitation need to be conditioned somehow on the seasonal forecast. Again there is a range of options for modelling this conditioning (Wilby et al. 2002). For example, if a Markov model is used, the probability of precipitation can be conditioned not only on the generated occurrence of precipitation on the previous day(s), but also on some aspect of the seasonal forecast, such as the predicted rainfall total exceeding some predefined threshold. As a simple example, Fig. 8.9 compares the probabilities during El Niño and La Niña years of daily precipitation during the winter months of January–March in San Diego exceeding various thresholds. Rainfall at all but the highest intensities typically occurs more frequently under El Niño conditions than under La Niña conditions.

Alternatively, the probability of precipitation could be estimated using a statistical model. This regression approach has the advantage of not dividing the degrees of freedom up by the repeated splitting of the dataset when calculating conditioned parameters, but does require the form of the relationship between the conditioning variable and the precipitation probability to be specified.

A more sophisticated approach to conditioning the generator on the seasonal forecast involves modelling the occurrence of precipitation on the basis of the predicted daily sequence of the large-scale atmospheric circulation. Since the daily

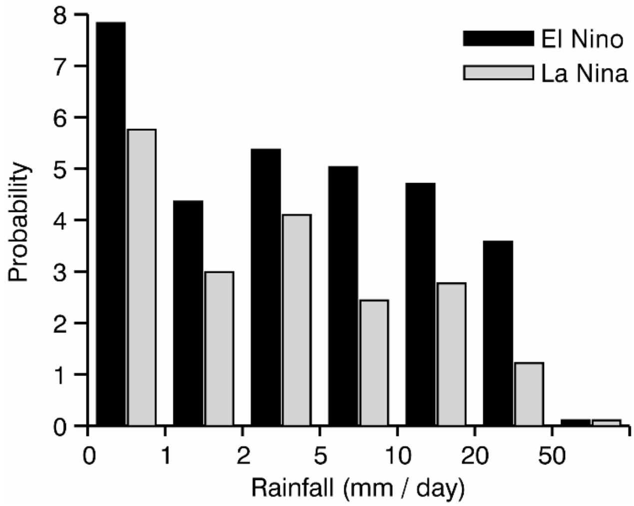

Fig. 8.9 Histograms of wet spells in San Diego commencing any time between 01 January and 31 March for the 50-year period 1951–2000, given that the January–March averaged NIÑO3.4 index was greater than +0.7 (i.e. El Niño conditions prevailed), and less than –0.7 (i.e. La Niña conditions prevailed)

atmospheric circulation has the weather persistence implicitly built in, there is no need to condition the precipitation on the previous days' weather (either by Markov or spell-length modelling). Because the weather persistence is implicit in this approach, these models are called "hidden Markov models" (Robertson et al. 2004). Hidden Markov models condition the precipitation probability by identifying specific weather patterns and then classifying each day into one of the patterns. The daily sequence of the atmospheric circulation over the period of the seasonal forecast would normally be provided by the GCM, and so the procedure is somewhat similar to the spatial downscaling procedures described in Section 8.4.1. Apart from obvious differences in the form of the statistical model used, and in the temporal resolution of the GCM output (daily compared to seasonal average), the procedures have this in common: the large-scale GCM output is statistically corrected to provide an estimate of precipitation (daily occurrence or seasonal total).

Precipitation intensity is modelled in a similar way to spell-lengths: the distribution of non-zero precipitation intensities is represented (frequently by a gamma or mixed exponential distribution), and random intensities are generated for days in which precipitation is specified to occur. Thus intensity is modelled subsequent to occurrence. Again, the intensity of precipitation can be conditioned upon some aspect of the seasonal forecast if there is evidence that seasonal variability is affected by changes in precipitation intensity. Since the inter-annual variability of precipitation can be affected by changes in precipitation frequency and/or intensity, weather generators can be designed to account for both/either effect.

Weather generators can be designed to model a suite of meteorological parameters in such a way that the relationships between the parameters are consistent with the relationship in the real. For example, in many parts of the world there is a relationship between precipitation occurrence and maximum temperature, and some applications of seasonal forecasts it may be important to retain this relationship. Generated temperatures (and other parameters) are conditioned upon the generated precipitation occurrence. In a similar way, it is possible to generate weather sequences at a range of locations so that the generated weather is spatially realistic by accounting for the spatial correlations in the meteorological parameters. This consideration may be important in hydrological modelling, for example, where the spatial distribution of precipitation across a river catchment is important in affecting runoff.

8.5 Using Ensembles

There are two primary motivations for generating an ensemble of predictions (whether from a single model or a set of models). One is that the average of a set of predictions more closely approximates the climate signal than the prediction from any single ensemble member. However, a second motivation is to obtain some indication of the uncertainty in the prediction.[6] Since the ensemble mean indicates only the central tendency of the predictions, a separate measure is required to indicate the uncertainty. However, it is not obvious how the ensemble members can be used to indicate forecast uncertainty, or even whether they are successful in doing so. In Section 8.5.1 how uncertainty in a forecast can be communicated is discussed. Then some methods for describing the forecast uncertainty using an ensemble are considered (Section 8.5.2). A description of procedures for assessing how well the ensemble can be used for indicating changes in forecast uncertainty is reserved for Chapter 10.

8.5.1 Forecast Uncertainty, Forecast Confidence and Forecast Probabilities

Given the inherent uncertainty in forecasting seasonal climate conditions, the forecaster needs to provide some indication of this uncertainty. A common way of communicating such uncertainty is by indicating the level of confidence to be placed in the forecast. This level of confidence is inversely related to the degree of uncertainty in the forecast: when uncertainty is large a low level of confidence in the forecast is communicated, whereas when uncertainty is reduced confidence increases. The distribution of possible outcomes defines the full extent of the uncertainty in the prediction, but this distribution is unknown and so has to be approximated somehow. Once approximated, the forecaster's confidence can then be defined. The confidence in the forecast can be communicated in a number of ways, and how the ensemble may be used depends on which format is adopted.

One of the simplest ways of indicating forecast uncertainty is to specify a range of values within which the observed value is expected to lie with a predefined level of confidence. Usually this level of confidence is kept fixed from forecast to forecast, and the varying uncertainty is reflected by adjusting the width of the interval. Thus, when uncertainty is large (small) the interval is made wide (narrow). For example, forecast A, which states that there is a 90% probability of a seasonal rainfall total being between 100 and 200 mm indicates greater uncertainty than forecast B, which states that there is a 90% probability of the total being between

---

[6] Here, and elsewhere in this Section, "uncertainty" relates to the range of possible outcomes for a specific target period, and is not the same as, the climatological uncertainty as defined by Murphy (1973a). Murphy's definition is independent of the forecasts themselves, whereas here, as discussed later, uncertainty is represented by the extent to which the forecasts of individual ensemble members for the same target differ. If the forecasts for all the ensemble members are similar, then forecast uncertainty is low, but if they differ substantially then forecast uncertainty is high.

125 and 175 mm.[7] This format is known as a prediction interval (see Chapter 9) and is not widely used in seasonal climate forecasting partly because such intervals are frequently misinterpreted.

An alternative approach to that is more commonly used in seasonal climate forecasting is to fix the interval and to allow the level of confidence to vary. The interval itself can be fixed to meet the user interests, although in practice it is most commonly defined from the terciles of the observed data as measured over a climatological period. The fixed intervals are normally called "categories", along with the unbounded categories either side of the interval. More than one interval can be specified, and in this respect quintiles are being used with increasing frequency. To illustrate: the interval of 100–200 mm used in forecast A above could be used for all forecasts; decreased uncertainty implicit in forecast B would then be communicated by increasing the probability that the seasonal rainfall total will be within this range rather than by narrowing the range. It should be noted that there is no simple relationship between the change in the probability assigned to an interval and the changing level of uncertainty in the forecast, as illustrated in Fig. 8.10. Two forecasts are shown in the figure; forecast A involves less uncertainty

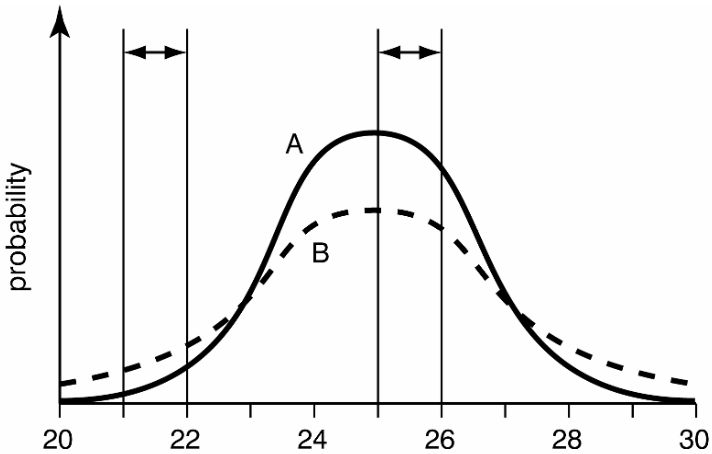

Fig. 8.10 Hypothetical example illustrating the complex relationship between forecast probability and forecast uncertainty. Forecast A (solid line) represents a forecast with relatively low uncertainty, and forecast B (dashed line) represents one with relatively high uncertainty. The narrow vertical lines indicate the limits of intervals for which forecast probabilities are desired. These probabilities are calculated by integrating the areas beneath the lines A and B within the range of the intervals

[7] If the confidence level is α, it would normally be assumed that the probability that the observed value will be less than the lower limit of the interval (125 mm in forecast B) is the same as the probability that the observed value will be greater than the upper limit (175 mm). This probability would be $1–2\alpha$ (5%) However, it is not necessary for the interval to be centred in this way, as long as the corresponding tail probabilities are then specified.

than forecast B. In the interval 21–22°C the probability increases with the more uncertain forecast, but decreases in the interval 25–26°C. This problem in interpretation can be avoided by specifying the probabilities for all categories, and comparing these probabilities to the climatological probabilities for the categories.

More detailed summaries of the distribution of possible outcomes, and thus of the uncertainty in the forecast, are possible by assuming a distributional form and describing this distribution by its parameter values (for example, the mean and variance of a normal distribution). These distributions can be used to estimate probabilities for any intervals, or intervals for any levels of confidence. Alternatively, some of the percentiles of the distribution (fitted or otherwise) can be specified. These options are discussed in more detail in the following sections.

8.5.2 Forecast Ensembles and Forecast Uncertainty

After correcting for systematic errors in the individual ensemble members (Section 8.3), their distribution is supposed to give an indication of the distribution of possible outcomes. The distribution of the ensemble members should therefore indicate the uncertainty in the forecast: in simple terms, if the various ensemble members are forecasting similar values then uncertainty is low, whereas if the values differ widely then uncertainty is high. However, with a finite ensemble size the distribution of the current forecasts is imperfectly sampled, and so the uncertainty implied by the forecasts has to be estimated.

One of the simplest ways of using the ensemble to indicate forecast uncertainty is to estimate the probabilities for categories by counting the proportions of the ensemble members indicating outcomes within each category. Errors in calculating these probabilities by counting can be derived from the binomial distribution, and can be substantial. The probabilities are more reliably obtained by fitting a distribution to the ensemble members using one of the methods described in Section 8.3.3 and then calculating the probabilities from the fitted distribution (Kharin and Zwiers 2003). Further improvements can sometimes be made by assuming a distributional form for the sampling errors associated with each ensemble member rather than for the ensemble distribution as a whole. Each ensemble member is “dressed” with a fitted distribution (Roulston and Smith 2002). One advantage of this approach is that the prediction errors of the forecasts can be accounted for to some extent. Alternatively, the probabilities could be estimated directly using a statistical model, such as a generalized linear model (Tippett et al. 2007). The statistical model would correct for both the predictive and the systematic errors in the model(s).

The use of a statistical model for estimating probabilities does not necessarily mean that the uncertainty implied by the ensemble distribution provides useful information. The most obvious candidates for predictors in the statistical model are the first few moments of the ensemble distribution, and virtually all of the usable information is in the ensemble mean. The ensemble mean communicates no information about the uncertainty in the forecast, which, instead, is derived from the error variance of the ensemble mean predictions. Errors in calculating the ensemble variance appear to be too large to derive much useful information in the ensemble spread (Kharin and Zwiers 2003). Alternative measures of spread, such as the inter-quartile range, could be used, but more detailed studies are required to identify how much of the variability in the ensemble spread beyond the sampling variability truly represents variability in forecast uncertainty. There has been minimal investigation into the information content of the shape of the ensemble distribution.

Instead of using the ensemble to estimate probabilities for predefined categories, they could be used to estimate the values associated with specific percentiles of the ensemble distribution. For example, the ensemble median is arguably more informative than the mean since the former is amenable to making a simple probabilistic forecast (there is an estimated 50% probability that the observed value will exceed the ensemble median, but the probability of exceeding the mean is unknown unless some distributional assumptions are made). The percentiles can be estimated either by fitting a distribution to the ensemble or to the individual ensemble members, or by treating the individual ensemble members as percentiles of the distribution. The latter approach is implicit when constructing ranked histograms, as discussed in Chapter 10. Effectively, such procedures are an extension to those used for defining prediction intervals since each end of the interval represents a fixed percentile of the forecast distribution.

8.6 Combining Forecasts

There is ample evidence that combining seasonal climate predictions from a suite of models provides an improved forecast over using even the best of the individual models (Doblas-Reyes et al. 2005; Hagedorn et al. 2005). The improvement is evident not only in forecasts of seasonal averages but also in some of the intraseasonal statistics such as storm frequencies. Similar conclusions can be drawn for forecasts at medium-range and shorter timescales, and multi-model approaches are being used increasingly in climate change work. At all timescales the improvement in the forecasts results from the improved representation of uncertainty arising from imperfections in model physics. In a single-model ensemble, uncertainty is represented only in terms of the initial conditions, and each ensemble member is subject to the same errors in the model physics, so that clustering of forecasts tends to occur. Alternative ways of accounting for the uncertainties arising from model errors include stochastic parameterization, and perturbed physics approaches, but the use of multi-models is likely to remain popular both in research and operations.

Simple averaging of predictions from different models is usually sufficient to improve the quality of a forecast, but it is tempting to weight the models by their respective skill levels. However, a major difficulty in assigning differing weights arises from the limited availability of hindcasts for which to assess relative model performances robustly. If the skill levels of the models cannot be definitively compared, it is then exceptionally difficult to outperform the simple average of the models' respective predictions (Kharin and Zwiers 2002). Further, as the number of models is increased, problems of over-parameterization arise, and so simple averaging of recalibrated model output again generally proves to be the most effective approach.

To illustrate, monthly predictions of the NIÑO3.4 index from three of the models that participated in the DEMETER experiment were combined using a range of methods. Each of the three models (ECMWF, CNRM, and the Met Office) had nine ensemble members, produced forecasts from four start dates, and generated predictions with lead-times of up to 5 months. Hindcasts were available for the 44-year period 1959–2002, and all results were cross-validated using a 3-year cross-validation window (i.e. 1 year either side of the predicted year was omitted). Results for all lead-times and seasons were pooled.

The details of the various combination schemes are not important, but include two Bayesian model weighting schemes, canonical variate analysis, generalized linear models, multiple regression, and stepwise regression. For all but the Bayesian schemes, the respective model ensemble means were used.[8] Simple model averaging (i.e. equal model weighting) was used as a benchmark level of skill. In all cases the forecasts were expressed as probabilities of the NIÑO3.4 index falling below the lower quartile, above the upper quartile, or within the inter-quartile range. The forecasts were evaluated using the quadratic score, which is a measure of the squared error in the probability assigned to the category that verified (Chapter 10). The score ignores the probabilities assigned to the other categories. Since it is an error score, a perfect set of forecasts would achieve a score of 0.0.

Since each of the schemes can be applied to the models individually to recalibrate the model output, the models can be combined in one of two ways: recalibrate each model individually, and then calculate a simple average of the recalibrated predictions ("recalibration"); apply the schemes to all the models simultaneously ("combination"). For the simple equal weighting, the combination and recalibration schemes will give identical results. The scores for the various schemes are illustrated in Fig. 8.11. Most of the schemes improve the forecasts of individual models, and in most cases the combined forecasts (using either combination approach) improve upon the forecasts from the best single model. The

[8] The use of the ensemble mean only generally gave the best results. Alternatives tried were: to include the ensemble variance (which can provide some marginal improvements in skill), and higher moments; to use all the ensemble members; to use the first few principal components of the ensemble members.

averaging of the recalibrated forecast consistently outperforms the combination method, presumably because of the over-parameterization of the latter.

The results shown in Fig. 8.11 are based on combining forecasts from only three models. Since many of the models used in seasonal climate forecasting are fairly closely related, the predictions from these models are often strongly correlated, potentially creating problems of multicolinearity (see Chapter 7, Section 7.4.1). Some success has been achieved in addressing multicolinearity problems by using procedures equivalent to truncated principal components regression. Such an approach will also help to reduce problems of multiplicity that arise from considering the skill of more than one model (Chapter 7, Section 7.4.1), and which are exacerbated when downscaling approaches are incorporated into combination algorithms. However, it is not clear that principal components regression is appropriate in the context of forecasts of precipitation, for example, where the assumptions of multivariate normality are often violated.

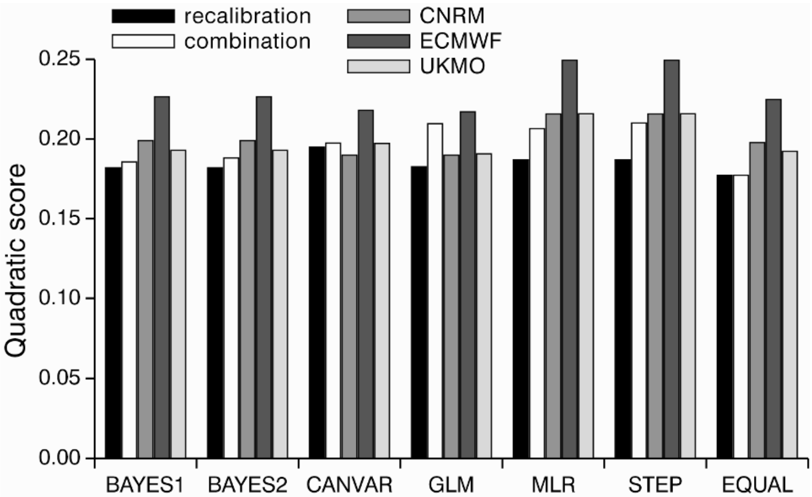

Fig. 8.11 Quadratic scores for monthly predictions of the NIÑO3.4 index, using various forecast combination schemes [Bayesian model weighting schemes (BAYES1 and BAYES2, canonical variate analysis (CANVAR), generalized linear models (GLM), multiple linear regression (MLR), stepwise regression (STEP), and equal weighting (EQUAL)]. The schemes are compared by combining the predictions using two procedures: attempting to account for differences in model skill ("combination"), and by simple averaging of predictions after recalibrating the individual models ("recalibration"). Results for the individual models are shown also

Problems of multicolinearity in forecast combination algorithms, while not unique to multiple regression, which has been a commonly used method for combining forecasts, are not an issue with some alternative combination methods. Canonical covariate analysis, for example, has some strong similarities to principal component regression, and explicitly addresses the inter-correlations of models (and of individual ensemble members, if used). In addition, the first few principal components of the model predictions could also be used in a wide range of alternative statistical schemes, such as generalized linear models (Chapter 7, Section 7.4.3). These approaches deserve further consideration, especially in the context of combining predictions of precipitation amounts, where the standard assumptions of multiple regression (multivariate normality) are sometimes invalid.

8.7 Summary

Producing a seasonal climate forecast from a dynamical model involves a great deal more than simply running the model, and viewing the results. The first problem is to decide which dynamical model(s) should be run given the practical constraints of computing resources. Assuming that dynamical models can represent the underlying physical processes correctly, fully coupled models theoretically should give the best predictions of seasonal climate if they can be initialized accurately, but this initialization can be problematic, and computing resources can be prohibitive. An alternative is to use uncoupled atmospheric models and to prescribe the SST forcing. In the latter case, the SSTs have to be predicted first, and so the uncoupled approach involves a "two-tiered" process.

Once model predictions have been made, they then need to be corrected for systematic errors. These errors result from consistent differences between the model and the observed climatologies, and can be identified by differences in the probability distributions of climate parameters for the model and the observed data. However, since one contribution to the systematic errors in the model is that the geography is distorted, simple gridpoint-by-gridpoint comparisons of model and observed climatologies can be inappropriate. Instead some form of spatial correction to the model output is desirable.

Even after systematic error and spatial correction, the model predictions may require further processing in order to be made relevant for specific locations. All dynamical models produce output that represents an averaged value over a gridded area typically of the order of between 10,000 and 100,000 km^2. Because local climate can vary considerably over fairly short distances, especially in areas of marked terrain, this gridded average may be unrepresentative of specific locations within the grid. The model prediction must therefore be "downscaled". Downscaling can also involve the conversion of a prediction for a gross summary of weather over a season, such as a 3-month rainfall total, to one containing more detailed information about the statistics of weather within the season.

After correcting the model output, the uncertainty in the forecast then needs to be communicated. Apart from the fact that an average of an ensemble of predictions is almost invariably a more accurate forecast than any single prediction, ensembles are a commonly used method of representing the uncertainty in the forecast. (The question of whether the ensemble does in fact provide a reliable indication of forecast uncertainty is deferred until Chapter 10.) If the various predictions from the ensemble are in close agreement then presumably we can place more confidence in the forecast than when the ensemble members predict widely different outcomes. There are a number of ways of assessing the level of agreement amongst the ensemble members. The most widely used approach is to count the proportion of ensemble members that predict an event of interest. However, more sophisticated procedures are available, and involve fitting a distribution to the predictions, which gives a more reliable indication of the model's forecast distribution, and using statistical models to correct the forecast distribution to account for model skill. Such procedures are discussed in more detail in the following chapter.

Just as an ensemble of predictions from one model provides a more accurate forecast than any single model prediction, so also forecasts obtained by combining predictions from a range of different models are an improvement upon forecasts derived from a single model. There have been numerous attempts recently to combine predictions from different models in ways that account for differences in the skill of the individual models. However, with the typically small sample sizes available for seasonal forecasts, it is difficult to estimate with sufficient accuracy the differences in the performances of the models, and so a simple average of the predictions from the various models is a high standard to beat.

After constructing a forecast, an indication of the reliability of the probabilistic information communicated needs to be performed by conducting a detailed assessment of the quality of a set of historical forecasts produced in a consistent way with the current forecast. The verification of these historical predictions provides an indication of the information content in the forecast, and relevant procedures are discussed in Chapter 10.

Chapter 9
An Introduction to Probability Forecasting

David B. Stephenson

This chapter reviews the basic probability concepts needed to understand probability forecasting and presents some simple Bayesian approaches for producing well-calibrated probability forecasts. Forecasts are inherently uncertain and it is important that this uncertainty is estimated and communicated to forecast users so that they can make optimal decisions. Forecast uncertainty can be quantified by issuing probability statements about future observable outcomes based on current forecasts and past observations and forecasts. Such probabilistic forecasts can be issued in a variety of different forms: as a set of probabilities for a discrete set of events; as probabilities for counts of events; as quantiles of a continuous variable; as interval forecasts (pairs of quantiles); as full probability density functions or cumulative distribution functions; or as forecasts for whole spatial maps. Since models predict the future state of model variables rather than actual real-world observable variables, probability forecasts need to be recalibrated on observations as an inherent part of the forecasting process. Rather than the (marginal) pro bability distribution of ensemble predictions, what forecasters should issue are estimates of the conditional probability distribution of the future observed quantity given the available sample of ensemble predictions.

9.1 Introduction

There is one thing in climate science that we can be certain about: weather and climate forecasts will always be uncertain. To deterministic ways of thinking, this uncertainty is rather annoying, but is a defining characteristic of many areas of modern science. How should we deal with this forecast uncertainty? One approach is to deny it and simply issue deterministic forecasts (e.g. “it will rain tomorrow”) with no estimate of the forecast uncertainty. This doesn’t mean that the forecaster thinks there is *no* uncertainty in the forecasts, but rather that an estimate of the forecast uncertainty is *not available*. Although simple to communicate, this

David B. Stephenson
University of Exeter

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

approach has several problems. Rather than be interpreted as the most probable future outcome, the forecast can be misinterpreted to be the only possible future outcome. This can lead the forecast user to lose confidence in the forecast provider when the forecast outcome fails to happen. Hence, probability forecasts can be more "believable" than are deterministic forecasts. Probability forecasts are also required for the user to be able to make optimal decisions – the predicted uncertainty of the forecasts is a key element in the decision-making process. In order to make optimal decisions, it is necessary for the forecast user to be able to quantify risk by having estimates of the probabilities of all the different possible outcomes.

Hence, the goal of weather and climate forecasting is to provide the most reliable probability estimates for future observed events given the available predictions and all available past information. Unless one has a perfect model identical to that of the real world initialised identically to the real world (!), model predictions should not be considered to be identical to forecasts of observable events. In practice, the probability of future observables has to be *inferred* from the set of available weather and climate predictions. The conditional probability distribution of the future observable is *not* simply the probability distribution of the ensemble predictions – it has to be estimated from the model data using a suitable probability model such as a regression model. This important process of mapping the forecasts back into observation space has been referred to as *forecast assimilation* by Stephenson et al. (2005).

Most seasonal forecasting centres *calibrate* their forecasts to look like reality by simply adding constants and/or multiplying by constants to correct for biases in the mean and variance (Chapter 8). This simple procedure is based solely on the mean and variance of past forecasts and past observations, and ignores information about the joint distribution of past observations and forecasts (e.g. the skill of the forecasts). A better approach is to *recalibrate* the forecasts using a model based on the regression, for example, of past forecasts on past observations. For example, the Swedish Meteorology and Hydrology Institute uses an adaptive Kalman filtering technique to adjust all the ensemble members based on daily updates of an error equation evaluated on the control forecast. Rather than perceive this as a simple post-processing step, one should realize that this is an inherently important aspect of the forecasting process that requires as much care and attention as invested in other aspects of the forecasting process such as coupled model development, data assimilation, ensemble generation, etc. The full potential of climate forecasts cannot be fully realized without these activities.

The aim of this chapter is to give a brief introduction into why we need probability forecasts, how they can be issued, and what needs to be done to make well-calibrated probability forecasts. Section 9.3 motivates why we should issue probability forecasts and discusses the different types of probability forecast. Section 9.3 gives a brief overview of the basic probability concepts needed to understand probability forecasts. Finally, Section 9.4 presents ideas and examples on how to produce well-calibrated probability forecasts.

9.2 Probability Forecasts: Why Issue Them and What are They?

9.2.1 Why Issue Probability Forecasts?

Forecasts can be either *deterministic* or *probabilistic.* Deterministic forecasts issue a specific value or category that is considered to be most likely to occur in the future. Unless issued together with skill measures such as mean squared forecast error, deterministic forecasts provide no indication of forecast uncertainty. In contrast, probabilistic forecasts do attempt to quantify the uncertainty by making clear probability statements about the chance of occurrence of future outcomes.

There are several important reasons why it is better to issue probabilistic forecasts rather than deterministic forecasts:

- The future state of a complex system such as climate cannot be predicted with certainty
- Probability forecasts allow different decision-makers (forecast users) to make their own optimal decisions, whereas deterministic forecasts are essentially a decision already made by the forecaster
- Probability forecasts are essential for quantitative assessment of risk
- It is dishonest and legally dangerous to claim that there is no uncertainty in the forecasts.

However, there are several difficulties when issuing probabilistic forecasts, such as:

- More information needs to be communicated so the forecasts can be difficult to communicate concisely (e.g. in short television broadcasts)
- The understanding and perception of probability and risk varies enormously from person to person
- Not all users want to make optimal decisions – they often prefer the forecaster to issue a definitive statement about what will happen (despite the fact that this is impossible!)
- The probabilities may be difficult to quantify reliably especially when uncertainties are due to unknown unknowns (e.g. missing processes) and cascade through several stages of the ensemble forecasting system.

In order to surmount these difficulties it is necessary for forecasters and users to work together at improving communication and understanding of what they are attempting to achieve.

9.2.2 Sources of Uncertainty

Uncertainty is endemic in forecasting because the future is not certain. There is a growing need in climate science to quantify the different sources of uncertainty. To help do this, it is useful to try and classify all the different possible sources of uncertainty.

Uncertainty arises from two main sources: aleatoric and epistemic. Aleatoric uncertainty is due to chance (from the Latin word aleator meaning a dice-player) whereas epistemic (or structural) uncertainty arises from our incomplete and incorrect knowledge of the world. Epistemic uncertainty can be summarized by the unknown unknowns in the famous quote by Donald Rumsfeld at the US Defence Department Briefing on February 12, 2002:

> Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tends to be the difficult one.

Below is a classification of the types of uncertainty that arise in climate science:

- Observational uncertainty
  - Sampling error in measurements
  - Systematic error in measurements (e.g. instrumental biases)
  - Inherent uncertainty in statistics caused by sampling of natural variability over finite periods (e.g. the historical record, the future forecast period, the 30-year base period used to define climatology)
- Model data sampling uncertainty (aleatoric uncertainty)
  - Finite length model simulations (e.g. climate time-slices)
  - Finite ensemble of all possible model runs
- Model parametric uncertainty (i.e. known unknowns)
  - Physical uncertainty in model parameters (e.g. cloud physics)
  - Sampling uncertainty in statistical estimates of parameters
  - Non-uniqueness in model parameters (model parameter degeneracy)
- Model structural uncertainty (i.e. unknown unknowns)
  - Incomplete knowledge of external factors (e.g. future emissions scenarios)
  - Misidentification of physical and/or statistical models (e.g. simplification of land-surface processes, weak but unaccounted for effects such as slow composition changes)
  - Numerical and representation error (e.g. grid boxes are not the same as single-site locations)

Various strategies have been developed in recent years in order to get estimates of these different sources of uncertainty – for example, ensemble and multi-model prediction and stochastic parameterization techniques. However, it is important to be aware that uncertainty is invariably underestimated due to the fundamental inability to quantify epistemic uncertainty caused by unknown unknowns (there can always be surprises – e.g. the appearance of the Antarctic ozone hole!).

9.2.3 Types of Probabilistic Forecast

Probabilistic forecasts are probability statements about future outcomes. They need not simply be issued as a probability for an event. Some examples of different probabilistic forecasts (see additional discussion in Chapter 8, Section 8.4.1) are listed here:

1. A set of probabilities for the discrete set of events that can occur. For example, probabilities of 0.2, 0.4, and 0.4 for the categories of below-normal, normal, and above-normal. For binary events (two mutually exclusive and complete categories such as rain and no-rain), effectively only one probability needs to be issued (e.g. probability of rain) since the two probabilities must add up to one (one of the events must happen!).
2. Probabilities for counts: e.g. the probability of 4 hurricanes making landfall in the U.S.
3. Interval forecasts in which one specifies a range of continuous values in which is likely to include the observation with a certain fixed probability. For example, temperature could be predicted to be in the range [23.7°C, 29.3°C] with 95% chance – this is known as a prediction interval in the statistics literature (not to be confused with confidence interval that refers to an interval on a distributional parameter not a random variable). Error bars (as often used by physicists) such as 10.0°C ± 1.2°C are constructed by adding and subtracting one standard deviation, and are examples of 68.3% prediction intervals for Normally distributed variables.
4. One or more quantiles of a continuous variable. For example, one could predict the 50th quantile (the median) of the distribution, in which case the observed value would be expected on average to exceed this value on half the occasions. Note that predicting the median is fundamental different to predicting the mean since in general one does not know a priori the probability of exceeding mean value. If one wants to be correct 50% of the time then one should use the median value rather than the mean – it is for this reason that median forecasts are popular in financial forecasting.
5. The full probability density function (p.d.f.) or cumulative distribution function (c.d.f.) of a continuous variable. In other words, one could try to predict a whole function each time. If the distribution always has the same form, then it

may be possible to summarize this distribution by fitting it to a known distribution function (e.g. Normal/Gaussian) and then quote the distribution's parameters (e.g. the mean and variance of the Normal distribution).

6. Probability forecasts for whole spatial maps (e.g. precipitation maps). These forecasts are p.d.f.'s of a 2-dimensional spatial function and methods have not yet been fully developed for doing such forecasts.

Despite many years of issuing weather and climate forecasts, there is still much work to be done on producing and assessing such types of forecast. There is also a real need to educate the forecast users as to why probability forecasts are preferable and how they can best use them.

9.3 Basic Probability Concepts

9.3.1 Interpretations

Probability is open to many different interpretations. It underpins modern statistics and is essential for rigorous scientific enquiry. The concept was formally defined by Pierre Simon Laplace in his 1812 treatise on the analytical theory of probabilities and has caused much debate ever since. The word "probability" is derived from the Latin word *probare* meaning to test/approve, which is rather paradoxical since only when the probability is exactly one or zero can anything be definitely proven!

The probability p = P(A) of an event A is a measure between 0 and 1 of whether the event is likely to happen. When p = 1 the event is certain to happen, when p = 0 the event is impossible, and when p = 0.5 there is a maximum uncertainty about whether or not the event will happen.

In 1933, Andrey Kolmogorov formulated three basic axioms that a number has to satisfy to be a probability:

1. All probabilities are greater than or equal to zero: $P(A) \geq 0$ for all events (i.e. no event is more unlikely than a zero probability event).
2. The probabilities of all events in event space always sum up to one (i.e. some outcome must happen!).
3. The probability of either one or other mutually exclusive events (i.e. events that cannot happen at the same time) is equal to the sum of the probabilities of each event alone. In other words, $P(A \text{ or } B) = P(A) + P(B)$ for all mutually exclusive events A and B.

Note that the axioms can be restated more generally in terms of conditional probabilities, e.g. the probability of event A given event C occurs, P(A|C), rather than in terms of absolute probabilities such as P(A).

A simple way of communicating probability is in the form of the odds of an event. The odds of an event is defined as the ratio of the probability of the event occurring to the probability of it not occurring, i.e. P(A)/P(not A). So an event with probability 0.001 has odds of 1/999 (or 999:1 against in gambling jargon). Odds can range from zero to infinity and are equal to one for events whose occurrence and non-occurrence are equally likely (known as evens by gamblers). Odds can be used to assess the total risk of a set of independent events by simply multiplying together the odds of the individual events.

There are several different ways in which probability can be estimated and interpreted:

- Number of symmetric ways

If an event A can happen in k ways out of a total of m equally likely possible ways, then the probability of A is given by P(A) = k/m. For example, the probability of getting an odd number when throwing a 6-sided die is given by 3/6 since there are three ways to get an odd number (i.e. numbers {1,3,5}) out of a total of six equally likely outcomes {1,2,3,4,5,6}.

- Relative frequency of an event in repeated trials/experiments

For repeated trials, probability can be estimated by the "long-run" relative frequency of an event out of a set of many trials. If an event occurs k times in n trials then the relative frequency k/n provides an unbiased estimate of the probability of the event. In the asymptotic limit as the number of trials n tends to infinity, the relative frequency converges to the true probability of the event (by the "Law of Large Numbers"). This approach to defining probability from repeated trials is known as the "frequentist" interpretation. Note that unlike laboratory experiments, individual weather and climate events are unique and so can never be truly repeated!

- Subjective approach

The frequentist approach has a number of disadvantages. Firstly, it cannot be used to provide probability estimates for events that occur once only or rarely (e.g. climate change). Secondly, the frequentist estimates are based entirely on the sample and so cannot take into account any prior belief (e.g. common sense or scientific knowledge) about the event. For example, an unbiased coin could easily produce three heads only when tossed ten times and this would lead to a frequentist probability estimate of 0.33 for heads. However, our belief in the rarity of biased coins would lead us to suspect this estimate as being too low. In other words, the frequentist estimate would fail to reflect our true beliefs. In such cases a more flexible approach to probability must be adopted that makes use of not only the available sample of data but also incorporates any prior information. The word subjective does not mean that this approach is less rigorous than the frequentist approaches – instead it means that the estimated probability of an event will not necessarily be the same number for everyone but will depend on what prior information each person has.

One way to elicit the subjective probability of an event A from a group of experts is to ask what price B they would pay for a fair bet. The subjective probability p = B/W is then given by the price they would be prepared to pay to bet divided by the amount W they would win if the event occurred. Fair means that neither you nor the betting partner would be expected to make any net profit, i.e. $p(W-B)+(1-p)(-W)=0$. To make a fair bet, all prior knowledge must be taken into account, e.g. any bias in the coins, the previous form of a horse in a horse race, etc. This is often achieved by using Bayes' theorem (see next section) and so the subjective estimation of probability is often referred to as Bayesian estimation/inference.[1]

9.3.2 Joint and Conditional Probabilities

We are often interested in the situation when two events happen at the same time. For example, to get snow falling on the ground, it is necessary that two events, {A = "precipitating cloud"} and {B = "boundary layer below freezing"} occur at the same time.

The joint probability $P(A \text{ and } B)$ of events A and B is the probability that the two events occur together. The conditional probability $P(A|B)$ of A given B, is defined as $P(A|B)=P(A \text{ and } B)/P(B)$ and gives the probability of A occurring given that B has occurred. For example, to estimate the conditional probability of rain during El Niño episodes, one would estimate the probability of rain only during El Niño events rather than over all events. The concept of conditioning is fundamental for understanding statistical models (see Section 9.3.4). The unconditional probabilities P(A) and P(B) are known as marginal probabilities and so the joint probability $P(A \text{ and } B)$ is the product of the conditional probability $P(A|B)$ and the marginal probability P(B).

These ideas can be illustrated by considering exclusive, exhaustive, and/or independent events:

- Exclusive events are events that cannot occur simultaneously so $P(A \text{ and } B)=0$, and $P(A|B)=0$ if $P(B)>0$.
- Exhaustive events are events that describe all the possible outcomes and so $P(A \text{ or } B)=1$. It can be shown that for such events $P(A \text{ and } B) = P(A)+P(B)-1$ (by using the probability identity $P(A \text{ or } B)= P(A) + P(B) - P(A \text{ and } B)$).

[1] For a good online introduction, see: http://en.wikipedia.org/wiki/Bayesian_inference

- Independent events are ones where $P(A \text{ and } B) = P(A)P(B)$ and so $P(A \mid B) = P(A)$. The probability of A is unaffected by conditioning on B. Events where P(A|B) differs from P(A) are known as dependent events since the occurrence or non-occurrence of event B affects the chance of event A occurring.

Events can share one or more of these properties, for example, the events heads and tails of a single coin toss provide examples of exclusive and exhaustive events. Exclusive independent events can only occur if one or more of the events have zero probability of occurring.

9.3.3 The Prosecutor's Falacy and Bayes' Theorem

To assume that $P(A \mid B) = P(B \mid A)$ results in a mistake known to the legal profession as the Prosecutor's fallacy. One form of the fallacy results from neglecting the a priori odds of a defendent being guilty – i.e. the chance of an individual being guilty absenting specific evidence is the gross incident rate of perpetrators in the general population. When a prosecutor has collected some evidence B (for instance a DNA match) and has an expert testify that the probability $P(B \mid A)$ of finding this evidence if the accused were innocent (event A) is tiny, the fallacy occurs if it is concluded that the probability of the accused being innocent $P(A \mid B)$ must be comparably tiny. The probability of innocence $P(A \mid B)$ would only necessarily be comparably tiny if the probability of innocence P(A) is comparable to the a priori presumption of guilt P(B).

By equating:

$$P(A \text{ and } B) = P(A \mid B)P(B) \text{ and } P(A \text{ and } B) = P(B \mid A)P(A),$$

one can derive the very useful identity known as Bayes' theorem:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{9.1}$$

Bayes' theorem provides a useful way of getting from the unconditioned prior probability P(A) to the posterior probability $P(A \mid B)$ conditioned on event B. A is the event to be predicted, P(A) are ones prior beliefs about A, and B is the sample of data available (e.g. numerical climate model predictions). In other words, by conditioning on newly available data, it is possible to update ones estimate of the probability of event A.

9.3.4 Regression as a Conditional Probability Model

Empirical forecasting and forecast calibration and verification rely heavily upon regression models. Although it might not seem obvious at first sight, regression models are models for conditional probabilities. Consider, for example, the linear regression of a response variable, Y, on an explanatory variable, X:

$$Y = \beta_0 + \beta_1 X + \varepsilon \,. \qquad (9.2)$$

The parameters β_0, β_1, and σ_ε can be estimated by minimising the sum of the squared errors, which is known as Ordinary Least Squares (OLS) estimation. This is equivalent to Maximum Likelihood Estimation (MLE) if one assumes that the random effects are independent of one another and of X, and are Gaussian (Normally) distributed with zero mean and a constant variance:

$$\varepsilon \sim N(0, \sigma^2) \,. \qquad (9.3)$$

The symbols $\sim N(.)$ here mean distributed as a Normal distribution with parameters (.). Modern statistical terminology uses upper-case Roman letters to denote random variables (e.g. X), lower-case Roman letters to denote specific or observed/measured values of random variables (e.g. x), and Greek letters to denote unknown population parameters (e.g. β1, σ, etc.). For example, the probability of a random variable being more than x units above the mean is denoted $P(X - \mu > x)$.

The two OLS equations above can be written more elegantly as the following conditional probability model:

$$Y \mid X \sim N(\beta_0 + \beta_1 X, \sigma^2) \qquad (9.4)$$

In other words, the values of Y for a given value of X are normally distributed with a mean value given by $\beta_0 + \beta_1 X$ and a constant variance of σ^2. Hence, linear regression can be understood as a probability model/distribution for Y that has distribution parameters (the population mean) which depend linearly on X. The joint, marginal, and conditional probability distributions are illustrated in Fig. 9.1.

By writing regression models as probability models, it becomes evident how the models can be extended. For example, for processes with varying amounts of variance (heteroscedastic processes) the variance can also be made to depend on X or to model non-normal responses one can use a different distribution to normal. It can also be clearly noted that unlike correlation, regression has a direction – Y is conditioned on X, which is very different to X being conditioned on Y (see the earlier discussion about the Prosecutor's fallacy).

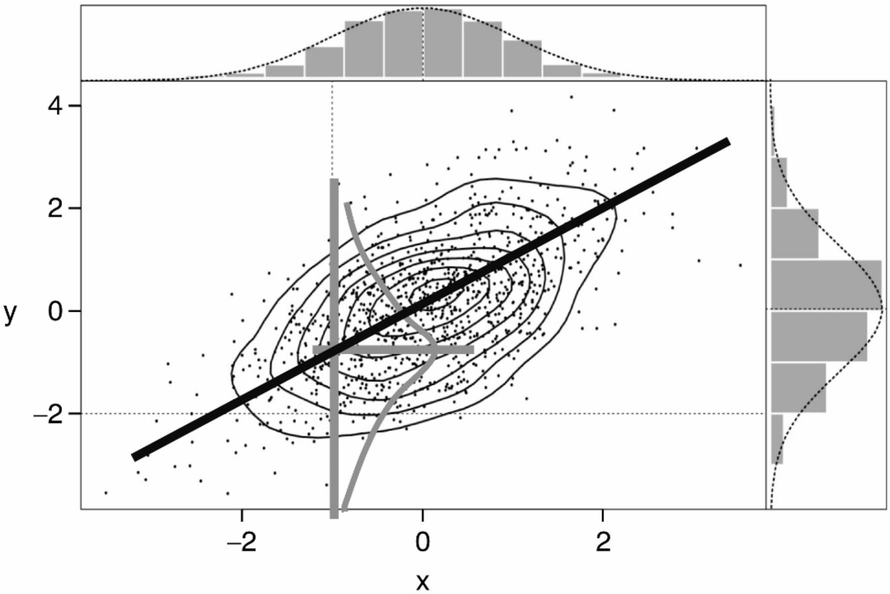

Fig. 9.1 Scatter plot showing marginal distributions (histograms at the edge of the plot) and isolines of the joint probability density together with a regression fit showing the conditional probability distribution at a specific value of x

9.4 Recalibration and Combination

9.4.1 Basic Ideas

Recalibration and combination are topics of fundamental importance in forecasting. There are essentially two main reasons why forecasts do not match observations:

- Forecasts are unable to discriminate between different observed situations
- Forecasts are poorly labelled, e.g. the forecasts are on average 5°C too warm

The ability of a forecasting system to discriminate between observed situations is known as forecast resolution, and its existence is a necessary yet not sufficient condition for forecasts to have any skill. For example, the probability forecasts of a day having rain rather than no-rain should on average be greater on observed rain days than on no-rain days. Forecast accuracy also depends on the good labelling of the forecasts, which is referred to as forecast reliability. For example, it should rain on average on 60% of the days when the forecaster issues probability of rain forecasts equal to 0.6. Forecast reliability can be improved by recalibration of the forecasts using previous pairs of forecasts and observations, whereas forecast resolution cannot generally be improved by recalibration. However, resolution can be improved by combining forecasts with other forecasts that are better able to resolve different situations.

Well-calibrated probability forecasts of future weather or climate variables can be produced in many different ways. One of the simplest methods for producing a probability forecasts is to estimate uncertainty on a deterministic prediction by fitting a probability distribution to a sample of past prediction errors (the differences between the past observed values and the corresponding predictions). Another more sophisticated approach is to develop a regression model of the observations on the predictions and then use the regression equation to make probabilistic predictions of future observables. This calibration approach using linear regression is known as Model Output Statistics (MOS) and has been widely used to improve deterministic predictions in the U.S. and elsewhere (see Glahn and Lowry 1972; Wilks 2005). Forecasts are improved by recalibration because numerical model predictions are only ever approximations to reality and so they will always have systematic prediction errors ("All models are wrong, but some are useful" – G. E. P. Box).

In order to quantify forecast uncertainty due to uncertainty in initial conditions, many operational forecasting centres are now producing ensembles of weather and climate forecasts rather than single deterministic forecasts. Since forecast users generally require well-calibrated probability forecasts, new synthesis methods have started to be developed for recalibration and combination of multi-model ensemble predictions. For example, the MOS methodology can easily be extended to multi-model predictions from several different numerical models by performing a multiple regression of the observations on the set of different predictors. There are many possible methods for combining forecasts, but no unique method can be prescribed that is ideal for all the types of weather/climate forecasting problems. However, there is a need to establish a framework that can incorporate the different approaches for combining weather and climate predictions in order to provide the most informative forecasts of future observables.

9.4.2 Conceptual Framework for Forecasting

Figure 9.2 shows a highly simplified low-dimensional schematic of the forecasting process. The state vector of the real atmosphere moves dynamically around q-dimensional observation state space whereas the model state vector moves around p-dimensional model state space (Stephenson et al. 2005). Three important steps are needed in order to find the probability density function $p(y_f \mid y_i)$ of a future observable variable y_f:

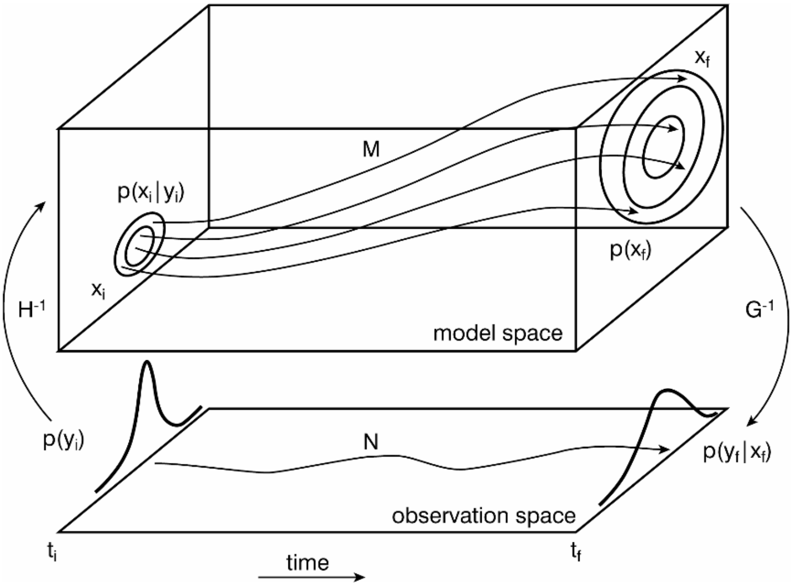

Fig. 9.2 Conceptual framework for forecasting. Note the duality between data assimilation that maps data from observation space into model space and forecast assimilation (calibration and combination) that maps model predictions back into forecasts of observable quantities (From Stephenson et al. 2005)

- Data assimilation to find $p(x_i \mid y_i)$. To initialize models with observations, information in observation state space need to be mapped into model state space (Bouttier and Courtier 1999).
- Ensemble prediction to find $p(x_f \mid x_i)$. The desired probability density is obtained by approximating the integral over all possible model states x_i by a finite sum of predictions from randomly generated initial conditions (Monte Carlo sampling). For this to be a good approximation, the initial ensemble states should be randomly sampled from the distribution $p(x_i \mid y_i)$. This condition is not generally satisfied in current operational ensemble forecasting systems (Stephenson and Doblas-Reyes 2000).
- forecast assimilation to find $p(y_f \mid x_f)$. A procedure for mapping the model predicted state back into observation space. This important final stage has been referred to as forecast assimilation by Stephenson et al. (2005) due to its analogy to data assimilation (apparent in Fig. 9.2). Forecast assimilation is a generic framework that incorporates all other post-processing approaches such as bias-correction, statistical downscaling, model output statistics, perfect prognosis, etc. (Wilks 2005).

It is often naïvely assumed that in the absence of systematic errors (Chapter 8, Section 8.2), predicted model variables are equal to forecasts of observations (i.e. $x_f = y_f$). This is mathematically incorrect since it ignores the important distinction between model state space and observation state space – the two spaces are not the same. Model variables (e.g. grid point variables) are only ever representations of observable variables (e.g. measurements at specific locations). This fallacy then leads to the probability distribution of ensemble predictions $p(x_f \mid y_i)$ being mistaken for the real quantity of interest: the posterior predictive distribution of the observables $p(y_f \mid y_i)$. Instead of equating model predictions with observables (i.e. $y_f = x_f$), model predictions should instead be considered as proxy information that can be used to infer the probability of future observables.

In order to be able to convert ensembles of model predictions into a probability forecast, forecast assimilation requires a probability model (e.g. regression) for linking x_f to y_f. Bayes' theorem can be used to estimate the conditional probabilities from the unconditional (uninformed) probability distributions. For example, variational data assimilation uses

$$p(x_i \mid y_i) = \frac{p(y_i \mid x_i)p(x_i)}{p(y_i)} \tag{9.5}$$

to update the prior (background) distribution $p(x_i)$ to obtain the posterior distribution $p(x_i \mid y_i)$. Similarly forecast assimilation can make use of

$$p(y_f \mid x_f) = \frac{p(x_f \mid y_f)p(y_f)}{p(x_f)} \tag{9.6}$$

to update the prior (e.g. the climatological) distribution $p(y_f)$ to obtain the more certain posterior distribution $p(y_f \mid x_f)$.

The following sections demonstrate these concepts with a few simple examples.

9.4.3 Forecasts of a Binary Event

The Bayesian approach is best illustrated using the simple example of forecasts of a binary event labelled by the random variable Y = 0 or 1 (e.g. no-rain/rain). Suppose that out of an ensemble of m forecasts, x forecasts predict that the event will occur and $m - x$ forecasts predict that it will not occur. The frequentist estimate for the probability $p = P(Y = 1 \mid X = x)$ is the relative frequency x/m. However, this estimate has several serious disadvantages:

- When $x = 0$ and m, the forecaster issues probabilities of 0 and 1, respectively. In other words, the forecaster states that the event is either completely impossible or completely certain to occur. It is unlikely the forecaster really believes this statement!

- The probabilities can take only the finite set of discrete values 0, 1/m, 2/m, ..., 1. Different size ensembles lead to different sets of probability values that can make comparison and interpretation difficult.

All these disadvantages can be avoided by adopting a more Bayesian approach. The Bayesian approach uses Bayes' theorem to update prior knowledge about p, described by the prior probability density function $f(p)$, using information contained in the model prediction data x:

$$f(p \mid x) = \frac{f(x|p) f(p)}{f(x)} \tag{9.7}$$

Note that uncertainty about the probability of the future event is now incorporated by using probability distributions rather than single point values of p. Provided the distribution of p is not multi-modal, the probability density function of the probability can be modelled quite flexibly using the two-parameter Beta distribution:

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \tag{9.8}$$

where $\Gamma(.)$ is the Gamma function is a normalising constant that is required to ensure that the integral of the probability density $f(p)$ from $p=0$ to 1 equals one. The probability is said to be Beta distributed as follows: $p \sim \mathrm{Beta}(\alpha,\beta)$. The mean and variance of the Beta distribution are given by:

$$\mu = \frac{\alpha}{\alpha+\beta}, \tag{9.9a}$$

and

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{9.9b}$$

Some examples of Beta distributions are shown in Fig. 9.3.

Using Bayes' theorem, it can be shown that when the prior is $p \sim \mathrm{Beta}(\alpha,\beta)$, and the number of predicted events is binomially distributed, $X \mid p \sim \mathrm{Bin}(m,p)$, with probability p, then the posterior distribution $p \mid (X=x) \sim$ $\mathrm{Beta}(\alpha+x,.$ $\beta+m-x)$ In other words, the effect of the model predictions is simply to update the Beta parameters describing the probability distribution for the probability of the event: $\alpha \to \alpha+x, \beta \to \beta+m-x$.

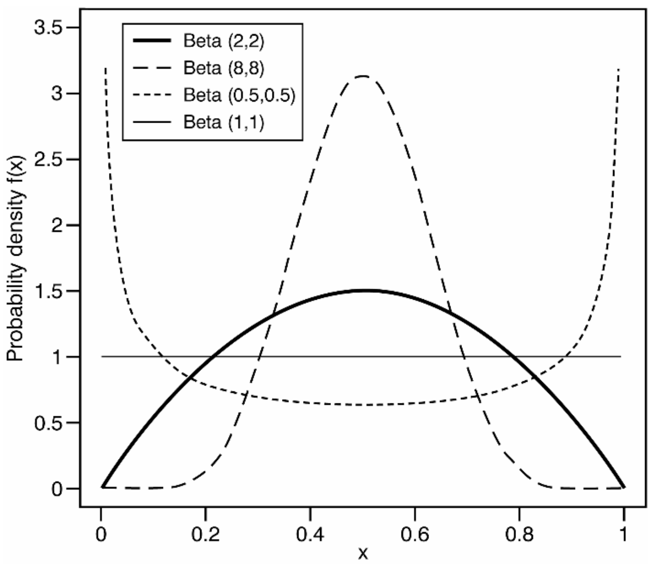

Fig. 9.3 Some examples of symmetric beta distributions: Beta(2.0,2.0), Beta(8.0,8.0), Beta(0.5,0.5), and Beta(1.0,1.0)

An example of this is shown in Fig. 9.4 where it is assumed that the prior mean probability of the event is 0.2 (e.g. climatology) and then this is updated by knowledge that $x = 0$ out of $m = 9$ ensemble predictions forecast the event. For the sake of argument, it is assumed here that information in the prior is equivalentto 6 ensemble forecasts and so $\alpha + \beta = 6$ and hence $\alpha = 1.2$ and $\beta = 4.8$. This kind of expert judgement about relative information in the data is needed in order to be able to define the prior. By comparing panels (a) and (b) in Fig. 9.4, it can be noted that the effect of the ensemble predictions in this example is to sharpen the distribution (reduce the uncertainty in the probability of the event) and to shift it towards zero. If a forecast user wanted a point summary for the probability forecast, one could issue the mean probability value $(\alpha + x)/(\alpha + \beta + m) = 0.08$ or the posterior mode (the most probable probability value):

$$(\alpha + x - 1)/(\alpha + \beta + m - 2) = 0.015\,.$$

The uncertainty in p can be summarised by the standard deviation of the posterior distribution:

$$\sqrt{(\alpha + x)(\beta + m - x)/(\alpha + \beta + m)^2(\alpha + \beta + m + 1)} = 0.068.$$

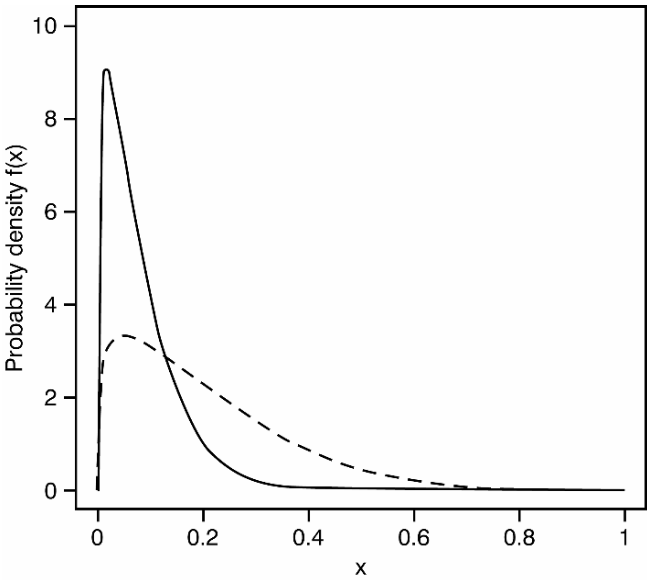

Fig. 9.4 Comparison of prior and posterior distributions of a probability forecast for a binary event: prior distribution Beta(1.2,4.8) (dashed line) and posterior distribution Beta(1.2,13.8) (solid line)

9.4.4 Forecasts of Normally Distributed Variables

The equations for forecast assimilation of normally distributed predictions are simply the dual of the equations used for variational data assimilation but with x and y interchanged. One assumes that the $q \geq 1$ observable variables y and the $p \geq 1$ model predictions x are (multivariate) normally distributed as follows:

$$y = y_b + \varepsilon_C \,, \tag{9.10a}$$

and

$$x = G(y - y_0) + \varepsilon_S \,, \tag{9.10b}$$

where y_b is the background observable state (e.g. the climatological mean value or a persistence forecast) and ε_C and ε_S are (multivariate) normally distributed errors with zero mean and background observable covariance C and forecast error covariance S, respectively. For generality, a bias term y0 needs to be included to take account of the mean forecast bias often found in model predictions. The $(p \times q)$ matrix G is the forecast likelihood operator that can be estimated by multivariate regression of the model predictions on the observed values. The equations can be rewritten more elegantly as the following probability models:

$$y \sim N(y_b, C)\,, \tag{9.11a}$$

and

$$x \mid y \sim N(G(y - y_0), S)\,. \tag{9.11b}$$

Then Bayes' theorem can be used to show that

$$y \mid x \sim N(y_a, D)\,, \tag{9.12}$$

with the forecast observable state ya and the forecast error covariance D given by

$$y_a = y_b + L(x - G(y_b - y_0))\,, \tag{9.12a}$$

$$L = CG^T (GCG^T + S)^{-1}\,, \tag{9.12b}$$

$$D = (I - LG)C = (G^T S^{-1} G + C^{-1})^{-1}\,. \tag{9.12c}$$

The $(q \times p)$ matrix L is the forecast gain/weight matrix that quantifies the relative contribution of the predictions to updating the prior background mean y_b. The model prediction data updates the background observable state to give an improved forecast of the observable. The forecast observable state is the mode of the posterior distribution $p(y \mid x)$, and is referred to as the Maximum A Posteriori Estimate (MAPE) that should not to be confused with the Maximum Likelihood stimate (MLE) which maximises $p(x \mid y)$. The MAPE maximises the probability $p(y \mid x)$ or alternatively minimises the cost function $-2\log p(y \mid x)$, which is given up to a constant by

$$J = (y - y_b)^T C^{-1} (y - y_b) + (x - G(y - y_0))^T S^{-1} (x - G(y - y_0))\,. \tag{9.13}$$

The cost function is the sum of two penalty terms: one that penalises departures from the background observable state and one that penalises departures from calibrated model predictions. This variational formulation of forecast combination and calibration is analogous to the variational formulation of data assimilation (e.g.

3-d VAR), and can be implemented in a continuous manner using Kalman filter and other state space approaches. Two examples are briefly presented to illustrate the power of this approach.

9.4.4.1 Example 1: NIÑO3.4 Index Forecasts

Coelho et al. (2004) used the above approach to recalibrate and combine multimodel hindcasts (past forecasts) of the NIÑO3.4 index in December starting from the preceding July. The hindcasts were freely provided by the European Union project DEMETER.[2] A similar Bayesian approach (but with the inclusion of ensemble spread information in the likelihood model) was also used by Coelho et al. (2003) to assess the additional skill provided by coupled model seasonal forecasts of the NIÑO3.4 index produced by different versions of the seasonal forecasting system at ECMWF.

A least squares regression of historical December values of the NIÑO3.4 index on preceding July values was used to define a simple prior probability forecast. Forecasts made in cross-validation mode (i.e. omitting the year to be forecast when estimating the regression parameters) generated by this empirical approach are shown in Figure 9.5a. The empirical scheme shows some skill in forecasting the observed values that arises because of the persistence during ENSO events. Note that all except one of the 13 observed values fall within the 95% prediction interval. The empirical forecast is, by definition, designed to be well-calibrated and so on average only 1 in 20 observations should fall outside the prediction interval.

This is certainly not the case for interval forecasts based on the ensemble mean and spread of the raw uncorrected 9-member ECMWF coupled model forecasts (Figure 9.5b). The coupled model forecasts give a narrower prediction interval that fails to contain the majority of the observations. This reliability problem is due primarily (but not entirely) to the coupled model forecasts being too cold. In addition to being too cold, the coupled model forecasts also have less variance than that seen in the observations.

Figure 9.5c shows the forecasts obtained by Bayesian combination of the statistical and coupled model forecasts. These forecasts have narrower and better calibrated prediction intervals than those of the statistical or coupled model forecasts alone. The combination and recalibration has helped to improve the precision and accuracy of these interval forecasts.

[2] See: http://www.ecmwf.int/research/demeter

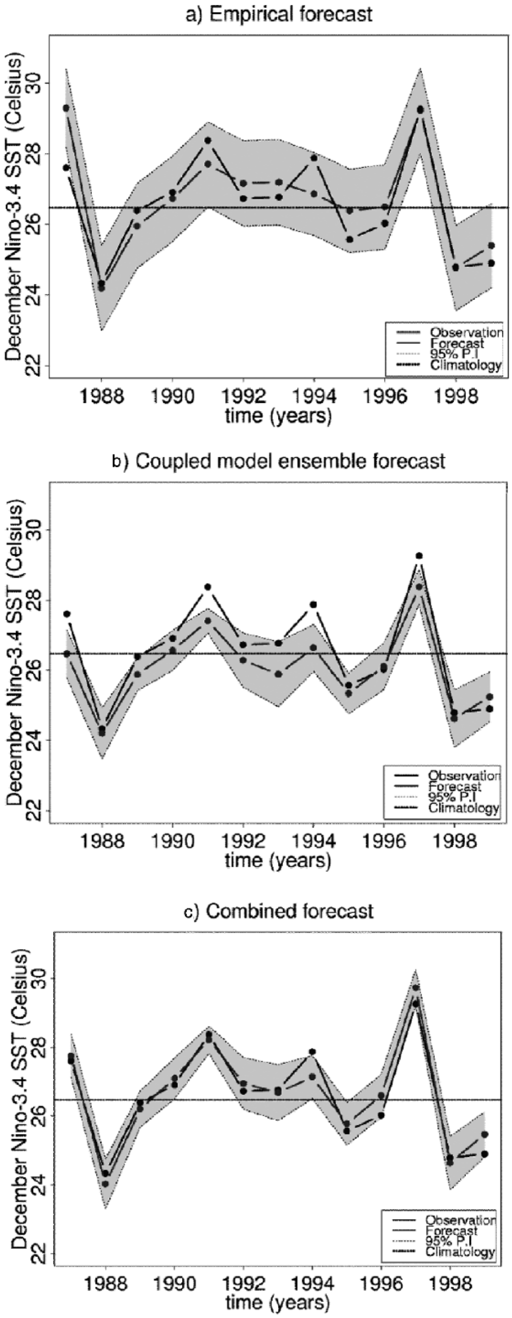

Fig. 9.5 Interval forecasts of December NIÑO3.4 index starting from preceding July: (a) empirical regression forecasts, (b) raw ECMWF ensemble mean coupled forecasts, and (c) the Bayesian combined forecasts. Observed values (circles), mean forecasts (circles at midpoint of grey shading), and 95% prediction interval (grey shading)

9.4.4.2 Example 2: Equatorial Pacific Sea Surface Temperatures

Inspired by the success of this approach, the forecast assimilation equations were then tested for multi-variable (i.e. grid point data) multi-model ensemble predictions. Stephenson et al. (2005) demonstrated the method on 6-month lead forecasts of equatorial Pacific sea surface temperatures.

Figure 9.6a shows longitude-time plots of probability forecasts for the binary event of SSTs being below their long-term mean value. The naïve multi-model approach calculated the probabilities by fitting Normal distributions to the seven model ensemble means at each of the 56 grid points. The forecast assimilation approach used the distribution of historical values as the prior and then combined this with the ensemble means from the seven coupled models taking care to calibrate these all together using a multivariate regression to estimate the likelihood parameters. From Fig. 9.6a, it can be seen that the forecast assimilation approach was able to shift the patterns eastward in order to get the correct sign of SST anomaly west of the dateline. Figure 9.7 shows the Brier score (see Chapter 10) as a function of longitude for both these combination approaches – the forecast assimilation gives better skill (lower Brier score) in both the eastern and western Pacific.

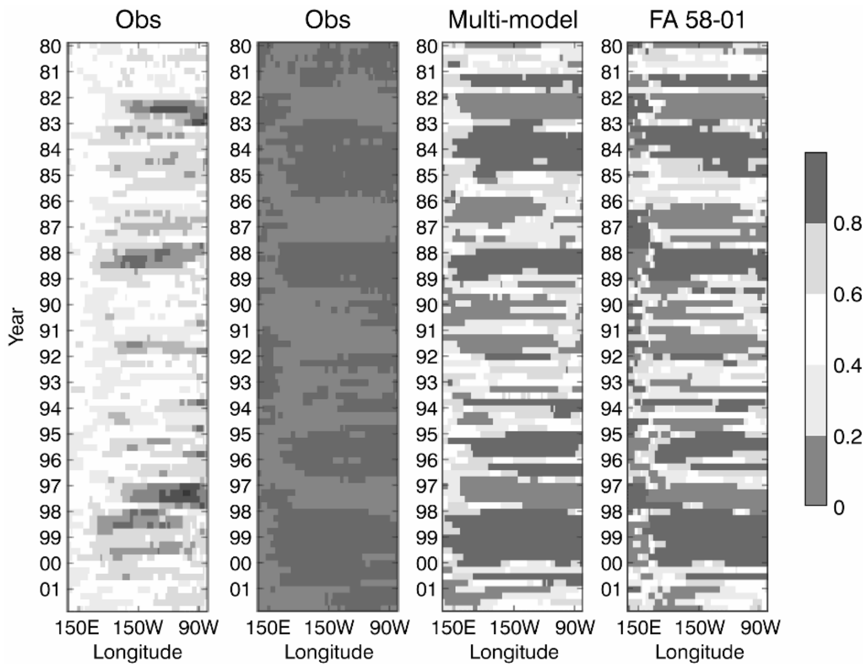

Fig. 9.6 Longitude-time plots of equatorial Pacific SST observations and forecasts: (a) observed anomalies 1980–2001, (b) binary event of observed anomaly less than or equal to zero, (c) multi-model ensemble mean and variance probability forecast of the event, and (iv) the forecast assimilation probability forecast (From Stephenson et al. 2005)

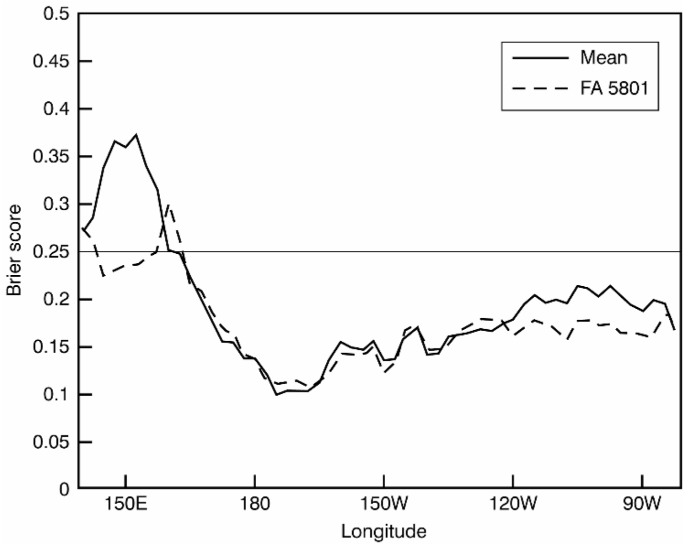

Fig. 9.7 The Brier score for the forecasts as function of longitude (From Stephenson et al. 2005)

This pilot study demonstrated that the forecast assimilation procedure could be successfully applied to multi-model grid point fields. Coelho et al. (2005) went on to apply the approach to improve hindcast predictions of seasonal mean precipitation over South America. The success of this approach has led to the EUROBRISA Project, a new transatlantic initiative, that will use this approach in real-time to improve seasonal forecasts for South America (see the EUROBRISA web site for more details[3]).

9.5 Summary

The main ideas in this chapter are that:

- Predictions of model variables are fundamentally different to forecasts of observables. Model predictions need to be mapped back into observations as an inherent part of the forecasting process (forecast assimilation).
- The probability distribution of ensemble predictions is not what we need to know. What we need are estimates of the conditional probability distribution of the future observed quantity given the available sample of ensemble prediction data.

[3] See: http://www.met.rdg.ac.uk/~swr01cac/EUROBRISA

Bayesian methodology allows one to estimate this distribution by simultaneously combining and recalibrating the ensemble predictions. It can be used to produce a reliable (well-calibrated) posterior distribution that avoids making unbelievable statements (e.g. such as issuing probabilities of 0 and 1).

Chapter 10
How Do We Know Whether Seasonal Climate Forecasts are Any Good?

Simon J. Mason and David B. Stephenson

When seasonal climate forecasts are expressed probabilistically, it is not possible to answer simple questions such as "how often are the forecasts correct?" The simpler attributes of forecast quality, such as "accuracy" or "correctness" are not applicable to probabilistic forecasts, and instead the main attributes of interest are: reliability, which defines whether the confidence communicated in the forecasts is appropriate; resolution, which defines whether there is any usable information in the forecasts; discrimination, which defines whether the forecasts are discernibly different given different outcomes (somewhat similar to the attribute of resolution); and sharpness, which defines the level of confidence that is communicated in the forecasts (regardless of whether that level is appropriate). How these attributes are measured depends on how the forecasts are expressed. In this chapter these attributes are explained in detail, and representation by various graphical procedures and scoring metrics is described. Partly because there is more than one desirable attribute to good probabilistic forecasts, it is argued that there is no single scoring metric that can adequately summarise forecast quality, and that in many cases graphical procedures also hide important aspects of forecast quality. The aim in this chapter is to provide some guidelines for interpreting and recognising the strengths and limitations of the most important verification tools as applied to seasonal climate forecasts.

Simon J. Mason
International Research Institute for Climate and Society

David B. Stephenson
University of Exeter

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

10.1 Introduction

Forecast verification is an essential part of atmospheric science: the science of meteorology is ultimately judged by the skill of its predictions. Forecast verification is a multi-disciplinary area of research that requires careful summary and interpretation of pairs of past forecasts and observations. A comprehensive overview of forecast verification is presented in Jolliffe and Stephenson (2003); only a brief summary of issues can be presented here, and so the focus is on topics that are not discussed at length there. Specifically, because of the probabilistic nature of seasonal climate forecasts, this chapter considers only the verification of probability forecasts and of ensembles of forecasts more generally.

The chapter only considers procedures for indicating the quality of forecasts as opposed to their value; "forecast quality" is concerned with how well the forecasts match the observations, whereas "forecast value" is concerned with the benefit (whether economic, social, or otherwise) that can be realised through decisions made in response to the forecasts. In focussing on questions of quality, the *potential* for forecasts to have value is addressed; whether the forecasts can actually be used to realise that value raises questions about the impact of the climate conditions that verify, and about the options available for mitigating such impacts. Using even the simplest of decision-making models it can be demonstrated that forecasts with high quality can have negative value. For example, one such model, namely the cost-loss model, posits a specific "loss" resulting from the occurrence of adverse climate conditions, and a specific "cost" that can be incurred to mitigate these costs entirely if action is taken in advance. Given a set of forecasts and observations, it is possible to compare the costs and losses that would be incurred with and without forecasts. Despite its over-simplicity, the model is useful in demonstrating that seasonal forecasts can have value only under certain conditions: the relative costs of taking some actions compared to the losses mitigated can result in dis-benefit, even with high quality forecasts. Readers interested in procedures for estimating forecast value should consult the book by Katz and Murphy (1997).

A primary theme of the current chapter is that just as forecast quality is a necessary, but not sufficient, condition for forecasts to have value, so also individual attributes of forecast quality are necessary but not sufficient for "good" forecasts. In the following section the complex nature of forecast verification is indicated. The impossibility of summarising the quality of a set of forecasts by a single number is emphasised; because of the multifaceted nature of forecast quality, any single metric inevitably hides important information about the quality of the forecasts. Some graphical procedures are detailed (Section 10.2.1) that provide a more comprehensive indication of quality than is possible using scores. Nevertheless, for good and bad reasons, scores remain popular, and since there are large numbers of verification scores that have been proposed, the properties of such scores for probability forecasts are considered in Section 10.3 so as to provide criteria for identifying which scores may be preferable to others. In Section 10.4, some examples of commonly used verification scores for probability forecasts are examined.

10.2 Attributes of Good Probability Forecasts

Perhaps the single most commonly asked verification question is "How often are the forecasts correct?" Although this question has intuitive appeal, when forecasts are presented as probabilities, questions about the "correctness" or otherwise of forecasts become unanswerable. Instead, probability forecasts are assessed on the basis of whether they reliably indicate changes in the uncertainty of the outcome: the forecasts are considered "reliable" when the forecast probability is an accurate estimation of the relative frequency of the predicted outcome (Murphy 1993).

Reliability, however, is not the only attribute of probability forecasts that is important. If the climatological probability of an outcome can be estimated accurately in advance, a set of forecasts that always indicate the climatological probability will be reliable, but will not provide any indication of the changing likelihood of the outcome from case to case. A second attribute, namely that of "resolution", is therefore important. Probability forecasts have good resolution when they can successfully distinguish cases in which the probability of an event is high from those in which the probability is low. Forecasts with good resolution will have varying probabilities from forecast to forecast, and the more these probabilities diverge from the climatological probability, the sharper the forecasts are said to be. From an alternative perspective, if forecasts are good, the discrimination between the forecasts will be clearly defined given different outcomes.

Good probability forecasts will have good reliability as well as high resolution (and, implicitly, high sharpness), and will be well-discriminated. How these various attributes are measured depends to a large extent on the format of the probability forecasts. In the following sections the definitions of these attributes are considered in more detail. In the following discussions various scores are mentioned that aim to measure only a specific attribute of the quality of a set of forecasts. In each case, with the exception of the ROC area (Section 10.2.3), these scores are distinct from scores that attempt to provide an overall summary of forecast quality. Discussion about the summary scores is reserved until Section 10.4.

10.2.1 Reliability

10.2.1.1 Definition

As discussed in Chapter 8 (Section 8.5.2), one objective in generating an ensemble of forecasts is to obtain an indication of the uncertainty in a forecast. However, it cannot automatically be assumed that the distribution of the ensemble members reliably indicates the true uncertainty: a decrease in the variance of the ensemble members does not necessarily mean that the outcome has become less uncertain. If the implied uncertainty in the forecasts is appropriate, the forecasts are said to be reliable or well-calibrated. Specifically, reliability is defined as consistency between the a priori predicted probabilities of an event and the a posteriori observed relative frequencies of this event. Reliability is measured in different ways depending on how the uncertainty in the forecast is indicated (see Chapter 9, Section 9.2.3 for an introduction to the different ways in which probability forecasts can be expressed).

10.2.1.2 Reliability of Interval Forecasts

Reliability is calculated most simply when forecast uncertainty is indicated using prediction intervals. In this case the forecast confidence is kept fixed, and so reliability can be assessed by comparing the coverage probability (sometimes called "capture rate": the proportion of times the observed value is contained within the prediction interval) with the confidence level for the intervals. If the observed value falls too infrequently (or frequently) within the range defined by the prediction intervals then the forecasts are over-confident (under-confident).

To illustrate, two sets of forecasts of the December values of the NIÑO3.4 index for 1981–2000 are shown in Table 10.1. The forecasts were obtained by simple linear regression using either the June or the September values of the index as predictors. The models were trained using data for 1951–1980. Prediction intervals were calculated based on the cross-validated error variance over the training period (Chapter 7, Section 7.3.3), and the widths of the intervals were set to define a 50% level of confidence (i.e. 50% of the intervals are expected to contain the observation). For both sets of forecasts, eight of the 20 years (40%) are contained within the prediction intervals. The intervals are therefore too narrow, and the forecasts are thus over-confident.

Although they have intuitive appeal, there are a number of problems with using coverage probabilities as measures of forecast quality. The first problem is that this measure of reliability does not distinguish between sets of predictions with similar coverage probabilities but different interval widths. For example, both sets of forecasts in Table 10.1 have equal reliability, but the forecasts from September have consistently narrower intervals than those from June, and so are more informative (the narrower intervals imply less uncertainty in the forecast). A related problem is that the correct coverage probability, p say, can be achieved by unskilful forecasts simply by making the prediction interval infinitely wide p% of the time, and infinitely narrow the remaining times. These problems point to the impossibility of adequately representing forecast quality by a single score. More specifically, reliability is a necessary but not a sufficient attribute of a good set of forecasts (Murphy 1991).

Table 10.1 Observed values and forecasts of the December 1981–2000 NIÑO3.4 index. The upper and lower 50% prediction intervals are indicated, and intervals that capture the observed value are shaded

Years	Obs	June	September
1981	–0.105	–0.391 (–0.807 to 0.0.25)	–0.191 (–0.494 to 0.112)
1982	2.590	2.019 (1.567 to 2.472)	1.998 (1.673 to 2.323)
1983	–0.464	1.343 (0.911 to 1.775)	0.044 (–0.258 to 0.347)
1984	–1.238	–0.811 (–1.231 to –0.390)	–0.017 (–0.319 to 0.286)
1985	–0.212	–0.725 (–1.144 to –0.305)	–0.253 (-0.556 to 0.050)
1986	1.261	0.283 (–0.133 to 0.699)	1.225 (0.914 to 1.536)
1987	1.167	2.218 (1.758 to 2.678)	2.421 (2.086 to 2.756)
1988	–1.892	–1.969 (–2.418 to –1.520)	–1.121 (–1.431 to –0.812)
1989	0.094	–0.812 (–1.233 to –0.391)	–0.207 (–0.510 to 0.095)
1990	0.491	0.323 (–0.093 to 0.740)	0.496 (0.192 to 0.800)
1991	1.756	1.569 (1.131 to 2.007)	0.769 (0.463 to 1.075)
1992	0.399	1.342 (0.910 to 1.775)	0.343 (0.040 to 0. 647)
1993	0.371	1.341 (0.908 to 1.773)	0.742 (0.436 to 1.048)
1994	1.272	0.807 (0.386 to 1.229)	0.903 (0.595 to 1.210)
1995	–0.785	0.249 (–0.167 to 0.665)	–0.451 (–0.755 to –0.148)
1996	–0.394	–0.127 (–0.542 to 0.268)	–0.155 (–0.458 to 0.147)
1997	2.629	2.272 (1.810 to 2.734)	2.955 (2.605 to 3.305)
1998	–1.366	–0.419 (–0.836 to –0.003)	–0.644 (–0.949 to –0.339)
1999	–1.408	–1.011 (–1.435 to –0.587)	–0.838 (–1.144 to –0.532)
2000	–0.695	–0.547 (–0.965 to –0.129)	–0.309 (–0.612 to –0.006)

10.2.1.3 Reliability of Probabilities for Categories

When forecasts are communicated as a variable probability assigned to a predefined category, reliability is effectively defined in the same way as for the prediction intervals: forecasts are reliable if the observation falls within the category as frequently as the forecast implies. The "observed relative frequency" (equivalent to the "coverage probability" for interval forecasts), has to be calculated for each distinct value of the forecast probability. For example, seasonal rainfall totals should be between 100 and 200 mm on 20% of the occasions in which the forecast probability for this interval is 20%, and on 40% of the occasions in which the forecast probability for this interval is 40%, etc.

The observed relative frequencies conditional upon the forecast probability can be plotted as reliability or attributes diagrams.[1] Although the diagrams are designed to show the reliability of forecast probabilities for a specific event (i.e. for a two-category system), because the definition of an event does not have to

[1] The attributes diagram is the same as the reliability diagram, but includes an additional line to indicate where resolution equals reliability (see Hsu and Murphy 1986; Mason 2004).

remain fixed, forecasts for multiple categories can be included in the calculations.[2] The interpretation of reliability diagrams may be facilitated by considering some idealised examples as shown in Fig. 10.1. If the forecasts are perfectly reliable then the observed relative frequency will equal the forecast probability for all values of the forecast probability, and so the reliability curve will lie along the 45° diagonal. In practice, even if forecasts have excellent reliability, sampling errors result in departures from the diagonal, and so some indication of how close the curve is to the diagonal may be required to assist in the interpretation of the curve (Bröcker and Smith 2007; Kumar 2007).

More typically the forecasts are not perfectly reliable anyway, and so the curve will lie off the diagonal illustrating one or more of the characteristics shown in Fig. 10.1. In Fig. 10.1a the forecast probabilities are consistently lower than the observed relative frequencies, indicating that the event always occurs more frequently than anticipated. In Fig. 10.1b the opposite is true, and the event occurs less frequently than anticipated. In these two cases the forecaster is under-/over-forecasting, respectively. For seasonal climate forecasts, the most common situation is indicated in Fig. 10.1c. Here the event occurs more frequently than indicated when the forecast indicates a decreased probability of the event occurring compared to climatology (to the left of the dotted line), but less frequently than indicated when the forecast indicates an increased probability of the event occurring compared to climatology (to the right of the dotted line). Although the forecasts correctly indicate increases and decreases in the probabilities of the events, the changes in probability are over-stated, and the forecasts are said to be over-confident. The greater the degree of over-confidence, the shallower is the slope of the curve. If the curve becomes horizontal there is no information in the forecasts: the relative frequency of the event equals the climatological probability regardless of the forecast probability, and the forecasts are said to have no resolution (Section 10.2.2). A fourth possibility is indicated in Fig. 10.1d, where the changes in the forecast probabilities understate the changes in the relative frequencies of the event, and the forecasts are said to be under-confident. In this case the forecasts have high resolution, but poor reliability.

An example of a reliability diagram is shown in Fig. 10.2. The diagram is based on 43 years of forecasts of NIÑO3.4 sea surface temperature (SST) anomalies, produced as part of the DEMETER project (Palmer et al. 2004). Forecasts from the ECMWF, Météo-France, and Met Office models for lead-times of 0–5 months and for four initialization seasons are included. For each model, nine ensemble members were available. The forecast probabilities were obtained by calculating the proportions of ensemble members forecasting temperatures in the coldest

[2] A separate multi-category reliability diagram showing the percentage of observations less than distinct percentiles of the forecast distribution has been proposed (Hamill 1997), but has not yet been widely adopted.

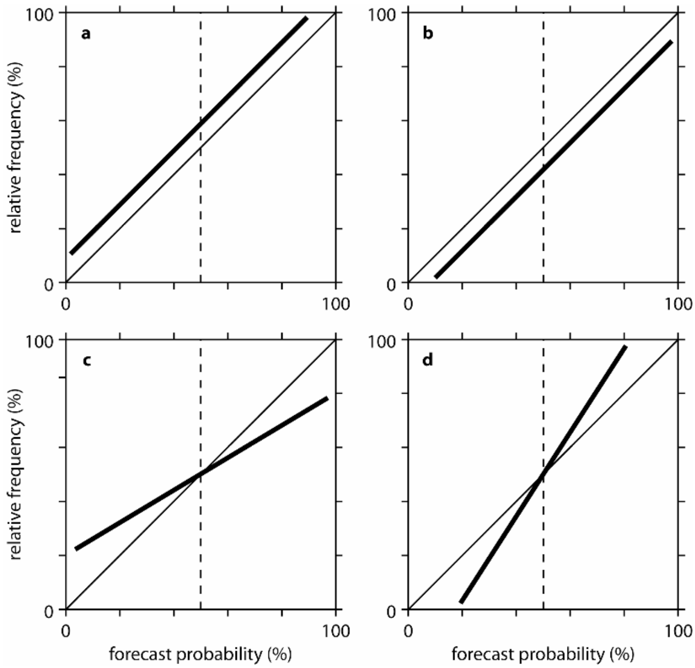

Fig. 10.1 Idealized reliability diagrams indicating cases of (a) under-forecasting, (b) over-forecasting, (c) over-confidence, (d) under-confidence. The vertical dotted line indicates the climatological probability of the event occurring, which in this case is set at 50%

(black) and warmest (grey) 25% of years, respectively. For seasonal forecasts it is standard to bin the forecast probabilities into 11 categories, with the first bin representing forecast probabilities of <5%, the second 5–15%, …, and the last ≥95%.[3] The frequencies with which forecasts in each bin occur are presented in a histogram. The reliability curves follow the 45° line closely indicating good reliability for forecasts of both anomalously warm and anomalously cold conditions. The histogram indicates that the forecast probabilities do not peak in frequency at the climatological probability of 25%, which is what would have been expected if the models had had little or no signal. Ideally, the forecasts should have high frequencies of probabilities close to 0% and 100%, whilst retaining reliability (i.e. the forecasts should be sharp), in which case the histogram would be *u*-shaped.

[3] Since it is possible to tweak the binning to optimize the impression of reliability, the WMO has recommended the procedure as adopted here. These recommendations are detailed in the Standardized Verification System for Long-Range Forecasts (SVSLRF). Further details about the SVSLRF, which contains a list of recommended verification procedures, are available from the WMO Lead Centre for Verification: http://www.bom.gov.au/wmo/lrfvs/

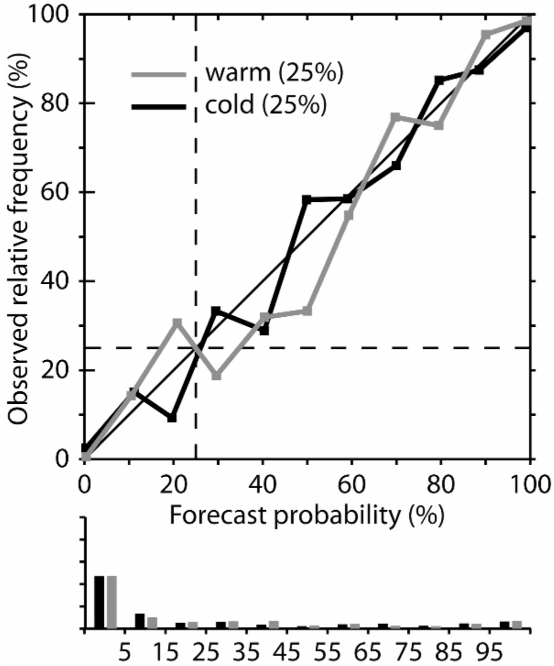

Fig. 10.2 Examples of reliability diagrams for ECMWF, Météo-France, and Met Office 0–5-month lead forecasts of NIÑO3.4 SST anomalies. The forecast probabilities were obtained by calculating the proportions of ensemble members forecasting temperatures in the coldest (black) and warmest (grey) 25% of years, respectively. The histogram indicates the frequency of forecasts of probabilities of <5%, 5–15%, etc.

However, the precise shape of the histogram of sharp forecasts depends on the climatological probability of the event (see Section 10.2.4).

A measure of the distance between the sample reliability curve and the diagonal is an intuitive measure of forecast reliability. A commonly used such metric is the reliability component of the Murphy (1973a) decomposition of the Brier score (Section 10.4.1). Assuming that there are m points on the reliability curve, and that the forecast probability for the kth point is p_k, the reliability score is defined as:

$$\text{reliability score} = \frac{1}{n}\sum_{k=1}^{m} n_k \left(p_k - \overline{o}_k\right)^2 , \tag{10.1}$$

where $\overline{o}_k$ is the observed relative frequency for the kth probability bin, and n_k is the number of forecasts in this bin.

The distances defined by Eq. (10.1) represent the differences between the various forecast probabilities and the corresponding observed relative frequencies (the probabilities that should have been assigned), and thus are measures of the average "error" in the forecast probabilities. Therefore small values of Eq. (10.1) represent good reliability (see the discussion on necessarily mean that the forecasts contain useful information; perpetual forecasts of the climatological probability have perfect reliability. Therefore, as discussed in Section 10.2.1.1, reliability is a necessary but not sufficient attribute of good forecasts.

Errors in calculating the observed relative frequencies for each forecast probability bin are binomially distributed with parameters n_k (the number of forecasts in bin k) and p_k (the average forecast probability for this bin). Given the limited sample sizes of seasonal forecasts, the number of forecasts in a given bin can be too small to give a meaningful estimate of the observed relative frequency (Bröcker and Smith 2007). To increase the sample size, pooling of forecasts is necessary, whether from different lead-times or seasons (as in Fig. 10.2), and or for different locations. Information about differences in the quality of the forecasts for the different lead-times, seasons, and locations is therefore masked, and there are good a priori reasons to expect the quality to differ (see Chapter 3). Pooling of forecasts for different locations is particularly problematic, not only because the forecast for proximate locations are unlikely to be independent (thus over-estimating the number of forecasts in each bin), but more specifically because it could be argued that the interpretation of the forecast probabilities is being changed. A (reliable) forecast probability of p% should imply that an event can be expected to occur on p% of the *occasions* a forecast with this probability is issued, but if forecasts are pooled for different locations the forecast is being verified with the interpretation that an event is expected to occur over p% of the *locations* at which a forecast with this probability is issued. For all these reasons, reliability diagrams have not been used extensively for seasonal forecasts, although when sufficient forecasts are available the use of the diagrams is promoted in SVSLRF.

10.2.1.4 Reliability of Ensemble Forecasts

Reliability diagrams are appropriate only for forecasts presented as probabilities of events (although the definition of an event does not have to remain fixed). A common method of indicating reliability when the forecast distribution is presented as percentiles is to use the ranked histogram (popularly called a Talagrand diagram). The ranked histogram is constructed by sorting the m ensemble members to form $m + 1$ bins, and then counting the numbers of times the observed value falls within each bin. If the forecast distribution reliably reproduces the distribution of possible outcomes then the observed value should be a random draw from this same distribution, and so should occur in each of the bins an equal number of times (Hamill 2001). The proportion of the total number of observations in each bin therefore should follow a uniform distribution, and a Cramér – von Mises test can be used to test for systematic errors (Elmore 2005).

Examples of ranked histograms are illustrated in Fig. 10.3. The histograms were constructed using 50 years of model simulations of September–November rainfall for Brisbane (Fig. 10.3a) and Kalgoorlie (Fig. 10.3b), Australia.[4] Nine ensemble members were considered, creating ten bins, and so if the ensemble distribution is reliable the observed rainfall would be expected to occur in each bin five times. For Brisbane (Fig. 10.3a), most of the observations are in the last bin, indicating that the observed rainfall is frequently more than the simulated rainfall of all nine ensemble members. The approximate upward slope of the histogram from left to right indicates that the simulated rainfall is negatively biased: the median observed rainfall over the 50-year period is about 180 mm compared to the median simulated rainfall of about 150 mm. In contrast, for Kalgoorlie (Fig. 10.3b) the mean bias is minimal (41 mm observed, 44 mm simulated), but the numbers of times that the observed rainfall is either less than or more than all nine simulated values (the first and last bins) is inflated. Inflated frequencies in the outermost bins indicate that the observed rainfall falls outside the range of the ensemble distribution too frequently, and the binned histogram is said to be *u*-shaped. A *u*-shaped histogram is often interpreted to be indicative of an ensemble variance that is too small (increasing the variance of the ensemble distribution would decrease the proportion of observations outside the ensemble range), but can also be a result of conditional bias (Hamill 2001). There is no significant correlation between the ensemble mean and the observed rainfall for Kalgoorlie, and so when dry (wet) conditions are forecast, the observed rainfall is likely to be more (less) than all the ensemble members.

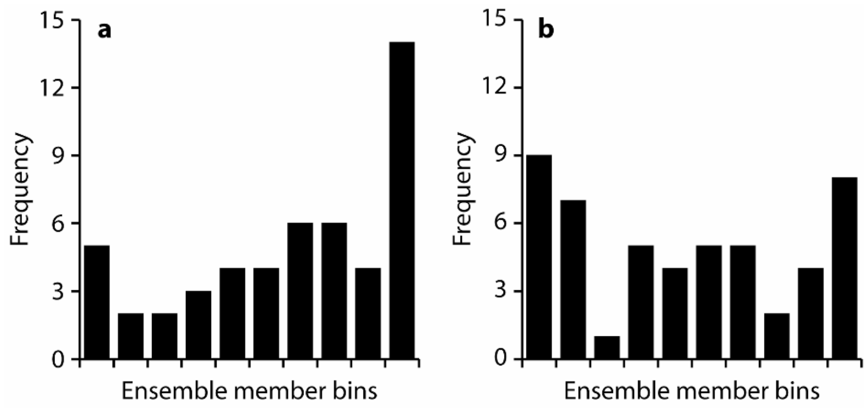

Fig. 10.3 Ranked histogram for ECHAM4.5 simulations of September–November 1951–2000 seasonal rainfall totals for (a) Brisbane and (b) Kalgoorlie, Australia

[4] The simulations are from the ECHAM4.5 atmospheric general circulation model (Roeckner et al. 1996), forced with observed SSTs. For this example the simulations for Brisbane and Kalgoorlie are taken simply as the model value for the grid containing 27.45°S, 153.03°E, and 30.78°S, 121.45°E, respectively.

Since it is possible to construct a perfectly uniform ranked histogram from forecasts that do not have good resolution (Hamill 2001), a modification to the histogram has been proposed to test that the probability of each of the bins remains constant regardless of the forecast. This probability, known as the conditional exceedance probability (CEP), is defined as:

$$P\left(x > \hat{x}_k \mid \hat{x}_k\right) = \frac{1}{1 + \exp\left(-\beta_{0,k} - \beta_{1,k}\hat{x}_k\right)}$$

where $\beta_{0,k}$ and $\beta_{1,k}$ are parameters to be estimated, x is the observed value, and $\hat{x}_k$ is the kth percentile of the forecast distribution. The CEP is useful for measuring whether the probability of the observation exceeding the ensemble median, for example, increases if the ensemble forecasts are all indicating anomalously dry conditions. If this probability is conditional upon the actual forecast values then it is argued that the forecasts from the ensembles are unreliable even if, over all the forecasts, the ensemble median is exceeded 50% of the time (Mason et al. 2007).

10.2.1.5 Multi-dimensional Reliability

Ranked histograms and CEP diagrams both consider the prediction of a single parameter at a single point. Since dynamical models produce predictions of multiple parameters at multiple points, there may be interest in verifying the joint distributions of these predictands. For example, it is possible that a model could produce reasonable forecasts of precipitation and of temperature, but produce unrealistic simultaneous forecasts of these two parameters. Recent developments in verification methodology have begun to address the need to assess a model's ability to predict reliable joint distributions of parameters. Two such procedures are considered here: minimum spanning trees and bounding boxes.

Minimum spanning trees are a multi-dimensional adaptation of the ranked histogram (Wilks 2004). An example of a minimum spanning tree is shown in Fig. 10.4a, showing the simulations (crosses) and observations (circle) of rainfall for Brisbane and Kalgoorlie for 2000. The tree is constructed by connecting each ensemble member with the nearest other member, and the total distance of all the connecting lines is then computed. This procedure is repeated replacing one of the ensemble members with the observed values, and thus treating the observation as if it were an ensemble member. The distances obtained using the outcome in place of each ensemble member are used to define the bins in the histogram. Then the distance obtained using all ensemble members without the outcome is binned. A histogram is constructed by repeating the procedure for all forecasts; as with the ranked histogram, if the outcome is indistinguishable from the ensemble members the minimum spanning distance for the tree constructed using all the ensemble members will be a random draw from the bins, and so the histogram should be level. In Fig. 10.4b, the histogram shows too many distances in the first bin, indicating that the replacement of an ensemble member with the observed values typically increases the total spanning distance (i.e. the ensemble members are not a good representation of the outcome). Downward sloping histograms can result from mean biases, an ensemble spread that is too small, and/or conditional biases (cf. ranked histograms, for which only the mean bias results in a sloping histogram in either direction depending on the sign of the bias). As with ranked histograms, a uniform minimum spanning tree histogram.

Bounding boxes are defined as the range of the ensemble predictions in k-dimensional space (Weisheimer et al. 2005). If the vector of observed values falls within the box then the outcome is interpreted as being consistent with the multi-dimensional distribution of the ensemble members, and thus is indistinguishable from an ensemble member. Bounding boxes are most commonly used with uncalibrated model output.

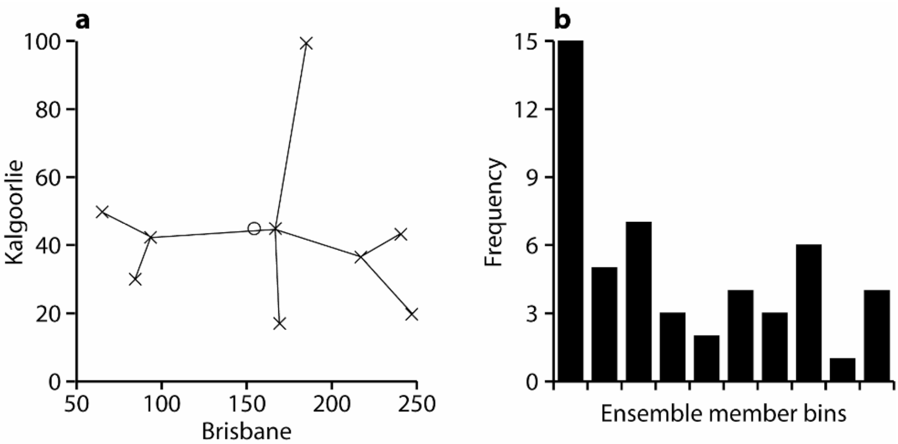

Fig. 10.4 (a) Minimum spanning tree for observed (circle) and ECHAM4.5 simulations (crosses) of September–November 2000 seasonal rainfall totals for Brisbane and Kalgoorlie, Australia. The tree is constructed for the case without the observation. (b) Minimum spanning tree histogram for 1951–2000

10.2.2 *Resolution and Discrimination*

10.2.2.1 Definition

It has been argued in the previous section that reliability is a necessary but not sufficient attribute of a good set of forecasts. Specifically, if the climatological probability of an event is known then a number of simple strategies can be devised to ensure that the forecasts are reliable, but which are otherwise uninformative.

What is required in addition to reliability is an ability to distinguish between cases when the probability of an event is inflated from cases when the probability is deflated. More precisely, forecasts have good resolution when the outcome is strongly conditioned upon the forecast.

10.2.2.2 Resolution Given Probabilities for Categories

Resolution is most clearly defined in the context of forecasts expressed as probabilities for categories: if the forecasts have good resolution and good reliability then the probability of an event occurring should increase (or decrease) when the forecast probability increases (or decreases). In this context the most commonly used measure of forecast resolution is, like that for reliability, based upon the Murphy (1973a) decomposition of the Brier score, and is closely related to the reliability diagram (Section 10.4.1). It is defined as:

$$\text{resolution score} = \frac{1}{n}\sum_{k=1}^{m} n_k \left(\overline{o}_k - \overline{o}\right)^2 , \qquad (10.2)$$

where $\overline{o}$ is the climatological probability of the event. Unlike the reliability score, the resolution score is not an error score, but can be interpreted as the weighted variance of the observed relative frequencies, with large variance representing good resolution (see Section 10.3.1 on score orientation). Note that because of the squaring in Eq. (10.2) forecasts that have an increase in the observed relative frequency with a decrease in the forecast probability will score equally well on resolution as forecasts that indicate the correct direction of change in the probability of an event. As with reliability, therefore, resolution is not a sufficient attribute of good forecasts.

10.2.3 Discrimination

10.2.3.1 Definition

The attribute of discrimination is similar to that of resolution, but considers the conditional distribution of the forecasts given the outcomes rather than of the outcomes given the forecasts. Whereas resolution is concerned with whether the expected outcome differs as the forecast changes, discrimination is concerned with whether the forecast differs given different outcomes. In the general framework for forecast verification introduced by Murphy and Winkler (1987), the first perspective is known as a calibration-refinement factorization, whereas the latter is called a likelihood-base rate factorization.

10.2.3.2 Discrimination Given Probabilities for Categories – the Relative Operating Characteristics (ROC)

The most commonly used method of identifying whether a set of forecasts is well-discriminated given different outcomes is the relative operating characteristics (ROC; sometimes called receiver operating characteristics) graph (Mason 2003). This procedure requires the outcome to be binary, just as in the case of a reliability diagram, and so separate results are usually calculated for each category if there are more than two categories. The rather horribly named ROC is actually equivalent to a non-parametric test commonly used for testing for differences in central tendency, namely the Mann-Whitney *U*-test[5] (Mason and Graham 2002). In the current context, the *U*-test can be applied to assess whether there is any difference in (i.e. discrimination between) the forecasts when an event occurs compared to when the event does not occur. The ROC is most commonly applied to probabilistic forecasts, in which case it indicates whether the forecast probability was higher when an event occurred compared to when not, but it can as easily be applied to deterministic forecasts (in which case it indicates whether the forecast rainfall, for example, is higher when rainfall is above-normal than when not).

As a simple example, Table 10.2a contains 30 years of retroactive probabilistic forecasts of above-normal (defined as observed rainfall being above the upper-tercile) December–February total rainfall for Lusaka, Zambia, obtained using a simple statistical model.[6] These forecasts can be ranked from most confident to least confident, and the ranks of the forecasts are shown in the third column. It seems reasonable to select the seasons with the highest probability (1973/74 and 1975/76, with forecasts of 65%) as the seasons in which one would be most confident about the observed rainfall being above-normal. Similarly, the season with the second highest probability (1975/76, with a forecast of 60%) would be the season in which one would be next most confident that observed rainfall was above-normal.

The table can be re-ordered so that the seasons are listed in order of the ranks of the forecasts rather than chronologically, as shown in Table 10.2b. The seasons for which one would be most confident observed rainfall was above-normal are

[5] More strictly, the Mann-Whitney *U*-test is used to test whether the probability that a sample from one population (e.g. a forecast for when rainfall is observed to be above-normal) has a value larger than that from another (e.g. a forecast rainfall is observed not to be above-normal) is 50%, but if assumptions are made about the distributions of the two populations (specifically, that they have similar shapes and variances) then the test can be used to compare the central tendencies (the medians) of the distributions (Sheskin 2007). These assumptions generally are irrelevant in the context of forecast verification.

[6] The data are based on the example in Chapter 7, Section 7.3.3. The retroactive procedure used an initial 10-year training period (1961/62–1970/71) to forecast the next year, and was updated each year so that the last training period for forecasting 2000/01 was 39 years long. Forecast probabilities were obtained from the error variance of the cross-validated predictions (see Chapter 7, Section 7.3.3), and then rounded to the nearest 5%.

then at the top of the table. If the forecasts are good, then the actual seasons in which rainfall was above-normal should be towards the top of the table. If the forecasts are effectively useless, the above-normal seasons will be randomly distributed through the table, and if they are bad these seasons will be towards the bottom of the table. The actual seasons of above-normal rainfall are marked by grey shading, and do appear preferentially to be towards the top of the table.

To construct the ROC, start at the top of the table and, treating each forecast as a prediction of an event (i.e. above-normal rainfall), count the proportion of correct forecasts (the hit-rate) and incorrect forecasts (the false-alarm rate). These scores are shown in Table 10.2b: for the highest ranking forecasts, rainfall was above-normal

Table 10.2 (a) Forecast probabilities and ranks (in descending order) for above-normal December–February 1981/82–2000/01 seasonal rainfall totals for Lusaka, Zambia, made using a simple statistical model. (b) Forecasts shown in order of descending rank, with corresponding hit and false-alarm rates. The 'events' (observed rainfall above the upper tercile) are indicated by grey shading

Year	(a) Forecast (%)	Rank	Rank	Year	(b) Hit rate	False-alarm rate
1971/72	45	4	1	1973/74		
1972/73	10	29	1	1975/76	1 of 10	1 of 20
1973/74	65	1	3	1988/89	2 of 10	1 of 20
1974/75	40	9	4	1971/72		
1975/76	65	1	4	1978/79		
1976/77	30	19	4	1995/96		
1977/78	30	19	4	1998/99		
1978/79	45	4	4	1999/00	4 of 10	4 of 20
1979/80	35	14	9	1974/75		
1980/81	35	14	9	1983/84		
1981/82	35	14	9	1984/85		
1982/83	20	23	9	1989/90		
1983/84	40	9	9	2000/01	7 of 10	6 of 20
1984/85	40	9	14	1979/80		
1985/86	35	14	14	1980/81		
1986/87	20	23	14	1981/82		
1987/88	15	27	14	1985/86		
1988/89	55	3	14	1996/97	9 of 10	9 of 20
1989/90	40	9	19	1976/77		
1990/91	25	22	19	1977/78		
1991/92	20	23	19	1992/93	10 of 10	11 of 20
1992/93	30	19	22	1990/91	10 of 10	12 of 20
1993/94	20	23	23	1982/83		
1994/95	15	27	23	1986/87		
1995/96	45	4	23	1991/92		
1996/97	35	14	23	1993/94	10 of 10	16 of 20
1997/98	5	30	27	1987/88		
1998/99	45	4	27	1994/95	10 of 10	18 of 20
1999/00	45	4	29	1972/73	10 of 10	19 of 20
2000/01	40	9	30	1997/98	10 of 10	20 of 20

in only one of the years and so one of the ten above-normal events were correctly identified, and one of the twenty non-events. The next highest ranking forecast (1988/89) was for a season that was above-normal, and so now two of the ten events have been correctly selected. Effectively Table 10.2b involves constructing a series of contingency tables in which forecasts of an event are issued using successively lower warning thresholds, *t*. Initially a warning is issued when the forecast probability is at least 65%, then the threshold is lowered to 55%, etc. So, at each point on the ROC curve, the hit and false-alarm rates, $HR(t)$ and $FR(t)$ respectively, are calculated as:

$$HR(t) = P\left(p \geq t \middle| x = 1\right), \tag{10.3}$$

$$FR(t) = P\left(p \geq t \middle| x = 0\right), \tag{10.4}$$

where p is the forecast probability, and $x = 1$ if the event occurs, and $x = 0$ otherwise.

The ROC graph is constructed by plotting the hit rates against the false-alarm rates. The graph for the example is shown in Fig. 10.5. The diagonal line on the graph indicates the line of no-skill. If the events were uniformly distributed through the table, the hit and false-alarm rates would accumulate at approximately the same rate, and so the ROC curve would follow the diagonal line. However, if the forecasts are good, the hit rate will accumulate faster, and so the graph will curve towards the upper left. In the extreme case of perfect discrimination, the curve will reach the top left corner. The example shows that the forecasts are well-discriminated.

Noting that the procedure for constructing the ROC graph is based only on the ranks of the forecasts, it should be evident that any monotonic transformation of the forecasts will not affect the graph at all. For example, if the forecast probabilities for all forecasts were increased (or decreased) by 10%, the graph would be unaffected. Alternatively if the forecast probabilities for all forecasts above 50% were increased by a fixed amount, and those below were decreased, the graph would again be unaffected. This insensitivity has been cited as a criticism of this verification procedure since the reliability of the forecast probabilities is ignored. The message is that a good ROC graph does not necessarily imply that the forecasts are well-calibrated.

The area beneath the ROC graph is increasingly used as a measure of discrimination, partly because of the inclusion of the ROC as a recommended verification procedure in the SVSLRF. For forecasts of no skill, for which the ROC curve lies along the diagonal, the area would be 0.5, and the maximum area of perfect discrimination is 1.0. The area is related to the *U*-statistic by a factor of the numbers of events and non-events, and can be interpreted as the probability of successfully

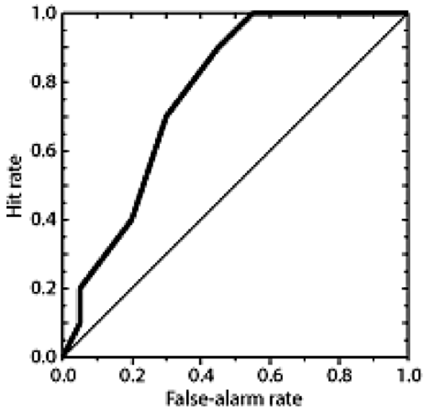

Fig. 10.5 ROC diagram for the retroactive forecasts of December–February 1981/82–2000/01 seasonal rainfall totals for Lusaka, Zambia

distinguishing an event from a non-event given a forecast of each (Mason and Graham 2002). For the example in Fig. 10.5, the area is 0.7675, implying that there is a greater than 75% probability that the forecasts can successfully discriminate an above-normal season from other seasons. The area is sometimes criticised as a summary measure of forecast performance because of its insensitivity to reliability. However, the ROC graph has an advantage over the reliability diagram in being less sensitive to sampling errors, and so can be more meaningfully constructed given the small sample sizes typical of seasonal forecasting.

10.2.4 Sharpness

10.2.4.1 Definition

Resolution, in the sense defined above, together with reliability, incorporate the idea of "sharpness". Although there is no formally recognised mathematical definition of sharpness,[7] the general concept is usually clear: sharpness refers to the degree to which the forecasts depart from the climatology. If forecasts are expressed as intervals, sharp forecasts are indicated by narrow intervals; if as probabilities of categories, sharp forecasts are expressed as probabilities that differ from the climatological probability, and are close to 0% or 100%; if as a forecast distribution, sharp forecasts are indicated as narrow distributions. Sharp forecasts

[7] The variance of the forecasts around the climatological probability is sometimes used to define sharpness, although arguably this definition makes sense only if the climatological probability is 0.5.

imply high confidence (see Chapters 8, Section 8.5.1), but do not necessarily imply good forecasts; as with reliability, sharpness is a necessary but not sufficient condition for high forecast quality.

Unfortunately, after appropriate recalibration (see Chapters 8 and 9) the sharpness of seasonal forecasts is typically much weaker than that of weather forecasts because of the large inherent uncertainty in predicting seasonal climate. In the extreme case of no predictability, the forecast probability should always be equal to the climatological probability.

A specific question of interest that is concerned with the sharpness of forecasts pertains to the ability of an ensemble to indicate changes in the uncertainty in the forecast. More specifically, does a sharper forecast mean that uncertainty is reduced? This question has received considerable attention in the context of forecast ensembles (including multi-model ensembles) where there is interest in the case-to-case variability in the ensemble spread: does this variability in the ensemble distribution contain any useful information? In other words, can one be more confident that the observed value will be close to the ensemble mean when the ensemble spread is small compared to when it is large? Two general approaches to the question have been adopted, both of which can be fraught with difficulties. One approach involves seeking a relationship between some measure of the spread in the ensemble, and some measure of accuracy in the central tendency of the ensemble distribution. These procedures are discussed in next section. The second approach attempts to measure the quality of the forecasts when the ensemble distribution is considered explicitly and to compare this with the quality when the variability in the ensemble distribution is ignored (Section 10.2.4.3).

10.2.4.2 Accuracy[8]–Spread Relationships

A common approach to the question of determining the information content of an ensemble distribution is to identify whether there is any relationship between the accuracy of the forecast, as measured by the "error" in the ensemble mean, and the uncertainty in the forecast, as measured by the ensemble variance (the "spread"). The theory behind this approach is that a larger ensemble spread implies greater uncertainty in the forecast, and hence larger errors in the ensemble mean can be expected. However, this theory is often based on a misconception of any accuracy-spread relationship that may exist. Assuming that the ensemble distribution is a reliable indicator of the true distribution of possible outcomes, the expected error is zero regardless of the uncertainty; it is the variance of the errors that should increase with increasing ensemble spread, not the expected error, as

[8] The term "skill" is often confusingly used instead of accuracy. Although "skill" is often used in a more generic sense than the definition provided in the glossary, in the current context it invariably refers to "accuracy", and so the latter term is preferred here.

indicated in Fig. 10.6. This misconception is not adequately resolved by defining the error in absolute terms (or by squaring the errors), and any standard form of regression between forecast error and some measure of forecast spread is poorly designed to identify any relationship that may exist.

10.2.4.3 Skill of the Ensemble Spread

Given the form of the relationship between accuracy and spread as indicated in Fig. 10.6, it is more helpful to reconstitute the problem as identifying whether there is any useful information in the case-to-case variability in the ensemble spread (or, more generally, the ensemble distribution[9]). An approach that offers more promise than seeking accuracy-spread relationships when sample sizes are small is to calculate whether the performance metric of the forecasts improves if the information in the spread of the ensemble distribution is considered compared to if the ensemble distribution is kept fixed. It is inferred that the variability in the ensemble spread does provide meaningful indications of changes in uncertainty if the measure of forecast quality is highest for the forecasts with varying ensemble spread.

There are numerous ways of implementing such a procedure. Perhaps the simplest is to assess the quality of the ensemble forecasts when using a counting procedure to obtain forecast probabilities (Chapter 8, Section 8.5.2), and then to reassess the quality after reducing the ensemble to the ensemble mean so that the forecast probabilities are either 0% or 100%. Although such an approach is unfair because most of the information in the ensemble mean is lost by converting it to a binary forecast, it has been used occasionally, most commonly when the ROC area is the verification metric of choice. Because of the unfair treatment of the ensemble mean as a single-member ensemble, such results are heavily biased in favour of finding useful information in the ensemble spread. A fairer approach using the ROC is to calculate the area based on the ranks of the ensemble means. However, the results can then be biased against finding information in the ensemble distribution because of the unsatisfactory calculation of probabilities by counting for the ensemble.

A more satisfactory procedure is to use a distribution-fitting approach to obtain probability forecasts from the ensemble (Chapter 8, Section 8.5.2). The question at hand can then be reformulated to: is there any information in the ensemble beyond the first moment of its distribution? The forecast probabilities are first obtained by fitting a distribution firstly with a fixed variance (or a variance

[9] For the sake of economy of phrase, in the rest of this section the term "spread" is assumed to incorporate changes in shape as well as variance.

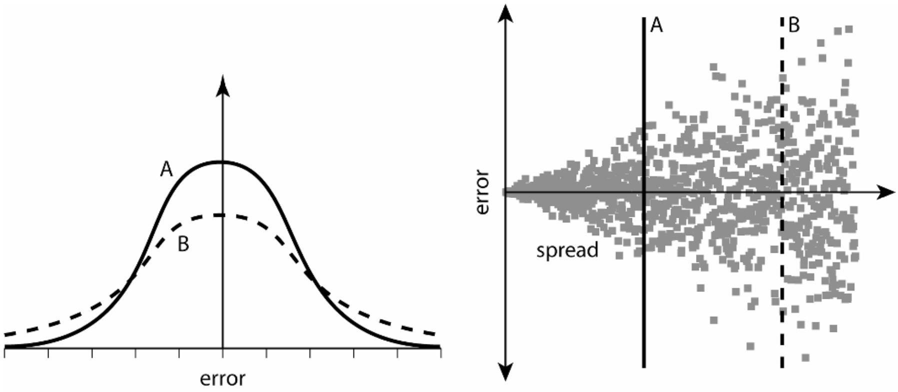

Fig. 10.6 (a) Distributions of forecast errors given high confidence forecasts with small ensemble spread (A) and low confidence forecasts with large ensemble (B); and (b) a hypothetical sample of ensemble mean error and ensemble spread measurements for a continuous range of spread values. The distributions in (a) are drawn from the points on the x-axis marked in (b)

that is a function only of the mean in the case of meteorological parameters with skewed distributions, such as precipitation), and secondly with a variance defined by the ensemble variance. If the forecast distribution does contain useful information the score should improve by allowing the variance to vary, and it can then be inferred that the ensemble provides useful information about the uncertainty in the forecast. It is thus possible to decompose the skill of the ensemble into signal and indications of uncertainty.

As an example, consider simulations of Brisbane September–November total rainfall for the 50-year period 1951–2000. The simulations are based on output from the ECHAM4.5 atmospheric general circulation model forced with observed SSTs. Fitting a gamma distribution to the 85 ensemble members, probabilities for the rainfall less than the lower tercile were calculated. The gamma distribution was fitted using a fixed and a varying shape parameter. Compared to climatology, a Brier skill score (see Section 10.4.1) of 0.081 was achieved given a fixed shape parameter, and of 0.060 given a varying shape parameter, indicating positive skill in both cases. However, the probabilities given the varying shape parameter scored slightly worse, and the negative skill score of –0.022 when measured against the fixed shape parameter indicates that there is no useful information in the ensemble shape.

10.3 Properties of Summary Measures for Probability Forecasts

Although it has been argued throughout Section 10.2 that forecast quality cannot be represented adequately by a single metric because there is more than one important attribute of good probability forecasts, sometimes it is desirable to quantify the forecast performance in a summary measure. Metrics that measure the individual attributes have been discussed briefly, but in all cases were found to be problematic as summary measures of overall quality: specifically, while a bad score does indicate bad forecasts, a good score does not necessarily indicate good forecasts. Before considering examples of summary measures of forecast quality it is therefore of value to consider the desirable properties such scores should have.

There is a wide range of performance scores, partly reflecting the need for different verification methods depending on how the probability forecasts are issued. Scores for cases in which there are only two categories are discussed in detail in Section 10.4.1, while scores for multiple categories are considered in Section 10.4.2. There are a few measures that apply to forecast distributions on a continuous scale, and these are covered in Section 10.4.3. Very few studies have addressed verification methods for probability forecasts of count data, interval and quantile forecasts, and probability forecasts of spatial distributions; these are areas of probability forecasting that merit more attention.

10.3.1 Score Orientation and Skill Scores

Before discussing the desirable properties of verification scores (Section 10.3.2), it is helpful to distinguish between positively and negatively oriented scores. Good forecasts achieve a high score if the score is positively oriented, but a low score if it is negatively oriented. Negatively oriented scores frequently are some measure of the error in the score: if the forecasts are good the errors will be small, and so the score will be low. Positively oriented scores, however, give credit to good forecasts, and so a high score is desirable. In this chapter all scores are presented in their negatively oriented versions unless indicated otherwise.

Skill scores are positively oriented scores with specific characteristics: they compare the quality of one set of forecasts with that of a second set, known as the reference forecasts (sometimes the second set is implied), and equal zero if the quality of the two sets of forecasts is identical. A commonly used formula for deriving a skill score, *SS*, from a positively oriented score is:

$$SS = \frac{S - S_{ref}}{S_{per} - S_{ref}}, \tag{10.5a}$$

where S is the score for the forecasts in question, S_{ref} is the score for the reference forecasts, and S_{per} is the score for a perfect set of forecasts. For a negatively oriented score (Murphy 1973b), $S_{per} = 0$, and so Eq. (10.5a) reduces to:

$$SS = 1 - \frac{S}{S_{ref}}. \tag{10.5b}$$

Skill scores usually have a maximum value of one (or 100%), when the first set of forecasts perfectly outperforms the reference set, but their lower limit depends on the score for the reference forecasts, which can make the interpretation and comparison of negative skill scores complicated. Using Eq. (10.5), a positive skill score can be interpreted as the fractional improvement in the score for the forecasts against the score for the reference forecasts. An alternative formulation for a skill score is

$$SS = S - S_{ref}, \tag{10.6a}$$

for positively oriented scores, and

$$SS = S_{ref} - S, \tag{10.6b}$$

for negatively oriented scores. Using Eq. (10.6), a positive skill score can be interpreted as the amount of improvement in the score for the forecasts against the score for the reference forecasts.

Because a skill score is a relative score, the interpretation of the score depends upon the choice of the reference forecasts. Commonly used reference strategies include climatology and persistence, but these strategies are not necessarily equally unskilful. For example, when forecasting SSTs, the slowly evolving nature of the oceans makes persistence a much harder standard to beat for short lead-time forecasts than climatology (Sections 3.4, in Chapter 3, and 5.4, in Chapter 5). In addition, because scores necessarily over-simplify the complex nature of forecast verification, a negative skill score against climatology and/or persistence does not automatically imply that the forecasts do not contain any useful information. This information may have been lost by the score (Mason 2004).

10.3.2 Desirable Properties of Probabilistic Forecast Verification Scores

Given the wide range of scores available for assessing the quality of probabilistic forecasts, it is helpful to define a set of criteria that can be used to identify which scores may be the most appropriate. Three criteria are considered: propriety, equitability, and locality (Murphy 1993).

10.3.2.1 Propriety

An important concept in verification is whether or not a score can be improved by hedging of the forecasts. The Oxford English Dictionary defines hedging as "the securing of, or limiting the possible loss on, a debt, bet, or the like". In the context of forecast verification, "hedging" occurs when a forecaster issues a forecast different to what (s)he truly believes. Certain non-mathematical targets of performance can encourage a forecaster to issue a forecast that is inconsistent with his/her true belief: for example, not wishing to cause excessive alarm. However, some verification scores can also be hedged: the forecaster is encouraged to modify the forecast in order to improve the expected value of the score. Hedging is undesirable because it encourages the forecaster to issue a forecast that may be inconsistent with his/her true beliefs simply in order to achieve a better score, and so it is best to choose scores that cannot be improved by forecasting anything other than the forecaster's true beliefs.

A *strictly proper score* is a probability score, S, for which the forecaster uniquely optimises the expected score by forecasting his/her true beliefs. So if the forecaster believes an event occurs with probability q then the expected score should be minimised when the forecast probability actually issued, p, equals q. The score will be minimised if there is a unique stationary point at which

$$\left.\frac{\partial}{\partial p}S(p,q)\right|_{q} = 0 \text{ at } p = q\,. \tag{10.7}$$

A score is proper, but not strictly proper, if Eq. (10.7) is true for more than one value of q. Unfortunately, most skill scores defined using Eq. (10.5) are not strictly proper unless the categories are equiprobable, and/or unless the forecaster has absolute certainty about the outcome (Murphy 1973b; Gneiting and Raftery 2007). Skill scores defined using Eq. (10.6) may therefore be preferable, although further research is required to investigate their properties.

10.3.2.2 Equitability

Another property of scores that sometimes is considered desirable is *equitability*; a score is *equitable* if it takes the same value for all unskilful forecasts that have no association with the observations (i.e. forecasts that have no resolution). In the context of probabilistic forecasts there are a variety of forecast strategies that can be adopted that have no resolution (e.g. random forecasts, and perpetual forecasts of constant probabilities, including of climatological probabilities). Although it may seem desirable that these various naïve strategies should score equally badly, it is impossible for a probabilistic score to be equitable and strictly proper, and the latter property is to be preferred (Jolliffe and Stephenson 2007). The differences in the scores of differing no-resolution forecast strategies can be attributed to differences in their reliability.

10.3.2.3 Locality

A skill score is local if it depends only on the probability assigned to the outcome. The desirability of this property of verification scores is disputed: two main arguments are presented against locality, but both arguments are inconclusive. The first argument is that non-local scores can be less sensitive to the categorization of the observed values than local scores; the more categories that are used the lower the score tends to be (Daan 1985). However, one could argue that a forecast system with many categories attempts to communicate more information than a system with only a few, and so a greater degree of precision is required. Another argument against locality is that it seems reasonable to account for “distance” (i.e. to credit forecasts that assign high probability close to the observed value as well as to the outcome itself). For, example, given two forecasts for three ordinal categories, A {0.2, 0.3, 0.5) and B {0.1, 0.4, 0.5}, forecast B would seem a better score if the third category verified since, while both forecasts assign the same probability to this category, forecast B assigns a higher probability to the adjacent category. Forecast B has more probability close to the verification, and so seems intuitively better than forecast A. Implicit in such reasoning is the assumption that because category 3 verified, category 1 was less likely to have occurred than category 2. However, we know only that category 3 occurred, and do not know what the relative probabilities of the other categories were. In fact, we do not even know that category 3 was the most likely. The reliability of the probabilities of all three categories can only be assessed by considering the categories individually. From one perspective, then, locality seems to be a desirable property, although if it is accepted as such, there are a number of non-local scores that are widely used, including the ranked probability score (RPS; see Section 10.4.2).

10.4 Summary Measures for Probability Forecasts

10.4.1 Some Scores for Binary Events

By far the most commonly used summary measure of the quality of probability forecasts of binary events is the Brier score.[10] The Brier score is analogous to the

[10] Strictly, the Brier score, being a special case of the probability score (Section 10.4.2) for when there are only two categories, is defined both for the event and for the non-event categories.

mean-squared error, but is defined in terms of the "error" in the probabilities rather than in the actual units of the observations, and is calculated using:

$$\text{Brier score} = \frac{1}{n}\sum_{i=1}^{n}\left(p_i - o_i\right)^2 , \tag{10.8}$$

where o_i is 1 if the event occurs in the ith case, or 0 otherwise. Murphy (1973a) defined a well-used algebraic decomposition of the Brier score consisting of reliability, resolution, and uncertainty. The first two terms have been discussed in Sections 10.2.1 and 10.2.2, respectively, where their relationships to the reliability diagram were explained. The uncertainty term, defined as

$$\text{uncertainty} = \overline{o}\left(1 - \overline{o}\right), \tag{10.9}$$

is independent of the forecasts, but because it depends on the relative frequencies of the observed events, it can complicate comparison of scores for different sets of verification data if these relative frequencies are not constant.

The skill score version of the Brier score [using Eq. (10.5)] is negatively biased when climatology is used as a reference strategy, so that forecast quality looks worse than it really is. For ensemble forecasts this bias is related to sampling errors in calculating the forecast probabilities (Müller et al. 2005), and can be corrected by adding an additional uncertainty term that accounts for these sampling errors. Without the additional uncertainty term the imperfectly estimated probabilities for the forecasts are compared to perfectly estimated (and therefore perfectly reliable) climatological probabilities, making for an unfair comparison. The correction term, d, is calculated as

$$d = \frac{\overline{o}\left(1 - \overline{o}\right)}{m} , \tag{10.10}$$

where m is the number of ensemble members (Weigel et al. 2007). The debiased skill score is calculate using

$$SS = 1 - \frac{S}{S_{ref} + d} , \tag{10.11}$$

Because the definition in Eq. (10.8) applies only to the event, it is more correctly called the half-Brier score. However, since the contribution from the non-event category is the same [$\left(p_i - o_i\right)^2 = \left(\left(1 - p_i\right) - \left(1 - o_i\right)\right)^2$], it is redundant to score them both, and for the sake of simplicity the Brier score is widely calculated using Eq. (10.8). Throughout this chapter, the phrase "Brier score" refers to the half-Brier score, unless specified otherwise.

and since S_{ref} reduces to Eq. (10.9) for climatological forecasts, Eq. (10.11) simplifies to

$$SS = 1 - \frac{S}{d(m+1)} . \tag{10.12}$$

As an alternative score, just as the mean absolute error is sometimes used in place of the mean squared error, a mean absolute score for probability forecasts has been proposed:

$$\text{mean absolute score} = \frac{1}{n}\sum_{i=1}^{n} \left| p_i - o_i \right| . \tag{10.13}$$

A third score is the logarithmic score, defined as:

$$\text{logarithmic score} = \frac{1}{n}\sum_{i=1}^{n} \left[-(1-o_i)\log(1-p_i) - o_i \log p_i \right], \tag{10.14}$$

which is the average of the negative of the logarithm of the probability assigned to the verifying category. Since the logarithm of 1 is 0, Eq. (10.14) is an "error" score, like Eqs. (10.8) and (10.13).

The Brier, mean absolute, and logarithmic scores can be generalised as the mean error over all cases:

$$\overline{S} = \frac{1}{n}\sum_{i=1}^{n} S(p_i, o_i) .$$

The Brier score and mean absolute scores are quadratic and linear "error" or loss functions, whereas the logarithmic score is the negative log likelihood for a set of n independent Bernoulli events. As loss functions, the scores are negatively oriented in that smaller scores imply more skilful forecasts. The loss function for the above three scores is symmetric since $S(p,0) = S(1-p,1)$ and so they are each completely defined by just the single loss function $S(p,0)$. This loss function is given by p^2, p, and $-\log(1-p)$ for the Brier, absolute, and logarithmic scores, respectively.

The expected value of the score for an event that has probability q of occurrence (sometimes confusingly referred to as the "true probability", and hereafter referred to as the "event probability") is given by:

$$E(S(p,X)) = (1-q)S(p,0) + qS(p,1) = S(p,q) .$$

Consistent with the scores being minimized when the forecast probability equals the event probability, the minimum values occur where $p = q$ (Jolliffe and Stephenson 2007). Note, however, that the score is a function of the event probability, q, and is smallest when $q = 0$ and 1 (i.e. when the observed event is least uncertain).[11] The Brier and logarithmic scores have a saddle point (maxima of a valley and minima of a ridge) at $p = q = 0.5$ whereas the mean absolute score has a flat ridge at $q = 0.5$. This saddle point defines the unique stationary point required for a score to be strictly proper [Eq. (10.7)], thus implying that the Brier and logarithmic scores are proper scores, but the absolute score is not (if the forecaster thinks that the event probability is 0.5, it does not matter what probability (s)he issues since the expected score will be 0.5 regardless of the probability issued).

The logarithmic score penalises much more heavily poor forecasts different from $p = q$ than does either the Brier or the absolute score. The penalty becomes infinitely large when a probability of 0% is assigned to an event that does happen, or when a probability of 100% is assigned to an event that does not happen. This apparently excessive penalty can be justified on the basis that the implied odds of the actual outcome were infinitely small.

10.4.2 Scores for Multi-category Forecasts

The probability score is defined as the average over n forecasts of the sum of squared probability "errors" for each category, j, of m categories:

$$\text{probability score} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(p_{ij} - o_{ij}\right)^2 . \tag{10.15}$$

The full Brier score is a special case of the probability score for when there are two categories. The probability score considers the probability assigned to all categories, and so it does not have the property of locality. Despite being non-local, since there is no implicit ordering in the categories, the probability score does not account for distance, and for this reason it is not widely used. Because of the failure to account for distance, coupled with the lack of locality, the score has some undesirable properties. Consider, for example, the two forecasts A = {0.45, 0.55, 0.00} and B = {0.40, 0.30, 0.30}. If the first category verifies, the score for B (0.5400) is less than (and hence better than) for A (0.6050), despite the fact that A issues a higher probability to the outcome, and a higher probability to the adjacent category. A simple modification to Eq. (10.15) reduces it to the half-Brier score [Eq. (10.8)] and thus would resolve these problems:

[11] In the case of the Brier score, this dependence on the observed probability is reflected by the uncertainty term in Murphy's (1973a) decomposition of the score.

$$\text{quadratic score} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} o_{ij}\left(p_{ij} - o_{ij}\right)^2 . \tag{10.16}$$

A commonly used alternative to the probability score is the ranked probability score (RPS), which addresses the problem of lack of effectiveness by considering distance. Instead of comparing the probabilities assigned to each category with that of a perfect deterministic forecast, the RPS compares the cumulative probabilities:

$$\text{ranked probability score} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(P_{ij} - O_{ij}\right)^2 . \tag{10.17}$$

where P_{ij} is the cumulative probability assigned to category j in the ith forecast, and O_{ij} is 1 if the ith observation is in any of the categories less than or equal to j, and is 0 otherwise [cf. Eq. (10.8)]. So, the example forecasts A and B above, would be expressed as A = {0.45, 1.00, 1.00} and B = {0.40, 0.70, 1.00}. Similarly, for the observations, the cumulative probabilities given a perfectly deterministic forecast are used (i.e. {1.00, 1.00, 1.00}, {0.00, 1.00, 1.00}, and {0.00, 0.00, 1.00} if the first, second, and third categories were to verify, respectively). The RPS for A (0.3025), given that category 1 verifies, is now less than for forecast B (0.4500), and so the RPS correctly identifies A as the better forecast. Despite being strictly proper, the RPS, by definition, does not have the property of locality.

The skill score version [Eq. (10.5)] of RPS is biased when climatology is used as the reference strategy for the same reasons as with the Brier score. The skill score can be debiased using Eq. (10.11), but d is defined as

$$d = \frac{k^2 - 1}{6mk}, \tag{10.18}$$

where k is the number of categories, as long as the categories are equiprobable. See Weigel et al. (2007) for corrections to unequal categories.

In the previous section, the linear probability score was rejected because it is not a strictly proper score. The score can be generalized for cases when there are more than two categories [in a similar way to Eq. (10.15)], but it still suffers from the same problem of lack of propriety as its two-category version.

The logarithmic score is defined as the average of the negative of the logarithm of the probability assigned to the verifying category, and so can be interpreted as a measure of probability "error":

$$\text{logarithmic score} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} -o_{ij}\log\left[p_{ij}\right]. \tag{10.19}$$

Just as with the binary version of the score, it penalizes large forecast errors much more severely than any of the other scores, and in the extreme case of a zero probability being assigned to a verifying event, the logarithmic score is infinite. Although not yet widely used, the logarithmic score has all the desirable properties of verification scores (ignoring equitability), and can be generalized to cases of continuous forecasts and still retain all properties (see Section 10.4.3).

10.4.3 Scores for Continuous Forecasts

Probabilistic scores for continuous forecasts are not widely used for seasonal climate forecasts, partly because full forecast distributions are rarely specified, and also partly because options for scoring such forecasts have not been discussed much in the climate literature. There is a version of the linear error in probability space (LEPS) suitable for probabilistic forecasts (Ward and Folland 1991). The LEPS score was derived to measure the error in a forecast in terms of distance measured by the climatological cumulative distribution rather than in terms of the original units of the forecast (so, for example, a forecast error of 1°C for a normally distributed variable would be penalised much more heavily if the forecast and observed value are close to the mean than if near the tails of the distribution). The version of the score for continuous probabilistic forecasts lacks propriety, and so its use should be discouraged (Mason and Mimmack 2002).

A preferable option is the continuous version of the ranked probability score (CRPS):

$$\text{continuous ranked probability score} = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\left[F_i(z) - O_i(z)\right]^2 dz\,, \qquad (10.20)$$

where $F_i(z)$ is the cumulative forecast probability for the ith forecast, and $O_i(z)$ is 0 if the ith observation is less than z, and is 1 otherwise [cf. Eq. (10.17)]. The score describes the average of the squared difference between the forecast and observed cumulative distributions, where the observed cumulative distribution is a step function represented by the cumulative distribution of a perfectly accurate deterministic forecast; it is calculated by integrating the squares of the vertical distances between the two curves,[12] as represented by the grey shaded areas in see Fig. 10.7. Note that the squaring in Eq. (10.20) is along the probability axis not along the x-axis, and so the score reduces to the mean absolute error if the forecasts are deterministic (Hersbach 2000).

[12] Compare the Kolmogorov-Smirnov test, discussed in Chapter 8, which is based on the largest vertical distance between two such cumulative distributions.

The linear error score defined in Eq. (10.13) has been generalised for ensemble forecasts (Wilson et al. 1999), but the score lacks propriety. However, it can be adjusted so that it is strictly proper:

$$\text{proper linear score} = \frac{1}{n}\sum_{i=1}^{n}\left[\int_{-\infty}^{\infty} f_i^2(z)\,dz - 2f_i(x)\right], \tag{10.21}$$

where $f_i(x)$ is the forecast probability density at the observed value, x (Bröcker and Smith 2007). The integral in Eq. (10.21) renders the score proper, and is a representation of the sharpness of the forecasts, but does make the score lack locality. The only score that has the propriety, and the locality properties is the continuous version of the logarithmic score (Roulston and Smith 2002):

$$\text{logarithmic score} = \frac{1}{n}\sum_{i=1}^{n} -\log\left[f_i\right]. \tag{10.22}$$

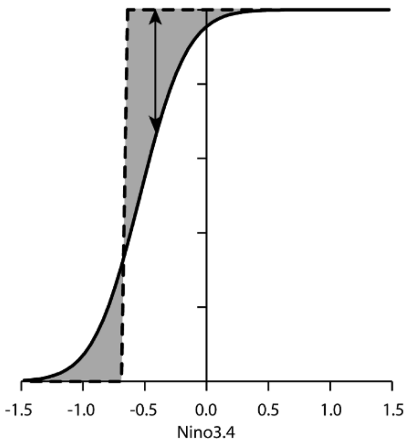

Fig. 10.7 Example of a forecast for the December 2000 NIÑO3.4 index (using the June predictor from Table 10.1) showing the "error" (grey shading) in the cumulative forecast probability, as measured by the squared distance between the forecast distribution (solid line) and the distribution for a perfectly accurate deterministic forecast (dashed line). The integral of the squared vertical distances between the two curves (shown by the arrow at an index value of –0.5) is the contribution to the continuous ranked probability score

10.5 Summary

Forecast verification, or the evaluation of the quality of a set of forecasts, is a multifaceted problem, and so there is no single metric that can comprehensively represent the quality of the forecasts. The problem is complicated in the case of probability forecasts because naïve attributes of forecast quality, such as "accuracy", are inappropriate – the question of whether a specific probability forecast is correct or incorrect is unanswerable. Specially designed verification procedures are therefore required for probability forecasts, but there are several different types of probability forecast, and each requires its own methods for verification. The main attributes of interest are:

- Reliability, whether the confidence communicated in the forecasts is appropriate
- Resolution, whether there is any usable information in the forecasts
- Discrimination, whether the forecasts are discernibly different given different outcomes (somewhat similar to the attribute of resolution)
- Sharpness, the level of confidence that is communicated in the forecasts (regardless of whether that level is appropriate)

The most commonly used graphical procedures for indicating forecast quality are the reliability diagram and accompanying frequency histogram (which together indicate reliability, resolution, and sharpness), and the relative operating characteristics graph (which indicates discrimination). Numerous summary measures of forecast quality have been defined, and so in choosing between them it is helpful to define a set of properties that verification scores should have. Perhaps the most important property is that only those verification scores should be used that cannot be improved by hedging the forecasts. Scores for probability forecasts that have this property are called proper scores. Unfortunately, the score that is arguably the easiest to interpret, namely the mean absolute probability error, is not strictly proper. Squared and logarithmic scoring rules are therefore generally preferred, although the logarithmic score is the only one that can be generalised to forecast of continuous probability distributions.

Part IV
Developing Successful Application Strategies

Chapter 11
Communicating Seasonal Forecasts

Mike Harrison and Jim B. Williams

Delivering, taking advantage of, and obtaining benefit from climate information, including predictions, are at least as substantial a challenge as producing the predictions in the first instance. It is also a challenge into which relatively limited resources have been invested so far by comparison to those devoted to the prediction problem. In part that contrast in resources use is underpinned by the relatively well-defined nature of the prediction problem as contrasted to the wide, multidisciplinary issues raised in terms of taking advantage and receiving benefit from climate information. It is out of the question for this book to delve into all of the issues involved, so extensive are these in terms of different sectors, individual countries, levels of decision makers, concerns/constraints regarding specific decisions, and so on, that we have attempted to provide only on overview in the hope that this will provide context against which individual issues might be considered. Specific examples of the use of climate information, and of the benefits derived therefrom, are provided in Chapters 12 and 13, whereas here the focus is on some of the fundamental issues underpinning climate services. To a certain, but not exclusive, extent the authors of this chapter have taken a perspective related to issues in developing parts of the world, issues that are thought to require a range of additional approaches to the straightforward end-to-end model appropriate to business uses. Nonetheless much of the chapter is relevant to the delivery of business information. Examples are included in Chapters 12 and 13 of both business and development activities. Consideration is given to the context of climate services within international development, the physical delivery of information (including delivery to remote communities), and difficulties in presentational delivery of information. Presentational delivery remains a major impediment to extracting the benefit from climate services, yet remains one to which minimal consideration is given in many instances. The focus here is on the pitfalls of oral information delivery; space precludes detailed coverage of visual information delivery, a further critical area.

Mike Harrison
Independent Consultant

Jim B. Williams
Consultant with the NRGroup

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

11.1 Learning to Manage Climate Risk in Development

Climate variability has received less attention than other development issues, in part because it has been considered one aspect of the environmental baseline that is not amenable to intervention. Climate clearly impacts development, for example on our ability to achieve the Millennium Development Goals (MDGs).[1] If future management of climate risks is to be improved then the need is to understand reasons for the relative lack of progress to date, and to incorporate lessons learned into future strategies.

The impacts of climate variability on development are twofold, direct and indirect.

1. **Direct**: Events such as droughts and flooding take a direct toll on lives, health, livelihoods, assets and infrastructure. Climate directly impacts food and fibre production, and the epidemiology of infectious diseases. Severe or repeated climate shocks can push vulnerable households into a persistent poverty trap when their individual coping responses usually involve divestment of productive assets, such as livestock or land. Without advanced warning, reactive societal safety nets are costly, and difficult to mobilize and target effectively (Benson and Clay 2004; this applies also in developed countries – Cutter et al. 2006).
2. **Indirect**: Although less visible than direct impacts, the indirect impacts of climatic uncertainty are equally serious impediments to development. Knowledge that a region is prone to climate events that endanger resources acts as a disincentive to investment, adoption of innovation and the success of other development interventions, particularly when approaches to the management and mitigation of those events are not available. For the risk-averse decision maker, climatic uncertainty necessitates short planning horizons and conservative risk management strategies that buffer against detrimental climatic events. But this is often achieved at the expense of inefficient resource use, reduced average productivity and profitability, and accelerated resource degradation due, for example, to under-investment in soil fertility inputs or conservation measures.

But in spite of its pervasive importance, and powerful new advances in climate-related science, opportunities have been largely underexploited in decision making not only in Africa, from where most examples are drawn in this chapter, but throughout the globe. Lip service is often paid to climate and its variability, but in practice, studies (Gibberd et al. 1995) have shown that climate information is

[1] See http://www.undp.org/mdg/abcs.shtml for a full list of the Goals, the Targets used to specify each Goal, the Indicators used to measure achievement of each Target, and the organisations (mainly UN) with ‘ownership’ of each Indicator.

rarely, if ever, incorporated into planning and management decisions as effectively as it could be.

Why should this be so? By far the majority of water supplies and agricultural production in Africa is dependent on rainfall, which fluctuates considerably from year to year. If water availability and food production are both in such short supply as to be a major brake on development, surely use of climate information to optimise use of these scarce resources should be a top priority?

Many reasons contribute to the low priority given to incorporating knowledge of climate variability into decision-making. It is most important that these be understood and addressed in concert because together they present a formidable obstacle to progress. Recognised reasons include:

1. **Whose responsibility**? In many instances the ability of national 'resource management systems' to respond to climate information is weak or non-existent. Resources are not so much 'managed' as collectively exploited. Such a system works adequately when population density is low in relation to the resource base, but as populations grow and resource use becomes more intensive, seasonal variation in water (rainfall) becomes ever more important to a growing multitude of stakeholders. Well-informed, coordinated and timely decision processes to proactively manage resources, involving meteorological services, disaster management organisations, agricultural departments and extension services, and so on, then become most important; but this is difficult if neither the information (in many cases in Africa due to collapse in observing and extension networks) nor the system and infrastructure and institutional architectures for making appropriate decisions exist, until an emergency is declared.
2. **Reluctance towards risk management**: In the past the occurrence of periodic drought was almost welcomed by mid-level government administrators. All sorts of project failures, from the colonial East African Groundnut Scheme onwards, were written off 'because of the drought', with no further questions asked (nevertheless it must be admitted that climate was far from being the only impediment faced by the Groundnut Scheme). It is important that collectively we become more proactive in resource management and include climatic risk in our calculations, amending management decisions as a continuous process rather than working to a fixed schedule. Of course one alternative to risk management is amelioration of resilience to climate variability (irrigation being one example); nevertheless risk management remains a major option in many parts of the world.
3. **Research-extension gap**: The agricultural research literature is populated with studies demonstrating the importance of climate variability, and how best to cope with it in different areas. Converting climate information into actionable decisions is not, however, always straightforward. Research systems, in addition, are often isolated from the communities they serve, and beneficial outcomes are difficult to achieve when messages are over-complex for an inadequately trained (or non-existent) extension service to devise appropriate

improvements to current practice in farming communities. New appropriate and user-friendly approaches to mass communication of such information are required avoiding the difficulties outlined later in this chapter.

4. **Insufficient climate data and resources available**: There has been a major decline in the number of climate observing stations, particularly but not uniquely in Africa, over the past 40 years. While this has been compensated, to some extent, by the development of operational weather satellites covering Africa, adoption and use of these new technologies in climate risk management has been slow in coming. In some areas the basic climate network has become so sparse that it will be difficult to detect and quantify climate change trends in these areas, let alone inter-annual variability. Improving the basic network in Africa is a current concern of the Global Climate Observing System (GCOS), based in WMO, among others. Additionally national weather services also often have limited personal and material resources.
5. **Institutional obstacles**: Climate risk management is an interdisciplinary activity, affecting many economic sectors and aspects of life, but it has no effective champion. In many Meteorological Services in Africa the application of climate knowledge in agriculture, health, water and other sectors is a low priority activity, poorly staffed and weakly focused. Relatively few Meteorological Services in Africa are fully engaged in their country's own national development and disaster risk reduction agendas. Supporting civil aviation services and participating in the global meteorological agenda have for many years been much higher priorities for these Services. As a consequence many climate observing networks are badly run down, and the data that exist are of limited utility and are analysed in a fragmented way, if at all. Some view the introduction of policies to commercialise meteorological and climate services or to seek cost recovery for data collection as an unmitigated disaster with regards to promoting the overall beneficial use of these data.
6. **Farmers know best**: People who have been working their fields for years in traditional manners have usually developed more or less successful coping strategies based on managing risk. They often have traditional 'seasonal forecasting' methods based on bird, animal and plant observations. However while traditional practices may be resistant to change, experience often demonstrates farmers' desires for 'other' knowledge systems that may be used alongside, and perhaps ultimately may displace, local practices. Complementary approaches, rather than replacement, offer a sympathetic way towards the introduction of new technologies, but in doing so it should be borne in mind that:

   - Adaptation will be facilitated if new forecasts are treated synergistically alongside traditional methods.
   - Traditional methods may not be able to cope with rapid population growth and land fragmentation, or systematic climate change (as in the Sahel during the 1970s and 1980s).

- Traditional farmers may resist change from a risk-reduction to a production maximisation strategy until they have an adequate safety net to support them through bad years; insurance safety nets are being piloted in several African countries.

7. **Negative perceptions regarding reliable climate forecast capabilities**: For many people the very idea of long-term or seasonal climate forecasting is considered an unrealisable dream, or one that challenges a 'divine prerogative'. While short-term weather forecasting has radically improved with the refinement of satellite coverage and global models, its forward view is restricted to a few days at most, so (they ask) how can one possibly give credit to a 3-month forecast? There is thus a significant credibility gap to be overcome, one that will be assisted by providing evidence that the new technology offers effective risk management options.

11.1.1 Towards Resolution: Improved Governance and Improved Science

All problems as outlined above can and must be overcome. In parts of Africa they are reflections of a wider suite of compounding problems over the last 25 years, including weak governance in some cases. Yet, elsewhere, application of climate science is moving forward rapidly, stimulated both by practical everyday requirements to optimise resource management as well as by the urgent need to understand global warming and climate change impacts.

11.2 Seasonal to Interannual Prediction: An Overview of its Role in Decision Making

11.2.1 The Management and Social Background to Applications of Seasonal to Interannual Prediction

There are numerous reasons for using seasonal to interannual predictions in decision making. Planning business processes, both national and international, has provided the motivation in many countries, but equally there is a wide range of potential applications in terms of public good. Included amongst the latter are national and international resource management (including water and food security), disaster preparedness and response, poverty reduction, protection of biodiversity, and the frequently discussed objective of sustainable development.

Sustainable development, with its agenda of maintaining and improving living standards in a sustainable manner, has been translated with respect to the developing world into actions codified as the MDGs. The underlying principle behind the MDGs is the generation of convergence of multi-country, multi-institutional activities and funding into defined and targeted, internationally agreed, activities. Of the eight MDGs, most, probably all, have a dependency on climate variability. Even in MDGs such as No. 2, "Achieve universal primary education", for which the climate variability link might not be immediately evident, there are indirect dependencies for which predictions might provide guides. For example, primary education requires, amongst other things, appropriate levels of health, which in turn are dependent upon adequate food and water supplies, and thus on environmental sustainability. Further, social stress, caused by environmental degradation and/or natural disasters, or political stress induced, say, by drought, might undermine the structural basis on which education is dependent. For MDGs such as No. 1, "Eradicate extreme poverty and hunger", and No. 7, "Ensure environmental sustainability", the links with climate variability are more immediately transparent. The view that environmental sustainability, and through that climate variability, underpins all MDGs (which in any case are interrelated and thus interdependent on climate) has been stated in various assessments, including:

- A healthy environment underpins human life and well-being by providing food, clean water, disease control, and protection from natural disasters – and is thus necessary to achieve each Goal (UN 2005).
- Ensuring environmental sustainability and access to energy services is key to achieving all of the MDGs (UNDP 2005).
- The Millennium Ecosystem Assessment and other global and regional studies have established beyond doubt the linkages between poverty, security and the environment – achievement of the MDGs and eradication of poverty will not be possible without taking on the issue of environmental sustainability (UNDP 2005).

The first MDG provides a useful exemplar of the complexity of issues surrounding decision making within the contexts of climate variability and sustainable development. Within a study of the MDGs in relationship to climate variability, scientists commissioned by the IRI have identified aspects of climate variability that might adversely affect achievement of the Goals – the following examples are relevant to Goals to "Eradicate extreme poverty and hunger" and those regarding health (IRI 2005):

- National loss of agricultural production through drought and flood → lack of food security
- Drought or flooding can lead to poor water quality → increased morbidity and mortality from diarrhoeal disease

- Flooding, or return to normal after drought, create favourable conditions for the spread of mosquito-borne diseases → higher infection rates, such as for malaria, and a reduced work force
- Loss of infrastructure through severe climate events, such as floods and storms → removal of infrastructure necessary to achieve the goal, and diversion of funding to replace infrastructure
- Climate variability acting as a disincentive → might affect investment, intensification, technological adoption, fertiliser use, high value agricultural enterprises
- Repeated hydrological disasters → stagnated economic growth
- As a consequence the poor might be trapped in a downward spiral of increasing poverty and asset loss

Accordingly information on climate variability, including predictions, might be used in a variety of ways to provide increased lead in early warnings relevant to the first MDG, and to manage opportunities and risks in years with both above and below-average agricultural production. It might also assist in stabilising crop stocks in terms of price and availability, and to adjust credit flows and production inputs to farmers. Insurance schemes designed to benefit poor farmers can also take advantage of the information. Further, advanced warning facilitates preparations for hydrometeorological disasters, and also helps in mitigating their consequences, through assisting planning to reduce losses in infrastructure and productive assets.

However, in order to achieve the above, climate information must be mixed appropriately with other information flows (as discussed in greater detail in Chapter 2), perhaps economic and productive assets, perhaps population statistics, perhaps infrastructural distribution, perhaps relevant policies and statutes, perhaps cultural approaches to decision making, and so on, in a manner that assists final regional and local decision making at all levels. There is a growing body of evidence to suggest that optimal decision making based on climate information is achieved cooperatively, through participatory processes in both vertical and horizontal senses, rather than through independent actions of individual stakeholders, or even individual businesses despite the normal propensity to seek business advantage. Participatory processes also help guard against possible detriment to uninformed individuals and groups. Thus a further desirable attribute towards optimising benefit is some form of coordinated decision making leading to harmonised responses to all information streams, including climate.

A practical example of the multiple information sources needed in one specific activity, agricultural production in a developed country, Australia, has been examined in an extensive operational research programme at the Agricultural Production Systems Research Unit (APSRU) in Queensland. Historical and current information streams are made available via the Internet to guide decisions at farm through to State levels. These information streams include (not a comprehensive list): local and national rainfall (in particular its stratification according to phases of the Southern Oscillation Index), soil conditions, local and national crop production and yields, fertiliser costs, and international crop prices. In order to combine all information streams with the aim of assisting decision processes at farm and higher levels, scientists at APRSU have developed a number of computer-based tools, including APSIM – the *A*gricultural *P*roduction *S*ystem S*im*ulator.[2]

Activities in Australia are greatly assisted by the availability of quality, well-distributed (spatially and temporally) data, not only on climate but also on the other information streams necessary for optimal decision making. Where certain data do not exist, such as some historical crop yield data, simulation techniques have been developed to fill in the record. Some transfer of the APSRU approach to other countries, under the *Res Agricola* (Farmers' Affairs) banner, including to some in the developing world, has been undertaken. But one problem in many countries is a lack of the data necessary to support relatively sophisticated approaches of this nature. Only proactive approaches will start addressing these fundamental data paucities.

Agriculture provides a fine exemplar of the various problems faced regarding decision making under the uncertainty imposed by climate variability. Decision making of this type takes place across a broad range of scales of various types and differing approaches, which include:

- Spatial
  - Local cropping decisions
  - National/regional food stocks and trade
  - International food security
- Temporal
  - Short-term logistics, such as planting and harvesting
  - Medium-term planning of crop types, sequencing and rotations
  - Long-term industrial decisions and land use
- Cultural
  - Commercial through to subsistence
  - Science-lead (based on ENSO predictions, for example) through to belief- or experience-lead (including indigenous knowledge)
- Institutional
  - Policies/regulations/statutes of any particular organisation or individual within those of each country

Within agriculture alone there is thus a broad gamut of outcomes sought, one that extends when other sectors are included. It may be difficult, perhaps impossible, to provide a single stream of climate information that is optimally tuned for decision making across all scales, and more defined solutions by objective are likely to

[2] See: http://www.apsru.gov.au/apsru/

be needed. There is a further consideration in that risk often only makes sense within the culture and psychology of individual decision makers. In that regard the approach currently adopted in general of setting up specific pilot projects may be well founded in practical terms, no generic approach likely to be derivable at this stage. The downside of this approach is that portability is often problematic, as solutions developed for specific applications at particular locations on the different scales may not necessarily translate readily to other locations and scales. For example, outputs from local-scale pilot projects may be difficult to scale up to national level or to be transferred to a different country.

Cultural and institutional issues often present some of the major impediments to incorporation of climate information into decision making. It may be an over-simplification, but where climate is immediately identifiable as representing a control on an activity (for example, societies have long recognised this control in terms of food security), then the use of climate information and the desire to have more information of higher quality is usually incontestable. On the other hand where the role of climate has been less well defined, or the perception exists that climate information is of insufficient quality to assist in decision making, as perhaps within the context of the education MDG mentioned earlier, then there may well be a relatively low, or even no, uptake of the information.

An excellent example of cultural/institutional impediments restricting the use of climate information is provided by the World Bank. The World Bank has just one major mission objective – to reduce global poverty. In doing that, issues of sustainable development, as now codified within the MDGs, loom large. Yet despite the fact that the Bank has played a major role in supporting international activities related to climate change, and is responsible for a major fund for activities in this area, climate does not yet appear within the guidelines for assessment of most projects to be funded by the Bank. Consideration of methodologies for assessing projects from a climate perspective, whenever appropriate, is now being made (Mathur et al. 2004; Burton and van Aalst 2004).

Thus if seasonal to interannual prediction, not to mention climate information in the broader sense, is to play a major role in decision making across the range of scales and potential sectors, then there is a need to mainstream the information into those organisations responsible for taking decisions. Mainstreaming is achieved through a sequence of approaches, amongst which needs to be the provision of convincing evidence that the information is credible, is relevant, and, most importantly, provides value. Or that, at the least, it has the potential to provide value. Demonstrating value is a complex task across the entire range of scales and sectors, and needs to be achieved from the perspectives of potential users with their individual needs rather than from the individual perspective of the climatologist. Without careful and sustained collaborative conversations there is the danger that the full benefits of the developing short-range climate prediction technology may not be attained.

One key issue is that any demonstration of value should be sensitive to existing and accepted decision processes – and that means that different approaches may be necessary for demonstrating value to subsistence farmers in Africa, to African national and regional bodies responsible for defining the institutional support and legal frameworks in which the subsistence farmers work, to the various NGO's, and to the World Bank in its activities to alleviate poverty in Africa. Any demonstration of value should also recognise that the time frameworks of decisions throughout all scales do not necessarily accord neatly with those common to current seasonal predictions, and that therefore climate information may need to be accommodated to each specific requirement. The current relative inflexibility of climate information, including seasonal predictions in terms of their somewhat fixed spatial and temporal scales, does little to aid the demonstration of value in many cases. Climatologists face a major challenge in developing a technological package that better addresses the requirements of the decision-making community at all levels and across all scales, a challenge necessarily addressed cooperatively through expanding interdisciplinary activities.

11.2.2 Delivery of Short-Range Climate Predictions to Users

One of the keys in demonstrating the value of seasonal predictions lies in the manner forecast information is communicated to the decision making community. Given the variety of potential decision makers, as outlined above, it seems unlikely that any single communication approach may satisfy all decision makers, and that some level of customisation is necessary.

Delivery, a major element of communication, breaks down into two components, dealt with separately in the following. The first of these components, technical delivery, has proved to be rather easier to progress than the second, presentational delivery.

11.2.2.1 Technical Delivery

In principle it is now possible to deliver forecasts, and other climatological information, to potential users almost anywhere on the planet within a brief period after production. One of the vital components of the delivery system, the Internet, was just coming into widespread use at the time of recognition in early 1997 that a major El Niño event was likely to be on the way. Throughout the middle part of that year predictions and interpretations were distributed worldwide from numerous centres, including universities and research institutes, using the new tool of the Internet. In hindsight that explosive use was not necessarily beneficial, given that many of the interpretations broadcast were based on the expectation that the canonical consequences of an El Niño would occur, which did not always prove to be the case, particularly across some regions surrounding the Indian Ocean.

The Internet continues, alongside and through the media, to be one prime channel of distribution for prediction information, and no doubt will carry on being so for the foreseeable future. Numerous prediction centres, including national meteorological centres, universities, and research institutes, now provide open or subscribed access to their latest prediction information, created using numerous numerical and/or statistical modelling approaches, through this channel. But despite its immediate and obvious benefits, use of the Internet raises a number of issues that restrict its usefulness:

- In principle it permits access to forecasts for all, regardless of their appreciation of the information presented.
- It allows contrasting and contradictory information to be broadcast without consolidation or guidance.
- It makes available information in a variety of formats that the lay user may be unable to integrate with ease.
- It carries no guarantees concerning the quality of specific information.
- While potentially an invaluable opportunity for broadcasting educational material to users, there is no certainty that any educative information will be absorbed; in fact much, although not all, educative material available through the Internet is written from the perspective of climatologists rather than that of a user, and aids little in decision processes.

On the positive side the Internet has many potential benefits, not least regarding responsiveness and flexibility. Information may be tailored, for example, to aid specific decisions; in fact, in principle, information could be delivered in formats specific to individual decisions. One further undoubted advantage, amongst many, of the Internet is its ability to facilitate information exchange between climate centres, permitting them to integrate information in such a manner as to provide more focussed information for the user. Two major such integration initiatives are outlined below, but there are numerous other regional activities in various parts of the world. Regional Climate Outlook Forums (RCOFs) were a progressive development over several years, which culminated in the pilot series in southern Africa covering the 1997/98 rainfall season in that region. Originally RCOFs were conceived without the Internet, although that facility has significantly aided their development. RCOFs were designed to be an approach to bring the benefits of seasonal predictions to governments and a variety of recipients for whom those benefits might be substantial but who might not receive that benefit through insufficient resources. Thus RCOFs bring together, in meetings often exceeding 100 attendees from a number of countries:

- Climatologists, whose roles are to create consensus forecasts from all available inputs (of which there are usually numerous, including from international and national centres), to interpret that consensus for users, and to educate users on issues climatological.

- Numerous types of user, from high-level decision makers, through intermediaries, to those at the working levels, whose roles are to inform the climatologists of their views and issues, and to work with the climatologists to understand the range of actions possible, given the consensus forecast.
- The media, in their role as essential intermediaries with users throughout the region.

Following the serendipitous initiation of the southern African pilot coincident with the 1997/98 El Niño event, RCOFs were constituted rapidly in other parts of the globe in order to help address expected forthcoming major climate anomalies. RCOFs continue to be held on regular schedules in various regions, have become valued components of the annual calendar by all involved, but are threatened in some locations through the high costs of holding these events; virtual forums, in which most information is transferred via the Internet, have been tested in some regions as a direct approach to restricting costs (IRI 2000).

The second initiative is the Regional Climate Centre (RCC) programme, a project of WMO. No formal RCCs have been instituted as yet, although proto-RCCs exist in a number of regions, such as ACMAD[3] for Africa, AGRHYMET[4] for West Africa, ICPAC[5] for East Africa, and SADC-DMC[6] for southern Africa. Recognising that climate expertise is often limited, especially in developing countries, RCCs have been designed to provide a regional expert resource base on which national meteorological services from all represented countries can call for advice on climatological matters. In regard to seasonal predictions the design of RCCs tasks them with creating consensus predictions from all Global Prediction Centres[7] as well as local sources, and to interpret and provide that prediction information through national institutes, such as the National Meteorological Services within the region.

Through a combination of the Internet, RCOFs and RCCs, information on climate variability can be transmitted to and received by a majority of individuals and institutes globally. But in the developing world that set of delivery approaches omits to include many who do not have access to the facilities of modern communications. An approach to conveying information to these populations, particularly those in rural locations, has been developed, originally around the distribution of

[3] See: http://www.acmad.ne/

[4] See: http://www.agrhymet.ne/eng/

[5] See: http://www.icpac.net/

[6] See: http://www.dmc.co.zw/

[7] See: http://www.wmo.int/pages/prog/wcp/wcasp/clips/producers_forecasts.html for the full list of Global Producing Centres.

wind-up radios that require no energy sources other than muscle power.[8] Satellite technology in the form of the RANET project[9] built around the First Voice International facilities using the Worldspace system of geostationary platforms,[10] tied in with relatively cheap digital radio receivers that include modems permitting data to be downloaded directly into personal computers, have now been combined with the use of radio broadcasts to deliver climate information into remote areas of Africa, Asia and the Western Pacific. RANET, in principle, permits the reception of timely seasonal prediction and other climate information over large parts of the globe that otherwise would have no access.

Other delivery systems not yet considered are possible. Prominent amongst these are the use of mobile telephones, a technology that has expanded rapidly in many developing countries. Innovative use of existing social structures, often untapped as yet, would assist with information delivery and education.

11.2.2.2 Presentational Delivery

Far more complex than physical delivery of information, and more important in determining the value of the transmitted information, is the communication of that information in the most effective manner for decision making. From the recipient perspective, whether an international manager or a subsistence farmer, optimising the input of climate information into decision making is the prime objective. It is not enough, therefore, merely to provide predictions, with related validation and verification information, but it is a responsibility to ensure that the entire package is presented such as to facilitate the decision processes of the ultimate recipients. Few of the current delivery channels offer information organised from the decision making perspective.

Communication separates into two aspects – visual and oral/written, both with their own specific pitfalls. The focus in the following is on oral/written, but first consider the presentation of predictions from RCOFs as an example of one issue of visual communication. The RCOF consensus predictions take a probabilistic form not purposely designed to assist any specific decision process, and are presented in tiered probabilities for each sub-region (Fig. 11.1). Yet in the background each region is coloured with an indication of the most probable outcome, an indication that readily may be taken to indicate a deterministic prediction. As an exercise, consider what message would be taken from this display by those

[8] The original wind-up radio, Freeplay, was designed by Trevor Baylis and initially manufactured in South Africa, later China.

[9] See: http://www.ranetproject.net/

[10] See: http://www.firstvoiceint.org/ and http://www.worldspace.com/

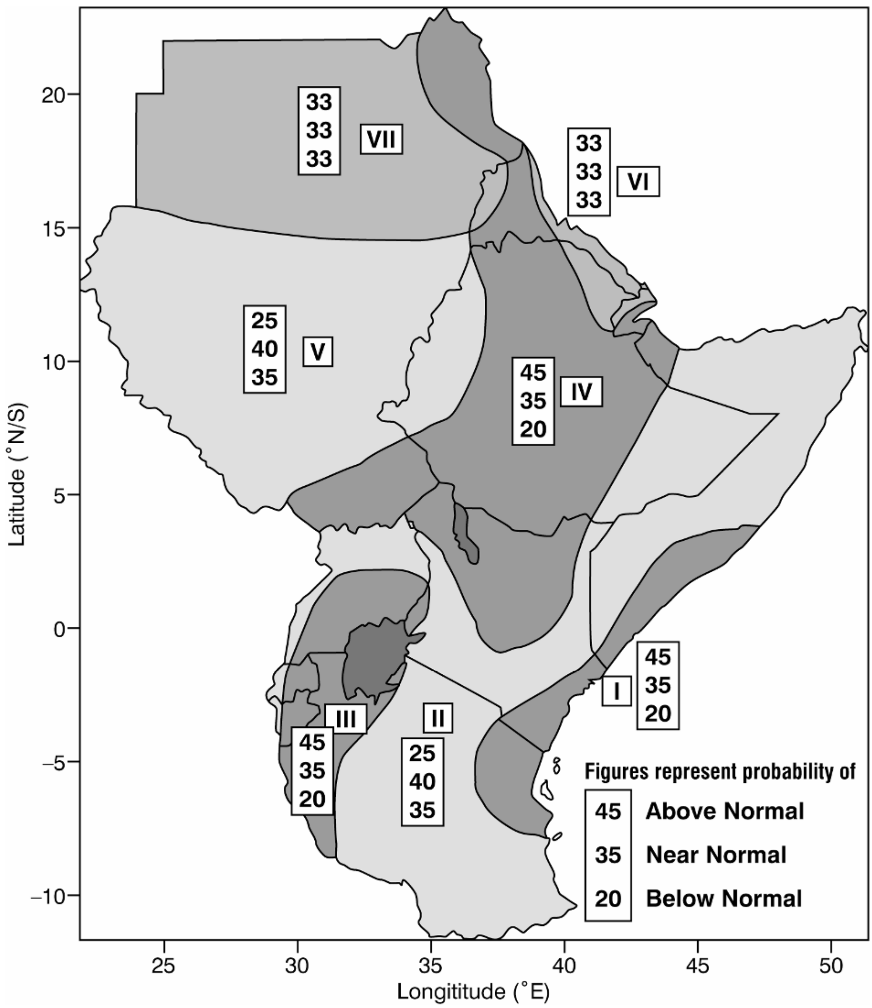

Fig. 11.1 Example of consensus prediction, for September–December 2006, from one of the East African Regional Climate Outlook Forums (courtesy of ICPAC). For each region tercile probabilities are indicated by figures, with wettest terciles at top. Background colours in the original, here converted to grey scale (for colour version see: http://www.icpac.net under "Forecasts"), indicate the most probable outcomes. Although present, but not clear in grey scale, these colours, which are prominent, often combine two similarly likely terciles into the form, say, "unlikely to be driest tercile"

unversed in its interpretation. Then consider how an untrained individual might interpret the visual information offered by the various alternate web sites providing predictions, each with their own visual presentational characteristics, for the same region. Finally, consider how that untrained individual might interpret the validation information presented on the various web sites, with their wide variations in selected metrics. The opportunities for delivering unintended messages through ill-considered visual presentations are endless, and, as is discussed further in Chapters 12 and 13, ideally all such presentations should be considered in full in collaboration with all recipients.

Written and spoken communications, as visual communications, offer climatologists numerous opportunities to fail in exchanging information with stakeholder communities. Perhaps one of the simplest examples is given by use of the word “normal”, a word that occurs frequently in the languages both of climatologists and of stakeholder communities, but one that translates imprecisely between individuals on numerous occasions. In fact during an RCOF in Kenya a few years ago there was most animated discussion of the interpretation of the word “normal” amongst over 100 delegates, with no consensus achieved. What might seem to be a “normal” rainfall season to a climatologist may appear anything but to, say, an agriculturist whose concerns extend well beyond rainfall totals themselves. Equally a “normal” agricultural season may be anything but in climatological terms. These two examples fail to illustrate by far all the contexts within which the word “normal” might lead to misinterpretation. Other apparently simple words may be equally readily misconstrued – “extremes” is a further example that arises regularly.

More complex miscommunications occur in a number of ways (affecting visual as well as written communication), including through the psychological processes that are sometimes referred to as ‘cognitive illusions’. The ‘framing effect’ offers a straightforward example – the two statements “there is a 30% chance of a drought this coming season” and “there is a 70% probability that rainfall will be adequate for cropping this coming season” effectively provide the same information, but the manners in which the statements are stated, or *framed*, invite possibly diverse decisions, a defensive approach against drought in the first case and a positive response to take advantage of possible beneficial climate conditions in the second. Further cognitive illusions are listed in the table below (based on Nicholls 1999), with examples given relating to events and experiences in RCOFs during 1997 and 1998 in southern Africa (when, given the impending substantial El Niño event, many presumed that rainfall-season drought, perhaps severe, was inevitable, although in reality adequate rain fell in many areas):

Cognitive illusion	Brief explanation	RCOF example
Framing effect	Framing of the same information in different ways invites dissimilar decisions. For example, '40% probability of above average rains' vs. '60% probability it will be average to dry'.	1997/98 worded forecasts were framed more from the perspective of drought (because of the *overconfidence* of some involved – see below) than from the perspective of possible average to above-average rainfall.
Availability	Biases originating either in information readily brought to mind or experienced recently, or in the relative availability of information, say through the media or Internet. More weight is given to recent or readily remembered information and/or to widely available information than to easily forgotten information or that with more restricted availability, however valid.	Media focus, often concentrating on the 1982/83 El Niño-related drought, preconditioned many to expect dry conditions in southern Africa because of the upcoming El Niño. Contrasting information that El Niño events are not always related to droughts in the region was not readily available. Many RCOF delegates made frequent reference to the 1991/92 drought, an event then clearly remaining in numerous minds.
Anchoring	Decisions may be based on, i.e. anchored to, familiar but perhaps unrelated and/or irrelevant information. On occasions this may be completely disconnected information, even if it is not perceived as such. But more often it may seem to be related but in reality offers limited benefit regarding current decisions.	Interpretation of the likely outcomes of the 1997/98 El Niño were frequently anchored on the 1982/83 and, especially, 1991/92 droughts. As background there was a 'weak' El Niño during 1991/92. The drought that year was one of the worst on record and was fresh in memory at the time but gave little or even no pertinent information regarding conditions in 1997.
Asymmetry between losses and gains	Related to *framing*, the perception gained that a particular line of action will yield the best outcome in terms of balancing possible losses and gains. Decisions tend to be made in either the most favourable or the most defensive manners according to individual perceptions of risk.	Wording of 1997/98 RCOF predictions, while carefully prepared, nevertheless promoted the concept of impending severe drought. This lead some to take drastic actions, such as not planting or selling/slaughtering all cattle, in the belief that such actions would minimise overall losses.
Ignoring base rates	Neglecting prior probabilities in coming to a decision – they can have considerable effects on decision processes based on a prediction. A classic example is for the UK	The prior probability of below-average rainfall across the full October to March rainfall season in South Africa, given an El Niño event in train, is only about 60%. This fact was usually ignored during 1997/98

(continued)

Table (continued)

Cognitive illusion	Brief explanation	RCOF example
	24-hour forecast, which has a claimed 'accuracy' of 83%. Given that, and assuming rain is predicted (deterministically), then the probability of *no* rain falling within any particular hour is not 17% but 70%. This offers a rather different perspective on the original forecast. The figure of 70% may be calculated (an enlightening exercise to undertake) using the base rate of 0.08 for hourly rainfall, i.e. the prior probability of rain in any hour is 0.08 (Matthews 1996).	when many scientists and others unconditionally linked El Niño events with drought (the quoted figure of 60% is a 'guestimate' derived from the South African Weather Service web site).
Overconfidence	Everyone tends to be overconfident in their beliefs and predictions, and climatologists, farmers, health practitioners, etc., are no exceptions. Only through regular critical feedback might overconfidence become less prevalent. Overconfidence tends to be greatest when accuracy is near chance levels.	Overconfident predictions for an El Niño-forced drought in southern Africa were made by many climatologists and others in 1997/98. Related equally overconfident expectations of drought impacts on farming and water resources in 1997/98 were assumed unconditionally in several southern African countries. Forecasters of necessity create individual seasonal forecasts relatively infrequently, and thus receive feedback equally rarely, leading to the potential for overconfidence.
Confirmation and hindsight bias	Biases that result from overconfident and imprecise recalls of past events (for example, forecasters tend to recall and promote preferentially their perceived previous accurate predictions while tending to discount or even ignore those 'less accurate'). Or biases based on presumed correct knowledge. The overconfidence is such that information that might disconfirm the bias is often not sought or is rejected.	Bias that El Niño events inevitably result in drought over southern Africa was widespread in 1997/98. Disconfirming evidence demonstrating that there is not a one-to-one association between drought and El Niño events was either discounted or, more generally, simply ignored or not even sought.

Table (continued)

Cognitive illusion	Brief explanation	RCOF example
Decision regret	Stronger potential regret in hindsight over decisions that lead to loss than those that lead to benefit. Thus there is often an irrational bias towards taking decisions so as to minimise possible regret should those decisions prove wrong.	Given that the 1997/98 southern African rainfall season was anticipated so negatively in deterministic terms by many, it is perhaps not surprising that no examples of decision regret immediately stand out from that RCOF experience – options simply did not appear to have been available *at the time*. Note that this is distinct from the regret felt *in hindsight* in southern Africa around decisions based on expectations of a drought some believed had been forecast unequivocally.
Inconsistent intuition	A preference to base decisions on personal intuitions rather than on objective methods. The former tend to be inconsistent while much evidence suggests objective approaches tend to produce the best results overall. In part this results from a lack of trust in automation and an inclination to make own judgements.	Intuitive modification by experts of probabilities when combining various numerical and statistical predictions during the course of agreeing a RCOF consensus forecast likely resulted in sub-optimal consensus predictions of probabilities.
Belief persistence	Either (a) first impressions are recalled preferentially over later ones, or (b) inertia in changing beliefs on the basis of later information received.	As an example of (a): the perceived ‘failed’ first RCOF forecasts for 1997/98 lead to later persistence of the belief that seasonal forecasts are ‘poor’ as a general rule (this perception required subsequent management attention). As an example of (b): the first forecast seen during the RCOF prediction consensus-building process is often weighted more highly than later-seen forecasts – in RCOFs these first forecasts tend to be each climatologist’s ‘own’, which that climatologist then tends to over-weight against other forecasts.
Group conformity	Group dynamics may lead to an erroneous biased consensus, perhaps through the dominance of one individual or of one well-presented opinion.	The process of producing an RCOF consensus prediction by a group of experts through subjective interpretation of various individual inputs usually results in a fine example of group conformity, as does the interpretation of that prediction by groups of users.

All of the above cognitive illusions may afflict both climatologists and stakeholders in a variety of manners, and all may lead to sub-optimal information delivery and processing, and hence decision making. Noting that the manner in which prediction information is employed may substantially affect the ultimate value obtained for any application, value that is often assumed, incorrectly, to be dependent upon prediction skill alone, then it becomes clear that there are numerous impediments to demonstrating, and maintaining, the value offered by a prediction service. Forecast quality, approaches to information delivery, and the manner in which information is processed and in which decisions are made, all influence separately but substantially the final outcome value achieved.

A third verbal issue, that stands alongside the interpretation of words and the manner in which words are used to communicate information, is the issue of the language used itself, often referred to as jargon. Normally – and that word is used according to this particular author's normal manner – climatologists and stakeholders use language in their own accepted community-wide ways, but with, unfortunately, limited overlap between those language groups. Effective communication is most assured when similar language is used on both sides, and in principle in order to feed the information most efficaciously into decision making the language used primarily should be that of the stakeholder rather than that of the climatologist. Climatologists thus face the extensive challenges of understanding the languages of the stakeholders they are attempting to serve and of focusing their information into the decision models of those stakeholders.

The policy and development community, for example, tends to use words such as 'vulnerability', 'resilience', capacity', 'development', 'poverty', 'equity' and 'adaptation'. Each of those words originates in conceptual models widely accepted and understood within the recipient community. These communities are seeking assistance and insights from climate experts that will directly help resolve issues as they interpret them, and it is into these models or issues that climatologists ideally should be projecting their information. Using language such 'El Niño', 'drought', 'above average cyclone frequencies', and the like, tends primarily to maintain, or even build, barriers between climatologists and recipient communities. Climate change jargon provides some excellent examples of the possible language/conceptual barriers raised between communities (adapted from Mathur et al. 2004):

- The IPCC considers climate change a pollution problem whereas the development community is concerned with the practical implications in terms of development, poverty and equity, and, in particular, with their management – information provided by the IPCC in general offers little to resolve the issues of the development community.
- While the climate community looks at science, scenarios and impacts, the development community is concerned with priorities, assistance strategies, and reductions in vulnerability.

- Climatologists discuss ‘future adaptation’, ‘top-down perspectives’ and ‘global assessments’, whereas the development community is concerned about ‘baseline adaptation’, ‘bottom-up perspectives’ and ‘national assessments’.

There are some fundamental disconnects in the above list, disconnects that apply in both directions between the development and the climate variability communities. Disconnects of this type need to be addressed if the full potential of seasonal prediction, and of climate information in general, is to be achieved.

It is not only with policy and development activities that language barriers exist, these barriers also being present in communication with expert sectoral forecast users, including those working within commercial contexts. The following table summarises language/perceptual barriers with regards to water management

Factor	Scientist’s perspective	Water manager’s perspective
Identifying a critical issue	Based on a broad understanding of the nature of water management	Based on experience of a particular system
Time frame	Variable	Immediate (operations) Long-term (infrastructure)
Spatial resolution	Defined by data availability or funding	Defined by institutional boundaries or authorities
Goals	Prediction Explanation Understanding of natural system	Optimisation of multiple conditions and minimisation of risk
Basis for decisions	Generalising multiple facts and observations Use of scientific procedures and methods Availability of research funding Disciplinary perspective	Tradition Procedure Professional judgement Training Economics Politics Job risks
Expectation	Understanding Prediction On-going improvement Statistical significance of results Innovations in method/theory	Accuracy of information Appropriate methodology Save money and time Protect the public Project jobs, agendas or institutions
Product characteristics	Complex Scientifically defensible	As simple as possible without losing accuracy Importance of context
Frame	Physical (atmospheric, hydrologic, etc.) conditions as drivers Dependent on scientific discipline	Safety and well-being Profit Consistency with institutional culture, policy, etc.
Nature of use	Conceptual	Applied

that need to be overcome to achieve maximum communication and information transfer and optimal decision making[11]:

The list in the above table extends the issue from language representing a barrier simply in terms of disciplinary outcomes and objectives to the institutional and cultural impediments built into language and communication. For example, as taken from the above table, any scientist might consider data issues primarily in terms of availability and funding while the water manager might be considering the specifics for a given system regardless of data availability. Similarly, while the scientist might be concerned with the statistical minutiae of predictions and their verification, the water manager would be most concerned with receiving and using accurate information that helps him/her meet his/her objectives.

Any ideal delivery system would overcome all visual, language, cognitive and psychological barriers discussed in this section – word selection, word usage/cognitive illusions, and disciplinary/cultural/institutional language use – and deliver information as required by each stakeholder ready for their processing directly into their particular decisions. To date, consideration of this objective within the context of seasonal to interannual prediction, an issue of far greater immediate importance than improving forecast quality *per se*, has been limited, and predictions continue in general to be provided in the eye of the climatologist rather than in the eye of the recipient. Focussed attention to the issue of communication is necessary if seasonal prediction is to deliver full benefit in all contexts; options raised frequently recently include the creation of 'bridging institutes' and the training of a cadre of communicators able to bridge the gaps between climatologists and recipients.

Acknowledgements The authors would like to thank Ms. Benedict Owuor (ICPAC) for kindly providing the original figure used to generate Fig. 11.1.

[11] Adapted from a personal communication from Pulwarty R (2003).

Chapter 12
Building National and Specialised Climate Services

John Bellow, Abdalah Mokssit, Jim O'Brien, and Rachid Sebbari

Beneficial application of information, whether this information is based on historical data or on predictions, is the ultimate objective of any climate service. The definitive measure of success of a climate service is in the value that climate information imparts in its final use. In earlier parts of this book we have examined the science of seasonal to interannual prediction itself, whereas here we are exploring the ways in which that information might be used. By contrast with the scope for the earlier sections of the book the possible issues to be considered here are substantial, and beyond any capability to provide a fully comprehensive treatment. In the previous chapter, there was an examination of structural and institutional issues requiring serious consideration in the establishment of climate services, whereas in this chapter and the following one there is a restricted, of necessity, review of some current activities in the field. It should be stressed that all examples in these two chapters follow the end-to-end model for applications rather than the integrated approach, a presentation consistent with the prevailing on-the-ground situation at the time of writing; in due course the balance between the two models may change. The two examples in this Chapter both relate to the building of climate services, in the first case by a Meteorological Service for a range of prospective users, and in the second by a University providing services mainly to the agriculture and forestry sectors. Developing world countries are frequently well-placed geographically to benefit from the maximum prediction skills available, and in the first section the full process taken by the Moroccan Meteorological Service to develop climate services to support planning for agriculture and water, including at high government levels, is described; thus in this example there are multiple users at different decision levels. A commercial approach for agriculture and forestry is described in the second section, where the potential user base is broad with a wide

John Bellow and Jim O'Brien
Florida State University

Abdalah Mokssit and Rachid Sebbari
Direction de la Meteorologie Nationale (Morocco)

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

range of requirements that cannot be serviced on a one-to-one basis, and therefore less personal communication processes, in this case built around the Internet, are needed.

12.1 The Genesis of Seasonal Forecast Needs in the Maghreb

12.1.1 Introduction

Around the world many countries are facing problems of maintaining water supplies in the face of long lasting droughts and other hazards, such as floods, wind storms and heat waves. Across the Mediterranean region the imbalance between water resource and demand within linked sectors, such as agriculture, is often critical, although there are large disparities between countries. Hence there is increasing need to provide seasonal climate predictions to enable countries to manage activities in sectors relevant for their development, such as agriculture and water resources, especially in cases when high impact anomalous climate events are expected.

12.1.2 Seasonal Forecasts for the Maghreb Region

Recent scientific achievements of the 20th century, based on enhanced observational capabilities, improved understanding of the physics of the climate system, and the advent of computer-based numerical climate forecasting of seasonal climate patterns, have enabled the international scientific community to begin predicting some components of the climate system for the next season, or even the next year, for some regions of the world.

Two fundamental approaches are used to make seasonal predictions. The first is based on statistical techniques, such as regressions, canonical component analyses, singular value decompositions, etc. (see Chapter 7). This approach can provide a useful initial understanding of the mechanisms that generate climate anomalies. It is mainly based on the identification of potential predictors, such as El Niño or other sea surface temperature anomalies, and the use of appropriate statistical methods to formulate quantitative predictor(s)-predictand relationships which represent the prediction model. Prior to developing a statistical model for Morocco, precipitation indices were computed based on a regionalization study using rotated principal component analyses, which led to identification of five coherent overlapping regions. A quantitative comparison of calculated versus observed values of the predictand provides hindcast verification of forecast performance, i.e. assessment of the skill of the model.

The second approach is based on numerical modelling (see Chapters 4–6). The prediction model is developed from basic atmospheric and oceanographic circulation theory rather than from past observations, as with statistical techniques. The basis of this approach is that tropical oceans and atmosphere behave as a coupled system. This approach is based on numerical modelling of the impacts of sea surface temperature (SST) anomalies on the (mainly) tropical atmosphere. Models are now able to produce realistic simulations of the major large-scale features of ENSO and to provide predictions of the future state of ENSO from information about the current state of the ocean, and also of Moroccan rainfall and temperatures. Additionally the two approaches, statistical and numerical modeling, can be combined to make predictions of Moroccan temperatures and rainfall through downscaling techniques (see Chapters 7 and 8).

12.1.3 Seasonal Forecasting – The Al Moubarak and El Masifa Projects

The Direction de la Météorologie Nationale (DMN) of Morocco has explored both statistical and dynamical approaches to making seasonal predictions of precipitation in Morocco through two major projects, Al Moubarak (based on statistical models) and El Masifa (based on both dynamical and statistical models).[1] These studies have led to adoption of a statistical model which uses SST anomalies over the tropical Pacific Ocean in October-November-December to make predictions of precipitation for February-March-April over Morocco. Also, the skill of the Arpège-Climat dynamical model from Météo-France has been evaluated, and this model is now running on the DMN supercomputer (an IBM) to make seasonal predictions every month using SST anomalies.

Nevertheless one has to recall that the purpose of long range prediction is to implement a procedure, in this case end to end, that produces information end users consider helpful in decision making. To achieve this, the following steps are normally adopted by DMN, as was the case in the El Masifa and Al Moubarak projects:

1. Produce a clear definition of the seasonal forecasting project objectives
2. Define a realistic project framework in terms of deliverables; all spatial and temporal scales to be agreed between the National Meteorological and Hydrological Service and the end user (all end users will have their own interpretations and understanding of each variable).
3. Complete preliminary studies to assess predictability in the targeted region while collecting all the necessary data: climate data, ocean data (including

[1] The two projects, Al Moubarak and El Masifa, are part of the same programme: Al Moubarak.

SSTs), data from existing numerical model simulations and prediction experiments, reanalysis data (Fig. 12.1).

4. Determine climatologically homogeneous regions and define appropriate climate indices for all (Fig. 12.2).
5. Determine sound project methodologies and procure tools as necessary (such as a Global Circulation Model from an advanced centre – this might be achieved in partnership with the advanced centre, in the Moroccan case Météo-France, although some international/organizational funding might be required).
6. Conduct an a priori evaluation of potential prediction skills (Fig. 12.3).
7. Agree exchange and evaluation protocols with targeted end users. In these protocols, all conditions and limitations on use of the prediction bulletin needs to be carefully defined, and end user commitment to providing feedback should be obtained.
8. Design format of an end user information bulletin: ensure terminology used is understood by end users. Agree with end users dissemination method for the information bulletin (consider collaborative end user participation in producing the bulletin, as is the case in the Regional Climate Outlook Forums, see Chapter 11).

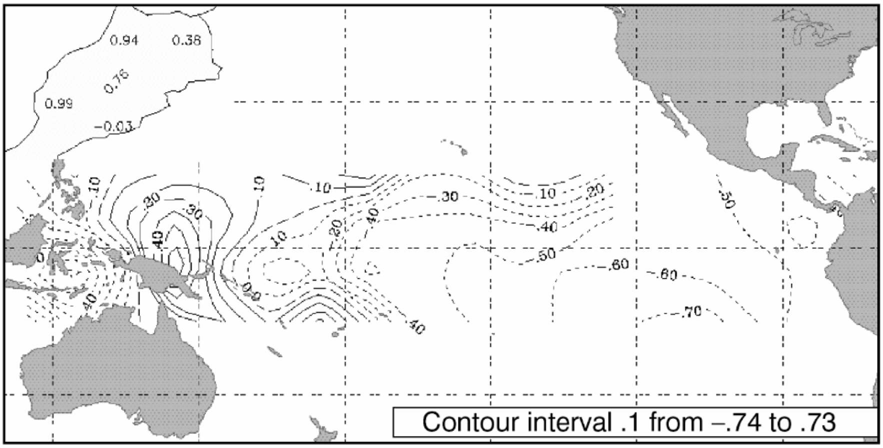

Fig. 12.1 Using a statistical technique known as Canonical Correlation Analysis (CCA) estimated patterns of correlation between two fields have been derived as an example of a preliminary predictability study. The diagram illustrates the most important pattern of correlation between Sea Surface Temperature anomalies in the tropical Pacific Ocean in the months October–December (OND) and February–May (FMAM) rainfall across Morocco (shown in the top left-hand corner, where the numbers represent correlations between regional indices of rainfall and the first rainfall canonical component). In this case SST Anomalies over most of the Pacific Ocean tend to be inversely correlated with Moroccan rainfall; the negative correlations in the Ocean are strongest in the region affected by ENSO. Overall there is a relatively high correlation between Pacific Tropical SST anomalies and late season rainfall at 4 months lag (the time lag is here measured taking the middle of the seasons considered)

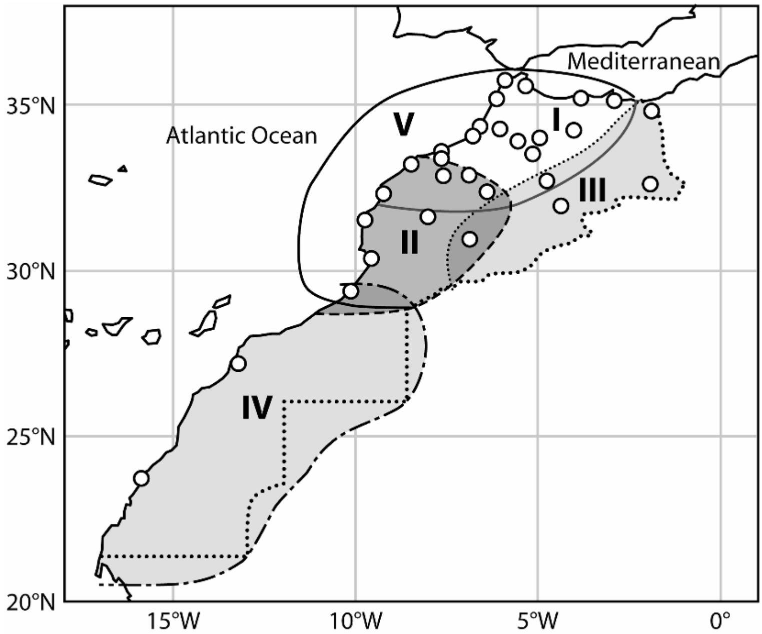

Fig. 12.2 Five overlapping rainfall regions for Morocco, derived using rotated principal component analysis based on precipitation data from many synoptic stations (circles). These regions were used in the El Masifa project

9. Develop evaluation/verification procedures agreed with end users.
10. Create a forum for communication with the end users.
11. Hold capacity building workshops to improve the overall system, including communications between scientists and end users.

Results of the preliminary predictability experiment presented in Fig. 12.3 were based on a 15-year Arpège-Climat experiment for the Maghreb region. In summary, key results from this experiment were:

- The quality of predictions from the ARPÈGE model decreases eastward, from Morocco to Tunisia.
- Statistical methods are more accurate for Algeria.
- The Analogue method is better than other linear methods.
- Tunisia is the country with lowest rainfall predictability (perhaps because of its distance from the Atlantic Ocean).
- Prediction skill is generally higher for seasons than for individual months.

To summarise this section, the operational end-to-end chain that goes from the raw climate data to the final decision process in a given sector is illustrated in Fig. 12.4.

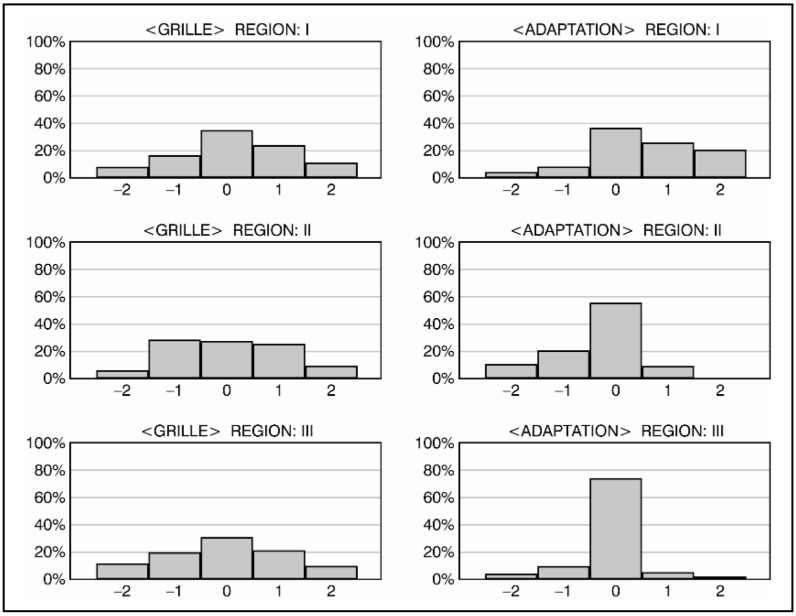

Fig. 12.3 Results of rainfall prediction hindcast experiments – differences between predicted rainfall terciles less observed terciles are shown based on direct grid point predictions from the Météo-France Arpège GCM on the left and following statistical downscaling of the model predictions on the right. The Regions are as in Fig. 12.2. In total 40 Arpège-Climat runs were used. The statistical downscaling used principal component analyses of predicted 500 hPa height fields across the period December 1978–November 1993, from which the ten highest-loaded components were used in the multiple linear regression downscaling model to district monthly rainfall indices

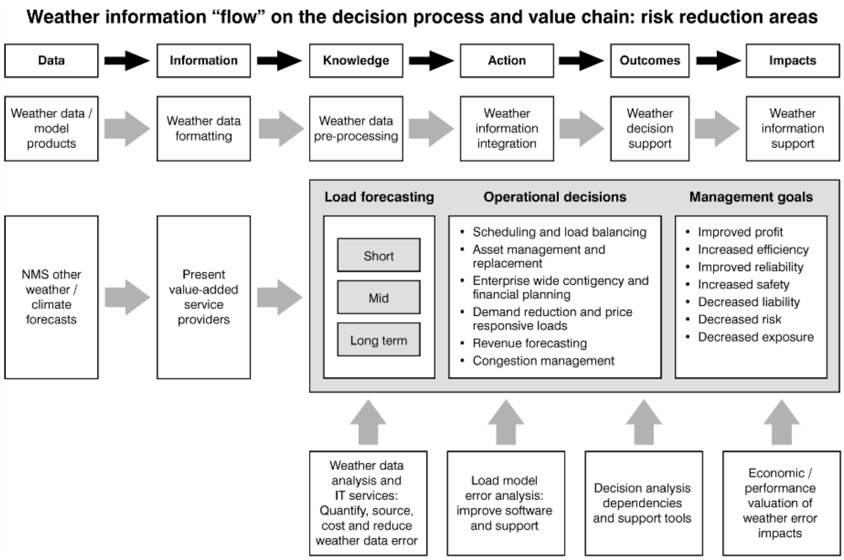

Fig. 12.4 Weather information “flow” in the decision process and value adding chain: Risk reduction areas

There is also a need for a dialogue between climatologists and end users regarding the limits of uncertainty in prediction, not only within a climate context but also within an end user context (Fig. 12.5).

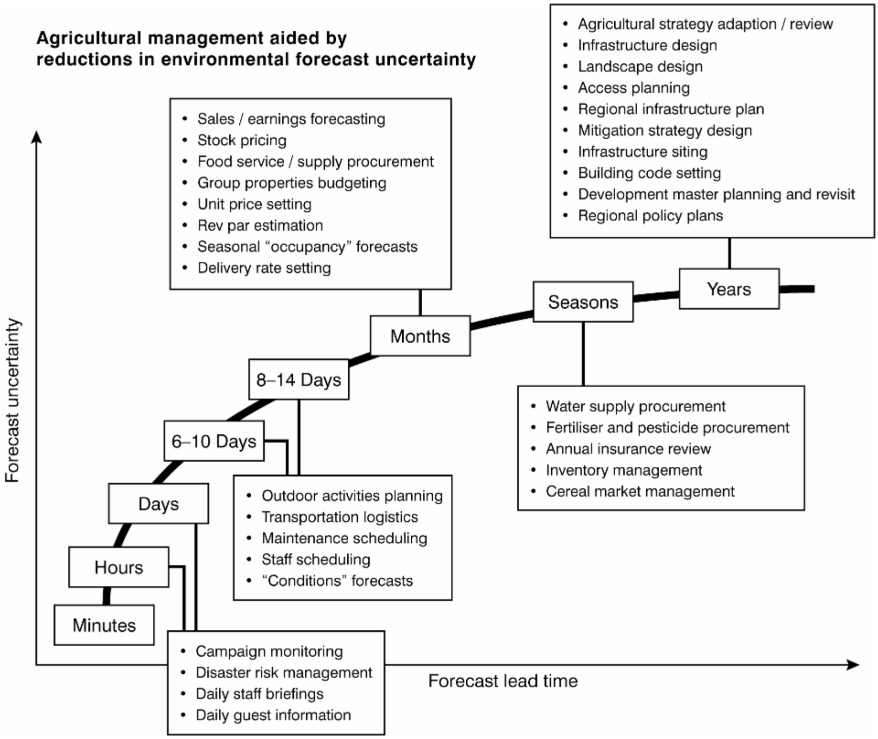

Fig. 12.5 Climate forecast uncertainty and management of agricultural applications

12.1.4 Correlation Studies of the North Atlantic Oscillation, Sea Surface Temperature Anomalies and Moroccan Rainfall

The North Atlantic Oscillation (NAO) is one of the atmospheric see-saws originally discovered, as was the Southern Oscillation, by Gilbert Walker. In this case it represents adjustments in surface pressure between the regions of Iceland and the Azores. When the oscillation is positive (relatively high pressure in the Azores region) then the winter Atlantic storms tend to take a more northerly track, bring rainfall to northern Europe but leaving southern Europe and Morocco relatively dry. These storms track further south, bring above-average rainfall to Morocco, when the oscillation is negative (relatively low pressure in the Azores region) (Fig. 12.6).

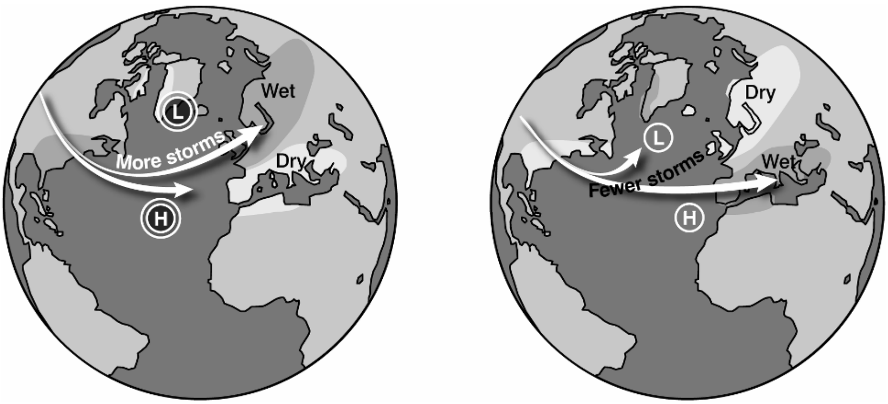

Fig. 12.6 Schematic of the positive (left) and negative (right) phases of the North Atlantic Oscillation (NAO) (adapted from Martin Visbeck and Heidi Cullen)

Results of numerical experiments to predict the two indices representing the North Atlantic and Southern Oscillations with the Arpège-Climat model at various lead times are shown in Fig. 12.7. Correlations remain positive between predicted and observed values of the Southern Oscillation Index for several months, whereas those for the NAO index become zero almost immediately. Unfortunately these

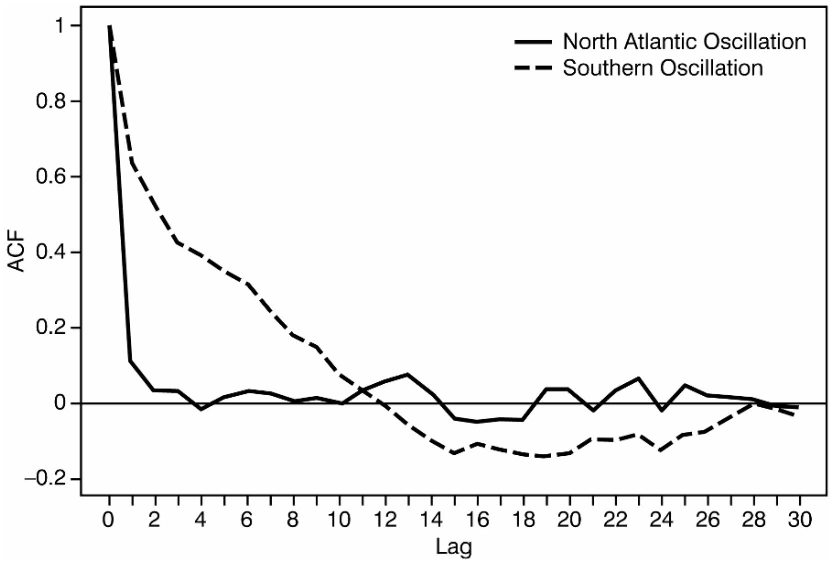

Fig. 12.7 Lagged correlations between observed and predicted values of the North Atlantic and Southern Oscillations indices based on experiments with the Arpège-Climat model

model experiments, as with many others, suggest that the predictability of the NAO index, together with onward prediction of Moroccan rainfall, is limited by this approach. More promising for prediction of Moroccan rainfall are approaches through the Southern Oscillation and/or Pacific Ocean sea surface temperature anomalies.

Correlations between the Arpège-Climat model ensemble mean and observed seasonal rainfall over Morocco reach about 0.4, predictability that presumably originates more in the South Pacific Ocean than it does in the North Atlantic Ocean (Fig. 12.8).

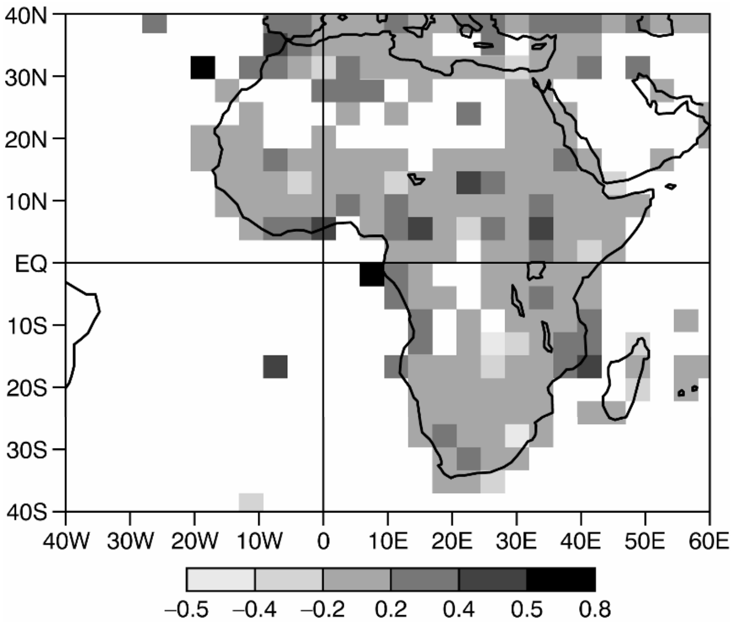

Fig. 12.8 Cross-correlation between observed precipitation values and the ensemble mean of three GCM ECHAM4 runs over 1961–1993 during February, March and April, with values around 0.4 for Morocco

12.1.5 Operational Seasonal Forecasts Using Global Climate Model Runs

The main method used in Morocco for prediction is statistical downscaling of numerical ensemble predictions from the Arpège-Climat model. In summary, the ensemble approach adopted is as follows:

- Nine runs (ensemble members) are started from different initial atmospheric conditions, and forced with the same predicted SST anomalies, created by making predictions for the 5 coming months using a simple multiple regression model in which the predictands are SST anomalies for the previous 2 months.
- A further nine ensemble runs are started from the same initial atmospheric conditions but forced with different predicted SST anomalies, created by adding small perturbations to the latest observations and then employing the same empirical model as described in the previous bullet.
- Initial atmospheric conditions are those for the last day of each month and for each of the previous 8 days.
- In practice predictions are substituted for analyses at DMN since DMN has no access to global analyses – this is not expected to cause any severe degradation in the quality of the predictions.

12.1.5.1 Direct Model Products

Two main direct outputs are available from the model runs (i.e. the two sets of nine members). The first is a deterministic prediction based on the ensemble mean of the nine members initialized with different atmospheric conditions but the same SST anomalies (Fig. 12.9a).

The second direct model product is a probability prediction of rainfall based on all 18 ensemble members (Fig. 12.9b).

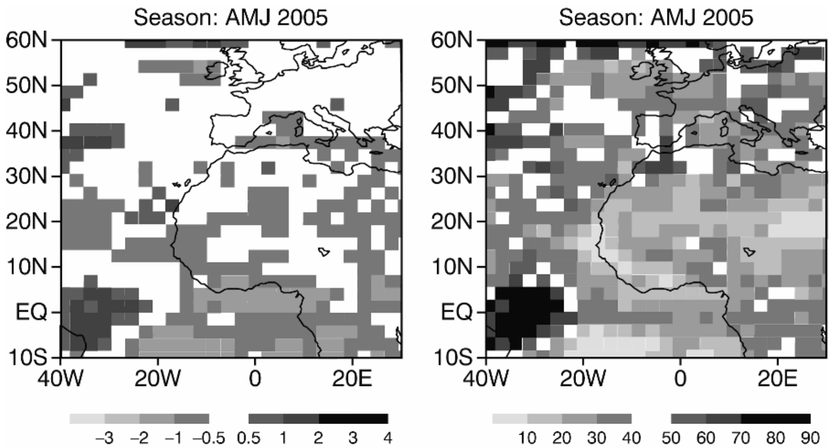

Fig. 12.9 (a) Nine-member ensemble mean deterministic prediction for April-May-June (AMJ) 2005 using February 2005 SSTs. The scale is calibrated in standard deviations. Moroccan rainfall in general was expected to be close to normal according to this prediction. (b) Probabilities that rainfall will exceed the average in AMJ 2005 according to the full 18-member ensemble: northern parts of Morocco perhaps have a better than 50% chance of seeing rainfall above average. Discrepancies in the two forecast maps stem from the use of different number of members (9 versus 18)

12.1.5.2 Spatial Downscaling to Regional Scales

There are three possible methods for spatially downscaling forecast information from global models to scales of interest to end users considered in Morocco. Two statistical approaches, regressions and development of analogues, have been assessed; the third, use of high resolution regional climate models, has not yet been adopted in Morocco. As the statistical approaches are based on 18-member Arpège-Climat model ensemble predictions, both deterministic (from the ensemble mean) and probabilistic (derived from all ensemble members) downscaled rainfall predictions are possible. Both statistical methods have been evaluated with simulations of nine member ensembles using the three European models that formed the DEMETER project and five homogeneous rainfall districts in northern Morocco (Fig. 12.2). These techniques will be applied operationally at DMN shortly.

12.1.5.3 Statistical Downscaling: Regression Method

The regression method used is based on principal component analysis of atmospheric fields including sea level pressure, 500 hPa heights and predicted rainfall. Multi-linear regressions have been created between observed district rainfall (over 20 synoptic stations over the period 1961–90 were used, see also Fig. 12.2) and the various atmospheric principal components. Loadings for these principal components are then calculated from Arpège-Climat model predictions, after which predicted district rainfall totals are derived from the regression equations. This regression method provides 18 predictions for each regional precipitation index on which to base a probability prediction.

12.1.5.4 Statistical Downscaling: Analogues Method

The analogue method works by searching for large scale patterns of a variable (such as sea level pressure or 500 hPa heights) in past years that are similar to predicted patterns of this variable from the Arpège-Climat model. Similitude is computed through a Euclidian distance or through correlation coefficients calculated over a selected number of principal components. The observed precipitation of each analogue year is then considered as the prediction. The number of predictions available from the analogue method is 18 times the number of principal components used, from which both deterministic and probabilistic forecasts may be calculated. Examples of the spatially limited direct model deterministic and probabilistic predictions are illustrated in Fig. 12.10, which also demonstrate the potential additional detail available from downscaling when compared to Fig. 12.9.

12.1.6 Calibration of Model Precipitation Indices

As an alternative to downscaling it is possible to use direct numerical predictions after these have been recalibrated (as in Fig. 12.10). The approach adopted in Morocco has been developed over the period since 1998 using a variety of different ensemble systems, beginning with three members in 1998 and extending to 18 members now; based on these ensembles, outputs can be provided in both deterministic (anomaly) or probabilistic (tercile) formats. In order to calibrate the model, in which means and standard deviations of specific variables may differ from those in the real world, it is necessary to derive reference climatologies for the model as well as for the observations. Model climatologies have been generated from retrospective runs performed in the same ensemble configuration as currently used in operations (notably for the SST statistical forecast scheme), and using the same number of ensemble members, across the 1979–1993 period. Observed climatologies have been based on either the ERA-15 data set or, for precipitation, on the Xie-Arkin data set.

The starting point for a deterministic forecast is the ensemble average, from which the downscaled normalized anomaly is converted to an anomaly recalibrated in terms of observational distributions through the following process (see also Chapter 8, Section 8.3). First, the average anomaly A_m from the model is calculated:

$$A_m = F - F_{clim,}$$

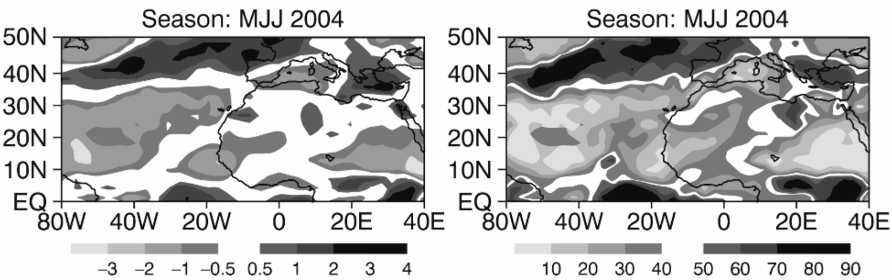

Fig. 12.10 (a) Deterministic precipitation predictions for two overlapping seasons over Morocco, May–July (MJJ) 2004 derived directly from Arpège-Climat model predictions. The scale is calibrated in standard deviations. Northern Morocco is likely to receive a little below normal rainfall. (b) As in (a), but for probabilistic predictions. The prediction suggests there is an increased likelihood of below-normal rainfall over northern Morocco. The scale is probabilities of exceeding normal

where F is the ensemble average forecast for month M and F_{clim} is the average forecast for the model climate reference experiments for the same month. Second, from this model average anomaly a normalized model average anomaly index, $I_{m,}$ is computed:

$$I_m = A_m/\sigma_{clim}$$

where σ_{clim} is the interannual standard deviation for the model climate reference experiments for that month. Third, the recalibrated predicted observed average anomaly, A_o, is calculated:

$$A_o = I_m * \sigma_{obs}$$

where σ_{obs} is the interannual standard deviation of the observations for that month; A_o is referred to as the 'anomaly' in the model product.

In the next stage seasonal anomalies $A_{o(season)}$ and $I_{m(season)}$ are calculated as simple arithmetic averages $A_{o(m)}$ and $I_{m(m)}$ across three consecutive months. Next a student *t*-test is performed to identify locations where anomalies differ significantly from model climatology anomalies:

$$t = (A_m/\sigma_{intra})*\sqrt{N}$$

where σ_{intra} is the within-ensemble standard deviation for the month or the season, N is the number of ensemble members and t is the so-called *t* value. Finally, terciles are calculated assuming a Gaussian distribution, where for the below normal tercile $I_m < -0.43$,[2] for the normal tercile ($-0.43 < I_m < +0.43$), and for the above normal tercile $I_m > +0.43$.

An example of the final prediction product used in the bulletin sent to end users is given in Fig. 12.11.

Subsequent to El Masifa, a number of prediction products from various modeling centres have been used to compile the current experimental prediction products Bulletin. These include the Bulletin El Masifa dynamical forecast from ARPEGE-Climat and the statistical forecast of SST anomalies, the NAO forecasts from CIMMS, and model predictions from ECMWF, the UK Met Office, the ECHAM3 dynamical model, the IRI and NCEP products. An example of the Seasonal Prediction Bulletin produced under the Al Moubarak programme and distributed every month since July 1996 within Morocco is provided in Fig. 12.12.

[2] The value 0.43 (of a standard deviation) corresponds to 16.5% of the Gaussian distribution, hence ($-0.43 < I_m < +0.43$) corresponds to 33% of the distribution.

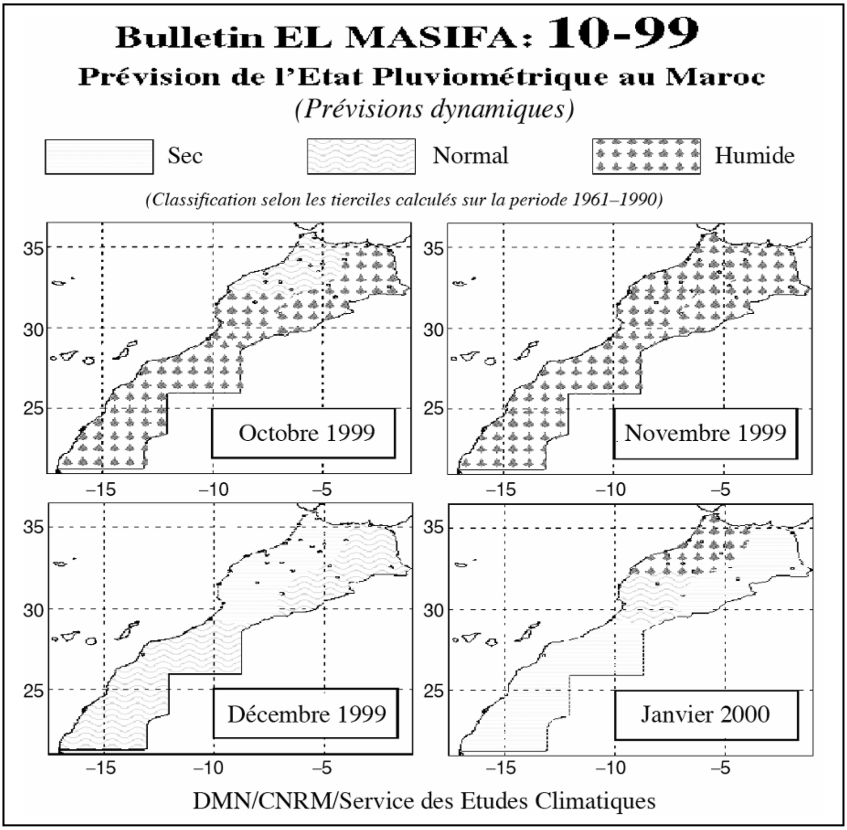

Fig. 12.11 An example of a prediction from the El Masifa project distributed during the winter of 1999–2000, and based on the five districts illustrated in Fig. 12.2. Predicted precipitation anomalies are categorized as "Above Normal" (Humide), "Near Normal" (Normal), and "Below Normal" (Sec)

12.1.7 Generating Interaction and Collaboration with Users

The most important step in developing a service based on seasonal predictions is to identify key sectors and decision makers that could benefit from seasonal forecasts in their management processes and to get in touch with potential users. The relationship with users could be set up in many ways but it is preferable to have collaborative relationships instead of commercial relationships. As an example, for Morocco the key sectors are high ministerial authorities, hydrological services and agriculture services.

Direction de la Météorologie Nationale

Centre National de Recherches Météorologique

Programme El Moubarak

Bulletin de prévision saisonnière des précipitations au Maroc

Echéance : **Octobre à Décembre 1999 et Janvier 2000.** *Issue le : 4/10/99*

Introduction :

Nous présentons-ci-après un ensemble de prévision saisonnière issues de centres Météorologique mondiaux différents et qui concerne le Maroc. Elles sont toutes *à titre expérimental*. La DMN n'assume aucune responsabilité sur les décisions prises sur la base de ces prévisions.

Prévision issue du CIMMS (université d'Oklahoma - USA -) :

Les prévisions du CIMMS sont élaborées à partir du mois de Novembre et portent sur le mois de Décembre, Janvier, Février et la saison Mars-Avril.

Prévision dynamique issue d'El Masifa: (base : Septembre 99)

Le modèle ARPEGE-Climat prévoit pour la moitié nord du Maroc (Régions Sahariennes exclues) ce qui suit:

Oct 99	Nov 99	Déc 99	Jan 2000
Normal	Humide	Sec	Normal

Oct-Nov-Déc (OND 99)	Nov-Déc-99 - Jan 2000
Humide	Normal

Prévision issue de l'U.K Met-Office: (base: Septembre 99)

Pour **OND**: Précipitation **autour de la normale** sur tout le Maroc légèrement **au-dessus de la normale** sur les provinces du sud. La prévision issue de ce modèle pour cette saison n'est pas assez fiable.

Prévision issue du CEPMMT (i.e. ECMWF in French): (base: Septembre 99)

Pour **OND**: Le modèle ne présente pas de signal significatif sur le Royaume.

Prévision issue d'autres modèles climatiques: (base: Septembre 99)

Pour **OND**: Dans l'ensemble le deux modèles ECHAM3 et CCM3 prévoient de conditions **au-dessous de la normale** sur les côtes atlantiques et des conditions normales partout ailleurs.

Conclusions:

L'analyse de sorties de l'ensemble de modèles, à part le modèle ARPEGE-Climat, montre que la saison **Oct-Nov-Déc** sera en général **Normale à Sèche** sur le Maroc. Le modèle ARPEGE-Climat prévoit quant à lui un état **légèrement humide** sur la moitié nord du Royaume.

Fig. 12.12 A sample page from the Al Moubarak project Forecast Bulletin that corresponds to the illustrations in Fig. 12.11

Collaboration should focus on the following objectives:

- Evaluate user needs
- Develop and demonstrate applications which address practical user needs
- Establish interactive dialogue with primary users
- Develop data/information delivery systems

Of course, in Morocco the agricultural and hydrological sectors are leading examples with which to develop collaborations. This could be done mainly through a Memorandum of Understanding (see example in Appendix, Section 12.3), in which the key elements to consider are:

- Experimental basis
- Seasonal forecast products and their presentation (e.g. Seasonal Forecast Bulletin)
- Target areas
- Delivery system
- Need for feedback: user evaluation of the forecasts

To ensure the usefulness of the Seasonal Forecast Bulletin, the information in the Bulletin should fulfil end user information needs. The producer of the seasonal forecasts should assess end user needs and propose a product which meets these needs and is easy to interpret. The forecast producer should consider the precise information requirements, as well as the appropriate time and space scales, as shown below in Table 12.1 for the agriculture sector.

The Bulletin producer should also explain to end users the uncertainty of seasonal forecasts. This could be done through explaining the use of the product. It is also very important to show the added value that the end user can benefit from. One way of estimating this information is by using skill scores such as the cost-loss model, or any other score presented in Chapter 10. Ultimately, value depends on the quality of the forecasts, the choice of probability threshold, the event, the user, and the manner in which decisions are made and the type of decisions being made (e.g. the same forecast may mean something very important for a wheat grower and very little to a citrus grower because of differences in crop phenology, technology employed, etc.).

Table 12.1 Time and space scales associated with various agricultural practices that might be assisted by a climate service

Activity	Information needed	Time and space scales
Cultivar selection	Average temperature, total rainfall	Season, field
Campaign monitoring, seeding period	Minimum temperature, freezing days, average temperature, total rainfall	1 day, month, season, field
Operations and manpower management	Minimum temperature, freezing days, average temperature, total rainfall	1 day, month, season, field

Ensemble prediction systems have been shown in many studies to improve the quality of seasonal forecasts when compared to deterministic systems. Further, it is important to use as many seasonal forecasts as possible to produce a multi-model ensemble, many of which are available through the Internet sites of meteorological centres. Each forecast should be weighted appropriately, perhaps using the skill of the model used to make the prediction.

The quality of probabilistic forecasts of regional Moroccan winter precipitation with the three models of Météo-France, the UK Met Office and ECMWF has been evaluated in the framework of the DEMETER project. The DEMETER results indicate that using multi-model ensembles is a pragmatic approach to the problem of representing model uncertainties in seasonal to interannual forecasts: probability forecast skill scores based on multi-model ensembles are generally higher than those from any single-model ensemble.

12.1.7.1 Presentation and Delivery of Information

Following discussions with end users the final seasonal forecast product may be defined. It has been decided from discussions between DMN and Moroccan Hydrological and Agricultural Services to provide predicted regional rainfall indices in terciles representing dry, normal and wet cases on a monthly basis. As a result, the Bulletin as illustrated in Figs. 12.11 and 12.12 is now issued to end users and sent mainly by mail at the beginning of each month.

12.1.7.2 Benefiting from Seasonal Forecast: Some Success Stories at DMN

Evaluation indicated that the Bulletins provide useful information for many users. Dissemination of a prototype experimental seasonal forecast for higher-skill areas is already operational in Morocco for many users, especially high authorities, the hydrological and agricultural services. An external evaluation has been done by the hydrological service that demonstrated that the predictions are useful and can be used in many cases to make strategic decisions, such as selling water to farmers for irrigation, planting trees in forests, etc.

The usefulness of seasonal predictions is indisputable. This climate information can be utilised by many users to assist in difficult decisions, such as those concerning the amount of water available to produce hydro-energy, for supply to the agriculture sector and to communities, or for water management in general, so as to be prepared in case of droughts or floods. Such decisions are becoming more important because of the continual increase of water demand from all vital sectors, such as agriculture, energy power production, industry, etc.

Examples of the use of seasonal information in Morocco in the water resource sector include two forecasts utilized in the management of the Al Wahda Dam. In order to inaugurate the dam, the reservoir had to be filled. A large rain event was forecast on the medium range timescale: this information was used to set the inauguration date. A large amount of rain indeed happened and the dam could be thus inaugurated, in January 1997. Once the reservoir was full, the decision to generate hydroelectric power, by activating the turbines, had to be taken: this decision was eventually based on seasonal forecast.

For the Basin Management Agency (BMA) a decision to supply water to ORMVA (the agricultural office) to sell onwards to farmers for irrigation was also taken using seasonal forecast information. During the winter of 1998, the launch of a reforestation operation in the middle Atlas Mountains (1,200,000 trees planted) was based upon seasonal forecast information.

12.1.7.3 Feedbacks from Users

In order to develop seasonal prediction services it is necessary to consider product evaluation/verification. Evaluation should be done not only by the producer to demonstrate the quality of the forecast, but also by the end user to determine the benefit (value) gained from the seasonal forecast information. Some evaluation results for end-users in Morocco follow.

External evaluations were performed by the Ministry of Public Works, Agriculture and Environment. In this framework two main interesting results emerged:

- Users have their own evaluation/assessment approaches, which may differ substantially from the standard climatological ones.
- Users own perception of the forecast information plays an important role in such an evaluation/assessment.

12.1.8 Conclusions and Perspectives

In summary, based on both internal and external evaluations, it has been demonstrated that the El Masifa forecasts are improvements over climatology with 46–50% good forecasts as compared to 33% for climatology (see Table 12.3). Curiously, the assessment carried out by the Ministry of Agriculture and Public Works shows that the percentage of good seasonal forecasts (i.e. $| F - O | = 0$) is higher at lead time 3 months (i.e. M+3) than at shorter lead times (see Table 12.2).

There is still a need for better and more robust models used for seasonal forecasts for the Maghreb region: the El Masifa project has certainly contributed to model improvements and has also given the opportunity to install a forecast system which is now a useful asset for DMN.

Table 12.2 Seasonal forecasts assessment performed by the Moroccan Ministry of Agriculture and Public Works mostly based on their perception of how the rainfall was predicted. The first column, | F – O |, represent the "distance" of the forecast from the observations; a value of 0 (first row) means that the forecasts fall in the same category as the observations, hence this is considered a good forecast. The following four columns, M+1 to M+4, indicate the lead time in months. Note that the sum of each column is 100%

	Ministry of Agriculture and Public Works			
Forecast	M+1	M+2	M+3	M+4
| F – O | = 0	29%	54%	63%	22%
| F – O | = 1	41%	26%	28%	60%
| F – O | = 2	30%	20%	9%	18%

Table 12.3 Summary of seasonal forecasts assessment performed by both the Moroccan Ministry of Agriculture and Public Works (second column) and the Direction de la Météorologie Nationale (third and fourth columns). Values as in Table 12.2, but now the lead times are condensed in a single value

Forecast	Ministry sample	DMN sample	DMN-region IV sample
| F – O | = 0	39%	46%	50%
| F – O | = 1	34%	33%	39%
| F – O | = 2	27%	15%	11%

In order to improve on these achievements, several aspects of the seasonal forecasting system of DMN will be looked at in the near future:

- Introduction of probabilistic predictions via an ensemble prediction system
- Evaluation of skill over Morocco when using other GCMs
- Implementation of a coupled GCM-ocean model
- More research in downscaling and regional modelling
- Evaluation of seasonal forecasts for specific applications via case studies to respond user needs (e.g. in the agricultural and water resource sectors)

12.2 Mitigating El Niño/Southern Oscillation (ENSO) Effects in the Southeastern USA for Agriculture and Wild Fires

The Southeast Climate Consortium (SECC) provides climate services to the Southeastern United States and has grown out of its early inception as the Florida Climate Consortium (FLC), comprised of Florida State University, University of Florida, and University of Miami. The consortium has expanded with the support of the Regional Integrated Sciences and Assessment (RISA) and funding from the Risk Management Agency (RMA) and the Cooperative State Research, Extension, and Education Service (CSREES) of the United States Department of Agriculture to include Auburn University, University of Alabama at Huntsville, and the University of Georgia. The scope of consortium activities was also expanded from services for agricultural producers to encompass the needs and interests of forest managers, state and local policy makers, and water resource management agencies. The SECC links the resources of the Alabama, Florida, and Georgia state climate offices and researchers in the biophysical and social sciences at the participating universities to apply research based information about seasonal climate variability to the needs of diverse stakeholders throughout the region. The SECC's mission is 'to use advances in climate sciences, including improved capabilities to forecast seasonal climate, to provide scientifically sound information and decision support tools for agriculture, forestry, and water resources management in the Southeastern USA.'

12.2.1 Delivery of Climate Services

Previous studies have suggested that the development and delivery of seasonal climate forecasts could be beneficially applied to the management process in agriculture (Lamb 1981; Sonka et al. 1987; Stern and Easterling 1999; Hansen and Jones 2000). The actual utility of the forecasts is impacted by the specific agricultural operation under question and both the flexibility and the willingness of the decision maker to consider the forecast information. Additionally, the forecast skill, the timeliness of the forecast with respect to when meaningful adaptations can be made, and the ease with which decision makers can acquire and understand the forecast, all contribute to the benefits that a forecast might produce. Hansen et al. (2004) suggest that a clear understanding of forecast skill by growers is of particular concern because opportunities to benefit from forecasts of good agricultural conditions as well as to prepare against forecast poor agricultural conditions will be missed if forecast accuracy is not understood. Both forms of missed opportunities may further serve to damage the credibility of the forecasting organisation and reduce the likelihood of future forecasts being fully considered.

12.2.1.1 ENSO Signal in the Southeast

Climate forecasting in the Southeastern United States derives most of its skill from teleconnections with sea surface temperatures in the equatorial Pacific, and the SECC has emphasized the use of the El Niño/Southern Oscillation (ENSO) index defined by the Japanese Meteorological Agency (JMA) in its forecasts. With this index, the monthly mean sea surface temperature anomalies are averaged for the area 4°N–4°S and 150°W–90°W. The index is a 5-month running mean of spatially averaged SST anomalies. Where index values are 0.5°C or greater for 6 consecutive months (including October, November and December), the ENSO phase for October through the following September is categorized as El Niño, La Niña (index values equal or less than –0.5°C), or neutral (between –0.5°C and +0.5°C). Previous research has demonstrated that ENSO exerts a substantial influence (Fig. 12.13)

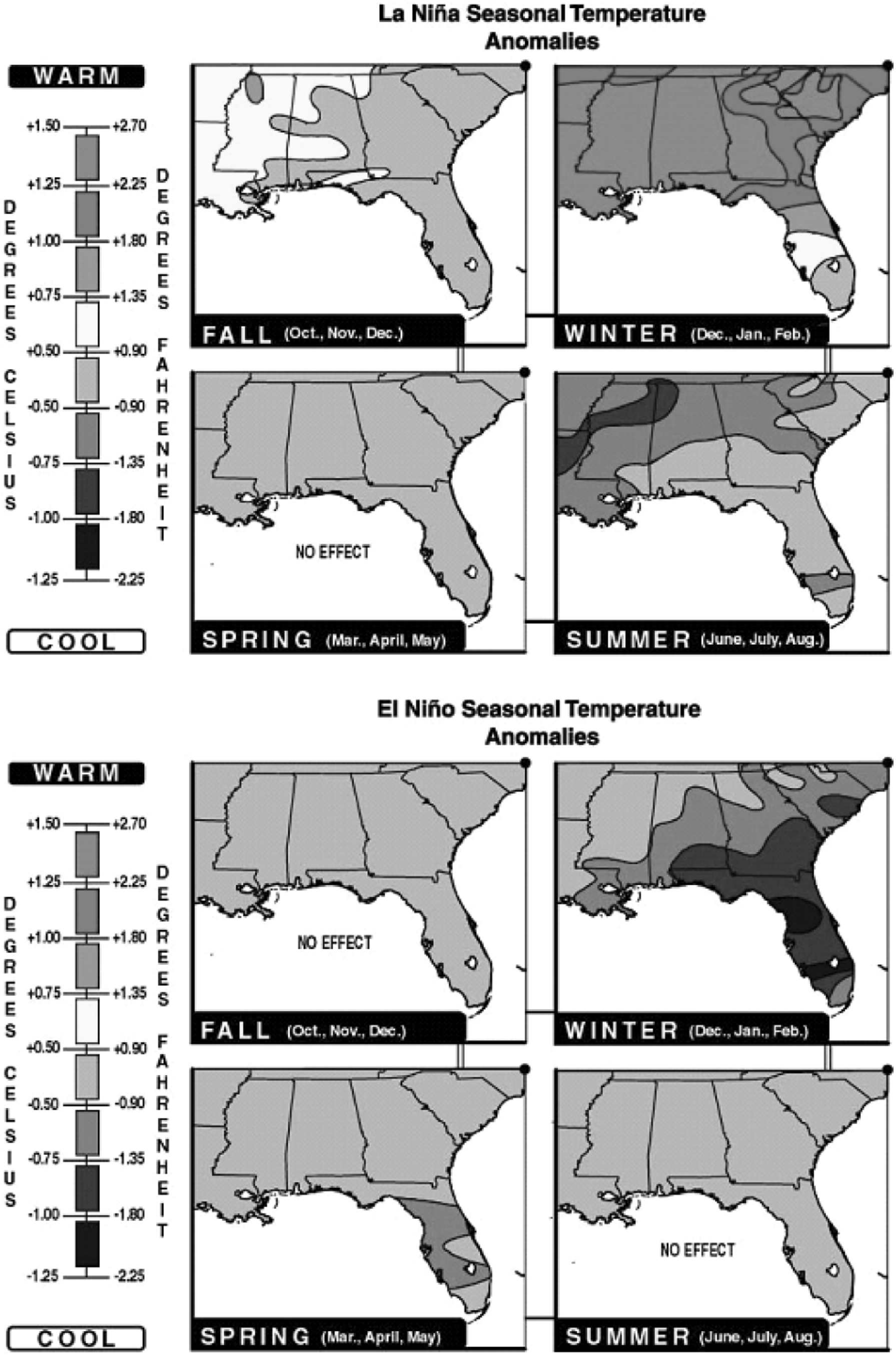

Fig. 12.13 Temperature anomalies in the Southeastern United States due to the two ENSO phases, El Niño (top) and La Niña (bottom)

whereby El Niño years tend to be cooler and La Niña years somewhat warmer between October and April (Kiladis and Diaz 1989; Sittel 1994). There is substantial spatial variability in this signal however, with a reduced amplitude in the southern Florida peninsula. Impacts on precipitation are also spatially variable, with rainfall greater than climatology generally observed during those months (Fig. 12.14) under the influence of El Niño conditions and reduced rainfall when La Niña conditions predominate (Sittel 1994).

In examining the impact of ENSO phase on regional and seasonal climate variability, historical temperature, and precipitation records from the NOAA national climate data center (NCDC) were categorized into El Niño, La Niña, or Neutral phases using the JMA definition given above. The presence of significant variability in seasonal climate that can be explained by shifts in sea surface temperature therefore provides the underlying basis for the development of climate services by the SECC.

12.2.1.2 Team Approach to Developing Climate Services

The SECC has emphasized a participatory approach with users of climate information to generating research questions and developing applications for climate information and climate forecasts. Ideally, the SECC acts as a nexus to bring together a range of stakeholders, including climate scientists and forecasters, researchers and managers of natural resources, decision and policy makers, and Extension agents and producers, and coordinates the development, formulation, and delivery of climate information appropriate to the range of stakeholder groups. This approach considers the framework of Pielke and others (Pielke et al. 2000) by explicitly including representatives involved in the generation of information/forecasts, those involved in the communication and dissemination of the information, and the end users at local and regional scales.

In practice, partnerships with stakeholders involve the SECC, the state Cooperative Extension services, who emphasize the summation and communication of current knowledge and advice to producers and resource users at the county scale, the state climatologists and climate offices who are considered official sources of climate information for state and local governments, and topic specific focus groups or advisory boards comprised of representatives of industry, producers, and Extension. These collaborations were initiated early enough to elicit potential end-user preferences for media and format of presentation and have contributed directly to early and rapid adoption by some user groups. These results also suggest that information presented through the Internet and Extension bulletins is a preferred way for stakeholders to access climate information in this region. End-users have limited time and interest in considering the full range of complexity and detail that can be communicated through terciles, deciles, or box-and whisker plots, and these early findings indicate that simple presentations, such

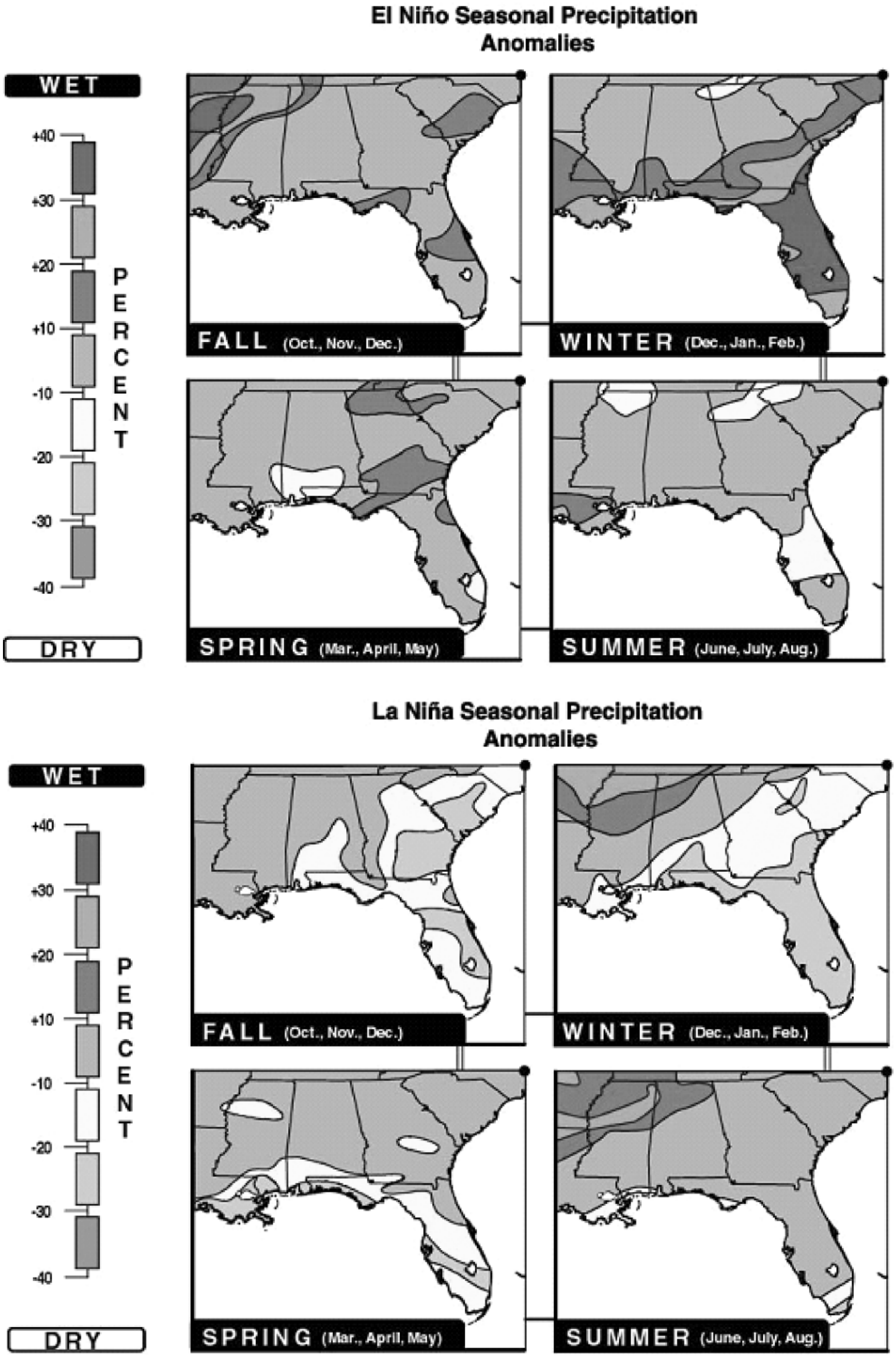

Fig. 12.14 Precipitation anomalies in the Southeastern United States due to the two ENSO phases, El Niño (top) and La Niña (bottom)

as the probability of exceeding a critical threshold, are greatly preferred, and that in many cases this information should be presented in comparison to historical or reference conditions to provide meaningful context.

The development of individual services or forecasts arises on a case by case basis as the outcome of interactions between SECC climate Extension specialists and other stakeholders. Climate Extension specialists are able to introduce to a group the current 'state of understanding of climate variability and climate forecasting' and initiate discussions about ways in which stakeholders' activities and decisions might be influenced by this knowledge. Stakeholders are able to identify their strategic management options that are influenced by variability in climate conditions and begin to identify the characteristic of a climate service product that could support their decision making process. Lead time and seasonality of the information, the accuracy or uncertainty in the information that would be acceptable, and the actual type of information that will support adaptive management, are some characteristics that stakeholders have provided that has subsequently been incorporated into prototype forecast products. Climate specialists and researchers then use this information from stakeholders to assess the full potential for developing a product with the desired attributes; whether the forecast product will have skill, if the available data supports delivery at the timescales and frequency that is required for successful adoption, and how the resultant forecast can be delivered and communicated to potential end users. Prototype decision support tools or forecasts are developed and shared with stakeholders, who are able to further contribute to its refinement by identifying weaknesses or areas for enhancement. Refined prototypes are ultimately incorporated into the AgClimate.org web site, where they become available to a broad audience. The AgClimate web site has a comments mechanism that allows additional user feedback to be solicited and incorporated into the products on an ongoing basis.

12.2.2 Examples of Applied Forecasts

12.2.2.1 Wildfire Forecast

The seasonal variation in temperature and precipitation across the southeast dramatically influences the potential for forest and grass fires, a phenomenon that historically has been determinative in the vegetation and ecology of the region at a landscape scale. The work of Brenner (1991) and Jones et al. (1999) in the southeastern USA, and that of Swetnam and Betancourt (1990) in the western United States, have shown that the ENSO phase is a significant predictor of the risk of wildfires occurring. However, in the Southeast, increasing development and population pressures, prescribed burning, and the conversion of forest and scrub lands have served to diminish the signal in terms of actual number of acres burned. In areas with extensive development and conversion, the occurrence of anomalously wet or dry conditions combined with cooler or warmer temperatures may not have the same implications for fire risk as in those areas with greater forest cover. Furthermore, due to differences in seasonal climate and plant phenology, absolute changes in precipitation or temperature levels are not directly related to changes in seasonal fire risk, as wildfire risk is magnified when the two act in concert.

Through interaction with regional foresters and land managers, a drought index widely used and accepted by stakeholders for assessing fire conditions in forests was chosen and an investigation made to examine the potential impacts of ENSO phase on this index rather than on precipitation and temperature alone. The KBDI, or Keetch-Byram Drought Index (Keetch and Byram 1968), is a cumulative index wherein the KBDI value for yesterday is modified by the maximum temperature and precipitation values for the past 24 hours to develop a new KBDI value for today. The value of KBDI for a location can then be used by managers and foresters, as well as state and local governments, to assess the fire risk in their areas. The values of KBDI range from 0 (saturated soil and fuel) to 800 (extremely dry soil and fuel). Four threshold values for KBDI, 450, 500, 550, and 650, were chosen to represent the categories of risk of interest to managers. The approach was to examine the range of KBDI values for the historic period of record 1948 through 2004 using National Weather Service Cooperative Observer Network (TD3200) data, which provide daily maximum and minimum temperatures, and daily precipitation totals. Because fire danger in the southeast is confined climatologically to the January through July period, the study emphasized monthly anomalies in the location specific KBDI values. During each of the years the evolution of KBDI values across the region was examined relative to the JMA ENSO index value. Examination of monthly KBDI values across the region indicated that fire danger is influenced by ENSO phase, with substantially increased risks during the La Niña or 'cold' phase and anomalously lower risks during the El Niño or 'warm' phase.

The next step is to develop this relationship into a probabilistic forecast for each of the threshold values at the county scale. Because of the limited number of ENSO events in the historic record, a bootstrapping approach was used to resample the historic record on a monthly basis. The bootstrap data set consisted of 1,000 years of resampled data for each station, and provided 250 years of climate data for each ENSO phase and for climatology (all years). The forecasts for each month at a location were then developed by combining the current KBDI conditions at the station with each of the resampled years from the appropriate ENSO phase. The probability distribution of the occurrence of each of the four threshold KBDI values for 7 days or more during the month was produced in this fashion. The resultant forecast was mapped to a county scale map of the three States for each threshold value, moderately dry (450), abnormally dry (500), severely dry

(550), and extremely dry (650). Probabilities were also color coded so that increasingly intense and darker reddish-brown colors were associated with increased probabilities of high KBDI values (Fig. 12.15).

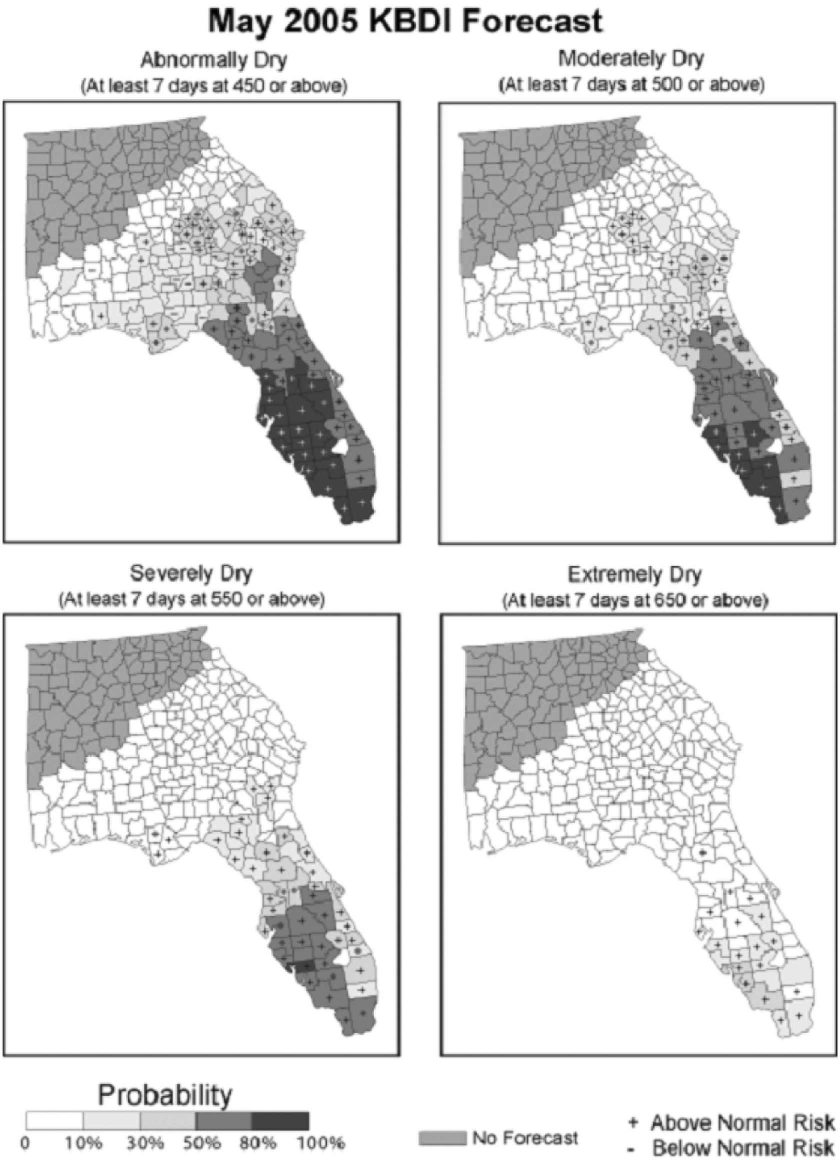

Fig. 12.15 Keetch-Byram drought index (KBDI) applied to the southeast United States using current conditions and an ENSO based precipitation forecast. KBDI drought forecasts are updated monthly with a maximum 6 month lead time

Finally, the forecast probability was compared to the probability for the central 60% of climatological values for all years, and counties where the probability of a threshold KBDI value being exceeded were marked with a plus sign '+' while counties where the probability for a threshold was found to be in the lower 20% tail were marked with a minus sign '-'. The forecast has been extensively reviewed by a broad range of stakeholders and is produced on a monthly basis during January through July. The tool (Fig. 12.15) can be viewed at http://agclimate.org. Users choose the monthly forecast period of interest from January to July and are provided with KBDI risk forecasts for four critical thresholds relevant to the southeast USA. Where high risk conditions are observed, managers may be able to reallocate resources and personnel or take preventative measures such as prescribed burning to manage the risk.

12.2.2.2 Yield Risk Tool

The earliest example in the SECC of using climate information to generate applied forecasts is for crop yield risk. The ENSO signals on precipitation and temperatures using the JMA index classifications suggest possible influences of climate anomalies on crop yields, and previous studies (Jones et al. 2000; Hansen 2002) have provided evidence supporting the hypothesis that crop yields show ENSO mediated anomalies. Participatory research on potential uses of climate information indicated that producers might respond to predicted variability in seasonal precipitation and temperature by altering the planting dates of their crops. Changes in planting dates may enhance the probability of greater crop yields either by avoiding suboptimal conditions early or late in the season or by capitalizing on expectations of good growing conditions. Crop simulation models, such as the DSSAT 4.0 (Jones et al. 2003) family of models, use daily weather variables (maximum and minimum temperature, daily total precipitation, and total incident solar radiation), previously validated parameters for individual crops and crop varieties, and location specific soil attributes, to predict crop performance and yield by simulating physiological processes driven by environmental conditions. Because these models have been relatively well validated within the southeast, as well as under a broader range of conditions, they provide a basis to examine the relative impacts of climate variability on crop yield.

The National Climate Data Center (NCDC) cooperative observer network data (TD3200) was used to drive crop simulations at the county scale during the historic period of record, 1948–2004. During the nominative cropping seasons for peanuts, cotton, tomato, and potato, yields simulations were made using the three dominant soil types within a county, at successive planting dates from the earliest possible dates to the latest for each year of the record. Using the historic JMA ENSO index, simulated yields were classified by ENSO phase and averaged within planting date and soil type. The completed yield values for relevant counties were incorporated into a simple tool that allows users to customize the output to their specific situations: crops, location, soil type, as well as average yields seen on their specific fields. They can then look at the expected impacts of changes in planting date based on the current ENSO forecast. Through further interactions with stakeholders across the region, the tool was modified to permit users to indicate availability of irrigation, to specify crop varieties where early or late cultivars exists, and fertilization levels for some crops.

The yield forecasts are presented as probabilities of achieving specific yield levels or of exceeding specific yields given the users proposed planting date and the current ENSO phase. In Fig. 12.16 a yield forecast for peanut yields in Mitchell County, Georgia is provided. After selecting the appropriate county, the user has also chosen Norfolk loamy sand from a choice of three dominant soil types for the county. The forecast yield distribution assumes a long-term yield of 2,400 lb acre^{-1}, a 14 May sowing date, and that the current ENSO phase is El Niño. A grower may use this tool to explore the impacts of the ENSO phase on changes in crop performance associated with alternate planting dates with a specific county and soil. The information may further support variety selection, field allocation and other strategic management activities.

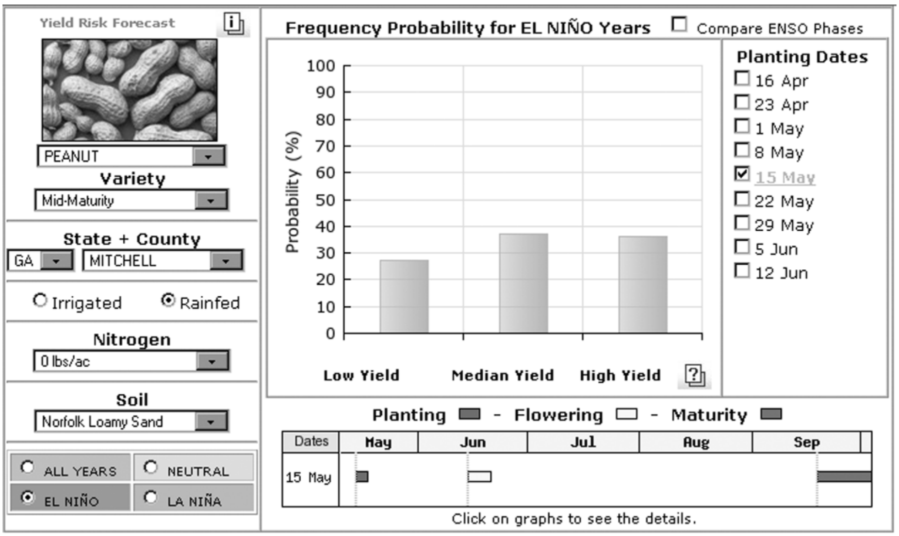

Fig. 12.16 Probabilistic crop yield forecast for Mitchell county, Georgia for Georgia Green variety peanut assuming a 14 May sowing date and a long term mean yield of 2,400 lbs acre^{-1}

12.2.2.3 Chilling Units and Growing Degree Day Tools

The observed impacts of ENSO on winter temperatures and freeze events naturally suggest a possible role for climate information in managing production risks associated with the production of perennial crops in the Southeast. Working with fruit breeders, growers, horticultural researchers, and fruit Extension agents, research was undertaken to examine the impacts of ENSO phase on winter chill accumulation throughout the region. Chill accumulation or chilling is a measure of time spent at or within certain temperature ranges, for example, hours during which the temperature is below 45°F. Different crops have specific temperature ranges that are most efficient in permitting chilling and eventually overcoming dormancy. Sufficient accumulation of chilling promotes improved yields of fruits, better growth by the plants, and more synchronous timing of harvests, while insufficient chilling tends to reduce fruit quality and yields and distorts both plant growth and the timing of development stages of the crop such as maturity dates of fruits. Growers and Extension agents believe that knowledge about expected chill accumulation may support decisions about marketing strategies, horticultural management such as pruning and fertilization, and applications of oils or other approaches to overcome insufficient chilling.

The TD3200 data set from NCDC was used to develop daily measures of chilling for a variety of perennial crops for the period 1948 through 2004. Totals for biweekly periods throughout the winter in each county were then analyzed with the historic JMA ENSO index to identify potential spatial or temporal signal in chill anomalies within the region. The results indicated a strong increase in chill accumulation over much of the region during El Niño events, and a milder though still significant decrease in chill associated with La Niña. Examination of 15 day periods throughout the season produced the unexpected result of indicating a temporal shift for when the majority of chilling takes place. El Niño events have historically been associated with a shift toward greater chilling in the early winter November to January (NDJ) as compared to climatology or La Niña. Growers found this particularly interesting, as it may promote earlier maturity and harvest dates and permit greater sales during lucrative early marketing windows.

The research results were then developed into two decision support tools that stakeholders can access via the Internet. The first tool provides forecasts for chill accumulation on a county by county basis across the region. Users select from three crop specific formulations for chilling, blueberry, peach, or strawberry, or a generic 45°F chill hours model; they also indicate their location. The tool initially provides expected average accumulations and anomalies of chilling based on the current SECC ENSO forecast (Fig. 12.17). Users can make comparisons based on alternate ENSO forecasts for further insight. Accumulation values are provided for biweekly periods from October through April, the nominal period for chill across the region. Additionally the seasonal total in each period is also displayed. Users

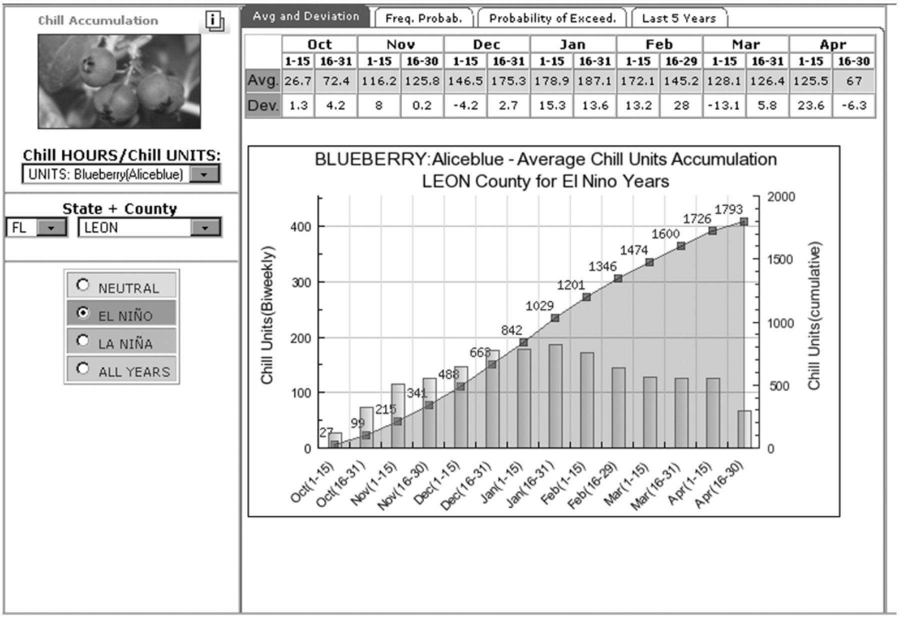

Fig. 12.17 SECC chill forecast tool customized for blueberry crops in Leon county, Florida. This presentation of the forecast emphasizes biweekly (histogram) and seasonal (area curves) chill accumulations based on an El Niño ENSO phase as defined by the JMA index

may examine records from the most recent 5 previous years at their location in order to integrate their prior knowledge and the individual performances at specific operations. The tool also provides a means for examining the complete probability distribution and probability of exceedance distributions for total chill accumulation at 2 week intervals throughout the season based on an ENSO forecast.

Alternately, the same research has been developed into a tool to provide rapid forecasts of spatial patterns of chilling. For individual crops or the generic chill hours formulation, users can look at ENSO phase-based forecasts for seasonal and biweekly periods throughout the winter. Two forms of the probability forecast are provided. The first map type indicates the probability of accumulating chill greater than the median accumulation for each county and indicates the counties expected to receive greater than 'normal' chill for the chosen time period (Fig. 12.18). The second map type emphasizes the tails of the probability distribution and shows the probability that accumulated chilling will be greater or less than the 20% tail. Chill maps are color coded as well with intense blue colors indicating high probabilities of more chilling (cool conditions) and intense reds indicating low probabilities (warm conditions).

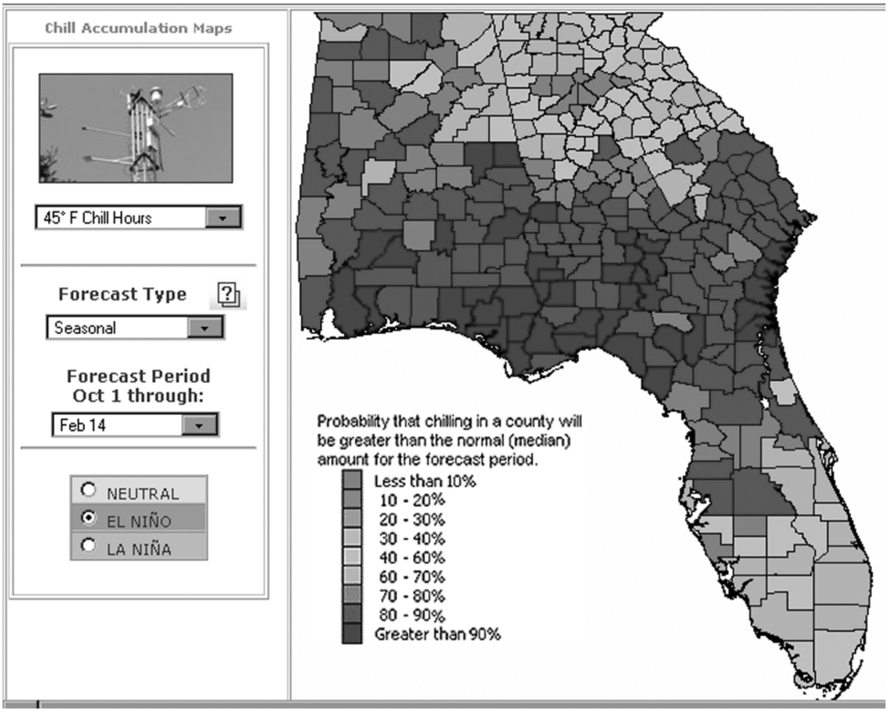

Fig. 12.18 The SECC chill forecast is also distributed as a spatial forecast based on the mapping of county level probabilities. The map provides the probability of experiencing greater chill accumulation levels than the median in each county based on the current ENSO phase forecast. The illustrated forecast has been customized to provide a seasonal forecast for the period 1 Oct through 14 Feb. The forecast provides an outlook for 45°F chill accumulation as well as crop specific chill accumulation for blueberry, peach and strawberry

When chill accumulations are forecast to be adequate or better for a crop, growers may more confidently make investments in management and marketing. Comparisons of forecasts and recent historical conditions may also permit growers to calibrate experiences specific to their operations in planning for timing of labor, pruning, market contracts. In situations where inadequate chilling is likely, growers may forego investments, curtail expenditures, or plan for compensating management such as the application of suitable oil sprays to overcome some chilling deficits.

12.2.2.4 Seasonal and Commodity Outlooks

Working closely with Extension agents and Extension specialists responsible for specific commodities or agricultural sectors indicated a need for a different sort of decision support tool from those previously discussed. In several cases a model to run simulation based forecasts for a specific commodity was not available, or the agents had concerns that had not been fully researched. During the process of repeated meetings and discussions about the potentials, limitations, and applications of climate information, it has become increasingly clear that there is a strong demand for interpreted climate forecasts or outlooks that are easily understood by specialists with limited previous exposure to climate and meteorological science and the statistical approaches. The interpreted forecasts also need to be easily and efficiently communicated by Extension agents to their clients, the farmers, who may or may not have access directly to the agclimate.org web site. The development and delivery of seasonal climate outlooks and commodity outlooks by the SECC grew as a response to this need.

The initial approach to this product involved four quarterly climate outlooks released in early September, December, March, and June. Commodity specific outlooks for winter pasture (September), Citrus (December) and Peanut (September and March) were also released. In order to develop the outlooks the first step is to examine current sea surface temperature trends within the equatorial Pacific and to fix the official ENSO phase forecast that will be used by all the forecast products that are available through the AgClimate web site. The selection of the appropriate ENSO phase classification is made by an experienced climatologist from one of the American Association of State Climatologists Recognized State Climate Office (ARSCO) offices within the region. One of the state climatologists then reviews the expected shifts from climatology and writes a brief one page summary of expected impacts during the following 3 months. The summary includes a synopsis of current ENSO conditions and trends written in simple laymen's terms. The document then contains a review of the expected impacts of the predicted ENSO phase on temperature and precipitation related variables. The next section of the summary reviews the current climatic conditions within the region and compares them with the predicted impacts. Finally, seasonally appropriate climatic impacts such as wildfire risk, freeze risk, or hurricane forecasts are included. This document is then approved by the state climatologists and becomes the basis for further interpretation.

Commodity specific outlooks are then developed through a collaborative process between climate Extension specialists in the SECC and commodity Extension specialists in the state cooperative Extension programs. The information from the seasonal climate outlook is discussed with the commodity specialists who consider the forecast and are able to ask detailed questions and to consider the implications of the forecast for the activities related to their commodity in the upcoming quarter. A new outlook that provides much less detail about expected climatic conditions, but which applies the climate forecast to management options that can be adjusted, is produced. For example, in the winter forage outlook which is released in early September, the principal climatic impacts of the expected ENSO phase are briefly discussed and then are followed by a discussion of appropriate forage crops that should or should not be planted, whether certain land features, i.e. sloping lands or lowlands, will require particular consideration, what types of field conditions might warrant adaptive management, and how the anticipated climatic conditions might effect strategic management decision such as planting dates, supplemental feed purchases, variety selection, fertilizer applications, etc. The outlook concludes with additional sources of information specific to both climatic and forage conditions in the region and a list of contacts who producers and county agents might contact for further details as needed. This document is then circulated between the climate Extension specialists and the commodity specialist in each state for revision and final approval, after which it is released directly to county Extension agents via email and made available for downloading at the AgClimate web site.

12.2.3 Future Directions

Currently, SECC climate services are based primarily on analysis of historic ENSO signals downscaled to the county scale, as well as the use of bootstrapped data sets derived from the same period of record. The use of other approaches to forecast derived climate variables that stakeholders find useful is an active area of research within the SECC. Within the Southeast USA, the summer months and ENSO neutral phases are times of limited predictability, yet these periods remain very important to producers. The use of numerical climate models for simulating crop yields is being explored, using both dynamical and statistical approaches to downscaling from large scale (200 km grids) to local scale (20 km grids, counties, or stations).

An alternate approach to improving the suite of forecast products, based on climate variability, is the incorporation of near real-time observational data into the forecast product. The SECC is partnering with the Florida automated weather network (FAWN) and the Georgia Environmental monitoring network (GAEMN) to explore approaches to integrating observational data into hybrid products. Here, the data for the season to date of the individual forecast product is drawn from the daily or hourly station data and is processed with the historical data sets (TD3200) representing the remaining seasonal interval. This has the advantage of removing successively more variation from the probability forecasts, but at the expense of reduction in forecast lead times and increased maintenance and production costs for each forecast update. The initial decisions to base current products on biweekly, monthly, and seasonal intervals suggest that biweekly updates may be warranted; however most current products were designed to be easily expanded by the addition of a ‘current season’ forecast, which would default to observational data at any update frequency combined with the forecast based on the current ENSO phase for the remainder of the forecast period.

Climate services for reducing risk in agriculture and natural resource management are more than an archive of historic climate data that has been processed into daily, weekly, monthly, seasonally, or spatial discrete averages. The work of the

SECC continues to support the idea that forecasts and the climate variables selected for analysis can be successfully identified through interactions with the proposed end users. When analyses are made and forecasts provided intentionally to address stated needs and interests, adoption can be substantially enhanced. When provided with alternate approaches to provide the same types of climate information, users can readily identify which formats are most accessible, as well as identify where the format and substance is lacking if they are to readily incorporate the information into their decision making processes. The high degree of interest in the prototype tools and decision support system has led to the expansion of many Extension programs in the region to include discussion of climate variability and climate forecasting alongside more traditional topics. Recent successes in developing and providing climate services proceed from interactions between a strong ENSO climate signal in the region and dialogue between the SECC and a relatively small segment of the potential end users of information on climate variability. Further work in disseminating information on climate variability and forecasting with new user groups will likely lead to further research and ultimately development of additional products to support newly identified activities where applied climate information can contribute to better decision making and reduced exposure to risk.

12.3 Appendix

12.3.1 Memorandum of Understanding (in French)

Royaume du Maroc
Secrétariat d'Etat auprès du ministère de l'Aménagement du Territoire,
de l'Eau et de l'Environnement, Chargé de l'Eau
Direction de la Météorologie Nationale

Convention de coopération avec
Dans le domaine de la prévision saisonnière

P. J: 1- Exemple de bulletin de prévision saisonnière
2- Spécimen de questionnaire d'évaluation de l'utilisation de la prévision

La Direction de la Météorologie Nationale s'est intéressée, depuis Juillet 1994, au développement de la prévision saisonnière pour répondre au grand besoin nationale en la matière. De nombreuses étapes ont été franchies jusqu'à présents et d'importants progrès ont été réalisés dont la production régulière, depuis Mars

1998, de bulletins de prévision saisonnière de précipitation. Ces bulletins sont élaborés à titre expérimental et sont envoyés aux autorités de décision dont le Ministère de l'Aménagement du Territoire, de l'Eau et du Climat, le Ministère de l'Equipement et du Transport, la Gendarmerie Royale, ...

Tenant compte des résultats encourageants obtenus et afin de **satisfaire au mieux les usagers** en :

- **adaptant** les produits de la prévision saisonnière à leurs attentes,
- **choisissant** parmi la multitude des paramètres produits par nos modèles de prévision saisonnière, ceux qui répondent le plus à leurs besoin,
- **les aidant** à cibler les prédicteurs potentiels basés sur les sorties de nos modèles et directement utilisables pour leurs prévisions et programmations (rendements agricoles, gestion des barrages, …),

la DMN s'engage à mettre, mensuellement de Septembre à Février, à la disposition de ………, un bulletin de prévision saisonnière de type celui en pièce jointe dans les conditions précisées par les articles 1 à 6 ci-dessous :

Article 1 : La durée de la mise à disposition du bulletin est fixée par la DMN.

Article 2 : La forme ainsi que le contenu du bulletin sont susceptibles de changer en fonction des progrès et besoins.

Article 3 : Les prévisions sont livrées gratuitement <u>**à titre expérimental**</u> et la DMN n'assume aucune responsabilité pour les décisions prises en se basant sur ces prévisions.

Article 4 : L'usager …. s'engage à faire un rapport trimestriel détaillé, type celui en pièce jointe, sur les utilisations faites des prévisions et sur les apports économiques qu'elle ont permis.

Article 5 : L'usager……. s'engage à ne pas diffuser ou donner ces prévisions à quiconque et sous aucune condition.

Article 6 : Des réunions régulières peuvent se faire, suivant le besoin, entre les services concernés par la prévision saisonnière des deux entités (Service de Etudes Climatiques de la DMN et ….de …. Usager) pour discuter des éventuelles nouveautés et/ou modifications concernant le produit.

Signé pour la DMN Signé pour l'usager ….

Chapter 13
Water, Health and Early Warnings

Yahya Abawi, Paul Llanso, Mike Harrison, and Simon J. Mason

Following on from the previous Chapter are three contributions that cover the remaining "classical" areas for applications (alongside agriculture) of hydrology and health. Water management is the focus in the first section, in which a number of projects are described whereby historical and forecast information is used directly in planning specific actions; in this case the forecaster-user chain is short and manageable at a personal level. Next is a detailed account of the steps required to establish climate services in the health area. Finally, Early Warning Systems are described. Early Warning Systems have not tended to use predictions until recently, traditionally having been built around historical observations. In that context Early Warning Systems provide an example of an application mainly designed for humanitarian benefit built solely using climate data alongside other information, but with growing use of predictions. A worked example is included establishing the impact of climate variability on disease incidence, the results of which provide a basis for incorporating seasonal climate forecasts into a Malaria Early Warning System in southern Africa.

13.1 Introduction

In this chapter experiences in communicating and applying seasonal forecasts in the areas of water management (Section 13.2), human health (Section 13.3), and in early warning systems generally (Section 13.4) are discussed. It is emphasised

Yahya Abawi
Queensland Climate Change Centre of Excellence

Paul Llanso
Consultant, ex-WMO

Mike Harrison
Independent Consultant

Simon J. Mason
International Research Institute for Climate and Society

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

throughout that not only do predictions form only one component of climate information that is of interest in these sectors, but that climate is only one of many factors that influence decisions. As argued in Chapter 2, seasonal climate predictions need to form one component of a broader climate service, involving information at a range of timescales, including monitoring and historical information. This chapter provides examples of the importance of understanding the context within climate information may be used in the respective sectors.

13.2 Application of Seasonal Climate Forecasting in Water Resources Management

13.2.1 Introduction

This section describes the application of seasonal climate forecasts in water resources management in three different environments: a large scale irrigation system in the Murray Darling Basin of Australia; a medium scale irrigation system in the Indonesian island of Lombok; and small water resources systems in the Pacific Islands. Each case study is described in terms of their unique characteristics and lessons learned from these studies, which may help overcome some of the barriers discussed.

Although advances in climate science and improvements in modelling provide a direct quantification of the benefits and risks of using seasonal climate forecasts in making management decisions, there remain significant barriers to widespread adoption of such forecasting tools. Nicholls (1999) identified the difficulty people have in estimating and dealing with probabilities, risk and uncertainty as being one of the primary barriers. Additional constraints include: understanding probability and probabilistic forecasts and their reliability, lack of integration and evaluation, political and institutional influences, relevance and timing. Most of these barriers to the use of climate predictions represent a lack of knowledge about the forecast and impact systems, a forecast delivery problem, or difficulties in users reacting to the forecasts (Nicholls 1999). One reason why the quality of seasonal forecasts remains an impediment is that decision makers are not only concerned with statistical validity of a forecast from a climatology perspective, but also how reliably a forecast predicts impacts on the systems that are being studied. In a study of subsistence farming in Africa, Hulme et al. (1992) found that individual farmers were unlikely to benefit from forecasts because of poor forecast quality, of having insufficient flexibility to respond, and of the fact that decisions are based on a range of factors of which climate is only one.

In water resources management there are additional impediments that limit the use of seasonal climate forecasts. These include:

- *Dealing with complex systems*: water management involves issues of supply, quality, allocation, distribution, reliability and competing demands between consumptive (production) and non-consumptive (environmental) users. Often the development of a hydrological model for a catchment takes several years to calibrate and validate and is unique to each catchment. This validation must be in place before the impact of seasonal climate forecasts can properly be assessed.
- *Relevance*: climate forecasts mainly focus on prediction of rainfall. In some cases this may be used sometimes to predict impacts on agricultural decisions, such as in subsistence farming where a direct relationship exists. However, the impact on water resources systems is a combination of multiple interacting and complex variables. Forecasts of rainfall may not be directly relevant to water allocation decisions, whereas forecasts of streamflow may have more relevance.
- *Lack of hydrological data*: hydrological data (e.g. streamflow data) are limited, particularly in developing countries, unlike rainfall data where long-term records, albeit with major gaps in some regions, exist for many locations around the world that are useful for developing climate prediction models. Even where hydrological data exist, they are highly impacted by the construction of dams, weirs and irrigation diversions. To assess the impact of climate variability on streamflow and water allocation decisions we need synthesised or 'natural' records of streamflow where the effect of infrastructure development and irrigation diversions has been removed. These could then be used to simulate various scenarios of development and operational rules within a system.
- *External constraints*: Legal requirements can often lead to inappropriate and inefficient water use. In many countries, including Australia, water authorities will not allocate irrigation water based on probabilistic forecasts of dam inflows for reasons of litigation. They use a no-risk or zero-inflow scenario when deciding on allocation even where, as discussed later, a strong and statistically significant relationship between ENSO and future dam-inflows exists, e.g. as in north-eastern Australia. This conservative approach to water allocation by water agencies may adversely impact on growers' decisions to increase planting area, resulting in lost opportunities.

Despite these difficulties, significant opportunities exist for applying seasonal climate forecasts in water resources decision making. Catchment scale responses, such as streamflow, can capture the global and regional effects of climate variability (e.g. ENSO) much better than point scale responses such as rainfall. Dutta et al. (2006) found that higher skill (as measured by the LEPS score) was associated with seasonal forecasts of streamflow as compared to rainfall for the same period within the same catchment. Skillful forecasts of streamflow were also possible at longer lead-times (up to 4 months) than for rainfall.

In a study of Columbia River hydropower in the USA, Hamlet et al. (2002) found that forecasts of streamflow with 6 months lead-time can facilitate improvements in the operating performance and can increase energy production by as much as 5.5 million MWh/year, resulting in an average increase in annual revenue of approximately $153 million per year.

Bates (2002) identified several key points that need to be taken into account by researchers when dealing with climate forecasting and its application in water resources management:

- The net benefits of using seasonal forecasts in water management have not been demonstrated. This demonstration must take place within operational settings so that the focus is on user needs rather than those of climate scientists.
- Users require objective, explicit and user-friendly forecasts at temporal and spatial scales appropriate to their needs. They also require applicable information about forecasting errors and uncertainty.
- Appropriate mechanisms for technology transfer of climate forecasts to end-users must be in place.
- Socio-economic aspects of forecasting are crucial. For example stakeholders need better education on climate issues and processes, forecasting techniques and probability. On the other hand, scientists need to recognise the effects of institutional, political, educational and cultural constraints on policy formulation and decision making.
- Problems and issues are more complex in developing than in developed countries, because of lack of funding, data paucity, institutional capacity, social capital and lower priority placed for meteorological services, and there is a need to strengthen local meteorological capacities.
- Uptake of seasonal climate forecasts in water resources management would be best achieved through a cooperative approach between forecasters and users.

The following section describes the application of seasonal climate forecasts in water resources management from large-scale irrigation systems in the Murray Darling Basin of Australia and the Indonesian island of Lombok, to small water resources systems in the Pacific Islands. It is not intended to give a detailed account of the models or results, but to highlight salient differences between the systems and suggest possible solutions to addressing some of the impediments to effective implementation of climate forecasting in water resources management.

13.2.2 Border Rivers Catchment

A project funded by the Australian Murray Darling Basin Commission examined the use of seasonal climate forecasts in irrigation management and water allocation decisions in the Border Rivers catchment (between 27°30′ and 30°2′ S; 148°39′ and 152°9′ E) in the northern part of the Murray Darling Basin (Fig. 13.1). The key objectives were to answer these questions:

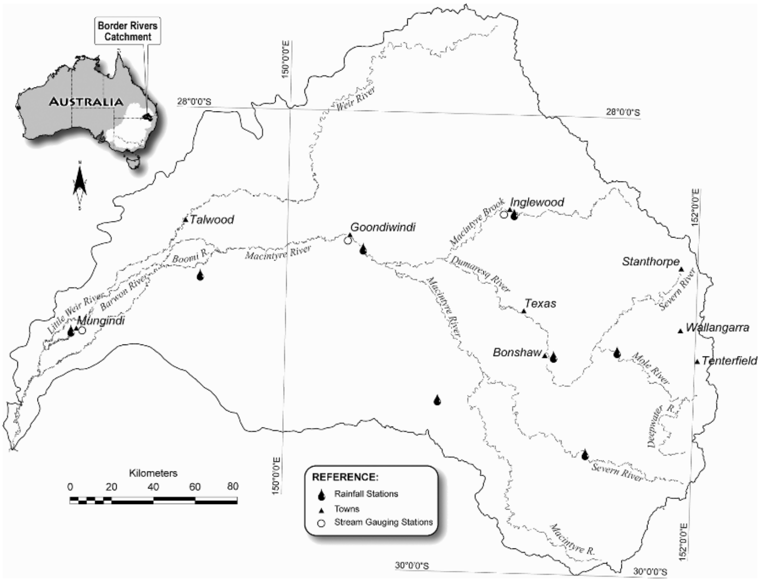

Fig. 13.1 Location of the Border Rivers catchment in the northern Murray Darling Basin, Australia

- Can seasonal climate forecasts lead to improved decision making, and thus higher profits, for growers, by guiding adjustments of areas planted to cotton in each season?
- What is the potential economic impact of seasonal climate forecasts if water managers incorporate climate forecasts in water allocation decisions?
- What is the likely uptake of climate forecast information in decision making by irrigators?

The results of these investigations have been reported by Abawi et al. (2005), Richie et al. (2004), Keogh et al. (2004), and Dutta et al. (2006). In this section the second question is briefly explored as this issue is seen as one of the barriers to the adoption of seasonal climate forecasts in irrigation decision making.

The Border Rivers catchment covers an area of 44,100 km^2 with an average annual rainfall of 800 mm in the east and of 550 mm in the west. Despite considerable infrastructure development during the last 2 decades, water remains limiting for irrigation due to the high climate variability associated with ENSO, which has a dominant influence on the climate of the region. Irrigated cotton accounts for 83% of the 53,900 ha developed for agriculture and is planted in early October and harvested in March.

Sources of irrigation water are from announced allocations (expressed as a percentage of an irrigator's licensed entitlement) and from water pumped directly from the river system during periods of high flow (off-allocation). Off-allocation pumping is only permitted when flows in the river exceed certain thresholds set for riparian needs and end of system flows. These thresholds are defined in the Border Rivers Water Resources Plan. For example, an annual target of 60% of mean annual flow is set for end of system flows to ensure downstream requirements are met. Allocations are determined by the water authority just prior to a cropping season and are based on a resource assessment taking into account: (a) the actual amount of water in State-owned storages; (b) high security requirements such as urban, industry, horticulture and dairying; (c) losses such as in delivery, seepage and evaporation; and (d) total licenses issued to irrigators. These allocations may be increased during the growing season if additional flow is received, but for legal reasons the water managers use a zero-inflow scenario when deciding on the allocations.

Risk-averse irrigators will make crop area decisions based on their initial allocation. However, a risk-preferring irrigator may take additional future inflows into account when making his/her decisions. Considering that in this catchment more than 60% of average annual flows occur after planting, knowledge of future inflow probabilities would be invaluable to farmers contemplating what area to plant. Unfortunately, no mechanism exists to communicate future inflow probabilities to the irrigators. This is a significant barrier to the adoption of seasonal climate forecasts, particularly in north-eastern Australia where the impact of ENSO on rainfall and streamflow is high and the skill in forecasts is moderate to high (Dutta et al. 2006).

To assess the impact of ENSO on dam inflows and water allocation decisions, daily 'natural' flows (1890–1997) for major streams and tributaries were obtained from the New South Wales (NSW) Department of Land and Water Conservation (DLWC) and used in a hydrological model (Integrated Quantity Quality Model – IQQM; DLWC 1998) to simulate dam inflows, allocation levels, water availability and other hydrological variables. These simulations were based on the current level of development and operating rules within the catchment.

Relevant hydrological data were extracted from the model output based on analogue years of El Niño, La Niña, and non-ENSO events. An analysis based on these events provided a useful reference of climate extremes. The presence or absence of these events is usually well known by June to September based on sea surface temperature (SST) anomalies in the Pacific Ocean, and can provide sufficient lead-time for decision making.

Using this approach Abawi et al. (2005) showed a median difference in annual inflow between the La Niña and the El Niño years in all state-owned dams within the catchment of 144 gigalitres (Gl). The median difference in inflow in the October–March period (inflows that water managers will not consider in the allocation decisions) was 55 Gl. Assuming that 40% of this water is lost due to evaporation, conveyance and other losses, the net additional volume of water (33 GL) translates to an irrigated cotton area of approximately 5,500 ha or 10% of the total area developed in the catchment. The additional value of this water based on the current price, yield and water requirements of cotton is about $16 million. The socio-economic impact of this additional water in employment and its multiplying effect on the regional economy could be in the order of two–three times or up to $45 million. The results presented here were based on the median flow, however, it can be repeated for the mean or other percentile values of the flow and could be used as a risk management tool.

To overcome the legal barrier faced by water managers, a possible solution would be to simulate allocations levels, based on seasonal climate forecasts, at the beginning and end of a cropping season and make these results available to irrigators. These results based on El Niño and La Niña events are given in Abawi et al. (2005), but could be reproduced for any climate predictors (e.g. SOI) and predictands (e.g. streamflow, allocation, rainfall). An example of the results from that study shows a 50% chance that during a typically dry El Niño year the allocation may be increased from an initial 28% to a final of 60% of the irrigator's license entitlement (corresponding results for La Niña years shows an increase from an initial 74% to a final of 100%). If appropriate dialogue can be established between water managers and water users, and the information on future inflows could be communicated on an all-care-but-no-responsibility basis, growers may incorporate such information into their decision making process. This shift of risk from water managers to water users would alleviate the legal issues faced by the water managers and pave the way for the adoption of climate forecasts in decision-making by the irrigators. Obviously the uptake of such information would ultimately depend on many factors including the type of enterprise, farm size, commodity price and future markets, financial position and an individual's attitude to risk.

To determine the factors that influence farmers' decisions and whether they would use seasonal climate forecasts in their decisions, a mail survey was sent to 931 irrigators involved in agricultural production from regulated water supplies in the Border Rivers, Gwydir, Namoi and upper Condamine catchments of the Murray Darling Basin. The questionnaire was designed to obtain information about general farm characteristics; irrigators' knowledge of the climate system including relationship between El Niño/Southern Oscillation and rainfall/streamflow; how irrigators make cropping and water decisions; who they consult; and information sources and use of computer technology. One hundred and seventy responses were received from the survey and the results described in (Keogh et al. 2004). The results show that almost 67% of irrigators access seasonal climate outlook information, but only 20% are sufficiently confident to apply this in their decisions. Almost 75% would change their crop area, and 43% their crop type, if given advanced information on probable water availability up to 4 months ahead of the irrigation season. The likelihood of climate-related decision making increased substantially with the size of the farm and type of enterprise (cotton farmers *vs* non-cotton farmers). Forecasts appear to be particularly useful for cotton growers; 80% of them were prepared to consider changing their crop area, compared to 45% of non-cotton growers. Farmers with more than 600 ha of irrigation are twice as likely to use climate forecasts in decision making than farmers irrigating less than 100 ha.

Risk-averse farmers are unlikely to incorporate seasonal climate forecasts in their decision making. The level of risk-aversion is inversely proportional to the level of wealth (Anderson et al. 1977), an observation which supports the findings of this survey that likely users of seasonal climate forecasts are large corporate irrigators and farmers.

To communicate results and to seek irrigators' and water managers' input to the research process, effective dialogue between irrigators, researchers and water managers was established through a steering committee, and regular meetings held during the progress of the project. This formative evaluation (evaluation during the life of the project) was used to gather information that may help identify factors that contribute to successful or low uptake of the research. In response to this feedback, regular monthly articles were published in local newsletters with updated information on streamflow forecasts, the general climate outlook and on understanding basic forecast terminology. The survey by Keogh et al. (2004) was conducted 3 years after the project commenced. The improved level of climate knowledge gained by irrigators is possibly a reflection of this cooperative approach.

13.2.3 Indonesian Island of Lombok

A similar study to that in the Murray-Darling Basin, funded by the Australian Centre for International Agricultural Research,[1] is being conducted on the island of Lombok in Indonesia to develop hydrological and crop models to assess the value of seasonal climate forecasts in water allocation and irrigation decisions.

Lombok lies in the eastern part of the Indonesian Archipelago and is situated between 8°12′ and 9°01′ S and between 115°46′ and 116°43′ E, covering a land area of approximately 4,800 km^2. The climate of Lombok is tropical and it is possible to grow three successive crops (usually rice, rice, secondary crops) each year provided water is not limited. Most rainfall occurs in the wet monsoon season '*Musim Hujan*' from November-March followed by the dry season from April-October. Seasonal and inter-annual rainfall variability is strongly influenced by ENSO. More than 90% of droughts in eastern Indonesia are associated with El Niño events. The influence of ENSO on the climate of Indonesia is described by Haylock and McBride (2001).

Rice is the dominant irrigated crop grown on the island. Whilst irrigation infrastructure is well developed water use efficiency is low, and large parts of the island experience water shortages due to inadequate supply and distribution of irrigation water, particularly in El Niño years.

---

[1] See: http://www.acair.gov.au

The approach used in assessing the utility of climate forecasts is similar to that described in the Border Rivers catchment. A significant difference in hydrology and water management issues in Lombok from the Borders River is that most rivers are un-regulated, with little storage upstream of irrigation areas. Therefore prediction of streamflows during the irrigation season is of higher importance for effective irrigation decisions and water allocations than in the Australian case.

Water availability throughout the irrigation system was modelled using the IQQM model and a Linear Programming Model (LPM) was developed to optimize cropping decisions under different climate, water, land and institutional constraints. A detailed description of the models, data issues and methodologies is given in Abawi et al. (2002). In this section some of the issues which are likely to impact on the adoption of seasonal climate forecasts in developing counties such as Indonesia are highlighted.

- Understanding the socio-economic culture in developing countries is crucial for successful adoption of climate information. Rice is grown more for its social value than for its economic value. The Indonesian Government encourages farmers to grow rice for self-sufficiency reasons. Farmers prefer to grow rice than any other crop because it provides the staple diet and can be stored on their farms for long periods, providing a buffer against crop failure in some seasons. Farmers' preferences, as well as those of the Government, must be incorporated in the development of decision support models.
- Understanding of markets, as well as of supply and demand issues, is very important. Rice requires twice or three times the amount of water to grow than other crops which may have a higher market value. In decision support models (e.g. LPM) if profit maximisation is the primary goal, crops such as chilies, vegetables and tobacco may be the preferred solution because of their higher value and lesser demand for water. However, in a closed market there is little opportunity for export, and market saturation will result in a sharp drop in prices causing financial losses to the farmer. Government regulations limiting the production of certain crops such as tobacco must also be taken into account as part of an overall solution.
- An average farm size in Lombok is about 0.25 ha (cf. 600 ha in the Border Rivers catchment). Most farmers are risk-averse, preferring less risky sources of income in order to reduce the possibility of loss. Therefore little visible opportunity exists for adjusting cropping patterns at the individual farm level based on climate forecast information. However, significant economic gains can be made through water allocation and cropping decisions at a scale that integrates the characteristics of many of these smaller farms. Water is allocated to each irrigation area by a committee of irrigators and government advisors. Input from government advisors with knowledge and access to climate forecasts is crucial to the success of these decisions.

- Strategies such as water pricing and water trading which are often adopted in developed countries in response to water shortages, such as those during an El Niño, are unlikely to be acceptable in Indonesia due to cultural and religious beliefs.

Other issues which may affect potential use of seasonal climate forecasts in agricultural decisions on the Indonesian Island of Lombok were identified in a survey by Sayuti et al. (2004) and include:

- Farmers have a low level of education; 75% have never been to school or have received only elementary schooling.
- A propensity for farmers to believe traditional forecasts (e.g. astrology and indigenous knowledge such as the flowering times of trees, movement of insects and wildlife) rather than scientific information such as forecasts based on ENSO.
- About half of the farmers in Lombok believe the advice of the government on water availability and crops to plant.
- 40% of farmers may change crop type if the government advises of an impending shortage or excess of water in the coming season.

Educational attainment levels influence the ability to adopt innovative processes and have significant implications in the development and implementation of action plans dealing with technology transfer. The survey in Lombok also showed a strong dependence by the farmers on advice from the government. This provides significant opportunities for working with government agencies, in addition to farmer groups, to successfully implement results, as government agencies also control water allocation and, in some cases, seed allocation. The role of government in the implementation of new programs has been used in Indonesia during the green revolution when the Indonesian army helped in the spread and adoption of new and improved rice varieties while farmers initially resisted the change.

In summary, social and cultural issues in developing countries are just as important as the science of seasonal climate forecasting. Forecasts must be reliable and skillful before anything can be done, but if the social, cultural and educational issues are ignored, prediction technologies are unlikely to be adopted.

13.2.4 Pacific Islands

Similar climate-related issues to those discussed for north-eastern Australia and Indonesia exist in the Pacific Islands. Drought is one of the major hazards facing Pacific Island nations, is strongly related to ENSO events on many islands, and can have severe impacts throughout the region. The drought impacts of the 1997–98 ENSO event have been well documented (Glantz 2001). Lessons learned from the 1997–98 droughts in the Pacific demonstrate the need for effective and timely forecasting and warning systems, drought response strategies, information on quantitative measures of drought, improved water management and improved crop and stock management.

Despite their vulnerability to the impacts of climate variability, most Pacific countries have limited meteorological service capacity to provide timely climate forecasts for their climate sensitive industries. In 2002 the Australian overseas aid agency, AusAid, funded a project to improve local meteorological capacity by developing climate forecasting software and by providing training for meteorological service staff and stakeholders from agriculture, water, health, fisheries, and disaster management. A further objective of the project was to evaluate the utility of seasonal climate forecasts in the management of water resources in selected countries.

Climate prediction software, called SCOPIC (Seasonal Climate Outlooks for the Pacific Island Countries[2]), was developed based on the operational seasonal climate forecasting system used by the Australian Bureau of Meteorology. SCOPIC produces forecasts based on the relationships between SST anomalies in the Pacific and Indian Oceans with rainfall or other hydro-meteorological variables on Pacific islands. Training in the use of the software, on climate issues and processes, and on forecasting techniques and basic statistical concepts such as probability and data analysis, was provided in a series of workshops in nine Pacific Island Countries (Fiji, Vanuatu, Tonga, Samoa, Cook Islands, Solomon, Kiribati, Tuvalu and Niue). Feedback from these workshops was used in the design of the software to meet the requirements of Meteorological Services and other stakeholders. Experiences from the workshops which may help in forecast preparation, delivery and the adoption of seasonal climate information, are discussed below.

Understanding the strengths and weaknesses of a forecasting system by users can engender user confidence in the forecasts. Unfortunately many climate forecasts issued by Meteorological Services to the public focus only on the probability of certain rainfall events, and errors and uncertainties of forecasts and forecast skills are not explicitly communicated. Visual presentations can help in the understanding of complex climate concepts, particularly when users do not have a detailed understanding of scientific issues. An example of such visual presentations is illustrated in Fig. 13.2, which shows the skill of seasonal climate forecasts in Kiribati (left) and in the Solomon Islands (right). Skills of forecasts for successive 3 month periods are expressed using LEPS scores (Linear Error in Probability Space; Potts et al. 1996) along the *x*-axis for different lead-times (*y*-axis). Shades of blue indicate more skill than climatology, while shades of red indicate less skill than climatology. This snapshot presentation is helpful in assessing when a forecast is useful and when it is not. For example, prediction of rainfall has moderate to high skill throughout the year in Kiribati with relatively long lead-times. On the

[2] See: http://www.bom.gov.au/climate/pi-cpp

other hand, in the Solomon Islands, there is good skill during November to March (wet season) but no skill from April to August (dry season). Understanding the strengths and weakness of a forecast can reduce the uncertainty associated with using climate forecasts in decision-making.

Users are more likely to appreciate the real levels of skill if these are explained using a time series of how the forecast would have performed in the past (hindcast), showing when they work and when they are unreliable. This approach may also help in communicating concepts such as 'probability' as most people understand frequencies (e.g. number of consistent forecasts from a total number of forecasts issued) better than probabilities. This is illustrated in Fig. 13.3 for the forecasts of June-July-August (JJA) circled in Fig. 13.2. It shows a time series of cross-validated hindcasts for the period in question. In cross-validated analysis, data for predicted periods are omitted successively from the model to avoid model bias. Here forecasts are expressed in terciles, i.e. the probability of rainfall being below-normal (tercile 1), normal (tercile 2) and above-normal (tercile 3). Although in a probabilistic forecast all eventual outcomes within a forecast pdf are possible, users' expectations of such forecasts are that the likely outcome would be within the tercile with the highest probability, particularly when these probabilities are significantly higher than climatology. To illustrate this, for each year of analysis, a blue bar is used in Fig. 13.3 to indicate that the observed value of rainfall was in the same tercile as the tercile with the highest probability (consistent forecast). A red bar indicates that observed rainfall was different by two categories from the most probable (inconsistent forecast), and yellow bars indicate that observed rainfall was in the neighbouring category (near consistent forecast). This illustration helps in understanding the frequency of observed rainfalls being within a user's expectation of a forecast. It also demonstrates that from the user perspective the utility of a forecast is not uniform and it is possible to a have a long sequence of 'consistent' forecasts (in the sense that observed rainfall lay in the predicted highest-probability tercile) followed by a succession of 'inconsistent or near consistent' forecasts.

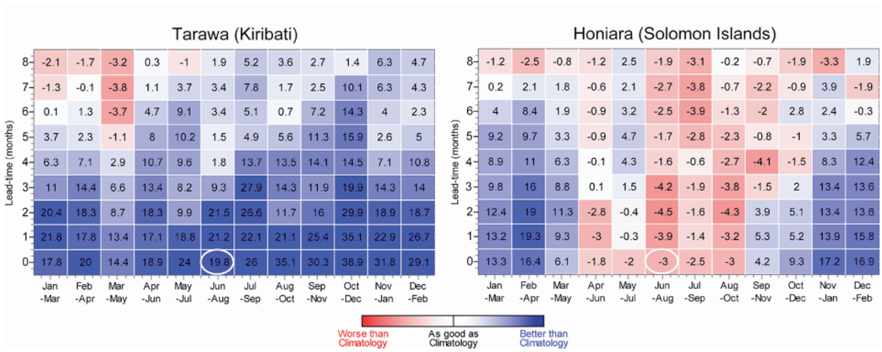

Fig. 13.2 LEPS scores for 3 month rainfall forecasts based on SST anomalies in the central eastern Pacific and the Indian Ocean. Kiribati (left); Solomon Islands (right)

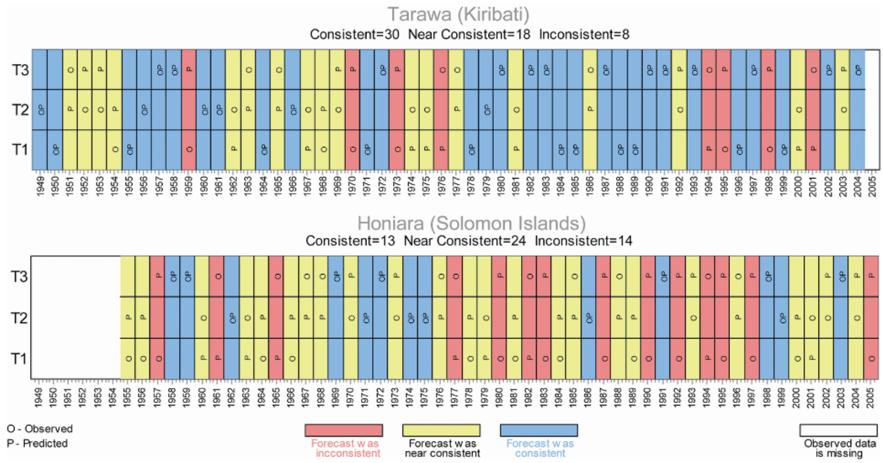

Fig. 13.3 Cross-validate hindcast analysis of a 3 month rainfall forecasts based on SST anomalies in central eastern Pacific and the Indian Ocean. Kiribati (top); Solomon Islands (bottom), T1, T2, T3 are rainfall terciles. P denotes tercile with highest forecast probability. O denotes tercile with observed value of rainfall

User expectations are often different from those of forecast providers. Forecasts produced by Meteorological Services usually offer a 3-month rainfall outlook based on current values of key climate predictors. In many applications, however, forecasts of different durations and lead-times are needed. For example, in Fiji monthly forecasts of rainfall are needed to plan effectively for fertiliser application to sugar crops. This mismatch between user needs and what is currently available from forecasting systems is likely to limit the use of seasonal climate forecasts in certain applications, and provides a real challenge for the developers of climate forecast systems to tailor forecasts that meet user requirements.

The second aim of this study in the Pacific Islands was to evaluate the utility of seasonal climate prediction in the management of selected water resources. On small islands, the main use of water is for domestic purposes and water is obtained either from rainwater tanks (e.g. Tuvalu) or shallow groundwater sources (e.g. Kiribati). In these simple hydrological systems there is a direct link between historical rainfall trends (ranging from a few months to a few years) and the status of water resources (such as the volume of water in tanks, groundwater salinity or the volume of shallow fresh groundwater lenses). Thus rainfall periods (totals) used in climate analyses depend on the 'hydrological residence time' of the selected water resource system. For example, rainfall totals over a 2- or 4-month period may correlate with water supplies in rainwater tanks (e.g. Funafuti in Tuvalu), while rainfall totals over a 30-month period may better reflect the volume of groundwater lens (e.g. Tarawa in Kiribati).

Analysis of past rainfalls can be used to monitor the status of these resources and actions taken once pre-determined thresholds are reached. White et al. (1999) used rainfall ranking (rainfall percentiles for different periods) to define ‘drought’ in a manner relevant to domestic water supplies on the Pacific Islands of Tuvalu and Kiribati. A time series of rainfall percentiles (over a time period appropriate to the system under study) can be used to warn of impending drought (e.g. when percentiles fall below 40%) or severe drought periods (e.g. when percentiles fall below 10%). Using these definitions historical droughts can be identified and related to climate episodes. Statistics on past correct warnings (i.e. when the 10% threshold is reached following the warning threshold of 40%) and false warnings (i.e. the threshold of 10% was not reached after the preliminary warning), in conjunction with the current status of ENSO, can be used to develop appropriate early warning and water management strategies. Past strategies to deal with the impacts of droughts on water resources have included restrictions on water supply, raising consumer awareness about the need for water conservation, and transportation of water or emergency use of desalination systems.

The drought identification method based on percentile ranking of rainfalls for different periods has been incorporated in the SCOPIC software, and the potential utility of this method as an early warning system is illustrated through an example of the 1997–98 El Niño in the Solomon Islands and Kiribati. The results are shown in Fig. 13.4 for both Kiribati and Solomon Islands, but are discussed here only in the context of drought for the Solomon Islands. In the Solomon Islands, a warm phase of ENSO (El Niño) is associated with drier conditions, whereas a warm phase of ENSO in the Kiribati is associated with wetter conditions because of its location further east in the Niño 3.4 pool of water in the central-eastern Pacific Ocean. This is illustrated in Fig. 13.4 by rainfall being out of phase in these countries for the El Niño of 1997–98 and La Niña of 1998–99.

In Fig. 13.4 is shown the time series of 6-month rainfall percentiles from 1995 to 1999, covering the El Niño event of 1997–98 and the La Niña event of 1998–99. The time series shows a sharp decline in rainfall percentiles from August 1997 (onset of El Niño) to February 1998, and for most of this period rainfall was below the 10% (severe-drought). Based on this information a drought warning could have been issued in August 1997 when the rainfall percentiles fell below 40%. Statistics on the success rate of past warnings in similar climate patterns, combined with forecasting of future rainfall, could then be used to implement appropriate drought management strategies. This is illustrated in Fig. 13.4 for SON, NDJ and JFM forecasts based on SST anomalies in the Pacific and Indian Oceans. The forecasts for the Solomon Islands (and similarly in Kiribati) capture the progress of the 1997–98 El Niño particularly towards the end of 1997 when

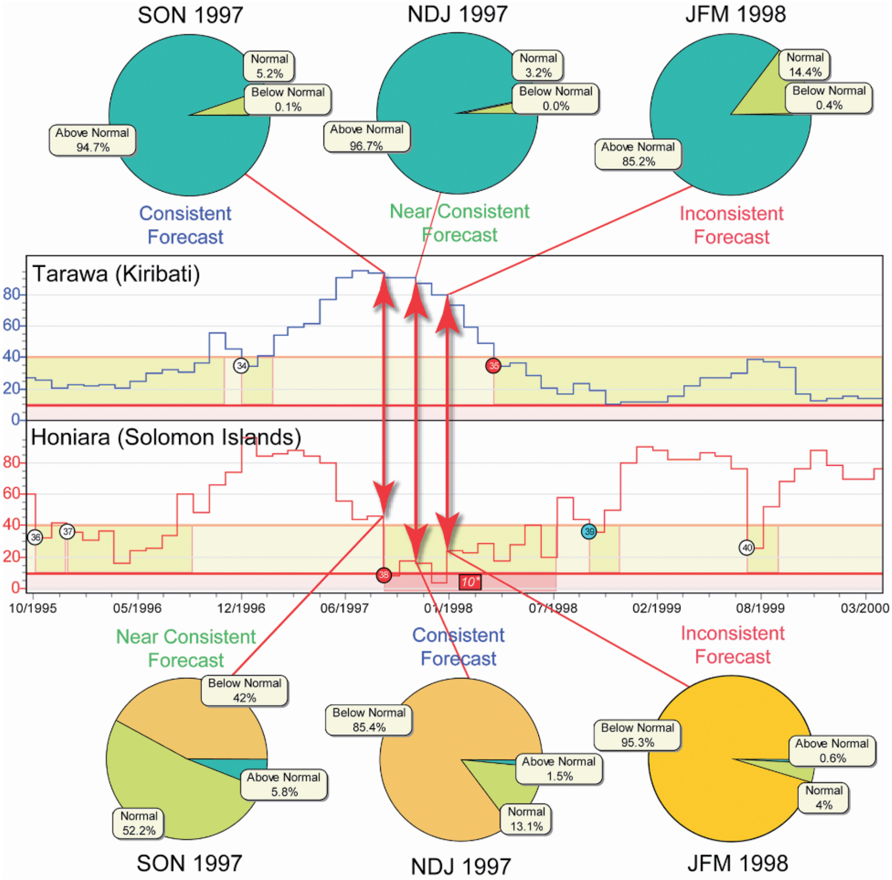

Fig. 13.4 Six-month rainfall percentile values and 3-month rainfall forecasts during the El Niño-event of 1997–98. Pie charts represent probability of above-normal (aqua), normal (yellow), and below-normal (orange) rainfall

ENSO is mature (see SST maps[3]). Tercile forecasts of rainfall, i.e. chances of below-normal:normal:above-normal for SON and NDJ were 42:52:6 and 86:13:1 respectively, which shows a high probability of below-normal rainfall particularly later in the year. Knowing the skill of forecasts during this period of the year (Fig. 13.2), decision makers could be confident on the basis of this knowledge of implementing appropriate strategies to deal with the hazards. Specific decisions will depend on the particular country and system. Research in various aspects of water management on the Pacific Islands is currently in progress.

[3] See: http://www.nrm.qld.gov.au/longpdk

13.2.5 Conclusions

Climate variability has a significant impact on water resource systems ranging from small water supplies for domestic consumption to large scale irrigation systems; however, opportunities for applying climate forecasts in water management decisions are not fully utilised. A number of impediments to the use of climate forecasts have been identified, including forecast accuracy, dealing with probabilities, risk and uncertainty, impact assessment, legal and institutional barriers, lack of integration, and communication. Possible solutions to some of these impediments have been suggested using the three case studies in the Asia-Pacific region. These include engagement of users in the research process, formative evaluation of research outcomes, communicating forecasts in a manner understandable to the users and demonstrating the benefits of forecasts relevant to the system being investigated. While it is clear that no single solution will fit all systems, indeed the treatment of all systems as homogenous has played a significant part in the lack of technology transfer and adoption in the past, prioritization of these issues (impedements) may reveal a starting point to increase the use of climate forecasts in water management decision making.

Clearly, forecast skill must exist before progress can be made, however, skill in a forecast system does not guarantee successful adoption. A perfect forecast will not have an impact unless it can lead to changes in decisions. Engendering user confidence in forecasts remains a high priority. For this to occur, the value of forecasts must be demonstrated through integration of forecasts with impact systems. Given the complexity of water resource systems, Einstein's quote "Make things as simple as you can, but no simpler" is a good guide to where we can begin with integration of climate systems, hydrological models and impact assessments.

The way forecasts are communicated influence their acceptance. Most users have difficulty in understanding *probability* but are comfortable if information is expressed in terms of frequency. The SCOPIC software was designed to overcome these difficulties in communication by providing a range of options for the user which has led to its wide acceptance in the Pacific region.

Building local capacity helps breakdown institutional barriers by raising profile of meteorological services in the community, establishes user confidence, and improves integration between science and management and the breakdown of communication barriers. This is a priority that needs to be addressed by leading institutions and governments around the world particularly in developing countries that are most vulnerable to impact of climate variability and climate change.

ENSO has a strong footprint in the region described by the three case studies which helps overcome the first barrier (forecast accuracy) and paves the way for impact assessment to be carried out. Forecast skill must exist for benefits to be demonstrated, but if the social, cultural and educational issues are overlooked these prediction technologies are unlikely to be successful in producing outcomes.

The three case studies have highlighted the importance of engaging stakeholders and the end-users in the research process, hence gaining ownership of the concepts, and helping in technology transfer.

13.3 Applications of Seasonal Forecasts to the Health Sector

We urge ministries of health and other ministries, as well as research institutions, to improve our understanding of the regional and national burden of disease due to weather and climate extremes and to identify effective and efficient interventions, such as early warning systems, surveillance mechanisms and crisis management. – WHO 2004a

This section examines how climate information is being applied to one of the most important areas of human existence: human health. In Section 13.3.1, some of the basic relationships between climate and health are identified. Section 13.3.2 then explores the issue of acquiring and preparing data for the study and modeling of the relationships as well as the conduct of routine operations that employ that knowledge. It concludes with recommendations on establishing operational services that incorporate the knowledge into public health services.

The goal of this section is not simply to describe the status and future prospects for the use of climate information in one important branch of human experience. Instead, it is to motivate climate modelers, climatologists, and those who are perfecting climate prediction to apply their knowledge and skills in creating concrete actions to diminish human misery.

13.3.1 Basic Relationships Between Climate and Human Health

As mentioned by Thompson and Perry (1997) and others, climate information is applied every day to many areas of human activities besides health, including energy, water management, agriculture and forestry, fisheries, urban and building, recreation and tourism, financial services, and transportation. But the focus in this chapter on human health makes the topic especially relevant to readers of this book, and the examples given are generally useful in understanding the procedures involved in the development and implementation of all climate applications. Applications in climate and health generally address four aspects of the human condition: disease, performance, comfort and attitude. While the last three areas are important and fit within a definition of health, most nations' health priorities are in the reduction of disease impacts.

This section deals with diseases that have been shown to have a relationship with climatic conditions and their changes that fall into two categories: infectious diseases such as malaria, dengue fever, meningitis, Lyme disease, West Nile virus, St. Louis encephalitis, and Murray Valley encephalitis; and, non-infectious diseases, which include heat stroke, skin cancer, allergic rhinitis, and some diseases of the eyes.

Malaria killed more than a million people in 2003, primarily in developing countries. Climate relationships are strongest in the life cycles of the vector – Anopheles mosquitoes – and of the parasite – *Plasmodium falciparum* being the most lethal (Thomson et al. 2004b). Lack of, or overabundance of precipitation can severely restrict the pools and stagnant ponds that are habitats for the mosquito's juvenile stages, and ambient air temperature directly affects the growth cycle of the parasite within the adult mosquito. The variations in climatic factors associated with El Niño events show some correlation to malaria outbreaks, and the scientific knowledge about the relationships is used in modeling the life stages of the parasite and the vector with respect to climate parameters. That, in turn, is one basis of a Malaria Early Warning System (Section 13.4.2) that can help to focus malaria control methods, which promise to reduce outbreaks and reduce the costs of control efforts.

Pathogens that incubate and develop outside the host (e.g. vectors such as the tick or mosquito) are usually susceptible to climatic conditions to some degree (WHO 2004b). The pathogen that carries malaria needs ambient temperatures of about 17°C to begin to develop within the mosquito. And, higher temperatures than that will cause the pathogen to develop more quickly, potentially allowing the mosquito's bites to infect more people.

Climatic variations will influence the distribution and development rate of the vectors, too, affecting such aspects and activities as the metabolic rate, egg production, and the rate of blood meals, as well as the characteristics of the vector itself. Rainfall extremes can affect the vector's habitat – too much may wash away habitat sites, but excesses that are somewhat below that amount can give the vector many more places that stay viable longer, permitting more eggs to develop. Decreased humidity usually means decreased vector life, and can be modeled for some vectors using saturation vapor pressure. Less rainfall usually means fewer useful habitats – small pools shrink to become unviable to maintain growth of the vector in its juvenile stages. While vectors that need conditions of fast moving waters may see their habitat diminish in periods of rainfall deficit, in a small number of cases other vectors that like still water may find their habitat increased, as the drying and slackening of streamflows may leave behind perfectly useful breeding pools. In general, though, more rainfall equals more useful habitats and vice versa.

Increased habitat can increase the geographical distribution of vectors. So, if rainfall can be correlated with specific vectors' habitat, and rainfall variations can be adequately predicted, projections can be made of what is likely to happen to vector distributions geographically and seasonally. While climate information is hardly the key to cures in infectious diseases, it can be a beneficial component of the control of the vectors and pathogens, and in some cases, in the treatment of the disease.

Climatic variations play an important role in some non-infectious diseases. The human body can withstand considerable external heat as long as it can dissipate the heat it generates internally, through thermoregulatory processes (WHO 2004a). It normally does this through convection, conduction, respiration, radiation and evaporation of sweat. But when the amount of incoming heat is high enough and the evaporative cooling produced by sweating is overwhelmed, the amount of body heat builds up, and heat illnesses can result. For example, heat related deaths rose dramatically across much of Europe in 2003, when the number of hot summer days was far above the long term average (WMO 2004). Climatic factors that most influence human health include air temperature, solar radiation, humidity, and wind speed.

13.3.2 Digging and Cleaning Datasets

"Enhanced planning and decision making is a fundamental capability, at all levels, for the prevention or mitigation of the negative impacts that are often associated with natural hazards. To that end, increased accuracy and reliability of information on weather, climate and water on a global scale and the free, unrestricted and timely access to that information, are some of the requisites for effective natural disaster risk assessment, vulnerability analysis, preparedness and response." – Michel Jarraud, Secretary-General of the World Meteorological Organization, in his statement to the World Summit on the Information Society.

Global health depends on the choices we make collecting and using information. Tools, methods and policies for managing information shape our ability to detect health problems, identify solutions and deliver effective interventions. As we leverage this new commitment [to invest in both the developing and industrialized worlds in strengthening data collection and management], we have learned several important lessons: first, there is an urgent need and opportunity to extract, analyse and use existing data across institutional and administrative boundaries; second, users must be enabled to interact and query their data instead of simply collecting volumes of printed reports; and, third, countries need help in communicating with politicians and the media to make it clear that better information is in the mutual interest of the government and its citizens. – Dr Sally Stansfield, Executive Secretary of the WHO-hosted Health Metrics Network (former Associate Director for Global Health Strategies of the Bill & Melinda Gates Foundation), introducing the HMN Strategy and Plan of Operations for 2007/2008.

Recognizing potential relationships between climate and other environmental components, on the one hand, and disease pathogens (and vectors) and humans on the other, depends on the availability of observations of the components and their resulting responses. Modelling those relationships depends on observations, as does the development and testing of operational services based on the models. Even the prediction of the evolution of a climate-sensitive disease, and related prediction-based early warning systems, require observations. And, the assessment of the quality and appropriateness of the outputs of those services depends on observations.

This section addresses the following questions:

- What observational data do you need?
- Where can you get it?
- What problems are there with it?
- What can be done about the problems?

13.3.2.1 What Observational Data do you Need?

In the context of human health, the data will be related to diseases, their impacts in humans, and related societal factors. Epidemiological data are needed to understand the causes of diseases and the factors that determine their distribution. Public health administrative data will describe the evolutions of epidemics, the clinical practices that are employed as well as the periods of the practices and results of treatments, and the control practices and their results. Data on the status and evolutions in a country's or region's infrastructure may figure into the understanding of the capacity and means that were or are employed in the control and treatment activities, as well as describing the variations and status of land use practices and other aspects of human activity that have a relationship with variations in vegetation, climate and diseases. And demographic data will provide information on population characteristics that affect vulnerability to diseases.

Climate and environment data will describe the weather, water, biological and terrestrial variations and status that may influence the diseases. Weather data will include traditional surface and upper air observations of elements including air temperature, rainfall and other forms of precipitation, humidity, wind, and solar radiation, and related elements such as visibility and suspended dust. Hydrological data will provide the status and trends of streamflows, rivers and lakes, and river basins. Oceanographic data will provide information on sea surface temperature and bathymetric profiles, as well as wind structure. Remotely sensed data will provide information on meteorological elements such as precipitation, cloudiness, humidity, snow and ice cover, solar radiation, wind, sea surface temperature and sea levels, and will also provide information on the status of vegetation. Forecast model output will furnish analyses and predictions of many of those variables, at varying grid resolutions. Geographic data will be needed to address the location, elevation, slope and aspect of study sites as well as the boundaries with water masses; and land use data will provide information on desertification, cultivation and irrigation practices, urbanization, and infrastructural changes that influence diseases.

13.3.2.2 Where Can You Get the Data?

Health-related data depend on the approaches to disease surveillance and the quality, quantity and completeness of the disease data. Data that describe diseases are often provided in the standard format of the International Classification of Diseases (ICD) and Related Health Problems. The best case will be in situations of notifiable diseases, especially where they are subjected to well-resourced surveillance programmes. The researcher in that case may find well documented, logically aggregated, consistent, and relatively complete data through the Ministry of Health. "In other situations, existing systems may need extensive modification, either in the way in which disease data are collected (e.g. diagnostics), or the manner in which data from individual health facilities are collected, aggregated and communicated to higher levels in the health system." – WHO 2004b. To capture information from disease outbreaks over small geographic areas or time periods, one may have to contact public health offices, municipal hospitals or individual clinics. Other sources include academic literature and WHO Regional Offices. The search for societal data can follow similar routes through respective national ministries at one end and literature searches at the other.

Climate and environmental data usually have fairly well centralized and structured archives and observational networks. The National Meteorological and Hydrological Services maintain the most comprehensive datasets for their nations' historical surface and upper air observations. They also participate in daily global data exchange of current observations through the WMO's Global Telecommunications System. The specifics of the available data and the method of acquiring it may be obtained from the WMO Permanent Representative (PR) of the country or territory where the disease is located, or the PR of the country or territory where the researcher is located. Aggregated data of lower resolution may reside at WMO Regional Specialized Meteorological Centres, and may be requested through the PRs noted above. The WMO also coordinates the international archive of globally exchanged hydrometeorological data (traditional as well as remotely sensed) through the World Data Centres in the United States (Asheville, North Carolina) and the Russian Federation (Moscow). Special datasets of very high resolution may also be acquired through academic researchers, scientific organisations and the scientific literature. Searches through Internet web sites often can find relevant datasets that may be downloaded directly, or which may be provided through correspondence with the owners. For example, the European Climate Assessment project makes available quality-controlled climatic datasets from within its region.[4]

[4] See http://eca.knmi.nl

13.3.2.3 What are Some Problems with the Data?

Researchers frequently encounter problems with observational data related to health, demographics, land use, infrastructure, environment, etc. The WMO generally classifies problems in hydrometeorological datasets into four categories (WMO 2003b):

- Inhomogeneity
  - Variations and unreal trends due to changes in instruments, sensors or processing equipment, observing or reporting practices, station locations, formulae, station environments (and changes in sensor algorithms or drift).
- Inaccuracy
  - Variations due to irregular staffing, deteriorated or failed equipment, imprecise reporting of location or elevation, improper encoding or decoding.
- Erroneous data
  - Errors in dataset identifiers, algorithms, transmission, transcription, encoding.
- Missing, incomplete or insufficient data
  - Missing due to limited observing programme, failed sensor platform or sensor, data exchange policy or practice
  - Incomplete due to non-digitized data, gaps due to civil conflict or disaster, inhomogeneity
  - Insufficient due to resolution in time, space, parameter set

Researchers frequently face similar problems with health data, compounded greatly by the absence of coordinated procedures for near-real time data collection and database management on all scales (sub-national through global scale).

13.3.2.4 What Can be Done About Data Problems in Climate- and Health-Related Datasets?

The following procedures are taken from the experience of WMO in handling traditional surface and upper air hydrometeorological data. They may also be applicable to other types of observational data that are needed in establishing the baseline relationships in climate and health.

The WMO promotes the overcoming of inhomogeneity in datasets by recommended practices known as “Direct” and “Indirect”. Direct practices are active as soon as they are adopted. Examples include the maintenance of a network of Reference Climate Stations, and adherence to the Global Climate Observing System’s Monitoring Principles. Indirect methods are applied on historical datasets, and include the use of metadata to re-quality control the data, using reference time series, breakpoint identification, and adjustment of data.

Procedures to limit inaccuracies in data, and to adjust inaccuracies in historical data, are described in the WMO's Guidelines on Climate Observation Networks and Systems (WMO 2003a) and Guidelines on Data Rescue (DARE). For erroneous data, the researcher can communicate with the dataset owner or manager to overcome errors in dataset identifiers, algorithms, transmission, transcription, encoding. It still may be necessary to apply DARE procedures.

Missing, incomplete or insufficient data can be the most frustrating problem for a researcher. The recommendations that follow cannot provide data for observations that were never taken, for example. If the data are missing due to a limited observing programme, data exchange policy or practice, seek the data through the WMO Permanent Representatives, through literature searches, or through institutional or academic networks. If they are incomplete due to their not having been digitized, or they have gaps due to civil conflict or disaster, or are questionable due to inhomogeneity, the National Meteorological and Hydrological Service may be able to resurrect the data through the application of procedures through the DARE project. It may be possible to have adequate results by substituting another dataset for the desired one. For example, Climate Change Detection indices can provide analyzed data that capture climatic extremes and trends (WMO 2003c). Data that are insufficient due to resolution in time, space, or parameter set may be approximated through use of remotely sensed data from Meteorological satellite, profiler, Doppler RADAR, etc. And model data that are needed at a specific location may be approximated through extrapolation of surrounding gridpoints' data.

To begin addressing health data collection and management problems, the World Health Organization began hosting the Health Metrics Network (HMN) in 2005 (WHO 2004b). The HMN operates through high level collaboration among countries, international agencies, donors and foundations, and technical experts. It was initiated with a large grant from the Bill and Melinda Gates Foundation. The objectives of the HMN are the following:

- Define and set standards for core health information platform designs, key indicators, data and analytic capacities and guidelines for intra- and international information use
- Accelerate and focus development and improvement of national Health Information Systems in developing countries
- Develop policies and strengthen systems and incentives that improve access to and use of information by local, regional, national and global constituencies

13.3.3 Examples of Climate Applications for Human Health

Our specific contribution . . . has been to add an understanding of climate variability as a driver of both land-use change and human health and thus as an important *confounding* factor in understanding land use-health interactions. – M. C. Thomson et al. (2004a)

This section explores applications in which knowledge of climate variability is used to enhance the understanding, control and treatment of a number of diseases.

13.3.3.1 Onchocerciasis (River Blindness) and Its Relation to Land Use Cover Change[5]

This disease is sensitive to climate variability, especially in the macro sense of the seasonal distribution of its vectors and the availability and viability of vectors' habitats. However, the purpose of discussing it here is to set the stage for an exploration of the complexities of the contributions and feedbacks among ecological and social systems that make the development of climate applications for health so interesting.

Onchocerciasis was a neglected disease, but it has become more devastating since the 1970s. Its effects are greatest on the rural populations of the West African savannah living near fast-flowing rivers. The vector is the blackfly which needs to have a particular type of underwater vegetation to lay its eggs. At the initiation of the Onchocerciasis Program (OCP) in 1974, some savannah villages close to river valley habitats of the blackfly vector were suffering severely, with 60% adults infected, and 3–5% already blind. Many villages had been completely abandoned. At the peak of the OCP control activities in 1986, the estimates were that 30 million were affected and 2.4 million were infected. The Programme's 20-year eradication programme put the disease under virtually total control in 11 countries in West Africa. Resettlement occurred rapidly in villages that had been abandoned.

There are two species of the fly, and two ways the pathogen works in humans. In the savannah areas, there is a prevalence for the blinding form of the disease. In the deep forest, however, there is more likely to be a skin disease, which is much less problematic than blindness. The causal agent is the *filarial* worm – a segmented, small worm. The vector is the savannah species of blackfly that transmits the filarial worm to humans through its bite, and is mainly controlled by insecticide sprayed on savannah rivers. In the epidemic areas, the vector control for humans is supplemented by an oral drug – ivermectin tablets – that act on the filarial worm.

The spraying could be made ineffective by changes in climate that would increase the flow of rivers sufficiently to wash away insecticide without washing out the fly's habitat. But, the control programme has benefited from climatic knowledge and prediction of the monsoon winds and associated rainfall patterns, which has been used to plan the timing, locations, and type of insecticide used (Thomson et al. 2004b).

---

[5] Drawing heavily on work reported in Thomson et al. (2004a).

Intentional deforestation to convert areas for agriculture and human habitation resulted in savannah flies moving to newly habitable areas and bringing the blinding form of disease. Between 1973 and 1990, the studied area experienced a 10% increase in urban habitat, 18% increase in savannah area, 11% decrease in degraded forest, and 17% decrease in dense forest. This caused significant changes in the habitat for the fly and the type of fly that brings the blinding disease, and these had serious impacts on the growing numbers of humans in the region. From 1975 to 1980, the river blindness variety of the disease was only prevalent in 10.5% of the area. By 1997, it had virtually doubled. The existence of the effective control program meant that human settlement could move into other, uncontrolled areas, i.e. previously vacant land, and it did so with vigor. In 1973 along the river, 5% of the area was cultivated. In 1983, 30% was cultivated; by 1993, 70% was cultivated, and with the increase in cultivation and human habitation along the rivers humans were being increasingly exposed. Humans were also changing how water got to the river, which affected the habitat sites for the vectors.

By the mid-1990s the method of control of the disease had swung from one that relied predominantly on spraying insecticide on the vector habitats (as had been done under the OCP), to the oral drug (ivermectin) treatment of humans (the method employed in the African Program for Onchocerciasis Control (APOC) that followed after the OCP). But, expansion of the control efforts into forested areas resulted in severe adverse reactions and death among patients taking the drug, which was associated with ivermectin. Distribution of the drug was impeded because control organisations did not want to increase the problems of the reaction of the drug.

Why was it happening, and what could be done? The path to the answer to the first part of that question lay beyond onchocerciasis, to Loa loa – another disease that is transmitted by a different vector. It was shown that people infected through bites from the chrysops fly and contracting Loa loa were the ones having serious reactions to the ivermectin. The drug became ineffective not because it couldn't handle the onchocerciasis, but because, in the presence of the Loa loa in the body (a less serious disease), it resulted in severe reaction.

13.3.3.2 Loa Loa[6]

Loa loa is now the major issue – 20% of the population that is subjected to both onchocerciasis and Loa loa is at risk of adverse reactions to ivermectin. That means that, in terms of onchocerciasis, medical professionals are having to analyze who will get severe reactions, and they are working on alternative methods of introducing the drug ivermectin or other treatments.

---

[6] Drawing heavily on work reported in Thomson et al. (2004a, c).

The APOC gave high priority to mapping the spatial distribution of loiasis, so that it could modify its approaches to treatment and enhance the surveillance methods. Looking at all of Cameroon, for example, meant a great area to have to cover. But by focusing on areas where a 20% or greater prevalence was predicted, they could find where the Loa loa presence was dramatically increasing. Extensive study began with mapping the distribution of environmental factors that influenced Loa loa distribution. The mapping showed that the Chrysops fly habitat was associated with fringe areas between the dense forest and its borders with cultivated areas, where the fly could live in the high crowns and the larval stages could grow in wet, organically rich and muddy low-lying areas. That is a vastly different habitat from that of the blackfly of onchocerciasis, which needs fast-moving water.

A risk map was developed using data from Cameroon. Considerable prevalence data was available, through the Centre Pasteur of French Research for Development organisation (IRD). Newly available environmental data included imagery from the SPOT sensor and the Synthetic Aperture Radar sensor (SAR). The available population data were extensive – over 14,000 individuals from 95 villages over period of 10 years – age, sex, presence or absence of Loa loa, and the amount of blood taken for examination – and were factors for the regression analysis. Data on the village – latitude and longitude from the ordinance survey map or a Global Positioning System – were used to anchor epidemiological data. Population density, normalized difference vegetation index data (NDVI) from SPOT and SAR, and altitude from the US Geological Survey's Digital Elevation Model were obtained. The NDVI satellite data show the seasonal and inter-seasonal variations in the vegetation, which provides information on the status of the vector habitat as it responds to climatic variations. The data also correlate well with the changes in the distribution of the vectors. But, the NDVI data has 1 km resolution, while some forest galleries that are important in looking for vector habitat are only 50 m wide. Synthetic aperture radar resolved to 100 m was used and the finer resolution gave more detail in looking at forest areas.

Using the environmental data and considerations of age and gender, a logistic regression model was used to map the spatial prevalence of the disease. Of particular concern are areas where the prevalence exceeds 20%, and in only a very few locations did the prevalence exceed this threshold when the model indicated lower levels of risk. The model has been applied to map risk of the disease in the Cameroon region, the Sudan, and Ethiopia.

13.3.3.3 Meningococcal Epidemic Meningitis[7]

Meningococcal Meningitis transmission is by direct droplet contact. This disease only exists in the nose and throats of humans. Twenty- to forty percent of the West

[7] Drawing heavily on work reported in Thomson et al. (2004a).

African populations that are affected are symptomless carriers – they have it in nasal cavities in membranes but exhibit no symptoms. The seasonality and inter-annual variability of the disease are related to the proportion of clinical infections that get strong enough for patients to begin seeking treatment, compared to the sub-clinical infections. It is the strength of the infection, and the cycle of spreading the disease to another, that is more important to researchers, rather than the behavior of the persons in transmitting the disease.

The disease is sensitive to climate variations, especially on the seasonal scale. Epidemics typically reach their peak at the height of the dry season and diminish once the rainy season commences. If vaccination starts before the peak, the disease can be managed better, but that is costly and depends on well maintained surveillance and response mechanisms. An alternative is to vaccinate ahead of the onset of cases, but while that may be cost-effective it is difficult to implement.

There is a concentration of geographical distribution and seasonal occurrence in the Sahel zone of Africa – the risk factors are dry, dusty conditions, which increase the risk of the disease. The main apparent contributor is land use and land use change, where cultivation stirs up the soil. The fine airborne dust (1 μ) is a factor in the disease, as particles can reach well into the lungs.

Modelling through the Meningitis Forecasting Project for Africa showed that absolute humidity and land cover are reliable indicators to distinguish between areas with high and low risk of epidemic events. Other important indicators include population density, dust and soil type. Meningitis epidemics have been shown to be influenced strongly by low absolute humidity and dusty conditions outside the tropics as well. Analysis by the Meningitis Forecasting Project for Africa shows that the risk of Meningitis epidemics in Africa is concentrated in certain specific regions, and that the environment and particularly the climate variations are strong influences on where and when they will occur. The high predictive value of the model developed through the Project showed that an environmental model can be instrumental for policy makers in understanding the distribution of meningitis epidemic risk across Africa.

The model also has the potential to be useful in a Meningitis Early Warning System. It is able to resolve environmental components to inter-annual variability with respect to the different times of the year in which they are important. However, much work remains to be done, especially on the ways that the environmental and other factors influence the onset and extent of epidemics.

Changes in the spatial and temporal distribution of diseases result from changes in population, climate, land use, economy, and social structure, among others. Health decision makers often require simple solutions to complex problems. It is important to work with policy makers to understand their decision frameworks, to initiate the dialogue as to the value of specific information and to develop their use of uncertain information.

13.4 Early Warning Systems

Early Warning Systems (EWS) are a common approach in many activity areas to preparing to deal with anticipated problems when there is still time available to respond to and to mitigate those problems. According to the Department of Early Warning and Assessment (DEWA) of UNEP the fundamental roles of Early Warning Systems are to analyse and assess trends, to provide policy advice, to provide early warning information on threats, and to catalyse international cooperation based on best-available scientific and technical capabilities. A scan of the web reveals Early Warning Systems for breast cancer, national development projects, Internet and credit card fraud, war and conflict outbreak, the energy and water balance of the earth's surface, paedophilia, volcanic eruptions and earthquakes, and many more. One evolving EWS that has attracted much recent media attention at the time of writing is the Indian Ocean Tsunami EWS, instituted following the 26 December 2004 event. Many of these systems listed above naturally involve no element of climate, and frequently in the past those that do incorporate climate as a component have issued warnings based only on climate observations and not on predictions *per se*, a situation that is now changing rapidly as the potential benefits of predictions are becoming recognised. Famine, drought, heat, and health-related EWSs are examples of those that include climate components.

All Early Warning Systems tend to follow the same basic model. Relevant events known to be associated through experience and observation with the issue of concern, say famine, are monitored and interpreted. Actions are then taken once pre-determined trigger points are reached. In some cases there may be only a single trigger point, such as with a tsunami system when an earthquake satisfying basic criteria and/or evidence of an existing tsunami will trigger an "evacuate" instruction, whereas in other systems several trigger levels may exist instigating progressively more urgent responses. In all cases the idea is simply to use knowledge of precursors, together with efficacious monitoring, to provide as much warning as possible that potentially adverse impacts, such as declining food stocks or conditions suitable for a malaria outbreak, are in the course of development. Early Warning Systems in general do *not* predict that specific adverse impacts *will* occur in due course, but merely extend preparation time compared to systems that respond only once impacts have been recognised, at which stage it is often too late to take remedial actions. Systems that maintain watches for on-the-ground signs that impacts, perhaps famine, are occurring in reality, but without providing warning, are known as Early Detection Systems (EDS); EDS's are frequently used to complement EWS's. There is a recent trend, however, to incorporate predictions, including weather and climate, within EWS's in order to help focus the warnings.

A new concept of early warning has been developed using the ideas of chemical theory, in particular the terms 'hotspot' indicating a reaction that might become unstable, 'flashpoint' at which the reaction begins but may still be reversed, and

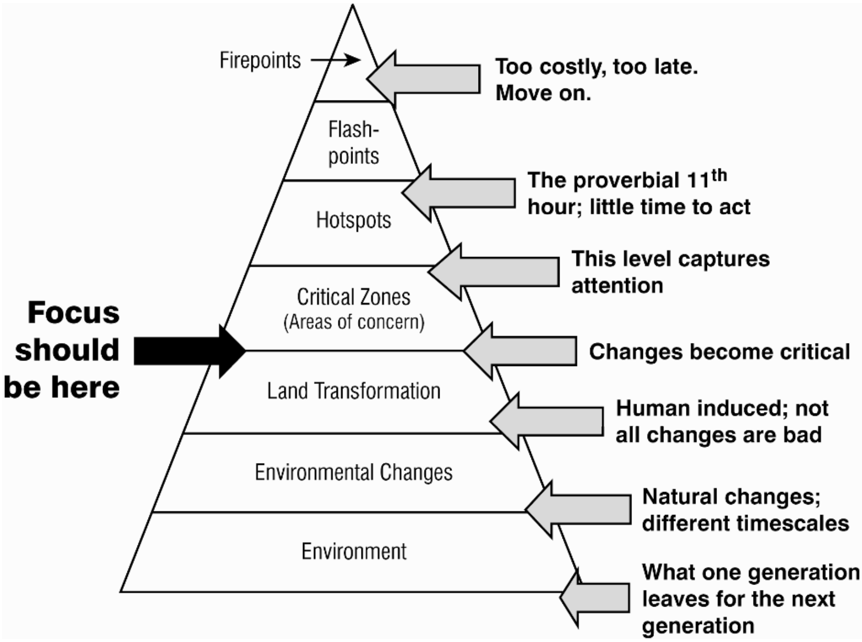

Fig. 13.5 Diagrammatic representation of the various stages through which environmental changes might translate to become firepoints. Preventative action is best undertaken at the 'critical zones' level rather than await the creation of hotspots and, later, flashpoints within them

'firepoint' at which the reaction cannot be reversed. Although Fig. 13.5 has been prepared on this concept within the context of environmental changes it is readily revised within any other context. Within the layers of Fig. 13.5 an EDS might be designed to operate within known critical zones, and in particular hotspots, to indicate that a flashpoint has been reached. An EWS would operate throughout to monitor the changing situation and to interpret information in terms of the likelihood that hotspots will translate to flashpoints.

At one time there was a remarkable proliferation of Early Warning Systems in developing countries, often with individual NGO's, International Development Agencies, UN Agencies and national Government Departments having their own systems, each with their own specific, but at times duplicative, objectives. Some consolidation has taken place in recent years, frequently under UN bodies including UNEP DEWA, such that duplication and differential triggering have been reduced. Examples of major Early Warning System activities include:

- The Humanitarian Early Warning System, HEWS,[8] organised by the UN World Food Programme on behalf of the IASC (Inter-Agency Standing Committee),

[8] See: http://www.hewsweb.org

which represents a large number of UN Agencies, various Inter-Government Agencies, NGOs and other Institutes – HEWS provides a one-stop web source for droughts, floods, storms and other weather events, locust invasions, seismic events, avian influenza, plus others on an as-needed basis.

- The Global Information and Early Warning System, GIEWS,[9] of the UN Food and Agriculture Organization, and the Famine Early Warning Systems Network, FEWS NET,[10] run by the US Agency for International Development together with several other US Government agencies, both with a focus on food security in the developing world.
- The Malaria Early Warning System, MEWS, coordinated by the UN World Health Organization, that developed from the Roll Back Malaria project.

Most Early Warning Systems have similar generic requirements:

- A good quality archive of historical data covering as many aspects relevant to the issue of concern as possible, in all necessary spatial and temporal detail; this archive should include all factors that relate to vulnerability in regard to the issue of concern and may therefore extend beyond information just on the events of interest to include, say, population data, land use data, economic data, climate data, and so on.
- An understanding of the historical data in terms of its relationships to, and hence its information content with regard to producing warnings for, the issue of concern.
- Agreed approaches to recognising trigger points and to responding to these in established manners, preferably with an experience-base assisting in establishing best practice.
- An adequate monitoring system for all relevant data at appropriate temporal and spatial resolutions.
- Systems for accessing monitoring data and for broadcasting warnings to decision makers within time constraints appropriate to the issue of concern.
- Institutional support at all necessary levels for maintaining the system and for responding to the warnings.

Each of the requirements in the above list presents its own specific challenges, particularly in regions where data are lacking, or data archiving is inadequate, or the necessary research is incomplete. But the final bullet in the above list is perhaps the most important, as without appropriate institutional support the potential benefits of any EWS may not be realised; care needs to be taken in any demonstration project that institutional constraints are recognised and addressed.

[9] See: http://www.fao.org/WAICENT/faoinfo/economic/giews/english/index.htm

[10] See: http://www.fews.net

Prediction *per se*, and seasonal to interannual climate prediction in particular, does not currently form a component of many Early Warning Systems, although on the climate side developments in the understanding of ENSO have enabled benefit to be gained from knowledge of the canonical expressions on temperatures, rainfall, storms, etc. during different phases of ENSO. However such knowledge can be used inappropriately when specific ENSO events produce non-canonical responses, as occurred around some parts of the Indian Ocean basin during the 1997/98 El Niño (see Figs. 1.1 and 6.10). Seasonal climate prediction, rather than use of climatology, is preferential, but so far has not been employed to the extent that it might. Perhaps one reason for the relative lack of use of climate prediction information, despite the evident advantages of its employment, is absence of clear demonstrations that predictions at their current state of development will benefit warnings. Further these predictions may lack the spatial and temporal specificities considered desirable to improve EWSs.

One area in which progress has been made in the incorporation of seasonal to interannual rainfall prediction is in MEWS, the Malaria Early Warning System, where additional Forums over and above the RCOFs have been trialled in southern Africa with regard to converting climate predictions into malaria outbreak predictions. As detailed in Section 13.4.2, there is a close link between both climate and climate variability with malaria incidence, both spatially and temporally, but it is in the epidemic areas where climate is only intermittently conducive to the spread of the disease that prediction would be of greatest benefit. Modelling studies have demonstrated that malaria incidence is, to an extent, predictable based on climate inputs alone, and thus prediction might be used either independently or within the structure of an EWS. Pilot studies within the southern African region have taken the approach of introducing predictions within the structure of a MEWS, as illustrated in Fig. 13.6.

The bottom row in Fig. 13.6 illustrates across 4 years the weekly incidence of malaria morbidity and mortality at a location within an epidemic zone, with the black line indicating historical averages. In the row above are shown rainfall observations for the 4 years, together with, in black, the climatology. Of interest, naturally, in this epidemic area is the fourth year, during which a malarial outbreak was preceded by above-average rainfall that established breeding sites for the vectors. Flag 3 (see top row), a trigger point obtained through an EDS, provides confirmation that an outbreak is in progress during this fourth year, but offers on its own rather limited preparatory opportunities. An EWS, that had not only suggested increasing vulnerability of the population to an outbreak through the earlier years (because of reduced resistance resulting from a period of limited mortality/morbidity accompanying rainfall around or below average – second row), but that also recognised the relatively high rainfall of the fourth year, might have offered a trigger point at Flag 2 at the onset of the heavy rains, with a few weeks' preparatory advantage over Flag 3.

Seasonal rainfall prediction, as in the third row, might offer the further months of preparatory time associated with the trigger point of Flag 1. In practice the

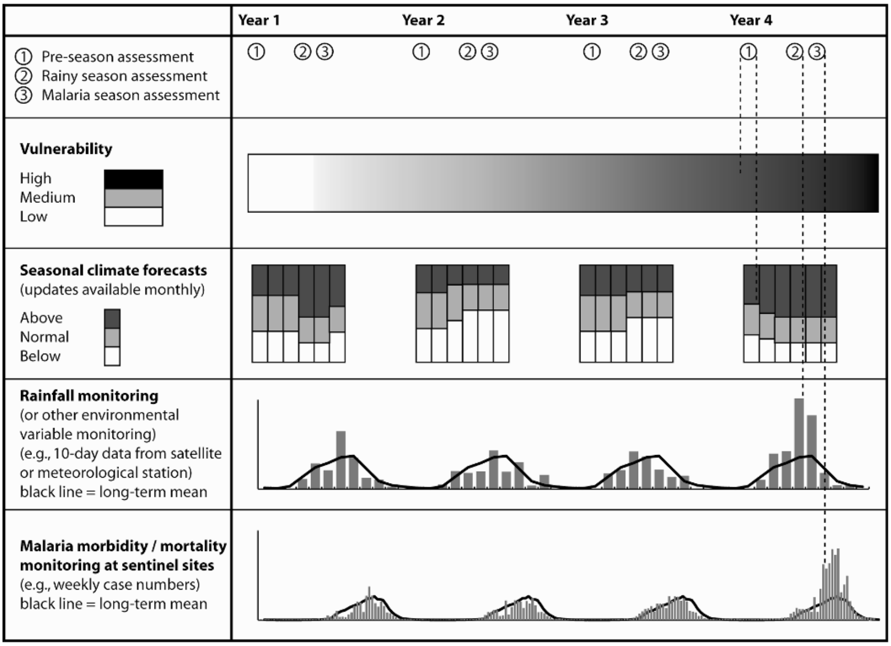

Fig. 13.6 Example of the manner in which a malaria Early Warning System might work using seasonal climate predictions (middle row) in an epidemic area. Malaria incidence is given in the bottom row with observed rainfall above; respective climatologies are shown as black lines. Vulnerability of the population is shown via a traffic light approach in the second row, while triggers are shown in the top row. See text for full details

decision making in this case is slightly more complex than is suggested in Fig. 13.6 as the probabilistic seasonal rainfall predictions shown have inflated levels of skill, and in reality the category with the highest probability is not likely to occur as frequently as indicated. Nevertheless this pilot study illustrates the potential benefits offered by seasonal to interannual prediction within the context of Early Warning Systems.

13.4.1 Health Early Warning Systems

13.4.1.1 Dengue Fever Early Warning in Indonesia

Dr Dana Focks (Focks 2003) reported on developments of an early warning system for dengue and dengue hemorrhagic fever in Southeast Asia. The early warning system predicted dengue prevalence in March to May in Yogyakarta, Indonesia. Using sea surface temperature data from meteorological satellite and data from the epidemic from 1987–1989 to develop a statistical model, the Dengue Early Warning system with 3 months' lead time was predicting correctly and had one error in the prediction for the year 1992 (a false positive): the model gave a probability of 0.64 of "epidemic" and 0.36 of "no epidemic". The WHO and PAHO formerly had promoted insecticide aerosols for control of mosquitoes – a government based approach that was costly and didn't work, and was abandoned. Then in the 1990s the new approach became "community-based source reduction", with the goal to clean or eliminate open containers (e.g. discarded tires, empty oil drums) that fill with water, which are a habitat for the mosquitoes that carry dengue. But, without a strong, direct government push, the recommendations were not followed through, which meant that there was no effective control with that approach either. The latest strategy is to use a climate-based Early Warning System to focus control on the containers most likely to provide adequate habitats, for example, a particular class; or, all those abandoned in selected sites, e.g. public lots and gathering rainwater, etc., to help focus the cleanup efforts.

13.4.1.2 Heat Health Warning Systems

Two major activities are underway to develop guidelines on heat health warning systems (HHWS). The World Meteorological Organization through its Commission for Climatology is studying universal thermal heat indices and heat health warning systems, with the goal of issuing guidelines that will help all WMO Members to establish warning systems to protect their populations from extreme heat events (WMO 2004). At the heart of most HHWS are forecasts of dangerous heat conditions, which can be based on:

1. *Single meteorological variables* such as air temperature or relative humidity. Relative humidity is often not used effectively, but temperature does contain information about the thermal environment.
2. *Simple thermal indices* (historic) as, e.g. the Heat Index. These are believed to have limited relevance and limited reliability.
3. *Weather classifications* (holistic approach). This approach has been shown to be successful in heat/health studies. The technique requires the development of a synoptic or weather type classification that can, depending on the level of sophistication, be data and analysis intensive. Furthermore, synoptic or weather types, as is the case for human energy-based biometeorological indices, can never be verified, as they are statistical or numerical constructs. This contrasts with conventional meteorological variables, as forecast values of these can be compared with actual observed values.
4. *Heat budget models*, such as the Universal Thermal Climate Index (UTCI). These are thermophysiologically relevant, consider the complete heat exchange

conditions, and are valid for all thermal environments (both heat and cold). Such procedures are able to fulfill the precondition that the same value of an index always means the same for the human body, independently from the combination of the single values of the meteorological input parameters.

The other major activity is the Watch Warning System work package of the Assessment and Prevention of Acute Health Effects of Weather Conditions in Europe (PHEWE) project and the associated EuroHeat project (considering the utility of seasonal climate forecasts), under the of the Fifth Framework Programme, funded by the European Union (WHO 2004a).

13.4.2 The Malaria Early Warning System: Malaria Incidence – Climate Relationships in Botswana

In this section, a step-by-step analysis is presented of the relationship between the annual incidence of confirmed malaria in Botswana, and a country-wide average of rainfall during the peak rainfall season. The results are then used to make a prediction of malaria incidence. The analysis is based on previous research by Thomson et al. (2005, 2006), but updates the data to 2005. The reader should be able to repeat, as an exercise, the analyses using the data presented in the tables.

Table 13.1 shows a set of annual total confirmed and unconfirmed malaria cases in Botswana for the 24-year period 1982–2005, together with the estimated population for the country. Any analysis of trends in malaria (or any other disease) should take account of changes in the population since the total number of people infected will almost inevitably increase if the total population increases. It is more informative to consider the trends in the proportion of people affected by a disease. For example, since the population of Botswana increased by approximately 80% over the 24-year period, the total number of malaria cases would have to increase by more than 80% to indicate that the disease had become more widespread. The standard way of indicating the proportion of people affected by the disease is to divide the number of cases by the total population. This number is known as the incidence. The incidence is often multiplied by 1,000 to indicate how many people out of a typical sample of 1,000 people were infected with malaria. The confirmed incidences per 1,000 are illustrated in Fig. 13.7.

Figure 13.7 indicates that outbreaks of malaria seem to have occurred at intervals averaging about 4 years. These outbreaks were then followed by exponential declines in incidence. This pattern is characteristic of a population's evolving immunity: after an outbreak, immunity is built up, and incidence declines, but with the consequent decreased exposure to the disease, immunity is lost, and the disease can re-occur at a potentially devastating extent.

Table 13.1 Annual confirmed and unconfirmed malaria cases in Botswana for 1982–2005, together with the estimated total population

	Malaria cases		
Year	Confirmed	Unconfirmed	Population
1982	85	332	1,019,690
1983	161	1,167	1,051,227
1984	320	1,831	1,083,739
1985	628	1,867	1,117,256
1986	1,437	2,994	1,151,811
1987	326	1,228	1,187,434
1988	9,013	21,587	1,224,159
1989	5,398	14,842	1,262,019
1990	1,916	8,457	1,301,051
1991	1,783	12,012	1,326,796
1992	415	4,293	1,358,554
1993	14,615	40,722	1,391,073
1994	5,335	24,251	1,424,369
1995	2,271	16,451	1,458,463
1996	25,641	80,004	1,493,373
1997	19,811	100,579	1,529,118
1998	5,810	59,623	1,565,719
1999	12,754	72,803	1,603,196
2000	8,056	71,555	1,641,570
2001	4,716	48,281	1,680,863
2002	1,283	28,907	1,721,096
2003	1,830	23,657	1,762,292
2004	3,453	22,404	1,804,475
2005	1,738	14,019	1,847,666

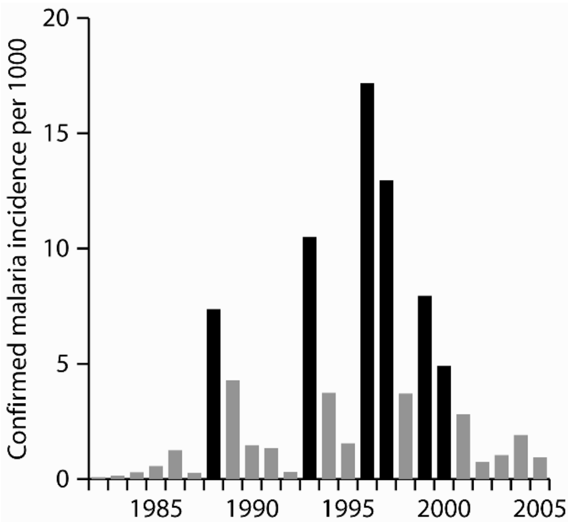

Fig. 13.7 Annual confirmed malaria incidences per 1,000 population in Botswana for 1982–2005. The black bars indicate epidemic years: those in which the incidence exceeded the upper quartile

If an outbreak is severe, it is called an epidemic. Epidemic years are formally identified as years in which incidence is higher than the upper quartile. The upper quartile defines the level of incidence that is exceeded on average once in every 4 years. The epidemic years are highlighted as black bars in the graph, and they occurred in 1988, 1993, 1996, 1997, 1999, and 2000. The worst epidemic occurred in 1996, and seemed to be part of an upward trend that has apparently been reversed. This upward trend is attributable, in part, to changes in the resistance of the malaria parasite to drugs, and represents an increase in vulnerability of the population to the disease. Similar vulnerability trends have been widely reported in other parts of the continent and beyond.

Changes in malaria control policy were implemented in 1996, and, if effective, would be evident in the data for 1997 and later. The policy intervention does appear to have been effective in halting, or even reversing the upward trend. However, the simplest way to test the effectiveness of the intervention is to consider its impact on the ratio of unconfirmed to confirmed malaria cases. Since the incidence of unconfirmed malaria consists of cases of malaria-type symptoms that could just be cases of fever, the ratio of unconfirmed to confirmed incidence is likely to have increased if the changes in malaria control have been effective. In Fig. 13.8 the confirmed and unconfirmed incidences are compared for pre- and post-policy change years. The incidences for 1997 onwards are shown in grey, and there is a clear increase in the ratio of unconfirmed to confirmed cases represented by the displacement of the grey markers to the right of the graph. The intervention therefore appears to have been effective.

Since most statistical tests assume that the data being analysed are normally distributed, the incidence data should ideally be transformed because the annual values are strongly positively skewed (skewness is approximately 1.8). Taking the

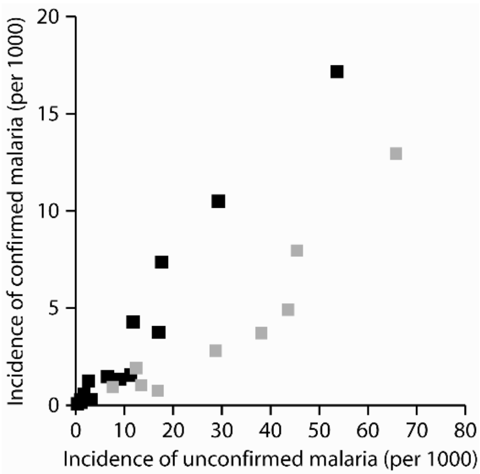

Fig. 13.8 Annual confirmed and unconfirmed malaria incidences per 1,000 population in Botswana for 1982–1996 (black) and 1997–2005 (grey)

logarithm of the incidence eliminates the positive skewness, and even introduces a slight negative skew (skewness is approximately −0.3). The log-transformed incidence is therefore a more appropriate value than the incidence for performing statistical analyses.

Epidemic malaria generally occurs in one of two areas: semi-arid areas where insufficient rainfall usually limits breeding areas for mosquitoes; and highland areas where cold temperatures can severely restrict the breeding cycle of the malaria parasite. In the "desert fringe malaria" areas, where rainfall deficiencies typically restrict the occurrence of malaria, occasional seasons of heavy rainfall can result in epidemic malaria because of the increase in mosquito breeding sites. Botswana is a semi-arid country, and most of the rainfall occurs in the summer months December–February (Fig. 13.9, black bars). The malaria incidence peaks about 2 months later (Fig. 13.9, grey bars). Rainfall varies considerably from year to year in Botswana (Table 13.2), and so, epidemic years may be most likely to occur after a good rainfall season.

Given the vulnerability trends already mentioned, as well as the effects of the policy intervention of 1996, a simple correlation between the rainfall and the log malaria incidence would not give an accurate indication of the strength of the effect of rainfall on epidemic risk. A regression model is to be used to estimate the influence of climate, but it is necessary to account for these known non-climatic influences in the model in order to estimate the climate's influence more accurately. Consider first the policy intervention. There are a number of possible effects of this intervention:

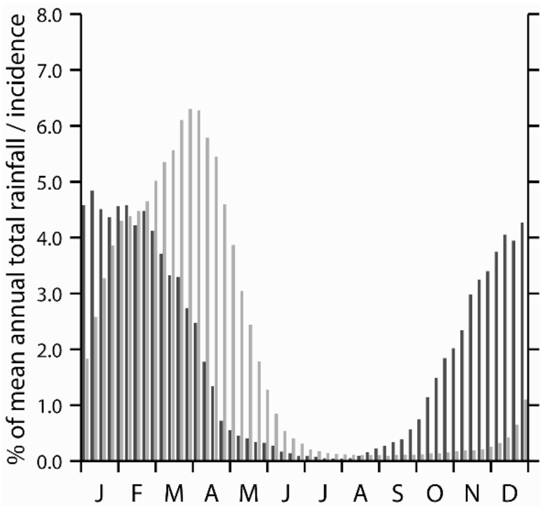

Fig. 13.9 Averaged weekly confirmed malaria incidences per 1,000 population in Botswana (black bars) and averaged weekly rainfall totals 1997–2005, as percentages of the mean annual totals. The weekly averages are filtered using a 5-week running mean

1. The background incidence (as measured by the mean) may been have changed
2. The vulnerability trend may have been changed
3. A combination of the two previous effects may have occurred
4. There may have been no discernible effect at all

To test these possibilities, additional variables need to be included in the regression model. Firstly, an indicator variable (a series of 0s and 1s) is defined that indicates years in which the intervention has occurred. This variable will therefore equal 0 up to 1996, and 1 from 1997 onwards. In a regression model, this variable can be used to represent a change in the background incidence (option a). An additional new variable is required to represent option b, but the vulnerability trend itself needs to be accounted for first. The vulnerability trend can be represented as a simple linear trend component by including the year in the regression model. The possibility of a modified trend can then be calculated by superimposing an additional trend component on the long-term vulnerability trend, starting only in 1997. This modified trend is incorporated by a variable that is obtained by multiplying the year by the indicator variable. Thus, three new variables are included that describe the possible effects of trends in vulnerability, and the effect of the policy intervention on both the trend and on the mean incidence.

Table 13.2 December–February rainfall (mm per day) averaged over Botswana for 1981/82–2002/03. The Climate Prediction Center Merged Analysis of Precipitation (CMAP) data were averaged across the 20 grid points between 17.5–27.5°S and 17.5–30.0°E as approximately representing Botswana

Year	Rainfall
1981/82	1.81
1982/83	1.81
1983/84	1.93
1984/85	2.46
1985/86	2.37
1986/87	1.89
1987/88	3.67
1988/89	3.93
1989/90	2.44
1990/91	2.98
1991/92	1.70
1992/93	2.48
1993/94	3.33
1994/95	1.90
1995/96	3.80
1996/97	3.56
1997/98	2.19
1998/99	2.66
1999/00	4.88
2000/01	2.31
2001/02	1.76
2002/03	2.23
2003/04	2.48
2004/05	2.37

Before proceeding, one additional new variable should be considered. In other areas excess rainfall has been found to impact mosquito populations negatively through the destruction of breeding sites. Independent research results indicate the existence of a quadratic relationship between rainfall and malaria incidence: an increase in the mosquito population, and hence the epidemic risk, with an increase in rainfall will only occur up to a certain point; if there is too much rainfall the risk of an epidemic may decline. To consider this possible effect of excess rainfall, an additional variable representing rainfall squared is included.

Five explanatory variables are therefore to be included in the model: rainfall, rainfall squared (to account for the possible effects of too much rain), year (to represent the vulnerability trend), year from 1997 onwards (to represent the effects of the policy intervention on the vulnerability trend), and an indicator variable (to represent the effects of the policy intervention on the background incidence). The results of the regression model are shown in Table 13.3. The partial *t*-statistics for all the regression parameters are well above 2.0, indicating that all five variables explain a significant proportion of the total variance at a 95% level of confidence. The coefficients for the two climate variables describe an inverted *u*-shape (because the parameter for rainfall squared is negative, –0.26), confirming that excess rainfall may result in a decrease in malaria incidence. The vulnerability trend was reversed by the policy intervention (the intervention parameter of –0.16 is stronger than the vulnerability trend of 0.07), but the positive coefficient for the effect of the intervention on the mean incidence (322.44) seems to imply that the incidence has increased. This paradox is simply a reflection partly of the fact that there are more years in the 24-year sample prior to the intervention than following it, and partly of a discontinuity in the trend line.

Table 13.3 Regression parameters, standard errors, and partial t-statistics for estimating log annual confirmed malaria incidence for Botswana for 1982–2005

	Intercept	Rain	Rain^2	Vulnerability	Intervention	
					Mean	Trend
Parameter	–148.85	2.00	–0.26	0.07	322.44	–0.16
Standard error	30.41	0.42	0.07	0.02	72.92	0.04
Partial t-statistic	4.89	4.77	3.96	4.76	4.42	4.42

Using the results in Table 13.3, the annual log incidences that can be attributed to the vulnerability trend and to the effects of the policy intervention can be calculated. The contributions from these non-climate variables are obtained by using the regression parameters from the Table for these variables only. These contributions can then be subtracted from the observed incidences so that the effects of the vulnerability trend and the policy intervention are removed. It is important to estimate the regression parameters for these non-climate variables while including the climate variables in the model because it is possible that part of the observed trend up to, and after, 1996, and any change in the mean after 1996 are partly an effect of climate trends over the same period. The log incidences after removing the non-climatic effects (and after centering so that the mean over the 24-year period is zero) are shown in Fig. 13.10. The "epidemic" years have been re-identified, and are shown in black. The new years are the same as shown in Fig. 13.7, except that 1989, is classified as an epidemic year instead of 2000.

If the regression is recomputed using only the climate explanatory variables to explain the incidence data shown in Fig. 13.10, the regression parameters (except that for the intercept) are essentially unchanged, but the strength of the influence of the climate can be estimated. The climate variables explain about 75% of the variance of the adjusted incidence data. The strength of this relationship is indicated in Fig. 13.11, where the quadratic nature of the relationship is evident. The quadratic relationship appears to be primarily a result of the 1 year with almost 5 mm/day of rainfall (the year 2000), but the quadratic curve shown was calculated without using 2000. There is an imperceptible change in the curve if 2000 is included. The curve indicates that the risk of an epidemic is maximized when there is about 3.75 mm/day.

A forecast of incidence can be made given a prediction of rainfall using the results in Table 13.4. The ensemble-mean rainfall prediction for December–February 2005/06 for Botswana from the ECMWF model was about 2.61 mm/day (after bias correction). This rainfall prediction converts to an estimated log incidence for 2006 of 0.14, or an incidence of 1.15 per 1,000. With a population of about 1.9 million, the estimated number of confirmed malaria cases would be a little

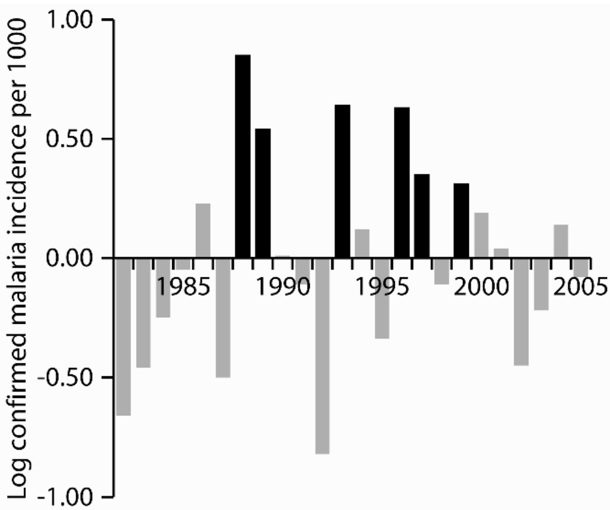

Fig. 13.10 Annual anomalies of confirmed log malaria incidences per 1,000 population in Botswana for 1982–2005, after removing the vulnerability trend and the effects of the policy intervention. The black bars indicate epidemic years: those in which the incidence exceeded the upper quartile

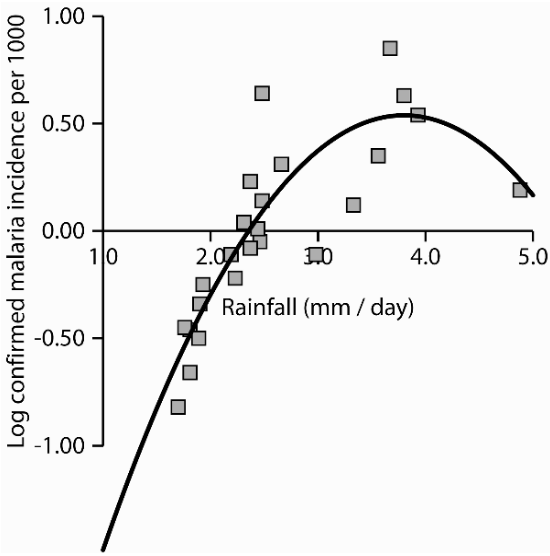

Fig. 13.11 Annual anomalies of confirmed log malaria incidences per 1,000 population in Botswana for 1982–2005, after removing the vulnerability trend and the effects of the policy intervention, and observed December–February rainfall for 1981/82–2004/05. The quadratic line was calculated without including 2000

Table 13.4 Regression parameters, standard errors, and partial t-statistics for estimating anomalous log annual confirmed malaria incidence for Botswana for 1982–2005, after removing the vulnerability trend and the effects of the policy intervention

	Intercept	Rain	Rain^2
Parameter	–3.24	2.00	–0.26
Standard error	0.53	0.37	0.06
Partial t-statistic	6.06	5.35	4.39

under 2,200, which is not enough to qualify as an epidemic, but certainly enough to indicate an increased threat against the background of the downward trend resulting from the policy intervention. Of course, the epidemic risk should not be estimated using only an ensemble-mean rainfall prediction. One possible approach would be to use the ensemble of predictions and to obtain incidence estimates from each member. The Malaria Early Warning System described above demonstrates the effective use of combining the monitoring of weather, providing seasonal to interannual prediction of weather with respect to climatological norms, understanding the vulnerability of population, and monitoring through sentinel sites the actual incidence of the disease, to provide effective early warning of potential malaria epidemics (WHO 2001).

13.4.3 Establishing and Furthering Operational Climate Services for Health[11]

Climate directly impacts food and fibre production, and the epidemiology of infectious diseases. Severe or repeated climate shocks can push vulnerable households into a persistent poverty trap when their individual coping responses involve divestment of productive assets, such as livestock or land. Without advanced warning, societal safety nets are costly, and difficult to mobilize and target effectively. – IRI 2005

This joint Communicable Disease Surveillance and Response, Protection of the Human Environment, and Roll Back Malaria publication was prepared with the understanding that climate based Early Warning Systems, when fully developed, do have the potential to provide increased lead times in which to implement epidemic prevention and /or control activities. Therefore their development should be encouraged, and both positive and negative experience of using such systems should be documented. – WHO 2004b

For an early warning system that incorporates climate information, we must understand the influences that climate has on human behavior, pathogens, and vectors. Certain human behaviour is strongly influenced by climate variability, and can determine the disease transmission pattern. Seasonal influences can change the balance of immunity or resistance. For example, with seasonal influenza in Europe, people tend to spend more time indoors in the winter, resulting in increasing exposures and contributing to the peak time for an outbreak of influenza. Similarly, gastroenteritis in developed countries can be associated with non-climatic factors, as people tend be outside more in the warmer weather, cooking and eating outside and sometimes not cooking food thoroughly, or putting cooked food onto the plates that had been used for the raw food, thereby introducing pathogens.

If one is interested in developing a climate early warning system for a disease, there are some confounders that must be addressed: population vulnerability, including the likelihood that a vector will find someone within the population that is infectable, and malnutrition (the immune system is depressed). In the Kenyan western highlands, epidemics may only occur when the population has become largely non-immune. For example, if it has been a while since the last epidemic, a high proportion of the young may not have been exposed.

Dynamics of the pathogen can be highly sensitive to climate, especially those pathogens that are born outside the final host. Thresholds of temperature frequently determine the viability of juvenile stages. But pathogen dynamics can also be confounders, such as in malaria where some pathogens are becoming resistant

[11] The bulk of this section draws from the report, Using Climate To Predict Disease Outbreaks: A Review (WHO 2004b).

to control and treatment drugs. If people are becoming infected at a higher rate, it may be due to climatic causes but it may also be due to changes in vulnerability. Climate factors such as temperature, rainfall and humidity strongly influence many of the patterns of geographical distribution and development of some disease vectors. However, the contribution of population movement and agricultural practices (e.g. deforestation and irrigation schemes) were shown to be very important in the discussion on onchocerciasis and Loa loa (Section 13.3.3): not only were people exposed to a higher risk of infection due to what was already in the savannah area, but the pathogen in the forest was converting to being a savannah type of pathogen.

It is important to rationalize the contributions of climate variability and of confounders in the variance of the aspect of disease that is the goal of an early warning system, to determine if it adds sufficient value and is practical to implement. The early warning system needs to look not only at the question of will there be an epidemic, but when an epidemic will occur. And it is important to know which of the factors will give the key information.

In developing a climate based early warning system, some perceptions may form obstacles and should be anticipated and addressed. That climate factors form only part of the set of the determinants can be reason enough for some to dismiss their value. The recipients may respect the climate factors, but not act on the early warning for other reasons. There may be insufficient interest at high levels in preventing a crisis. And, "In most cases purpose of early warning is undermined, because relief arrives too late due to poor organization at the donor level" (WHO 2004b).

What is the phase in an early warning system that is most crucial in determining its success? It is the phase following the issuance of an early warning. "Early warnings are of little use if the capacity to respond is not present" (WHO 2004b). The system must be budgeted for, the resources must be there, and all multidisciplinary collaboration must already have taken place and been successful. A preparedness plan and the organisation to apply it must exist, and each responding organisation must know its tasks, understand how their tasks fit in a well-integrated response, and have the will and commitment to implement their specific tasks.

Perhaps the most important phase of the development of an early warning system is the one that precedes it: the process of identifying the principal disease or diseases of most concern and interest, and of securing funding for the activity. The goals of an early warning system should be set in close collaboration among the climate scientists and the public health community.

13.4.3.1 Framework for Developing a Climate Based Early Warning System

From its study of existing early warning systems, the WHO determined that the framework consists of four preliminary phases, evaluate the epidemiological potential, the early warning system itself, and the response and evaluation.

The first of the preliminary phases is to evaluate the epidemiological potential: defining what constitutes an epidemic, and looking at how the disease progresses – the progression of pathogen and the vector, and human behaviour. Next, identification of the geographical location of epidemic areas is conducted, using such means as remote sensing and sentinel sites. Climatic and non-climatic disease risk factors are determined next. Then, the link between climate variability and climate predictors is quantified, and the models are constructed, taking care to account for confounders. Biological models and models of the processes in the pathogen, vector or host are incorporated to the extent that they explain how the disease aspects are affected by incremental increases or decreases in various climatic factors.

The early warning phase depends on disease surveillance, which differs from prediction in that surveillance is expressly concerned with detecting the incidence of the disease. The early warning system also incorporates the monitoring of the disease risk factors – what is the population doing, what is happening with drug resistance, what is the state and trend of the climatic factors? Monitoring requires datasets from earth observation and meteorological satellites, as well as traditional surface and upper air instruments. The last step comprises preparing model forecasts and running the health forecast model in an operational way at appropriate times.

The response phase concerns the treatment and control activities, and is tailored to the geographical and disease characteristics. It involves the treatment of people or control of vectors, and also informs the relief or containment activities, which are mostly the role of governments. The response should follow a predetermined preparedness and response plan, developed by a multi-disciplinary team – an integration of climatologists, operational meteorologists, researchers in the science of prediction, climate analysts with expertise in remote sensing, technical specialists to advise on the layers of information and analysis through GIS; and experts in epidemiology, treatment and control, population movement, land use, and policy makers. Funding may require extensive international involvement, so some representation of international funding experts may be helpful at some stages in the team.

The evaluation phase is open-ended, starting with an initial assessment and continuing with periodic assessments. Through the use of a questionnaire, interviews, or other procedures, it seeks answers to questions such as: is the early warning easy to use, are the predictions accurate, is the process cost effective, what is the best way to spend the resources for this problem? The developers and operators have to collaborate with users/stakeholders to get the answers – they are the ultimate experts in applying the control, or working with the treatments in clinics.

Securing the funding for collaborative work is complex: who takes the lead in initiating a project like this – is it the government, the climate community, control people, epidemiologists? That depends on whether it is a short-term project, or a sustainable activity, which has implications for the source of the resources. In many cases, the sustainability of such projects has depended greatly on the continuation of donor funding.

What would be the responses that are expected once an early warning is issued? That would be specific to what the early warning was designed to do. In most cases an early warning system provides input information into work already being done, rather than mandating unique actions. The likely result is that the kinds of work necessary to reduce the impact of an epidemic will have a better likelihood of succeeding. And, it will contribute to the reduction of death and human suffering.

Part V
The Future of Seasonal Climate Forecasting

Chapter 14
A Way Forward for Seasonal Climate Services

Mike Harrison, Alberto Troccoli, David L.T. Anderson, Simon J. Mason, Michael Coughlan, and Jim B. Williams

The enthusiasm for engaging the challenges of Seasonal to Interannual Prediction, both within the disciplines of physical and social sciences and at their interface, was well demonstrated through the energetic engagement of all during the May to June 2005 NATO ASI course, upon which this book is based. Several panel sessions were held during the course, which permitted everyone to offer views within an informal setting; some, not reflected in the main body of the book, are incorporated in this chapter. Little stays stationary in such a fast-developing field, and so, to provide the most advanced position at publication, this summarising chapter has included some of the latest development to supplement the material drawn from the course presentations and the panel discussions. Additionally, a view to the future is offered so as to provide further stimulation to those interested in the fascinating field of Seasonal to Interannual Climate.

14.1 The Science

In many communities throughout the world, economic, social and environmental development is rather more directly dependent on seasonal climate and its inter-annual variability than on day-to-day weather events, disasters apart. Yet, within

Mike Harrison
Independent Consultant

Alberto Troccoli and David L.T. Anderson
European Centre for Medium-Range Weather Forecasts

Simon J. Mason
International Research Institute for Climate and Society

Michael Coughlan
Australian Bureau of Meteorology

Jim B. Williams
Consultant with the NRGroup

A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk.*

the meteorological community much greater effort and expense goes into producing the daily weather forecast. At the same time, many communities are left to cope with the vagaries of climate with little effort being directed at providing the best-available climate information. Emphases are beginning to change, however, consequent on growing concerns over climate change and the recent advent of seasonal forecasting with its promising future developments, including increasing levels of skill. Seasonal forecasting not only makes climatology a 'living and very practical science', but also provides a most useful context for considering and valuing the daily weather forecast. It also provides a practical first step for coping with climate change: an inability to cope with climate variability as it is, does not augur well for how society might cope with the variability of some future climatic regime.

Two interrelated major scientific developments have made the progress in practical climatology possible: a vastly enriched knowledge of the physics of climate variability, achieved through enhanced global scale observing systems combined with the development of analytical and interpretative techniques; and the creation of computer-based models of the climate system – simple through to complex – that use observations to generate forecasts on timescales of a season or longer.

14.1.1 Understanding Climate Variability

Central to the enhanced understanding of climate variability lies the ever increasing knowledge of how the oceans and the atmosphere interact. In simple terms, the ocean provides a 'long-term memory' for the atmosphere while in turn the atmosphere helps drive the slow variations in the ocean. Furthermore, the 'memory' imposed by the ocean is distributed rapidly by the atmosphere, teleconnected to distant parts of the globe. The ENSO phenomenon, with its opposing maxima of El Niño and La Niña, is the strongest known modulator of climate variability on the global scale (other than the annual cycle of the sun) and provides predictability. Despite the recurrence of an El Niño phase every 2–7 years, ENSO is not an oscillatory phenomenon as such: an El Niño is not necessarily followed by a La Niña. The reason for the non-periodicity is not yet understood but several theories have been put forward, all revolving around the hypothesis that ENSO can be approximated by oscillatory systems. These theories can be divided into two main categories: (i) ENSO is a self-sustained oscillator; (ii) ENSO is a damped oscillator. In (i), the oscillator possesses a natural frequency which is perturbed by chaotic processes (weather) to be irregular, whereas in (ii), the oscillator requires some external forcing to keep the system going. The role of non-linearity and noise is markedly different in each case. Despite the attempts to provide unified theories for ENSO, the cause of the irregularity is still an open research topic.

One of the critical factors in these theories is the understanding of how ENSO events are initiated. Once an ENSO event has started, models – and therefore theories – do a reasonably good job at forecasting the subsequent evolution of the event, with lead-times up to several months. An atmospheric phenomenon called the Madden-Julian Oscillation (MJO), an intraseasonal oscillation of about 40–60 days, likely plays a major role in the initiation process and is currently the leading candidate under investigation. There is little doubt that weather can influence the evolution of ENSO events: the strengthening of the link between the weather and the seasonal to interannual climate communities is likely to be a fruitful path for research and future progress at both timescales.

Within this book there has been a focus on ENSO and its effects. ENSO is not alone, however, in forcing interannual seasonal variability. Both the tropical Indian and the Atlantic Ocean basins host processes related to rainfall variations in parts of surrounding continental masses, while the evidence for pertinent roles for extra-tropical oceans is also growing. None appears to exert the global-scale influences of the Pacific centred ENSO but nevertheless their effects are undoubtedly critical in some regions, and further understanding will lead almost certainly to improved predictions for these areas.

Not all seasonal variability is attributable to atmosphere-ocean interactions, and evidence is mounting that other sources of predictability exist. These sources include amounts of soil moisture across the continental masses, the distributions of continental snow and polar ice, atmospheric aerosol distributions, and even stratosphere/troposphere interactions.

There will remain always some seasonal variability not attributable to the phenomena discussed above; some climatologists might place this unexplained portion of the variability, which varies according to region, in the bin labeled ‘noise’, stating that it is not predictable – at least, with current models. Hopefully research breakthroughs, currently unforeseen, will prise some of the ‘noise’ from that bin and place it into a category bearing promise of improved predictions, thereby improving the predictability of the climate system.

14.1.2 Models

There are two basic approaches to seasonal to interannual predictions – statistical and dynamical – and both have progressed substantially over the past decades.

Dynamical models of the climate system are appealing tools for learning about the climate and for attempting to predict it. The appeal stems from the fact that these models are based on physical principles, as expressed by their mathematical formulation. Moreover, since model resolution may be modified with minimal effort and also the level of complexity can be usually adjusted in a modular way (e.g. a model of sea ice can be included or not depending if one is interested in high latitude phenomena), such models can be all the more attractive, and versatile.

The drawback is the large cost, both in financial terms (including computational, in a proportional way to resolution and complexity) and human resource terms and, therefore, their development is normally the prerogative of major research centres. These models have allowed considerable improvement in the quality of operational seasonal forecasts and will most likely contribute to further improvements in the future. However, even the most sophisticated model only gives an approximate representation of the very complex climate system. For example, resolution of current operational seasonal forecast models is not normally sufficient to resolve important phenomena such as the MJO in a satisfactory way. More crucially, however, atmospheric convection in the tropics, upon which also the MJO is dependent, is a weak point of most, if not all, models. Thus, the full potential of numerical models is far from being achieved and a vigorous model development phase is still underway through: improvement in the representation of critical physical processes such as atmospheric convection and oceanic coastal upwelling; increase in the level of model complexity; increase in model resolution; representation of model uncertainties via the use of multi-models and/or the implementation of stochastic processes.

Although dynamical models have numerous advantages over statistical models, and offer greater prospects for long-term improvements in performance, statistical models remain in wide use, and are likely to do so for years to come. Much of the popularity of statistical models comes from practical considerations, such as their minimal demands on computational resources and their relative simplicity. For these reasons statistical models are used extensively in developing countries. However, even in countries such as the United States, statistical models constitute an important input to the mix of operational forecasting systems for the simple reason that they continue to outperform dynamical model forecasts in some instances. With the establishment in November 2006 of Global Producing Centres, dynamical model predictions are being made increasingly available to forecasting centres that do not have the resources to run their own models, and so an upsurge in the application of statistical models to downscale and recalibrate dynamical model predictions is beginning. This process should enable most countries to take advantage of both approaches.

The output of dynamical and statistical models is increasingly used in a variety of decision making frameworks. Model enhancements, as well as increases in dynamical model resolution, will advance the science of seasonal forecasting in the long term. Additional practical benefit will be gained from greater flexibility in the interpretation of, and enhanced information supplied by, all types of models. For example, more detailed predictions in both the spatial and temporal senses, is the *sine qua non* for all who prepare and use seasonal forecasts, with information regarding the start of rains in seasonal regimes being prominent amongst the latter. Modellers so far have given limited attention to these problems, mainly on the grounds that these details fall into the area of unpredictable noise, but also because of resource limitations. Nevertheless, there are encouraging signs that advances will be made, not only through higher resolution models, but also through post-processing of the global model forecasts, either through empirical means or by embedding regional climate models capable, in principle, of providing information at higher spatial and temporal scales. It is likely that development and beneficial use of this information will need close coordination between providers and users.

14.1.3 Assessment of the Skill of the Models

Because seasonal forecasts are expressed as probabilities, they cannot be assessed as ‘right’ or ‘wrong’ in any simplistic way. While it is possible to assess the accuracy of deterministic forecasts, for probabilistic forecasts other attributes such as reliability and resolution are more appropriate. None of these attributes can be communicated in a single score, although regrettably there continues to be undue reliance on scores without recognition of the limited information that such scores can communicate. There are detailed diagnostic techniques that have been devised to provide comprehensive assessments of the quality of the forecasts, but a major limitation is that large sample sizes are generally required. Sample sizes of seasonal forecasts are small compared to those for weather forecasts, for example, and so robust estimates of forecast quality are lacking. Very few assessments of the quality of seasonal forecasts have been performed for the simple reason that these forecasts (and hindcasts) have been produced only since the early 1990s, providing sample sizes of only about 15 years.

Partly to address the problem of limited sample size, considerable effort has been invested in generating hindcasts, and projects such as DEMETER have been invaluable in obtaining realistic estimates of operational performance. Perhaps the main conclusion from these forecasts/hindcasts is that although seasonal forecasts of parameters such as the NIÑO3.4 index or of 200 hPa heights can be predicted with impressive skill, parameters of more direct interest to potential users of such forecasts, such as near-surface air temperatures and precipitation, are much harder to predict. Nevertheless there is considerable information content in the forecasts of temperature and precipitation at certain times of year and for some areas, mostly within the tropics and sub-tropics. Temperature forecasts are notably better than precipitation, although recent attempts to focus on the frequency of precipitation, rather than on total precipitation, are yielding promising results.

14.1.4 Conversion of Model Forecasts into Useable Form

Despite the major investment required to generate a prediction from a dynamical model (whether a coupled ocean-atmosphere model, or an atmosphere-only model), generating the model output is only one step in an involved process for generating a seasonal forecast. Dynamical models are far from perfect, and the differences between model and observed climates can result in substantial errors in forecasts. Because these errors are systematic, they can usually be removed using relatively simple statistical procedures. However, best results are obtained when systematic spatial displacements of the model's climate features are considered, and these spatial corrections require more advanced procedures.

Even without systematic errors, some form of post-processing of the model output is usually required to make forecasts relevant at spatial scales and locations of interest to specific users. Predictions straight out of a dynamical model are generally representative of large spatial averages, and "downscaling" procedures are required to translate the resolution of the forecasts to a more practically useful scale. Downscaling procedures have also been developed to provide statistics about the intraseasonal characteristics of weather. The dynamical models are useful for providing predictions of seasonally averaged conditions, but their current representations of weather variability are insufficiently realistic to be used directly. Downscaling (both temporal and spatial) can be performed using either statistical techniques or with limited-area, high-resolution dynamical models. The latter are expensive to run, and despite some promising results, there have still been no clear indications that they can outperform the statistical procedures.

Regardless of the temporal and spatial resolution of the forecasts desired, there is overwhelming evidence that the best forecast can be made by considering outputs from a suite of models. This multi-model approach can be justified on the basis of improved representation of the uncertainties in the forecast arising from imperfections in the models. Multi-model approaches are effective whether the individual models are dynamical, statistical, or a combination of both. However, there is still some debate about the best ways to combine the predictions from the different models. It seems intuitively appealing to weight the better models more highly than the ones with less skill, but in practice the limited sample sizes available make it virtually impossible to estimate differences in the skill of models robustly. As a result, simple averaging of the predictions from different models remains a very competitive procedure.

14.2 Communication and Integration

Developments in the underlying science need to be matched by improvements in the way climate information is communicated and integrated into societal structures. The current focus of seasonal to interannual prediction research is on the development of forecast systems, and particularly on dynamical models. Coupled models, perhaps the most promising long-term solution for predictions at the global scale in the maximum possible detail and with the highest quality, are complex and expensive to develop and maintain, as are the observing networks required to support them. Rightly, there is continuing investment in these models. But commensurate investment is required also in all downstream aspects, including delivery systems, interpretation and decision making approaches, management and mainstreaming, which ultimately determine the levels of societal benefit achieved. Benefit, or value, is not achieved through forecast quality alone. More consideration is required in building the case for increasing the funding of research in these latter areas.

Technology for information delivery is in reasonably good shape. Modern communications systems, satellites, the Internet and mobile phones offer the opportunity for rapid dissemination of information of all forms on the global scale. Even in the more remote areas where most advanced communications systems have not as yet penetrated, there are developments such as the RANET[1] project providing viable options that are progressively extending into more geographical areas. There is no technological reason in principle why, within the foreseeable future, even the most remote user might not have rapid access to some form of past, current and future climate variability information, should they so wish. The physical difficulties of delivery are being surmounted.

Delivery, however, is not simply a matter of providing the necessary technical facilities, but covers also the information delivered, its content and the manner in which it is presented. Communication with users is a key component of the forecasting system and particular focus should be devoted to this aspect in the near future. People often tend to think and view life in more or less deterministic terms (B is a result of A, or A causes B) but coping with unavoidable uncertainty or 'risk' demands more complex thought processes and a greater degree of prudence in order to cope with the possibility of making what might be seen in hindsight as a 'wrong' decision. Major efforts are required, possibly using familiar instances of probabilistic forecasting (as betting on horse racing and other uncertainties) to find ways of raising awareness in people towards managing seasonal risk, so that they can make best use of seasonal forecasts.

Ideally, the entire delivery system should be designed to assist decision processes, however individual. Here advances have also been made in recent years. An example is the approach adopted by APSRU[2] in delivering information in a utilisable manner, information that not only covers past and future climate variability but also, within the Australian farming context, multiple related information streams presented in a form that assists decision making. The Australian Bureau of Meteorology has also led in the production of user-friendly web sites.[3]

Whereas the APSRU approach is predominantly a service delivered through the Internet, face to face communication was pioneered in the Regional Climate Outlook Forums (RCOFs). Since their initiation during 1997, these Forums have

---

[1] RAdio and InterNET for the Communication of Hydro-Meteorological and Climate Related Information. See: http://www.ranetproject.net/

[2] Agricultural Production Systems Research Unit. See: http://www.apsru.gov.au/apsru/

[3] See: http://www.bom.gov.au/silo/

continued in many parts of the developing world.[4] And they continue to flourish with the first such event for continental Asia having been held in April 2005 in Beijing. The Forums, in their original form, are expensive to run, and in some regions there are concerns over sustainability. Regardless of sustainability, RCOFs have most certainly provided a stimulus to the introduction of climate services, including seasonal predictions, in many parts of the world. The nascent Regional Climate Centres, a project of WMO, will likely support future communication with stakeholders.

Finally, numerous pilot projects have probed the difficulties and the value obtainable from climate services. Within the current volume are examples of leadership by the National Meteorological Service of Morocco, and by Florida State University. There are other examples from Australia, Africa, the USA and the South Pacific. Various projects run under the banner of climate change, such as some within AIACC,[5] in practice tend to address climate variability rather than climate change *per se*.

Hence there has been a wealth of activity over recent years to promote the dissemination and uptake of the forecasts. But, despite this progress, there are still few clear demonstrations of consistently achievable value obtainable through the use of the developing prediction technology. What has been progressively recognised over the last few years are the outstanding issues of delivering services, within all of the technological and cultural contexts that that entails, as discussed below.

As has been argued in numerous places in this book, central to achieving value is both the decision process itself and the delivery of information appropriate to each decision. The decision process is the pivot around which information and knowledge are converted into value. The decision process is a growing area of research within the context of seasonal forecasts and requires substantially further attention. Improved understanding of decision processes, especially those processes that lie at the nexus of multiple information streams (such as those with economic, environmental and social components) will provide substantial benefits in designing climate information to achieve optimal benefit. A start was made in 2006 at the WMO International Conference on "Living with Climate Variability and Change".[6] More is needed, however.

---

[4] See: http://www.wmo.int/pages/prog/wcp/wcasp/clips/outlooks/climate_forecasts.html

[5] Assessments of Impacts and Adaptation to Climate Change. See: http://www.start.org/project_pages/aiacc.html

[6] See: http://www.livingwithclimate.fi

14.3 Getting There

It is essential that the objectives for any service delivery intended to provide value are recognised and incorporated as that service is created. According to one leading institute that works to build bridges between climate scientists and climate information stakeholders, the IRI,[7] the prerequisites for future services built on seasonal to interannual predictions include:

- Recognising stakeholders' needs, both real and perceived
- Identifying viable decision options that are sensitive to climate variability and to forecast content
- (in reflection of the preceding) Focussing on those aspects of stakeholders' activities with viable decision options
- Building effective and appropriate communication
- Generating sustained support by institutions and favourable policies (including at government level)

This is a valuable opening list of essential prerequisites; each is fundamental and each raises its own challenges. However it could be argued that "effective and appropriate communication" is the most fundamental aspect of all. Communication is necessary between all involved, throughout the forecaster to decision maker chain, to recognise stakeholder needs, to identify viable decision options, and to develop and deliver forecasts that address viable decision options. Communication is necessary also to build institutional commitment and to introduce the conditions suitable for the creation of favourable policies, including government policies, where they do not exist. To date, it is debateable whether there has been effective universal broadcasting of the benefits potentially but realistically available from the forecasts, and equally whether there has been effective communication between providers and stakeholders in all contexts. It is certainly debateable whether many products currently available freely through the Internet provide the level of communication of climate information in all regards that is necessary.

It is in regards to communication that the most significant advances may be made in the next few years. Compared with the steady evolution expected in forecasting systems, short-term benefits are readily available through improved communication and decision making. Most certainly there will be advances in our understanding of climate processes, our ability to observe the environment in numerous regards, and our ability to model and predict the environment, and benefits from these will unquestionably reach those attempting to manage under climate variability. But the most tangible stakeholder benefits are most likely to originate first in the delivery of information through protocols more amenable to

[7] http://iri.columbia.edu/outreach/publication/report/06-01/report06-01.pdf

stakeholder understanding and to incorporation into decision processes than is often the case today.

One of the main difficulties at present is in converting climate information directly into information of assistance in decision making in terms of agricultural production, malaria incidence, water resource levels, and so on. And few, if any, of these applications possess linear transfer functions from climate information to the application-related response. Approaches have been tested whereby model output is fed subsequently into sectorial models, such as for crop prediction or dam management. These approaches might be referred to as 'two-tier sectorial' models whereby coupled models provide the climate input, or 'three-tier sectorial' models when separate ocean and atmosphere components are used. Increasingly comprehensive 'single-tier sectoral' climate models are being developed, currently more specifically in regard to climate change, where such additional factors are being incorporated directly with the atmospheric model itself. This approach ensures climate feedbacks are simulated dynamically and guarantees overall consistency of predictions. These more comprehensive models most likely will find a role in ensemble seasonal prediction, and will later be extended to embedded regional climate models. Inevitably adoption of this approach will raise new issues regarding the validation of these extended models within the framework of interannual climate variability, and introduce new contexts of predictability, and challenges with linking to decision processes.

Regarding the interface between producers and stakeholders, one issue is that only a relatively small number of decision processes are neatly aligned to the timescales and lead-times associated with current prediction technology. There is an argument to adjust technical development away from the focus on improvement of predictions within the known window of predictability, towards development of information requirements dependent upon timescales appropriate to the decision processes themselves. That approach would require more imaginative use of climate information (which at times is neglected in the rush to use forecasts), more creative interpretation of the predictions themselves, and incorporation of other pertinent non-climate information, in order to provide a focussed complex designed to facilitate individual decision processes. Such an approach would focus all available information onto the specific requirements of individual decisions. The importance of decision processes is being recognised increasingly, with conferences such as "Living with Climate Variability and Change" mentioned in the previous section. One object of this conference was to transfer some research focus from 'skill/quality' to 'decision processes'. This and similar conferences may well assist in guiding the design of future research and operational programmes.

Regarding melding producer and stakeholder perspectives, the approach of developing full information and decision making packages for specific applications adopted by, for example, the IRI and APSRU is likely to begin to replace the original end-to-end concept that was taken at the outset, not least by CLIPS and the IRI, and that is still predominantly in use. That is not to say that end-to-end processes are inappropriate on all occasions – business uses are one example where the end-to-end model may in general be best (notwithstanding the fact the coordinated and cooperative decision making can lead frequently to optimal outcomes, even in the business world). The new comprehensive approach should be encouraged, but it retains nevertheless the disadvantage of producing solutions that tend to be culturally, sectorally and geographically specific.

Scientific advances will only produce benefit provided there is conversion of new information into value. A more coordinated approach would be beneficial in achieving this in regard to seasonal to interannual prediction, and the engagement of some form of international process would stimulate the coordination needed in all regards of defining research needs across the board, incorporating stakeholders at all levels, mainstreaming into policy, and delivering improved decision processes.

14.4 Goodbye Cinderella, Hello "Seamless Future"

For many years climate was the preserve of geographers, statisticians and historians. The emergence of meteorology during the mid-20th century as a 'hard', scientific discipline steeped in the mathematics of thermodynamics and fluid dynamics saw climate take on an almost 'Cinderella' role. Climatologists were more or less relegated to the task of archiving the data gathered for the sole purpose of predicting tomorrow's weather. During the latter part of the century that retreat to the background begun to reverse, with the mass of data collected beginning to reveal that climate, hitherto thought of as static except over very long periods, was anything but static. The 30 years needed to 'define' the climate of a locality did little more than define the climate of that 30 year period, with the climate of successive 30 year periods differing, and often markedly so. Hence the notion of climate variability was born.

As knowledge of the causes of climate variability grew, led by the rush to understand El Niño, then so too did the capacity to model the climate system. By coupling components replicating processes in the oceans and over the land surface to what was already being modelled in the atmosphere, one could begin to model the whole climate system, and indeed that broadening path continues with notions of 'earth system' modelling.

The imperative to understand the human imprint on climate began to rise around the same time, and indeed for a period has subsumed to a large extent the importance of modelling and predicting 'natural' climate variability. In reality the distinction is somewhat artificial since any effort to predict the climate of the future must perforce take into account all processes in play, both human and natural, to the extent that they are significant on the timescale of the prediction. Thus climate variability is now a prime area in which to develop a career, and its importance to society and the contribution that its science can make are now mostly recognised.

Yet the study of climate variability has not yet fully shrugged off its ‘Cinderella’ image, as the quality of seasonal predictions, the putative rationale for research in the area, are perceived as failing to live up to the expectations set by short range weather forecasts. As we have seen throughout this book, for reasons that have as much to do with understanding human perceptions as with understanding the fundamental physical science, bridging the weather-climate predictability gap is not a trivial task.

Nonetheless, the notion of a ‘seamless’ forecasting system with lead times from minutes to centuries is an attractive concept and possibly one that will eventuate as a reality in time. Already we are beginning to see ‘unified’ models that are capable of being run in various timescale modes and spatial resolutions, along with the application of ensemble predictions and multi-model schemes to the challenges of weather forecasting, seasonal prediction, and climate projections.

In reality, however, the journey has barely started with decision making still mostly compartmentalised on the supply side by practical distinctions between weather and climate forecasting activities, and on the demand side by a host of factors that have little to do with weather and or climate. Seamless forecasting systems promise as yet little information that is not already available in the separate formats. So the paradigm remains unfulfilled, viz. that of a ‘seamless forecasting system’ linked into a ‘seamless decision making system’, with clear challenges remaining for both sides of the divide.

Undoubtedly, a major driver for progress lies in what the science of climate variability offers by way of an opportunity for learning to adapt to climate change, with the seasonal forecast models providing a basis for validating climate change models, as well as offering a bridge to weather forecasting models. Adaptation and modelling together, in both seamless decision making and forecasting contexts still seems a logical path forward.

Acronyms

Common acronyms used throughout the book are defined here. Additional details on asterisked terms are provided in the Glossary of Terms.

3D-Var	Three-dimensional Variational (Data Assimilation)
4D-Var	Four-dimensional Variational (Data Assimilation)
AGCM*	Atmospheric Global Circulation Model
ACMAD	African Centre of Meteorological Applications for Development
AGRHYMET	Centre Regional de Formation et d'Application en Agrométéorologie et Hydrologie Opérationnelle
AIACC	Assessments of Impacts and Adaptation to Climate Change
AMIP	Atmospheric Model Intercomparison Project
APSIM	Agricultural Production System Simulator
APSRU	Agricultural Production Systems Research Unit
AR	Autoregressive
ARSCO	American Association of State Climatologists Recognized State Climate Office
AU	African Union
BMA	Basin Management Agency
BMRC	Bureau of Meteorology Research Centre
BoM	Bureau of Meteorology
CC*	Climate Change
CCA	Canonical Correlation Analysis
CDF*	Cumulative Distribution Function
CEP	Conditional Exceedance Probability
CGCM*	Coupled Global Circulation Model
CIMMS	Cooperative Institute for Mesoscale Meteorological Studies
CLIMAG	Climate Prediction and Agriculture
CLIPER	CLImatology and PERsistence
CLIPS	Climate Information and Prediction Services
CNRM	Centre National de Recherches Météorologiques
COPES	Coordinated Observation and Prediction of the Earth System

CV*	Climate Variability
DA*	Data Assimilation
DEMETER	Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction
DEWA	Department of Early Warning and Assessment
DJF	December, January, February
DMC	Drought Monitoring Centre
DMN	Direction de la Météorologie Nationale
DNA	Deoxyribonucleic acid
DSP	Dynamical Seasonal Prediction
ECHAM	ECWMF-Max Plank Institute Hamburg
ECMWF	European Centre for Medium-Range Weather Forecasts
EDS	Early Detection System
ENACT	ENhanced ocean data Assimilation and ClimaTe Prediction
EnKF	Ensemble Kalman Filter
ENSO*	El Niño/Southern Oscillation
EOF*	Empirical Orthogonal Function
ERA-15	ECMWF Re-analyses (for the 15-year period 1979-1993)
ERA-40	ECMWF Re-analyses (for the 40-year + period 1958-2001)
ERS	European Remote Sensing
ESSP	Earth System Science Partnership
EUC*	Equatorial Under Current
EUROBRISA	EURO-BRazilian Initiative for improving SOuth American seasonal forecasts
EUROSIP	EUROpean multi-model Seasonal to Inter-annual Prediction
EWS	Early Warning System
FAO	Food and Agriculture Organization
FAWN	Florida Automated Weather Network
FEWS NET	Famine Early Warning Systems Network
FLC	Florida Climate Consortium
FMAM	February, March, April, May
GAEMN	Georgia Environmental Monitoring Network
GCM*	Global Circulation Model
GCOS	Global Climate Observing System
GDP	Gross Domestic Product
GERB	Geostationary Earth Radiation Budget
GHRSST	GODAE High Resolution Sea Surface Temperature

GIEWS	Global Information and Early Warning System (FAO)
Gl	Gigalitres (10^9 litres)
GLM	Generalised Linear Model
GMT	Greenwich Mean Time
GODAE	Global Ocean Data Assimilation Experiment
GPC	Global Producing Centre
Ha	Hectare (10,000 square metres)
HEWS	Humanitarian Early Warning System
hPa	hectoPascal (a measure of pressure)
IASC	Inter-Agency Standing Committee
IBM	International Business Machines
ICPAC	IGAD Climate Prediction and Applications Centre
ICSU	International Council for Science
ICTZ*	Inter Tropical Convergence Zone
IGAD	Intergovernmental Authority on Development
IGBP	International Geosphere-Biosphere Programme
IHDP	International Human Dimensions Programme
iid	Identically and Independently Distributed
IMD	Indian Meteorological Department
IOC	Intergovernmental Oceanographic Commission
IPCC	Inter-Governmental Panel on Climate Change
IQQM	Integrated Quantity Quality Model (hydrological)
IRI	International Research Institute for Climate and Society
JAS	July, August, September
JFM	January, February, March
JJA	June, July, August
JMA	Japanese Meteorological Agency
KBDI	Keetch-Byram Drought Index
KF	Kalman Filter
LAD	Least absolute deviation
LEPS	Linear error in probability space
LPM	Linear Programming Model
MAPE	Maximum A Posteriori Estimate
MCA	Maximum Covariance Analysis
MDG*	Millennium Development Goal
MEWS	Malaria Early Warning System

MJJ	May, June, July
MLE	Maximum Likelihood Estimate
MOS	Model Output Statistics
MoU	Memorandum of Understanding
NAO*	North Atlantic Oscillation
NCEP	National Centers for Environmental Prediction
NDJ	November, December, January
NGO	Non-Governmental Organisation
NMHS	National Meteorological and Hydrological Service
NMS	National Meteorological Service
NOAA	National Oceanic and Atmospheric Administration
NCDC	National Climate Data Center
NSWDLWC	New South Wales Department of Land and Water Conservation (Australia)
NWP	Numerical Weather Prediction
OCN	Optimal Climate Normal
ODA	Official Development Assistance
OGCM*	Oceanic (or Ocean) Global Circulation Model
OGP	Office of Global Programs
OI	Optimal (or Optimum) Interpolation
OLS	Ordinary Least Squares
OND	October, November, December
ORMVA	Offices Régionaux de Mise en Valeur Agricole
PCR	Principal Components Regression
PDF (or pdf)*	Probability Density (or Distribution) Function
PNA	Pacific–North America
POP	Principal Oscillation Pattern
PROVOST	PRediction Of climate Variations On Seasonal to interannual Time-scales
QDPI	Queensland Department of Primary Industries
RANET	RAdio and InterNET for the Communication of Hydro-Meteorological and Climate Related Information
RCC	Regional Climate Centre
RCOF	Regional Climate Outlook Forum
RISA	Regional Integrated Sciences and Assessment
RMS (rms)	Root-Mean-Square
RMSE (rmse)	Root-Mean-Square Error
SADC	Southern African Development Community

SCOPIC	Seasonal Climate Outlook for the Pacific Island Countries
SECC	Southeast Climate Consortium
SINERGEE	Simulations from a Numerical weather prediction model to Exploit Radiation data from a new Geostationary satellite, Explore radiative processes and Evaluate models.
SIP	Seasonal to Interannual Prediction
SO*	Southern Oscillation
SOI*	Southern Oscillation Index
SSA	Singular Spectrum Analysis
SSH	Sea Surface Height
SST	Sea Surface Temperature
SSTA	Sea Surface Temperature Anomaly
START	SysTem for Analysis, Research and Training
SVD	Singular Value Decomposition
TAO*	Tropical Atmosphere Ocean
TOGA	Tropical Ocean Global Atmosphere
UN	United Nations
UNCED	United Nations Conference on Environmental Development (1992)
UNEP	United Nations Environment Programme
UNFCCC*	United Nations Framework Convention on Climate Change
USDA	United States Department of Agriculture
USDA-CSREES	USDA Cooperative State Research, Education, and Extension Service
USDA-RMA	USDA Risk Management Agency
UTC	Coordinated Universal Time
WB	World Bank
WCP	World Climate Programme
WCASP	World Climate Applications and Services Programme
WCDMP	World Climate Data and Monitoring Programme
WCIRP	World Climate Impact Assessment and Response Strategies Programme
WCRP	World Climate Research Programme
WFP	World Food Programme
WHO	World Health Organization
WMO*	World Meteorological Organization
XBT	eXpendable BathyThermographs
Z	Zulu: abbreviated form of time equivalent as far as meteorological practice requires to GMT or UTC

Glossary of Terms

Alberto Troccoli, Simon J. Mason, Mike Harrison, and David L.T. Anderson

Abstract: No, this is not the definition of abstract! There is still some confusion about meaning of some terms in the climate field. This "glossary of terms" chapter aims at reducing the level of uncertainty of these key terms. Note that when terms in this glossary are used within the definition of other terms we refer to them by italicising the terms.

Accuracy: In the context of forecast *verification*, the magnitude of the error(s) in a single or a set of forecasts: an "accurate" forecast is one with a small error. The accuracy of a set of forecasts is usually calculated by averaging the error metric over the individual forecasts. In the context of data quality, accuracy refers to the difference between the recorded values and the observed values.

Adaptation: Change to accommodate to new circumstances. In the *UNFCCC* this is change of activities to accommodate to *Climate Change*, but can refer more broadly to any changes required to address *Climate Variability*. Can take many forms, e.g. anticipatory/proactive, autonomous, planned, private, public, reactive, future, baseline. Hence also 'Adaptation Strategy'. 'Adaptation Capacity' denotes the level of ability to adapt.

Analogue (US spelling: Analog): In the context of seasonal forecasting, an "analogue year" is a year in which the season in question is considered similar to the *target* season. The similarity is usually defined in terms of the state of the *predictors*, and the outcome is therefore expected to be similar. A commonly used example of applying analogue years is to use the observed climate during past occurrences of *El Niño* conditions to forecast the climate during a subsequent *El Niño*.

Analysis [Field]: The result of the combination of observations with model data for a specific space-time interval, performed with a *data assimilation* method. In the context of forecasts, it provides the initial conditions for coupled *integrations*.

Anomaly: The difference between an observed value of a meteorological variable (e.g. seasonally averaged temperature) for a single period (e.g. JFM 2000) and its long-term average (e.g. JFM 1961–1990). In the case of seasonally averaged temperature, for example, a positive anomaly occurs when the temperature for the season in question is higher than average, and a negative anomaly occurs when the season is unusually cold.

Anomaly Correlation: The correlation between two sets of data in which distinct mean values are first subtracted from different data points to remove, for example, a seasonal cycle, or a spatially varying *climatology*. These seasonal and spatial effects would otherwise dominate the variances in the data resulting in misleadingly strong correlations.

Application: An activity which makes use of *climate information* such as a seasonal forecast.

Aqua-planet: An idealised configuration used in *Atmospheric Global Circulation Models* (AGCMs) in which the lower boundary is simply represented by a water covered world. This simplified environment is used to investigate atmospheric processes such as the distribution and variability of convection in the tropics and of the storm-tracks in mid-latitudes.

Assimilation Cycle: The sequence of operations necessary to produce an assimilation *analysis*, normally carried out at regular time intervals (typically order of a few hours for the atmosphere and few days for the ocean).

Assistance Strategy: As used for determining aid for developing countries by international aid agencies.

Atmospheric General Circulation Model (AGCM): A *Global Circulation Model* for the atmosphere.

Background (or First Guess) [Field]: A reference model state that is combined with observations to generate an *analysis* [field] using a *data assimilation* method. Background is the best estimate of the system prior to the use of the observations, which might be the direct model output, in which case it is also called *fore cast field* or, more generally, the combination of model output with other (pre-interpolated) data (e.g. a *climatology*). The use of the background field ensures that the *analysis* provides a smooth field from regions with good observation coverage to those with no or sparse observations.

Bias: A measure of how far the average statistic lies from the parameter it is estimating; i.e. the expected error that arises when estimating a quantity. In the context of forecast *verification* and model *validation*, the difference is between the average of the forecasts or simulations and the average of the observed values. Forecasts of rainfall, for example, are positively biased if, on average, they indicate conditions that are wetter than observed.

Biodiversity: The spectrum of life forms, and the width of this spectrum.

Bootstrap: A self-help start-off approach, in want of any standard starting procedure, using whatever information is to hand. Originates from “picking oneself up by the bootstraps”, and is used in many basic start-up contexts (including starting computers after a power down).

Bottom Up: A management approach that examines and resolves issues at the lowest, working levels, then ripples solutions upwards to the management and policy levels (cf. *top down*).

Calibration: The correction of model output for *systematic errors*. Model calibration usually involves a correction only for a *bias* in the mean value, and sometimes for the variance, but more sophisticated procedures can be used (Chapter 8, Section 8.3.3). See also *Recalibration*. Calibration can also refer to the training of a statistical model (Chapter 7, Section 7.3.3.5).

Capacity: The resource (human, technological, environmental, management) to complete activities and/or achieve goals. Hence also Capacity Building.

Chaos: A mode of behaviour of certain non-linear dynamical systems in which the most relevant characteristic is its sensitivity to initial conditions (the “butterfly effect”): small variations in the initial conditions of two dynamic systems in otherwise identical states will lead to a dramatic divergence in the behaviour of the systems over time. As a result of this sensitivity, systems that exhibit chaos often appear to be random. The randomness, however, is only apparent as such systems are indeed deterministic: they are in fact described by well defined mathematical expressions that do not contain random parameters.

Chilling Units: See *Degree-Days*.

Climate: The description in totality on all timescales of the atmosphere, the oceans, the land and the cryosphere within which *weather* sits at the short timescale. Relates also to the manner in which atmospheric and/or oceanic and/or cryospheric processes are experienced, personally, environmentally, or through process outcomes.

Climate Affairs: All aspects of policy and management related to processes to which *climate* is pertinent.

Climate (or Climatic) Change: Any real or perceived adjustment in any aspect of the *climate*. Changes can occur over numerous timescales and through a variety of *forcing* mechanisms. However the definition of Climate Change used by the

UNFCCC covers anthropogenically forced change only, on whatever timescales. Change occurring through natural causes on whatever timescale comes under the heading of *Climate Variability* according to the *UNFCCC*.

Climate Extreme: In the statistical sense, a climate event that sits towards the outer limits of observed distribution, such as temperatures in the coldest 5% or rainfall in the wettest 1% of occurrences. However, often used to indicate an undesirable societal and/or environmental consequence in which *climate* is known or is perceived to have played a dominant role. This latter use tends to ignore both any association with the statistical definition and also any other factor(s) that might have been present, however key.

Climate Forcing: Any external or internal mechanism that determines in part the form of the *climate*. Solar radiation and the earth's rotation represent the two major external factors. Internal factors include distributions of sea surface temperatures, of snow, etc.

Climate Forecasting/Prediction: Official terminology of *WMO* for predictions beyond 2 years; includes *Climate Variability* predictions that cover interannual, decadal and multi-decadal climate *anomalies* and Climate Predictions that cover future *climate* resulting both from natural variability and from anthropogenic causes. Elsewhere often used in the sense of predicting any and all changes in *climate* from all causes over the next century (this is the sense used by the IPCC, whereas the *UNFCCC* uses it for anthropogenically forced change only), and thus is broader than the official *WMO* definition. Normally does not imply seasonal or interannual predictions, for which the terms 'Short-Range Climate Prediction' and 'Seasonal to Interannual Prediction' are used.

Climate Information: Any information on *climate* available to be used in informing decisions. Frequently used in the sole context of historical records of observation data, but can include predictions or any other information produced using models.

Climate Risk Management: The process of mitigating the consequences of possible future, perhaps predicted, climate events on timescales of interest in advance of those events, taking into account as many information streams as feasible.

Climate Services: Any service that provides raw and/or interpreted *Climate Information* to *users*. Normally part of the function of a National Meteorological and Hydrological Service (NMHS), but independent and/or commercial climate services exist.

Climate Shock: The consequences of an unexpected climate event, strictly regardless of timescale but often used in the sense of relatively short timescale events.

Climate Uncertainty: That aspect(s) of future *climate* that cannot be predicted with certainty; often used also to indicate the perception of lack of knowledge of the future.

Climate Variability: Differences in *climate* between any two, or a sequence of, periods; often used in the interannual sense. Variability can occur over any time-scales and through a variety of *forcing* mechanisms. However, the definition of Climate Variability used by the *UNFCCC* covers natural, i.e. not anthropogenically forced, variability on whatever timescales. Variability occurring through anthropogenic causes on whatever timescale comes under the heading of *Climate Change* according to the *UNFCCC*.

Climatology: The description and scientific study of *climate* in all its aspects. Often the term is used to refer to the observed distribution of a meteorological parameter, or set of parameters, over a number of years (typically a 30-year period).

Cognitive Illusions: Perceptual difficulties that enable individuals to draw fully or partially incorrect conclusions while remaining convinced of the certainty of their arguments.

Cold Phase: The *La Niña* stage of *El Niño/Southern Oscillation* (ENSO*)* during which sea surface temperatures over the eastern equatorial Pacific Ocean are below average.

Coping Range/Strategies: The ability to handle *climate variability*, the process(es) of doing so, and the limits of *climate* beyond which it is not possible to manage.

Cost-Loss Model: A simple economic model covering the costs of protecting against an adverse event and the benefits (losses reduced) gained from that protection.

Coupled Model: Any two or more models which work independently but are linked at set time intervals to provide mutual feedbacks. In the case of seasonal to interannual forecasting this normally refers to the coupling of an atmospheric model to an ocean model and possibly to other important component models such as a land model and a sea-ice model.

Coupled General Circulation Model (CGCM): See *Coupled Model* and *General Circulation Model*.

Cumulative [or Probability] Distribution Function (CDF): A function that describes the probability distribution of a continuous or discrete variable by defining the probability that all possible values of the variable will be exceeded.

Data Assimilation: The combination of observations and model data, with the objective to achieve the 'best' description of the system being modelled. The 'best' description is normally the one in which the end result (also called *analysis*) is as close as possible to the observations but with the constraints imposed by the system (e.g. its dynamics), as given by the model data as well as by the statistics employed to relate observations and model data (*error covariance*). Many data assimilation methods are available as described in Chapter 5.

Decile: Each part of a distribution that divides the data into ten equal parts.

Decision Maker/Taker: See *User*.

Decision Process: The methodology through which information is assessed and a decision taken. Numerous factors, including cultural and policy factors, determine each process; thus there is no universally standard decision process in any context.

Degree-Days: A method of assessing the overall effect of temperature across a period. Growing degree-days (or units), for example, are calculated from the multiple of the number of hours the temperature is above a value necessary for growth in a particular plant with the excess temperature above that value; heating degree-days (or units) similarly are for those below a value at which heating is required.

Determinism (deterministic): In the context of predictions, to forecast specific values. A deterministic prediction offers a statement of an expected future with no likelihood attached.

Downscaling: The translation of a forecast from one spatial and/or temporal resolution to a finer resolution. In spatial downscaling, the term is frequently applied to the translation of a forecast from a gridded average to a local point.

Drift (also Climate Drift): The tendency for the solution of a *dynamical (or numerical) model* to move away from the observed state of the system being simulated, due to the presence of *systematic model errors*. Normally the drifting solution eventually reaches a new equilibrium, the so-called model climate.

Drought: A shortfall in water supply for an extended period below a threshold that is appropriate within each specific context (e.g. agriculture).

Dynamical Modelling: The process of simulating a system, such as the atmosphere, by solving, usually on a computer, the basic equations of state (the dynamics and the energy) for that system, in a numerical way.

El Niño: The *warm phase* of *ENSO*.

El Niño/Southern Oscillation (ENSO): A complex system of interaction between the atmosphere and the oceans, specifically across the equatorial Pacific Ocean. The strongest known internal *forcing* mechanism of *climate variability* through atmospheric *teleconnections* to many parts of the globe.

Empirical Modelling: See *Statistical Modelling*.

Empirical Orthogonal Function (EOF): See *Principal Component*.

End-to-End: The mono-directional approach frequently used in the delivery and use of meteorological products. At one end sit the raw data, in the middle are data processing, perhaps forecasting, and processing, while at the other end is the *user* and their *application*.

End User: See *User*.

Ensemble: A set of predictions (each referred to as a *member*) for a specific *target period* designed to test the sensitivity of a forecast to various differences, such as in the type of model, the initial state (*chaos*), model physics, etc. Typically uses one or more *dynamical models* but *statistical models* can also be used independently or alongside *dynamical models*. A correctly created ensemble defines an estimated distribution of future states providing a full set of the range of possibilities and their associated probabilities.

Ensemble Member: One of a set of forecasts for the same *target period*. The phrase is usually, but not necessarily, applied to refer to a single prediction from a specific model.

Equity: Used in the development community to indicate fairness in sharing of resources.

Equatorial Under Current (EUC): A jetlike ocean current flowing just below the sea surface toward the east and within a few degrees of the equator, especially in the Pacific and Atlantic oceans. It can reach speeds of more than 1 metre per second at a depth of about 100 m.

Error Covariance [Matrix]: The statistical relationship between a variable and another variable (including itself), in space and time, used to determine the relative weight of these variables when they are combined to produce an *analysis*. Error covariance is central to *data assimilation* as it determines the relative importance of, for example, observations and model, or more generally *background*, data. Hence ‘Observation Error Covariance’ and ‘Background Error Covariance’.

Event: In forecast *verification*, an event is an observation, during the *target period* of a forecast, of a specific outcome of interest. The outcome is explicitly binary: either an event occurs during the *target period*, or it does not occur. For some meteorological variables, occurrence is inherently binary (precipitation occurrence, for example), but for continuous variables, an event can be defined if the observed value lies within the limits defining the category of interest (temperature above 30°C, for example, defines an event for a category with no upper bound).

Extended-Range Weather Forecasting: Official terminology of the *WMO* to indicate predictions from 10 to 30 days (in general provided as averages across periods of several days).

Extension Service: Normally used regarding agriculture, typically a government service that provides expert advice to farmers.

First Guess [Field]: Same as *Background*.

Forcing: The source of the disturbance of a dynamical system, which normally appears on the right-hand-side of the equation of the system. For example, *wind stress* is one of the forcings of the ocean (models). See also *Climate Forcing*.

Forecast [field]: In *data assimilation*, model output used to form the *background* field, so-called because it is usually the result of a model *integration* started from an *analysis* produced in the previous *assimilation cycle*.

Forecasting System: Ranges in meaning from a comprehensive view integrating all the components that go into making a forecast – from the generation of initial conditions of the *dynamical models* used for the forecasts, to the running of the *coupled dynamical models*, to the *calibration* and assessment of the model output, to the application of forecast products to specific *users* – to contiguous parts of this comprehensive definition. The meaning should be clear by its context.

Gambler's Ruin: Risk management under uncertainty leads to a sequence of gains and losses that on overage should produce a net gain. Gambler's Ruin occurs when catastrophic losses are taken that prevent further activity.

General Circulation Model (GCM): A set of equations describing the three-dimensional evolution of the system to be modelled (e.g. the atmosphere) in a numerical form. The equations include those of the dynamics and energy of the system, as well as those of any other relevant process (e.g. chemical reactions).

Global Producing Centre: A *WMO* designation conferred from November 2006 on Centres that produce and distribute a minimum suite of global seasonal forecast products, typically using some form of *GCM*.

Group Velocity: The rate at which wave energy propagates. For nondispersive waves, such as a Kelvin wave, the phase and group velocities are the same. For dispersive waves, such as Rossby waves, the group and phase velocities are generally different.

Hindcast: See *Re-forecast*.

Homoscedasticity: The property of homogeneity of variance. A set of data has the property of homoscedasticity if there is equality in the variances of subsets of the data defined a priori (e.g. by time or by value of a second parameter).

Impact: Often used in the sense of the consequence(s) of a climate event, frequently a *climate extreme* or *climate change*, sometimes without due consideration of any other factors that may be present.

Increment: See Observation Increment.

Indigenous Knowledge: That knowledge built over centuries, and passed down through generations, that defines a society's learnt response(s) to events.

Initialisation: All the steps required to prepare a *coupled model* (for an *integration*) normally performed via a *data assimilation* method. Sometimes it refers only to the final step of ensuring dynamical balance intra- and inter-components of a *coupled model*, required to avoid jumps (also called *shocks*) in the solution of the model in the early stages of the *integration*.

Integration [of a Numerical Model]: The advancement in time of the solution of the numerical equations which constitute the model. When the integration refers to future times it is often called a forecast, whereas for past times it can be referred to as *re-forecast* or *hindcast*.

Internal Waves: Waves that can propagate through a fluid because it is stratified. In the tropical ocean, the speed of internal waves is 3 m s^{-1} or less. Internal Kelvin and Rossby waves are important in equatorial processes such as El Niño. They can travel large distances in the ocean (up to 10,000 km). Internal Kelvin waves also exist in the atmosphere where speeds are somewhat higher than in the ocean.

Inter Tropical Convergence Zone (ITCZ): A belt of high rainfall near the equator. It is formed by the vertical ascent of warm, moist air converging from the north and south. It is usually found a few degrees to the north of the equator but moves north and south with the seasons.

La Niña: The *cold phase* of *ENSO*.

Lead-Time: The time between the earliest moment at which the forecast could be released and the starting-time for which the forecast applies. The lead-time is not the same as the amount of advanced warning that is provided by the forecast: a forecast released at the beginning of the *target period* has a lead-time of zero, but does not imply that there is no advanced warning.

Long-Range Forecasting: Official terminology of the *WMO* to indicate predictions from 30 days to 2 years, usually provided as averages across a sub-period; includes Monthly Outlooks, 3-month or 90-Day Outlooks, and Seasonal Outlooks.

Madden-Julian Oscillation (MJO): A tropical atmospheric phenomenon characterised by an oscillation on the intraseasonal timescale (about 40–60 day). Some aspects of the MJO may have predictability beyond 10 days (the *Medium Range*).

Medium-Range Weather Forecasting: Official terminology of the *WMO* to indicate predictions from 72 hours (3 days) to 240 hours (10 days).

Member: *See Ensemble Member*.

Metadata: Information about the data themselves. In the context of meteorological observations, metadata typically are data about the instruments and their location, and the recording procedures.

Millennium Development Goals (MDGs): A set of eight Goals, originally agreed at the UN Millennium Summit in New York in 2000, that provides a time-bound (2015) international coordination framework for development activities. For all Goals there are certain quantifiable targets, each with sets of indicators.

Millennium Project: A project to assess approaches to achieving the *Millennium Development Goals*.

Mitigation: To reduce the consequences of an adverse event. In the *UNFCCC* mitigation is used in the context of reducing anthropogenically forced *Climate Change* through the reduction of greenhouse gas emissions.

Neutral Phase: That phase of *ENSO* that lies between the *warm phase* and the *cold phase*.

NIÑO3.4: One commonly used measure (or metric) of the state of *ENSO* based on sea surface temperatures over the tropical Pacific Ocean (similarly NIÑO3, NIÑO4, etc.). Specifically, the metric is calculated as the spatially averaged sea surface temperature over the domain 5°N–5°S, 170–120°W.

Normal: Used variously by climatologists and stakeholders, and thus prone to misinterpretation. Statistically it is the average value of a distribution. In seasonal forecasts, it is often used to indicate the central (i.e. inter-*tercile*) category (i.e. the middle third of data); thus also in this context above-normal and below-normal.

North Atlantic Oscillation: An atmospheric see-saw of pressure across the North Atlantic Ocean with two standard 'centres of action', one over Iceland and the other on the Azores. Swings from one phase to another produce large changes in the mean wind speed and direction over the Atlantic. Influential on European and North African *climate*.

Nowcasting: Official terminology of the *WMO* to indicate a description of current *weather* conditions and predictions out to 2 hours.

Numerical Modelling: The utilisation of a set of mathematical equations solved by means of computational procedures. In principle it should refer to either *Dynamical Modelling* or *Statistical Modelling*, but in practice it is more often used as synonymous with the former.

Observation Increment [or just Increment]: The difference between an observation and the interpolated model *background* value at the location of that observation. In other words, it determines the strength of the correction due to the *data assimilation* process. Normally, the first step in any *data assimilation* procedure: in the limiting case of an increment being equal to zero (i.e. model datum identical to the observation), no further calculations are required.

Oceanic (or Ocean) General Circulation Model (OGCM): A *Global Circulation Model* for the ocean.

Outlier: A datum that is numerically distant from other data in the same dataset. Outliers may be indicative of observational errors in which case they should be corrected or omitted from analyses, but when they represent *climate extremes* they may have an undue influence on the analysis.

Parameterisation: Normally utilised in dynamical models for simulating physical processes at scales smaller than those resolved by the model: radiative processes, clouds, convection of various kinds, large-scale latent heat release, etc. – the cumulative statistical effects of these processes are represented in terms of the model variables themselves rather than being represented explicitly.

Percentile: Each part of a distribution that divides the data into 100 equal parts.

Poverty: The social condition of having access to insufficient resources to maintain a satisfactory basic state of living. Poverty reduction is the main focus of the World Bank, and is included in the *Millennium Development Goals*.

Prandtl Number: The ratio of kinematic viscosity to thermal diffusivity. Small Prandtl number means heat diffuses more rapidly compared to momentum.

Predictability: The extent to which future states of a system may be predicted based on knowledge of current and past states of the system. Since knowledge of the system's past and current states is generally imperfect, as are the models that utilize this knowledge to produce a prediction, predictability is inherently limited. Even with arbitrarily accurate models and observations, there may still be limits to the predictability of a physical system.

Predictand: Sometimes called a "target" or "response" or "dependent" variable, a predictand is a variable for which a forecast is to be made. Common predictands in seasonal climate forecasting are 3-month rainfall totals and 3-month average air temperature.

Prediction System: Synonymous with *Forecasting System*.

Predictor: Sometimes called an "explanatory" or "independent" variable, a predictor is a variable which is used to make a forecast, and in many cases is some measurement of the forcing mechanism that makes seasonal prediction possible. Common predictors in seasonal climate forecasting are monthly averaged sea surface temperatures, and the *Southern Oscillation Index*.

Principal Component: A new variable, calculated similarly to a weighted average of an original set of variables, with the property that as much of the variance of the original variables as possible is represented, and with the total squared weights adding to one. Additional new variables, or principal components, can be defined that represent as much of the remaining variance as possible. Principal components exploit correlations between the original variables, and thus can act as efficient summaries of large datasets: much of the total variance of the original data may be represented by only a few principal components. Principal components are sometimes called empirical orthogonal functions (EOFs), although, strictly speaking, the EOFs are the weights that define the principal components.

Probability Density Function (PDF or pdf): A function that describes the probability of the value of a continuous variable being within any interval that must be calculated by integration. For discrete variables the probability is described by a probability mass function.

Probability Distribution: A complete mathematical description of the probabilities of all measurable subsets of a variable.

Projection: An estimate of a future state, or a series of envisaged possible future states, obtained from expert interpretation of available information.

Quality: In the context of predictions used as a generic term to indicate the technical level of excellence of the forecasts; e.g. *deterministic* forecasts that are consistently accurate are of relatively high quality. Compare with *Skill*. There is no necessary direct correlation between *skill* and *value*, *value* being dependent on the actual use of a prediction and the manner of that use.

Quintile: Each part of a distribution that divides the data into five equal parts.

Random Error: The imprecision in a given process due to the hard-to-control (sometimes uncontrollable) nature of some elements of the process. For example, repeated measurements of the same quantity are bound to yield different values (if precision of the instrument is sufficiently high) because the conditions of measurement vary, even if ever so slightly, from one measurement to the other. To sample such error, *coupled models* are often started from initial conditions that differ by small amounts.

Reanalysis: Use of the very latest numerical models and all available observations to create new *analyses* of the state of the atmosphere and/or oceans over past years. This approach produces consistent data sets over a number of years of higher quality than previously available, and that are invaluable for meteorological and climatological research.

Recalibration: The statistical adjustment of model output to improve the similarity between the model simulations/forecasts and the observed climate (Chapter 8, Section 8.3.3). See also *Calibration*.

Reference Climatology: A standard *climatology* for a model generated through *integrations* over many years; changes, such as *climate changes*, may be detected against this in late work with the model.

Re-forecast: A model *integration* over past times. Implementation of such *integrations* serves several purposes, two of which are highlighted here: (i) to assess and/or *calibrate* model *integrations* for future times (i.e. forecasts) and (ii) investigate performance of latest models over past events (e.g. the 1997–98 *El Niño*).

Regional Climate Model: A *numerical model* working over a smaller geographical region than a *Global Circulation Model* but at much higher resolution in order to provide enhanced spatial and temporal detail over that region. It typically employs boundary conditions taken from a *Global Circulation Model*.

Resilience: A measure of the extent to which a society or a system, such as an agricultural system, is capable of withstanding the deleterious consequences of some adverse event. Hence 'building resilience'.

Richardson Number: The ratio of the static stability squared to the wind shear squared. It gives a measure of the likelihood of a fluid to develop instabilities.

Risk: Uncertainty for which the *probability distribution* of an outcome (e.g. the objective of a *decision making process*) is known.

Risk Assessment: A study to determine the outcomes of decisions along with their probabilities.

Scenario: One vision of the future created through expert interpretation of available information.

Seamless Decision Making System: The implementation of a sequence of related or interlinked decisions across climatic timescales, adjusted as timescales change (cf. *Seamless Forecasting System*). Those decisions may be related only to a specific portion of the timescales, but that portion may readily extend within, beyond or across the artificial timescale boundaries set up by the scientists.

Seamless Forecasting System: The implementation of unified procedures aimed to provide information which appears transparent to the *user*. In weather/climate forecasting it usually refers to the use of a single *dynamical (coupled) model*, which is *integrated* over different time and space scales, e.g. from a few days to many decades, with an essentially fixed setup. Currently, artificial timescale boundaries between *medium-range*, monthly, seasonal, decadal and *climate change* are common in the production of forecasts. Given the absence of such boundaries in nature, Seamless Forecasting seems a natural approach. Note that purely from a *user* perspective, a Seamless Forecasting System is not strictly necessary if model output were provided in a seamless way.

Shock: A jump in the numerical solution of a model *integration* caused by an abrupt change in the initial conditions. Hence 'Coupling shock' or, more generally, 'Forecast shock'.

Short-Range Weather Forecasting: Official terminology of the *WMO* to indicate predictions from 12 to 72 hours (3 days).

Singular Value Decomposition (SVD): The decomposition of a matrix into a diagonal matrix and two orthonormal (orthogonal and with unit length) matrices. Somewhat confusingly, “SVD analysis” is often used to refer to maximum covariance analysis (Chapter 7, Section 7.4.3), which is one of many possible uses of SVD; for example, SVD is often used to calculate *principal components* in multiple regression and canonical correlation analyses.

Skewness: A measure of the asymmetry of the frequency distribution of records of a single variable. If there are more large positive than negative departures from the median then the data are positively skewed. Conversely, if there are fewer large positive than negative departures from the median then the data are negatively skewed. Skewness can affect the results of some statistical tests, and so distributional assumptions always should be considered.

Skill: A relative measure of the *quality* of a series of forecasts taken against an alternate forecast approach, usually a much simpler and cheaper approach such as guessing, or use of *climatology*, or persisting recent observations. Calculated such that a skill of 0% indicates the forecasts and the alternates are of identical *quality*, and positive values are desirable with 100% indicating perfect forecasts.

Solid Body Rotation: When a fluid rotates as if it were a solid.

Southern Oscillation (SO): An atmospheric see-saw of pressure across the Pacific Ocean with two standard ‘centres of action’, Darwin in Northern Australia and Tahiti in the central South Pacific Ocean. At these centres of action long-term sea-level atmospheric pressures are strongly inversely correlated such that when pressure is higher than normal at one it is lower than normal at the other. The atmospheric component of a major atmosphere-ocean interaction across the Pacific, part of which is the *El Niño* – hence *ENSO*.

Southern Oscillation Index (SOI): A measure of the state of the Southern Oscillation (q.v.); various approaches used.

Stakeholders: See *User*.

Statistical Modelling: In the broad sense using statistics to represent linkages within a system, but used within this book to indicate the employment of statistical relationships between *predictors* and *predictands* to create a predictive equation.

Statistical Significance: A standard procedure for interpreting the strength of a test statistic, such as the correlation between a set of forecasts and the corresponding observations. Statistical significance tests estimate how likely it is that a score that is at least as large as the calculated score on the test statistic could have been achieved purely as a result of a sampling accident. If this probability, known as a *p*-value, is sufficiently low (typically <5%, although other thresholds are used), then the result is said to be "statistically significant".

Sustainable Development: Probably the most controversial of the terms in this glossary. Here a few alternatives are given: pick your choice.

1. Development that meets the needs of the present without compromising the ability of future generations to meet their own needs. Report of the Brundtland Commission, Our Common Future (1987).
2. The management and conservation of the natural resources base, and the orientation of technological and institutional change, in such a manner as to ensure the attainment and continued satisfaction of human needs for present and future generations.
3. Within a country or region, gradual change characterized by economic growth, increased social equity, constructive modification of ecosystems, and maintenance of the natural resource base.

Systematic Error: Any difference between the observed and model climates caused by non-*random sampling errors*. The most commonly considered systematic error is the *bias* in the mean climate of the model.

Target Period: The period for which the forecast applies.

Teleconnection: A simultaneous or successive association between climate *anomalies* in separate parts of the globe. The climate *anomalies* in the disparate regions are related by having a common *forcing* mechanism broadcast over distances through a mechanism generically referred to as a teleconnection. The best known example of a teleconnection is the *Southern Oscillation*, which involves opposite tendencies in sea-level pressure in the western and east-central tropical South Pacific Ocean, and which is then transmitted to distant parts of the globe.

Tercile: Each part of a distribution that divides the data into three equal parts. Note that there are two terciles (the upper and lower), which define three categories. Sometimes, somewhat confusingly, the "tercile" is used to refer to the categories defined by the division (e.g. "above-normal tercile" refers to the category above the upper tercile).

Top Down: A management approach that examines and resolves issues at the highest policy and management levels, then ripples solutions downwards to the working levels (cf. *bottom up*).

Tropical Atmosphere Ocean (TAO) Array: An array of moored buoys across the tropical Pacific Ocean. These relay latest oceanic and atmospheric data via satellite links that are used in many seasonal prediction models.

Type I Error: The rejection of a null hypothesis when the null hypothesis is true. A type I error is made when the result of a statistical test is accepted as *statistically significant* when the strength of the result is only a result of sampling error. In *statistical modelling*, for example, type I errors include accepting spurious predictors, which results in operational forecasts being of lower quality than anticipated.

Type II Error: The acceptance of a null hypothesis when the alternative hypothesis is true. A type II error is made when the result of a statistical test is accepted as statistically insignificant when the weakness of the result is a result of sampling error or insufficient data. In *statistical modelling*, for example, type II errors result in actual forecast quality being less than the potential *predictability*.

Useful: Often used by climatologists in the context of their perceived *value* of their predictions, but frequently unfounded until proven to have *value*.

User: Many terms have been used to name those to whom seasonal predictions are directed – decision makers, decision takers, recipients, stakeholders, end-users, and so on. There is no clear distinction between these terms, and each may be interpreted differently at distinct stages of the delivery chain. Hence 'user' is employed here in a generic sense to cover the interpretation intended in this book. A 'user' is anyone that makes use of *Climate Information* available at any stage of the (comprehensive) *forecasting system*. As a consequence, there are different levels of users including intermediate or end/final users.

United National Framework Convention on Climate Change (UNFCCC): One of the so-called Rio Conventions that emerged from UNCED in 1992. The UNFCCC provides the framework under which intergovernmental climate change negotiations are held, and from which the Kyoto Protocol emerged.

Validation: The evaluation of the ability of a model to make good forecasts and/or to reproduce observed features of the *climate* system. Sometimes used as synonymous with *Verification*.

Value: Value is provided when information, such as a prediction, is employed actively to adjust an existing decision or to illuminate or confirm a new decision; predictions, of whatever quality, that are not used in, or do not contribute towards, any *decision processes* have no value. It is also used to indicate a measure of the benefit achieved through *climate information* use. Contrast with *quality*.

Verification: The measurement of the *quality* of a forecast or of a sequence of forecasts. The term is sometimes used to refer to the eventual outcome to which the forecast was targeted; thus a forecast is compared to its verification in assessing the *quality* of the former.

Very Short Range Weather Forecasting: Official terminology of the *WMO* to signify predictions from 2 to 12 hours.

Vulnerability: A measure of the extent to which a society or a system might be affected adversely by an event; hence vulnerability reduction.

Warm Phase: The *El Niño* stage of *ENSO* during which sea surface temperatures over the eastern equatorial Pacific Ocean are above average.

Weather: The day-to-day evolution of the atmosphere, as measured by variables such as temperature, wind and precipitation. Given its shorter averaging time, weather undergoes larger fluctuations than *climate*, which therefore acts as a weather smoother or integrator. Thus, at any given location, daily temperature anomalies of many degrees are normal whereas seasonal temperature anomalies are usually only a few degrees.

Wind Stress: Force exerted on surfaces with small irregularities (typically the oceans) by the atmosphere, due to pressure differences and viscosity. It is proportional to the square of the wind speed and it is the main driver of upper ocean circulation.

World Meteorological Organization (WMO): A Specialised Agency of the UN, based in Geneva, responsible for the international management concerning the atmosphere and oceans.

References

All the references cited throughout the book are listed below. The number in bold following the reference indicates the chapter(s) in which it was cited.

Abawi GY, Dutta S, Zhang X and McClymont D (2005) ENSO-based streamflow forecasting and its application to water allocation and cropping decisions – An Australian experience. In: Wagener T, Franks S, Gupta H, et al. (eds) *Regional Hydrological Impacts of Climatic Change – Impact Assessment and Decision Making*. IAHS Publication 295, Wallingford, pp. 346–354. **[13]**

Abawi GY, Yasin I, Dutta S, et al. (2002) Capturing the benefits of seasonal climate forecasts in agricultural management. ACIAR Final report. **[13]**

Alves O, Balmaseda MA, Anderson DLT and Stockdale TN (2004) Sensitivity of dynamical seasonal forecasts to ocean initial conditions. *Quart. J. Roy. Meteorol. Soc.*, **130**, 647–667. **[5]**

Anderson DLT and Gill AE (1975) Spin-up of a stratified ccean, with applications to upwelling. *Deep-Sea Res.*, **22**, 583–596. **[4]**

Anderson DLT, Sarachik ES, Webster PJ and Rothstein LM (1998) The TOGA DECADE: Reviewing the progress of El Niño research and prediction. Special Issue of *J. Geophysic. Res.*, **103**. (Preface at page 14,167) **[3]**

Anderson JL and Ploshay JJ (2000) Impact of initial conditions on seasonal simulations with an atmospheric general circulation model. *Quart. J. Roy. Meteorol. Soc.*, **126**, 2241–2264. **[8]**

Anderson JR, Dillon JL and Hardakar JB (1977) *Agricultural Decision Analysis*, Iowa State University Press, Ames, Iowa, 344 pp. **[13]**

Apel JR (1987) *Principles of Ocean Physics*, Academic, New York, 634 pp. **[5]**

Barnett TP, Latif M, Graham NE, Flügel M, Pazan S and White WB (1993) ENSO and ENSO-related predictability. Part I: Prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean-atmosphere model. *J. Clim.*, **6**, 1545–1566. **[8]**

Bates B (2002) Joint international workshop on applicable methods for use of seasonal climate forecasts in water management. IRI, Palisades, New York. **[13]**

Bengtsson L, Schlese U, Roeckner E, Latif M, Barnett TP and Graham NE (1993) A two-tiered approach to long-range climate forecasting. *Science*, **261**, 1026–1029. **[8]**

Benson C and Clay EJ (2004) *Understanding the Economic and Financial Impacts of Natural Disasters*, World Bank, Washington, DC, 119 pp. **[11]**

Bergthorsson P and Doos B (1955) Numerical weather map analysis, *Tellus*, **7**, 329–340. **[5]**

Bjerknes J (1966) A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, **18**, 820–829. **[4]**

Bouttier F and Courtier P (1999) Data assimilation concepts and methods. Lecture Notes, ECMWF, 59 pp. Available at: http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/ASSIM_CONCEPTS/Assim_concepts.html **[9]**

Branković Č and Palmer TN (2000) Seasonal skill and predictability of ECMWF PROVOST ensembles. *Quart. J. Roy. Meteorol. Soc.*, **126**, 2035–2067. **[8]**

Brenner J (1991) Southern Oscillation anomalies and their relationship to wildfire activities in Florida. *Int. J. Wildland Fire*, **1**, 73–78. **[12]**

Bröcker J and Smith LA (2007) Scoring probabilistic forecasts: The importance of being proper. *Weather Forecasting*, **22**, 382–388. **[10]**

Bryan K, Manabe S and Pacanowski RC (1975) A global ocean-atmosphere climate model. Part II: Oceanic circulation. *J. Phys. Oceanogr.*, **5**, 30–46. **[6]**

Busalacchi AJ and O'Brien JJ (1981) Interannual variability of the equatorial Pacific in the 1960's. *J. Geophy. Res.*, **86**, 10901–10907. **[6]**

Cane MA and Sarachick ES (1977) Forced baroclinic ocean motions: II. Linear equatorial bounded case. *J. Mar. Res.*, **35**, 395–432. **[4]**

Cane MA and Patton RJ (1984) A numerical-model for low-frequency equatorial dynamics. *J. Phys. Oceanogr.*, **14**, 1853–1863. **[6]**

Chang P, Yamagata T, Schopf P, et al. (2006) Climate fluctuations of tropical coupled systems – The role of ocean dynamics. *J. Clim.*, **19**, 5122–5174. **[3]**

Charney JG, Fjörtoft R and von Neumann J (1950) Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237–254. **[5]**

Chelton DB, Esbensen SK, Schlax MG, et al. (2001) Observations of coupling between surface wind stress and sea surface temperature in the eastern tropical Pacific. *J. Clim.*, **14**, 1479–1498. **[6]**

Chen D, Cane MA, Kaplan A, Zebiak SE and Huang D (2004) Predictability of El Niño in the past 148 years. *Nature*, **428**, 733–736. **[4]**

Cherchi A and Navarra A (2007) Sensitivity of the Asian summer monsoon to the horizontal resolution: Differences between AMIP-type and coupled model experiments. *Clim. Dyn.*, **28**, 273–290. **[6]**

Coelho CAS, Pezzulli S, Balmaseda M, Doblas-Reyes FJ and Stephenson DB (2003) Skill of coupled model seasonal forecasts: A Bayesian assessment of ECMWF ENSO forecasts. ECMWF Technical Memorandum No. 426, 16 pp. **[9]**

Coelho CAS, Pezzulli S, Balmaseda M, Doblas-Reyes FJ and Stephenson DB (2004) Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Clim.*, **17**, 1504–1516. **[9]**

Coelho CAS, Stephenson DB, Balmaseda M, Doblas-Reyes FJ and van Oldenborgh GJ (2005) Toward an integrated seasonal forecasting system for South America. *J. Clim.*, **19**, 3704–3721. **[9]**

Collins WD, Bitz CM, Blackmon ML, et al. (2006) The community climate system model version 3 (CCSM3). *J. Clim.*, **19**, 2122–2143. **[6]**

Conover WJ (2001) *Practical Nonparametric Statistics*, Wiley, New York, 584 pp. **[8]**

Cressman GP (1959) An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374. **[5]**

Cutter SL, Emrich CT, Mitchell JT, et al. (2006) The long road home: Race, class, and recovery from hurricane Katrina. *Environment*, **48**, 8–20. **[11]**

Daan H (1985) Sensitivity of verification scores to the classification of the predictand. *Mon. Wea. Rev.*, **113**, 1384–1392. **[10]**

Daley R (1991) *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, 457 pp. **[5]**

Davey M, Huddleston M, Ingleby B, et al. (2006) Multi-model multi-method multi-decadal ocean analyses from the ENACT project. CLIVAR Exchanges, No. 38, **11(3)**, 22–25. See also: http://www.ecmwf.int/research/EU_projects/ENACT/index.html **[5]**

Dee DP (2005) Bias and data assimilation. *Quart. J. Roy. Meteorol. Soc.*, **131**, 3323–3343. **[5]**

Delworth TL, Broccoli AJ, Rosati A, et al. (2006) GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Clim.*, **19**, 643–674. **[6]**

Doblas-Reyes FJ, Hagedorn R and Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. Part II: Calibration and combination. *Tellus*, **57A**, 234–252. **[8]**

Dutta SC, Ritchie JW, Freebairn DM and Abawi GY (2006) Rainfall and streamflow response to El Niño Southern Oscillation: A case study in a semi-arid catchment, of Australia. *Hydrol. Sci. J.*, **51**, 1006–1020. **[13]**

Eisenman I, Yu L and Tziperman E (2005) Westerly wind bursts: ENSO's tail rather than the dog? *J. Clim.*, **18**, 5224–5238. **[3]**

Eliassen A (1954) Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. Videnskaps-Akademiets Institutt for Vaer-og Klimaforskning, Report No. 5, 11 pp. **[5]**

Elmore KL (2005) Alternatives to the chi-square test for evaluating rank histograms from ensemble forecasts. *Weather Forecasting*, **20**, 789–795. **[8, 10]**

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10143–10162. **[5]**

Evensen G (1997) Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, **125**, 1342–1354. **[5]**

Fisher M (2003) Background error covariance modelling, Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, 8–12 Sept 2003, Reading, UK, 45–63. **[5]**

Focks D (2003) Impact of anticipated climate change on dengue in the Caribbean based on the new ocean/atmosphere-coupled Hadley climate model version 3 (HadCM3). In: Aron JL, Corvalán CF and Philippeaux H (eds) *Climate Variability and Change and their Health Effects in the Caribbean: Information for Adaptation Planning in the Health Sector*. WHO Conference and Workshop 21–25 May 2002. **[13]**

Fukumori I, Lee T, Cheng B and Menemenlis D (2004) The origin, pathway, and destination of Niño-3 water estimated by a simulated passive tracer and its adjoint. *J. Phys. Oceanogr.*, **34**, 582–604. **[4]**

Galanti E, Tziperman E, Harrison M, Rosati A and Sirkes Z (2003) A study of ENSO Prediction using a hybrid coupled model and the adjoint method for data assimilation. *Mon. Wea. Rev.*, **131**, 2748–2764. **[5]**

Gandin LS (1965) Objective analysis of meteorological fields, Hydrometeoizdat, English translation by Israel Program for Scientific Translations, Jerusalem, available from NTIS as N66–18047. **[5]**

Gent PR and Cane MA (1989) A reduced gravity, primitive equation model of the upper equatorial ocean. *J. Computational Phys.*, **81,** 444–480. **[6]**

Ghil M, Cohn S, Tavantzis J, Bube K and Issacson E (1981) Applications of estimation theory to numerical weather prediction. In: *Dynamic Meteorology*, Springer, New York, pp. 139–224. **[5]**

Gibberd V, Rook J, Sear CB and Williams JB (1995) Drought risk management in southern Africa: The potential of long lead climate forecasts for improved drought management. A Natural Resources Institute report for the UK Overseas Development Administration and the World Bank, NRI, Chatham, UK, 38 pp. **[11]**

Gill AE (1982) *Atmosphere-Ocean Dynamics*. Academic, New York, 662 pp. **[4]**

Glahn HR and DA Lowry (1972) The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211. **[9]**

Glantz M (2003) *Climate Affairs*, Island Press, Washington, DC, 184 pp. **[2]**

Gneiting T and Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.*, **102**, 359–378. **[10]**

Goddard L and Dilley M (2005) El Niño: Catastrophe or opportunity. *J. Clim.*, **18**, 651–665. **[3]**

Gong XF, Barnston AG and Ward MN (2003) The effect of spatial aggregation on the skill of seasonal precipitation forecasts. *J. Clim.*, **16**, 3059–3071. **[8]**

Goodwin P and Wright G (2003) *Decision Analysis for Management Judgment*, 3rd edn., Wiley, New York, 477 pp. **[2]**

Grabowski WW (2001) Coupling cloud processes with large-scale dynamics using the Cloud-Resolving Convective Parameterisation (CRCP). *J. Atmos. Sci.*, **58**, 978–997. **[6]**

Graham RJ, Gordon M, McLean PJ, et al. (2005) A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction general circulation model. *Tellus*, **57A**, 320–339. **[8]**

Gualdi S, Navarra A, Guilyardi E and Delecluse P (2003) Assessment of the tropical Indo-Pacific climate in the Sintex CGCM. *Ann. Geophysics*, **46**, 1–26. **[6]**

Guérémy J-F, Déqué M, Braun A and J-P Piedelièvre (2005) Actual and potential skill of seasonal predictions using the CNRM contribution to DEMETER: Coupled versus uncoupled model. *Tellus*, **57A**, 308–319. **[8]**

Hagedorn R, Doblas-Reyes F and Palmer TN (2006) DEMETER and the application of seasonal forecasts. In: Palmer TN and Hagedorn R (eds) *Predictability of Weather and Climate*. Cambridge University Press, Cambridge, pp. 674–692. **[3]**

Hagedorn R, Doblas-Reyes FJ and Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concept. *Tellus*, **57A**, 219–233. **[8]**

Hamill TM (1997) Reliability diagrams for multicategory probabilistic forecasts. *Weather Forecasting*, **12**, 736–741. **[10]**

Hamill TM (2001) Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560. **[10]**

Hamlet AF, Huppert D and Lettenmair DP (2002) Economic value of long-lead streamflow forecasts for Columbia River hydropower. *J. Water Resour. Plng. and Mgmt.*, **128**, 91–101. **[13]**

Hansen JW and Jones JW (2000) Scaling-up crop models for climate variability applications. *Agric. Syst.*, **65**: 43–72. **[12]**

Hansen JW, Marx S and Weber E (2004) The role of climate perceptions, expectations, and forecasts in farmer decision making: The Argentine Pampas and South Florida. Technical Report 04–01. The Earth Institute at Columbia University. **[12]**

Harper S (2000) Thermocline ventilation and pathways of tropical-subtropical water mass exchange. *Tellus*, **52A**, 330–345. **[9]**

Hasselmann K (1976) Stochastic climate models. Part I: Theory. *Tellus*, **28**, 473–485. **[3]**

Haylock M and McBride J (2001) Spatial coherence and predictability of Indonesian wet season rainfall. *J. Clim.*, **14**, 3882–3887. **[13]**

Held IM and Suarez MJ (1994) A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.*, **75**, 1825–1830. **[6]**

Hersbach H (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecasting*, **15**, 559–570. **[10]**

Hosking JRM (1990) L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Statist. Soc.*, **B52**, 105–124. **[8]**

Hoskins BJ and Hodges KI (2002) New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**, 1041–1061. **[6]**

Howard RA (1990) From influence to relevance to knowledge. In: Oliver RM and Smith JQ (eds) *Influence Diagrams, Belief Nets and Decision Analysis*. Wiley, New York. **[2]**

Hsu WR and Murphy AH (1986) The attributes diagram – A geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecasting*, **2**, 285–293. **[10]**

Huang B and Liu Z (1999) Pacific subtropical-tropical thermocline water exchange in the National Centers for Environmental Prediction ocean model. *J. Geophys. Res.*, **104**, 11065–11076. **[4]**

Hulme M, Biot Y, Borton J, et al. (1992) Seasonal rainfall forecasting for Africa. Part II: Application and Impact Assessment. *Int. J. Environ. Stud.*, **40**, 103–121. **[13]**

Ineson S and Davey MK (1997) Interannual climate simulation and predictability in a coupled TOGA GCM. *Mon. Wea. Rev.*, **125**, 721–741. **[8]**

IPCC (2001) Climate Change: Synthesis Report: A contribution of Working Groups I, II and III to the Third Assessment Report of the Intergovernmental Panel on Climate Change. In: Watson RT and the Core Writing Team (eds). Cambridge University Press, Cambridge, 398 pp. **[2]**

IRI (2000) Coping with the climate: A way forward. Workshop Report: A multi-stakeholder review of Regional Climate Outlook Forums. Pretoria, South Africa, International Research Institute for Climate Prediction. See: http://iri.columbia.edu/outreach/publication/ irireport/PretoriaSumRpt2.html **[11]**

IRI (2005) Sustainable development in Africa: Is the climate right? International Research Institute for Climate Prediction. (IRI Technical Report 01–05). Available at: http://iri.columbia.edu/outreach/publication/report/05-01/report05-01.pdf **[2, 11, 13]**

Jin F-F (1997) An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *J. Atmos. Sci.*, **54**, 811–829. **[4]**

Johnson GC and McPhaden MJ (1999) Interior pycnocline flow from the subtropical to the equatorial Pacific Ocean. *J. Phys. Oceanogr.*, **29**, 3073–3089. **[4]**

Johnson GC, Sloyan BM, Kessler WS, and McTaggart KM (2002) Direct measurements of upper ocean currents and water properties across the tropical Pacific during the 1990s. *Prog. Oceanogr.*, **52**, 31–61. **[6]**

Jolliffe IT (2007) Uncertainty and inference for verification measures. *Weather Forecasting*, **22**, 637–650. **[7]**

Jolliffe IT and Stephenson DB (2003) *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Wiley, Chichester, 240 pp. **[7, 10]**

Jolliffe IT and Stephenson DB (2007) Proper scores for probability forecasts can never be equitable. *Weather and Forecasting*, in press. **[10]**

Jones CS, Shriver JF and O'Brien JJ (1999) The effects of El Niño on rainfall and fire in Florida. *Florida Geographer*, **30**, 55–69. **[12]**

Jones JW, Hansen JW, Royce FS and Messina CD (2000) Potential benefits of climate forecasting to agriculture. *Agric. Ecosys. Env.*, **82**, 169–184. **[12]**

Jones JW, Hoogenboom G, Porter CH, et al. (2003) The DSSAT cropping system model. *Eur. J. Agron*. **18**, 235–265. **[12]**

Jung T (2005) Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quart. J. Roy. Meteorol. Soc.*, **131**, 1045–1073. **[6]**

Kalman RE (1960) A new approach to linear filtering and prediction of the ASME. *J. Basic Eng.*, 35–45. **[5]**

Katz RW and Murphy AH (eds) (1997) *Economic Value of Weather and Climate Forecasts*, Cambridge University Press, Cambridge, 222 pp. **[10]**

Keetch JJ and Byram GM (1968) A drought index for forest fire control. Res. Pap. SE-38. Asheville, NC: US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station. 35 pp. **[12]**

Keogh DU, Abawi GY, Dutta SC, et al. (2004) Context Evaluation: A profile of irrigator climate knowledge, needs and practices in the northern Murray-Darling Basin to aid development of climate-based decision support tools and information and dissemination of research. *Aust. J. Exp. Agric.*, **44**, 247–257. **[13]**

Keppenne CL, Rienecker MM, Kurkowski NP and Adamec DA (2005) Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction. *Nonlin. Proc. Geophys.*, **12**, 491–503. **[5]**

Kharin VV and Zwiers FW (2002) Climate predictions with multimodel ensembles. *J. Clim.*, **15**, 793–799. **[8]**

Kharin VV and Zwiers FW (2003) Improved seasonal probability forecasts. *J. Clim.*, **16**, 1684–1701. **[8]**

Kiladis GN and Diaz HF (1989) Global climatic anomalies associated with extremes in the Southern Oscillation. *J. Clim.*, **2**, 1069–1090. **[12]**

Killworth P, Stainforth D, Webb D and Paterson S (1991) The development of a free-surface Bryan-Cox-Semtner ocean model. *J. Phys. Oceanogr.*, **21**, 1333–1348. **[6]**

Kirtman BP and Schopf PS (1998) Decadal variability in ENSO predictability and prediction. *J. Clim.*, **11**, 2804–2822. **[4]**

Klinker E and Sardeshmukh PD (1992) The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608–627. **[6]**

Klinker E, Rabier F, Kelly G and Mahfouf JF (2000) The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results and diagnostics with operational configuration. *Quart. J. Roy. Meteorol. Soc.*, **126**, 1191–1215. **[5]**

Knaff JA and Landsea CW (1997) An El Niño – Southern Oscillation Climatology and Persistence (CLIPER) forecasting system. *Weather Forecasting*, **12**, 633–652. **[7]**

Kössler W (1999) Rank tests in the two-sample scale problem with unequal and unknown locations. *Stat. Papers*, **40**, 13–35. **[8]**

Kumar A (2007) On the interpretation and utility of skill information for seasonal climate predictions. *Mon. Wea. Rev.*, **135**, 1974–1984. **[10]**

Lamb PJ (1981) Do we know what we should be trying to forecast – climatically? *Bull. Amer. Meteor. Soc.*, **62**, 1000–1001. **[12]**

Lander J and Hoskins BJ (1997) Believable scales and parameterisations in a spectral transform model. *Mon. Wea. Rev.*, **125**, 292–303. **[6]**

Latif M, Pohlmann H and Park W (2006) Predictability of the North Atlantic thermohaline circulation. In: Palmer TN and Hagedorn R (eds) *Predictability of Weather and Climate*. Cambridge University Press, Cambridge, pp. 342–364. **[3]**

Latif M, Timmermann A, Grötzner A, Eckert C and R Voss (2002) On North Atlantic interdecadal variability: A stochastic view. In: Pinardi N and Woods J (eds) *Ocean Forecasting*. Springer, New York, pp. 149–178. **[3]**

Ledwell JR, Watson AJ and Law CS (1993) Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment. *Nature*, 364, 701–703. **[6]**

Leeuwenburgh O (2005) Assimilation of along-track altimeter data in the tropical Pacific region of a global OGCM ensemble. *Quart. J. Roy. Meteorol. Soc.*, **131**, 2455–2472. **[5]**

Lemos MCD (2003) A tale of two polices: The politics of climate forecasting and drought relief in Ceará, Brazil. *Policy Sci.*, **36**, 101–123. **[2]**

Loewenstein GF, Weber EU, Hsee CK and Welch N (2001) Risk as feelings. *Psychol. Bull.*, **127**, 267–286. **[2]**

Lorenc AC (1986) Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteorol. Soc.*, **112**, 1177–1194 **[5]**

Lorenc AC and Rawlins F (2005) Why does 4D-Var beat 3D-Var? *Quart. J. Roy. Meteorol. Soc.*, **131**, 3247–3257. **[5]**

Lorenz EN (1963) Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, **20**, 130–141. **[2]**

Lorenz EN (1993) *The Essence of Chaos*, University of Washington Press, Washington, DC, 227 pp. **[2, 3]**

Luyten JR, Pedlosky J and Stommel H (1983) The ventilated thermocline. *J. Phys. Oceanogr.*, **13**, 292–309. **[4]**

Malanotte-Rizzoli P, Hedstrom K, Arango H and Haidvogel DB (2000) Water mass pathways between the subtropical and tropical ocean in a climatological simulation of the North Atlantic ocean circulation. *Dyn. Atmos. Ocean.*, **32**, 331–371. **[4]**

Manabe S and Bryan K (1969) Climate calculations with a combined ocean-atmosphere model. *J . Atmos. Sci.*, **26**, 786–789. **[6]**

Manabe S, Bryan K and Spelman MJ (1975) A global ocean-atmosphere climate model. Part I. The atmospheric circulation. *J. Phys. Oceanogr.*, **5**, 3–29. **[6]**

Manabe S, Bryan K and Spelman MJ (1979) Global ocean-atmosphere climate model with seasonal variation for future studies of climate sensitivity. *Dyn. Atmos. Oceans*, 3, 393–426. **[6]**

Mason IT (2003) Binary events. In: Jolliffe IT and Stephenson DB (eds) *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, England, pp. 37–76. **[10]**

Mason SJ (2004) On using "climatology" as a reference strategy in the Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **132**, 1891–1895. **[10]**

Mason SJ and Graham NE (2002) Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteorol. Soc.*, **128**, 2145–2166. **[10]**

Mason SJ and Mimmack GM (2002) Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. Clim.*, **15**, 8–29. **[10]**

Mason SJ, Galpin JS, Goddard L, Graham NE and Rajaratnam B (2007) Conditional exceedance probabilities. *Mon. Wea. Rev.*, **135**, 363–372. **[10]**

Mathur A, Burton I and Van Aalst M (eds) (2004) *An Adaptation Mosaic: A Sample of the Emerging World Bank Work in Climate Change Adaptation*, A World Bank pubblication. Available at: http://www-wds.worldbank.org **[11]**

Matthews RAJ (1996) Base-rate errors and rain forecasts, *Nature*, **382**, 766. **[11]**

Maurer EP, Wood AW, Adam JC, Lettenmaier DP and Nijssen B (2002) A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States. *J. Clim.*, **15**, 3237–3251. **[8]**

Maurer EP, O'Donnell GM, Lettenmaier DP and Roads JO (2001) Evaluation of the land surface water budget in NCEP/NCAR and NCEP/DOE reanalyses using an off-line hydrologic model. *J. Geophys. Res.*, **106**, 17841–17862. **[8]**

McCreary JP and Anderson DLT (1984) A simple model of El Niño and the Southern Oscillation. *Mon. Wea. Rev.*, **112**, 934–946. **[3]**

McCreary JP and Anderson DLT (1991) An overview of coupled ocean - atmosphere models of El Niño and the Southern Oscillation. *J. Geophys. Res.*, **96**, 3125–315. **[3, 4]**

McCreary JP and Lu P (1994) Interaction between the subtropical and equatorial ocean circulations: The subtropical cell. *J. Phys. Oceanogr.*, **24**, 466–497. **[4]**

Meehl GA, Gent PR, Arblaster JM, et al. (2001) Factors that affect the amplitude of El Niño in global coupled climate models. *Clim. Dyn.*, **17**, 515–526. **[6]**

Meinke H, Nelson R, Kokic P, et al. (2006) Actionable climate knowledge: From analysis to synthesis. *Clim. Res.*, **33**, 101–110. **[2]**

Michaelsen J (1987) Cross-validation in statistical climate forecast models. *J. Clim. Appl. Meteor.*, **26**, 1589–1600. **[3]**

Miyakoda K (1986) Assessment of results from different analysis schemes. In International Conference on the Results of the Global Weather Experiment and Their Implications for the World Weather Watch, GARP Publications Series No. 26, (Vol. 1), WMO/TD No. 107, pp. 217–253. **[5]**

Mjelde JW, Peel DS, Sonka ST and Lamb PJ (1993) Characteristics of climate forecast quality: Implications for economic value to midwestern corn producers. *J. Clim.*, **6**, 2175–2187. **[2]**

Moore DW and Philander SGH (1977) Modelling of the tropical ocean circulation. In: Goldberg ED, McCave IN, O'Brien JJ and Steele JH (eds) *The Sea*, pp. 319–361. **[4]**

Müller WA, Appenzeller C, Doblas-Reyes FJ and Liniger MA (2005) A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *J. Clim.*, **18**, 1513–1523. **[10]**

Mun J (2004) *Applied Risk Analysis: Moving Beyond Uncertainty in Business*, Wiley Finance, New York, 481 pp. **[2]**

Murphy AH (1973a) A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600. **[10]**

Murphy AH (1973b) Hedging and skill scores for probability forecasts. *J. Appl. Meteor.*, **12**, 215–223. **[10]**

Murphy AH (1991) Forecast verification: Its complexity and dimensionality. *Mon. Wea. Rev.*, **119**, 1590–1601. **[10]**

Murphy AH (1993) What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather Forecasting*, **8**, 281–293. **[10]**

Murphy AH and Winkler RL (1987) A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338. **[10]**

Navarra A and Tribbia J (2005) The Coupled Manifold. *J. Atmos. Sci.*, **62**, 310–330. **[6]**

Neelin D, Battisti DS, Hirst AC, et al. (1998) ENSO theory. *J. Geophys. Res.*, **103**, 14,261–14,290. **[2]**

New South Wales Department of Land and Water Conservation (1998) Integrated Quantity-Quality Model (IQQM) Reference Manual **[13]**

New M, Hulme M and Jones P (2000) Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.*, **13**, 2217–2238. **[8]**

Nicholls N (1999) Cognitive illusions, heuristics, and climate prediction. *Bull. Amer. Meteor. Soc.*, **80**, 1385–1397. **[11, 13]**

NOAA NODC World Ocean Atlas (2005) Levitus S (ed) NOAA Atlases NESDIS 61 and 62, US Government Printing Office, Washington, DC, 182 pp. **[4]**

Pacanowski RC and Philander SGH (1981) Parameterization of vertical mixing in numerical models of tropical oceans. *J. Phys. Oceanogr.*, **11**, 1443–1451. **[6]**

Palmer TN (2001) A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterisation in weather and climate prediction models. *Quart. J. Roy. Meteorol. Soc.*, **127**, 279–304. **[6]**

Palmer TN (2006) Predictability of weather and climate: From theory to practice. In: Palmer TN and Hagedorn R (eds) *Predictability of Weather and Climate*. Cambridge University Press, Cambridge, pp. 1–29. **[3]**

Palmer TN and Anderson DLT (1994) The prospects for seasonal forecasting: A review paper. *Quart. J. Roy. Meteorol. Soc.*, **120**, 755–793. **[5]**

Palmer TN, Buizza R, Molteni F, Chen Y-Q and Corti S (1994) Singular vectors and the predictability of weather and climate. *Philos. Trans. Roy. Soc.*, **348A**, 459–475. **[5]**

Palmer TN, Branković Č and Richardson DS (2000) A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Quart. J. Roy. Meteorol. Soc.*, **126**, 2013–2033. **[8]**

Palmer TN, Alessandri A, Andersen U, et al. (2004) Development of a European multi-model ensemble system for seasonal-to-inter-annual prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872. **[8, 10]**

Palmer TN and Hagedorn R (eds) (2006) *Predictability of Weather and Climate*. Cambridge University Press, Cambridge, 702 pp. **[3]**

Panofsky HA (1949) Objective weather map analysis. *J. Meteor.*, **6**, 386–392. **[5]**

Pedlosky J (1996) *Ocean Circulation Theory*. Springer, New York, 464 pp. **[4]**

Peters H, Gregg MC and Toole JM (1988) On the parameterization of equatorial turbulence. *J. Geophys. Res.*, **93**, 1199–1218. **[6]**

Peterson DP (1968) On the concept and implementation of sequential analysis for linear random fields. *Tellus*, **20**, 673–686. **[5]**

Philander SG (2004) *Our Affair with El Niño*. Princeton University Press, Princeton, NJ, 275 pp. **[3]**

Pielke Jr RA, Sarewitz D and Byerly Jr R (2000) Decision making and the future of nature: Understanding and using predictions. In: Sarewitz D, Pielke Jr RA and Byerly Jr R (eds) *Prediction: Science, Decision Making and the Future of Nature*. Island Press, Washington, DC, 361–388. **[12, 14]**

Potts JM, Folland CK, Jolliffe IT and Sexton D (1996) Revised "LEPS" scores for assessing climate model simulations and long-range forecasts. *J. Clim.*, **9**, 34–53. **[13]**

Richardson LF (1922) *Weather Prediction by Numerical Process*. Cambridge University Press, Cambridge, 236 pp. **[5]**

Ritchie JW, Abawi GY, Dutta SC, Harris TR and Bange MP (2004) Risk management strategies using seasonal climate forecasting in irrigated cotton production: A tale of stochastic dominance. *Aust. J. Agric. Resour. Econ.*, **48**, 65–93. **[13]**

Richman MB (1986) Rotation of principal components. *J. Climat.*, **6**, 293–335. **[7]**

Robertson AW, Kirshner S and Smyth P (2004) Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. *J. Clim.*, **17**, 4407–4424. **[8]**

Rodgers KB, Blanke B, Madec G, et al. (2003) Extratropical sources of equatorial Pacific upwelling in an OGCM. *Geophys. Res. Lett.*, **30**, 1084. **[4]**

Rodwell MJ and Hoskins BJ (1996) Monsoons and the dynamics of deserts. *Quart. J. Roy. Meteorol. Soc.*, **122**, 1385–1404. **[4]**

Roeckner E, Arpe K, Bengtsson L, et al. (1996) The atmospheric circulation model ECHAM-4: Model description and simulation of present-day climate. MPI-Rep. 218, MPI für Meteorologie, Hamburg, 90 pp. **[8, 10]**

Rothstein LM, Zhang R-H, Busalacchi AJ and Chen D (1998) A numerical simulation of the mean water pathways in the subtropical and tropical Pacific Ocean. *J. Phys. Oceanogr.*, **28**, 322–343. **[4]**

Roulston MS and Smith LA (2002) Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, **130**, 1653–1660. **[8, 10]**

Sayuti R, Karyadi W, Yasin I and Abawi GY (2004) Factors affecting the use of climate forecasts in agriculture: A case study of Lombok Island, Indonesia. ACIAR Technical Report No. 59, 52 pp. **[13]**

Schlesinger ME (1979) Global ocean-atmosphere model with seasonal variationsfor future studies of climate sensitivity. *Dyn. Atmos. Ocean*, **3**, 427–432. **[6]**

Schopf PS and Cane MA (1983) On equatorial dynamics, mixed-layer physics and sea surface temperature. *J. Phys. Oceanogr.*, **13**, 917–935. **[6]**

Schopf PS and Suarez MJ (1988) Vacillations in a coupled ocean-atmosphere model. *J. Atmos. Sci.*, **45**, 549–566. **[4]**

Schopf PS and Suarez MJ (1990) Ocean wave dynamics and the timescale of ENSO. *J. Phys. Oceanogr.*, **20**, 629–645. **[4]**

Schopf PS and Loughe A (1995) A reduced gravity isopycnal ocean model: Hindcasts of El Niño. *Mon. Wea. Rev.*, **123**, 2839–2863. **[6]**

Sheskin DJ (2007) *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall/CRC, Boca Raton, FL, 1736 pp. **[8, 10]**

Shongwe ME, Ferro CAT, Coelho CAS and van Oldenborgh GJ (2007) Predictability of cold spring seasons in Europe. *Mon. Wea. Rev.*, in press. **[3]**

Shukla J, Anderson J, Baumhefner D, et al. (2000) Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593–2606. **[8]**

Shutts G (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteorol. Soc.*, **131**, 3079–3102. **[6]**

Sittel MC (1994) Marginal probabilities of the extremes of ENSO events for temperature and precipitation in the Southeastern United States. Tech. Rep. 94-1, Center for Ocean-Atmospheric Prediction Studies. Florida State University, Tallahassee, FL. [**12]**

Sonka ST, Mjelde JW, Lamb PJ, Hollinger SE and Dixon BL (1987) Valuing climate forecast information. *J. Clim. Appl. Meteor*, **26**, 1080–1091. **[12]**

Stephenson DB and Doblas-Reyes FJ (2000) Statistical methods for interpreting Monte Carlo ensemble forecasts. *Tellus*, **52A**, 300–322. **[9]**

Stephenson DB, Coelho C, Doblas-Reyes F and Balmaseda MA (2005) Forecast assimilation: A unified framework for the combination of multi-model weather and climate predictions. *Tellus*, **57A**, 253–264. **[3, 9]**

Stern PC and Easterling WE (1999) Making climate forecast information more useful. In: Stern PC and Easterling WE (eds) *Making Climate Forecasts Matter*. National Research Council, National Academy Press, Washington, DC, pp. 63–94. **[12]**

Stern PC and Easterling WE (eds) (1999) *Making Climate Forecasts Matter*. National Research Council, National Academy Press, Washington, DC, 175 pp. **[2]**

Stockdale TN, Anderson DLT, Alves JOS and Balmaseda MA (1998) Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392**, 370–373. **[3, 6]**

Swetnam TW and Betancourt JL (1990) Fire-Southern Oscillation relations in the Southwestern United States. *Science*, **249**, 1017–1020. **[12]**

Thompson PD (1968) Reduction of analysis error through the constraints of dynamical consistency. *J. Appl. Meteor.*, **8**, 738–742. **[5]**

Thompson RD and Perry AH (eds) (1997) *Applied Climatology: Principles and Practice*. Routledge, London, 352 pp. **[13]**

Thomson MC, Ericksen PJ, Ben Mohamed A and Connor SJ (2004a) Land-Use change and infectious disease in West Africa. In: DeFries RS, Asner GP and Houghton RA (eds) *Ecosystems and Land Use Change*. Geophysical Monograph Series 153, American Geophysical Union, pp. 169–187. **[13]**

Thomson MC, Connor SJ, Ward N and Molyneux D (2004b) Climate and disease in west Africa. *EcoHealth*, **1**, 138–150. **[13]**

Thomson MC, Obsomer V, Kamgno J, et al. (2004c) Mapping the distribution of Loa loa in Cameroon in support of the African Programme for Onchocerciasis Control. *Filaria J.*, **3**, 7. **[13]**

Thomson MC, Mason SJ, Phindela T and Connor SJ (2005) Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. *Am. J. Trop. Med. Hyg.*, **73**, 214–221. **[13]**

Thomson MC, Doblas-Reyes FJ, Mason SJ, et al. (2006) Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **439**, 576–579. **[13]**

Tippett MK, Anderson JL, Bishop CH, Hamill TM and Whitaker JS (2003) Ensemble square-root filters. *Mon. Wea. Rev.*, **131**, 1485–1490. **[5]**

Tippett MK, Barnston AG and Robertson AW (2007) Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. *J. Clim.*, **20**, 2210–2228. **[8]**

Toth Z and Kalnay E (1997) Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319. **[5]**

Troccoli A and Haines K (1999) Use of the temperature-salinity relation in a data assimilation context. *J. Atmos. and Oceanic Tech.*, **16**, 2011–2025. **[5]**

Troccoli A, Balmaseda MA, Segschneider J, et al. (2002) Salinity adjustments in the presence of temperature data assimilation. *Mon. Wea. Rev.*, **130**, 89–102. **[5]**

Turner AG, Inness PM and Slingo JM (2005) The role of the basic state in the ENSO-Monsoon relationship and implications for predictability. *Quart. J. Roy. Meteorol. Soc.*, **131**, 781–804. **[6]**

Uppala SM, Kållberg P, Simmons AJ, et al. (2005) The ERA-40 reanalysis. *Quart. J. Roy. Meteorol. Soc.*, **131**, 2961–3012. See also ECMWF Newsletter 101: http://www.ecmwf.int/publications/newsletters/ **[6]**

UN (2005) Investing in Development: A Practical Plan to Achieve the Millennium Development Goals. Millennium Project Report to the UN Secretary-General, p90. Available at: http://www.unmillenniumproject.org **[11]**

UNDP (2005) The Sustainable Difference: Energy and the Environment to Achieve the Millennium Development Goals. Available at: http://www.undp.org/energyandenvironment **[11]**

Valcke S, Caubel A, Vogelsang R and Declat D (2004) OASIS3 user's guide prism_2–4, PRISM Report Series No. 2, 5th ed., CERFACS, Toulouse, France, 60 pp. **[6]**

van Oldenborgh GJ (2000) What caused the onset of the 1997-98 El Niño? *Mon. Wea. Rev.*, **128**, 2601–2607. **[3]**

van Oldenborgh GJ, Balmaseda MA, Ferranti L, Stockdale TN and Anderson DLT (2005) Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15-year period. *J. Clim.*, **18**, 3250–3269. See also corrections Vol. 18. **[3]**

Vecchi G, Wittenberg A and Rosati A (2006) Reassessing the role of stochastic forcing in the 1997–8 El Niño. *Geophys. Res. Lett.*, **33**, L01706. **[3]**

Vialard J, Weaver AT, Anderson DLT and Delecluse P (2003) Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part 2: Physical validation. *Mon. Wea. Rev.*, **131**, 1379–1395. **[5]**

von Storch H and Zwiers FW (1999) *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, 484 pp. **[7, 8]**

Wang C and Picaut J (2004) Understanding ENSO physics – a review. In: *Earth's Climate*, Geophysical Monograph series 147, American Geophys. Union, 10.1029/147GM02. **[3]**

Ward MN (1998) Diagnosis and short-lead prediction of summer rainfall in tropical North Africa at interannual and multidecadal timescales. *J. Clim.*, **12**, 3167–3191. **[8]**

Ward NM and Folland CK (1991) Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea surface temperatures. *Int. J. Climatol.*, **11**, 711–743. **[10]**

Washington WM, Semtner AJ, Meehl GA, Knight DJ and Mayer TA (1980) A general circulation experiment with a coupled atmosphere, ocean and sea ice model. *J. Phys. Oceanogr.*, **10**, 1887–1908. **[6]**

Weaver AT, Deltel C, Machu E, Ricci S and Daget N (2005) A multivariate balance operator for variational ocean data assimilation. *Quart. J. Roy. Meteorol. Soc.*, **135**, 3605–3625. **[5]**

Weaver AT, Vialard J and Anderson DLT (2003) Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part 1: Formulation, internal diagnostics and consistency checks. *Mon. Wea. Rev.*, **131**, 1360–1378. **[5]**

Weber EU (2001) Personality and risk taking. In: Smelser NJ and Baltes PB (eds) *International Encyclopedia of the Social and Behavioral Sciences*. Elsevier Science, Oxford, pp. 11274–11276. **[2]**

Weigel AP, Liniger MA and Appenzeller C (2007) The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124. **[10]**

Weisheimer A, Smith LA and Judd K (2005) A new view of seasonal forecast skill: Bounding boxes from the DEMETER ensemble forecasts. *Tellus*, **57A**, 265–279. **[10]**

White I, Falkland A and Scott D (1999) Droughts in small coral islands: Case study, South Tarawa, Kiribati. UNESCO-IHP Humid Tropics Programme. **[13]**

WHO (2001) Malaria Early Warning Systems: Concepts, indicators and partners. Geneva, Roll Back Malaria Cabinet Project, World Health Organization (WHO/CDS/RBM/2000.32). **[13]**

WHO (2004a) Heat-waves: Risks and responses. WHO Regional Office for Europe. **[13]**

WHO (2004b) *Using Climate to Predict Infectious Disease Outbreaks: A Review*, World Health Organization, Geneva. **[13]**

Wilby RL, Conway D and Jones PD (2002) Prospects for downscaling seasonal precipitation variability using conditioned weather generator parameters. *Hydrol. Processes*, **16**, 1215–1234. **[8]**

Wilks DS (2004) The minimum spanning tree histogram as a verification tool for multidimensional ensemble forecasts. *Mon. Wea. Rev.*, **132**, 1329–1340. **[10]**

Wilks DS (2005) Statistical Methods in the Atmospheric Sciences: An Introduction. Academic, New York, 627 pp. **[7, 8, 9]**

Wilks DS and Wilby RL (1999) The weather generation game: A review of stochastic weather models. *Prog. Phys. Geogr.*, **23**, 329–357. **[8]**

Wilson LJ, Burrows WR and Lanzinger A (1999) A strategy for verification of weather element forecasts from an ensemble prediction system. *Mon. Wea. Rev.*, **127**, 956–970. **[10]**

WMO (1999) The 1997–1998 El Niño event: A scientific and technical retrospective. World Meteorological Organization, Geneva (WMO 905). **[1]**

WMO (2003a) Guidelines on climate observation networks and systems. World Meteorological Organization (WCDMP No. 52). **[13]**

WMO (2003b) Guidelines on climate metadata and homogenization.World Meteorological Organization (WCDMP No. 53). **[13]**

WMO (2003c) Report of the joint WMO/CLIVAR expert team on climate change detection, monitoring and indices. World Meteorological Organization (WCDMP No. 54). **[13]**

WMO (2004) Proceeding of the meeting of experts to develop guidelines on heat/health warning systems. World Meteorological Organization (WMO-TD No. 1212). **[13]**

Wolfram S (1994). Cellular Automata and Complexity: Collected Papers. Addison-Wesley, Reading, MA, 596 pp. **[6]**

Wu RG and Kirtman BP (2005) Roles of Indian and Pacific Ocean air-sea coupling in tropical atmospheric variability. *Clim. Dyn.*, **25**, 155–170. **[8]**

Wu RG, Kirtman BP and Pegion K (2006) Local air-sea relationship in observations and model simulations. *J. Clim.*, **19**, 4914–4932. **[8]**

Xu Q-S and Liang Y-Z (2001) Monte Carlo cross validation. *Chemom. Intell. Lab. Syst.*, **56**, 1–11. **[7]**

Yu ZJ and Schopf PS (1997) Vertical eddy mixing in the tropical upper ocean: Its influence on zonal currents. *J. Phys. Oceanogr.*, **27**, 1447–1458. **[6]**

Zebiak SE and Cane M (1987) A model El Niño Southern Oscillation. *Mon. Wea. Rev.*, **115**, 2262–2278. **[3, 6]**

Suggested Further Reading

A list of references for further reading of interest to both specialists and non-specialists is given here. The number in bold following the reference indicates the chapter(s) to which the reference is most appropriate.

Abawi GY, Smith RJ and Brady DK (1995) Assessment of the value of long range weather forecasts in wheat harvest management. *J. Agric. Eng. Res.*, **62**, 39–48. **[13]**

AMS (2003) Improving responses to climate predictions, Atmospheric Policy Program Report. *Amer. Meteor. Soc.*, 34. **[2, 11]**

Anderson JL (1996) A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Clim.*, **9**, 1518–1530. **[10]**

Avery, Olsen, Aber, et al. (2003) Policy discussions at the AMS policy forum on improving responses to climate predictions. *Bull. Amer. Meteorol. Soc.*, **84,** 1697–1702 **[2, 11]**

Barnett TP, Preisendorfer RW, Goldstein LM and Hasselmann K (1981) Significance tests for regression model hierarchies. *J. Phys. Oceanogr.*, **11**, 1150–1154. **[7]**

Barnett TP and Preisendorfer R (1987) Origins and levels of monthly and seasonal forecast skill for United States surface air temperature determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850. **[7]**

Barnston AG, Mason SJ, Goddard L, DeWitt DG and Zebiak SE (2003) Multi-model ensembling in seasonal climate forecasting at IRI. *Bull. Amer. Meteor. Soc.*, **84**, 1783–1796. **[8]**

Bell MJ, Martin MJ and Nichols NK (2004) Assimilation of data into an ocean model with systematic errors near the equator. *Quart. J. Roy. Meteorol. Soc.*, **130**, 873–893. **[5]**

Blais AR and Weber EU (2001) Domain-specificity and gender differences in decision making. *Risk Decision and Policy*, **6**, 47–69. **[2]**

Bretherton CS, Smith C and Wallace JM (1992) An intercomparison of methods for finding coupled patterns in climate data. *J. Clim.*, **5**, 541–560. **[8]**

Burton I and Van Aalst M (2004) Look before you leap: A risk management approach for incorporating climate change adaptation in World Bank operations, A World Bank pubblication. Available at: http://www-wds.worldbank.org **[11]**

Cash D and Buizer J (2005) *Knowledge–action systems for seasonal to interannual climate forecasting: Summary of a workshop*, report to the Roundtable on Science and Technology for Sustainability, Policy and Global Affairs. National Academies Press, Washington, DC. Available at: http://books.nap.edu/catalog/11204.html **[1, 14]**

Clemen RT (1991) *Making Hard Decisions: An Introduction to Decision Analysis*, PWS-Kent, Boston, MA, 664 pp. **[2, 11]**

Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *Int. J. Forecasting*, **5**, 559–583. **[9]**

CLIPS Curriculum web site with lectures on the Science and Applications of Climate Prediction: http://www.wmo.int/pages/prog/wcp/wcasp/clips/modules/clips_modules.html

D'Andrea FS, Tibaldi S, Blackburn M, et al. (1998) Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979–1988. *Clim. Dyn.*, **14**, 385–407. **[6]**

da Cunha E (1995) *Rebellion in the Backlands*, University of Chicago Press, Chicago, IL, 751 pp. **[2]**

David FN (1998) *Games, Gods, and Gambling: A History of Probability and Statistical Ideas*, Dover, New York, 320 pp. **[9]**

Davis M (2001) *Late Victorian Holocausts: El Niño, Famines, and the Making of the Third World*, Verso Books, London, 464 pp. **[3]**

DeGroot MH and Schervish MJ (2001) *Probability and Statistics*, Addison-Wesley, Reading, MA, 816 pp. **[9]**

Draper NR and Smith H (1981) *Applied Regression Analysis*, Wiley, London, New York. 706 pp. **[9]**

Elsner JB and Schmertmann CP (1994) Assessing forecast skill through cross validation. *Weather Forecasting*, **9**, 619–624. **[7]**

Fan Y, Van den Dool HM, Lohmann D and Mitchell K (2006) 1948–98 US hydrological reanalysis by the Noah land data assimilation system. *J. Clim.*, **19**, 1214–1237. **[5]**

Feddersen H, Navarra A and Ward MN (1999) Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Clim.*, **12**, 1974–1989. **[8]**

Finan T (1999) Drought and demagoguery: A political ecology of climate variability in Northeast Brazil. Bureau of Applied Research in Anthropology. Paper from workshop on: 'Public Philosophy, Environment, and Social Justice'. Carnegie Council on Ethics and International Affairs, Merrill House, New York. **[2, 14]**

Folland CK, Owen JA, Ward MN and Colman AW (1991) Prediction of seasonal rainfall in the Sahel region of Africa using empirical and dynamic methods. *J. Forecasting*, **10**, 21–56. **[7]**

Glantz MH (ed) (2001) *Once Burned, Twice Shy? Lessons learned from the 1997–98 El Niño*, United Nations University, Tokyo, 294 pp. **[1, 2]**

Glantz MH (ed) (2005) Usable science: El Niño early warning for sustainable development in Pacific Rim countries and islands. Report of workshop held 13–16 September 2004. ISSE/NCAR, Boulder, CO. **[13]**

Gordon G and Pressman I (1978) *Quantitative Decision-Making for Business*, Prentice-Hall, London, 546 pp. **[2, 11]**

Graham RJ, Evans ADL, Mylne KR, Harrison MSJ and Robertson KB (2000) An assessment of seasonal predictability using atmospheric general circulation models. *Quart. J. Roy. Meteorol. Soc.*, **126**, 2211–2240. **[8]**

Greenfield RS and Fisher GM (2003) Improving responses to climate predictions – An introduction. *Bull. Amer. Meteorol. Soc.*, **84**, 1685–1685. **[2, 11]**

Grimmett GR and Stirzaker DR (1992) *Probability and Random Processes*, Clarendon, Oxford, 608 pp. **[9]**

Hall N (ed) (1991) *The New Scientist Guide to Chaos*, Penguin, Harmondsworth, 223 pp. **[2]**

Hammer GL, Nicholls N and Mitchell C (eds) (2000) *Applications of Seasonal Climate Forecasting in Agriculture and Natural Ecosystems: The Australian Experience*, Kluwer, Dordrecht, The Netherlands, 469 pp. **[2, 12]**

Hammer GL, Holzworth DP and Stone RC (1996) The value of skill in seasonal climate forecasting to wheat crop management in a region with high climatic variability. *Aust. J. Agric. Res.*, **47**, 717–737. **[13]**

Hansen JW and Sivakumar MVK (2006) Advances in applying climate prediction to agriculture. *Clim. Res.*, **33**, 1–2. **[12]**

Harrison MSJ (2005) The development of seasonal and interannual climate forecasting. *Clim. Change*, **70**, 201–220. **[7]**

Hellmuth ME, Moorhead A, Thomson MC and Williams J (eds) (2007) *Climate Risk Management in Africa: Learning from Practice*, International Research Institute for Climate and Society (IRI), Columbia University, New York. **[2]**

Horel JD (1984) Complex principal component analysis: Theory and examples. *J. Clim. Appl. Meteor.*, **23**, 1660–1673. **[7]**

Hosmer DW and Lemeshow S (1989) *Applied Logistic Regression*, Wiley, New York, 307 pp. **[7]**

Huberty CJ (1994) *Applied Discriminant Analysis*, Wiley, New York, 466 pp. **[7]**

IFRCRCS (2001) World Disasters Report. International Federation of Red Cross and Red Crescent Societies, London, Eurospan. **[1, 14]**

IRI (2006) A gap analysis for the implementation of the global climate observing systems programme in Africa. IRI Technical Report No IRI-TR/06/01, 47pp. Available at: http://iri.columbia.edu/outreach/publication/report/06-01/report06-01.pdf **[2]**

Jacobs K (2003) *Connecting Science, Policy, and Decision-Making: A Handbook for Researchers and Science Agencies*. NOAA Office of Global Programs/University Corporation for Atmospheric Research. Available at: http://www.ogp.noaa.gov/mpe/csi/ doc/hdbk.pdf **[14]**

Jagger TH, Niu XF and Elsner JB (2002) A space-time model for seasonal hurricane prediction. *Int. J. Climatol.*, **22**, 451–465. **[7]**

Katz RW and Brown BG (1991) The problem of multiplicity in research on teleconnections. *Int. J. Climatol.*, **11**, 505–513. **[7]**

Keating BA, Godwin DC and Watiki JM (1991) Optimising nitrogen inputs in response to climate risk. In: Muchow RC and Bellamy JA (eds) *Climate Risk in Crop Production Models and Management for the Semiarid Tropics and Subtropics*, CABI, Wallingford, pp. 329–358. **[13]**

Kharin VV, Zwiers FW and Gagnon N (2001) Skill of seasonal hindcasts as a function of the ensemble size. *Clim. Dyn.*, **17**, 835–843. **[8]**

Kim K-Y and North GR (1999) EOF-based linear prediction algorithm: Examples. *J. Clim.*, **12**, 2076–2092. **[7]**

Kim K-Y and Wu Q (1999) A comparison study of EOF techniques: Analysis of nonstationary data with periodic statistics. *J. Clim.*, **12**, 185–199. **[7]**

Kirtman BP, Fan Y and Schneider EK (2002) The COLA global coupled and anomaly coupled ocean-atmosphere GCM. *J. Clim.*, **15**, 2301–2320. **[8]**

Krishnamurti TN, Kishtawal CM, Zhang Z, et al. (2000) Multimodel ensemble forecasts for weather and seasonal climate. *J. Clim.*, **13**, 4196–4216. **[8]**

Kumar A and Hoerling MP (1995) Prospects and limitations of seasonal atmospheric GCM predictions. *Bull. Amer. Meteor. Soc.*, **76**, 335–345. **[8]**

Kumar, A, Barnston, AG and Hoerling, MP (2001) Seasonal predictions, probabilistic verifications, and ensemble size. *J. Clim.*, **14**, 1671–1676. **[8]**

Large WG and Gent PR (1999) Validation of vertical mixing in a equatorial ocean model using large eddy simulations and observations. *J. Phys. Oceanogr.*, **29**, 449–464. **[6]**

Latif M, Sperber K, Arblaster J, et al. (2001) ENSIP: The El Niño simulation intercomparison project. *Clim. Dyn.*, **18**, 255–276. **[3]**

Lee PM (2004) *Bayesian Statistics: An Introduction*, Hodder Arnold, London, 368 pp. **[9]**

Lien R-C, Caldwell DR, Gregg MC and Moum JN (1995) Turbulence variability at the equator in the central Pacific at the beginning of the 1991–1993 El Niño. *J. Geophys. Res. Oceans*, **100**, 6881–6898. **[6]**

Livezey RE and Smith TM (1999) Considerations for use of the Barnett and Preisendorfer (1987) algorithm for canonical correlation analysis of climate variations. *J. Clim.*, **12**, 303–305. **[7]**

Livezey RE and Chen WY (1983) Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59. **[7]**

Manly BFJ (1994) *Multivariate Statistical Methods: A Primer*, Chapman & Hall, Boca Raton, FL, 215 pp. **[7]**

Mason SJ, Goddard L, Graham NE, et al. (1999) The IRI seasonal climate prediction system and the 1997/1998 El Niño event. *Bull. Amer. Meteor. Soc.*, **80**, 1853–1873. **[8]**

Matheson JE (1990) Using influence diagrams to value information and control. In: Oliver R and Smith JQ (eds) *Influence Diagrams, Belief Nets and Decision Analysis*, Wiley, New York. **[2]**

McCown RL, Wafla BM, Mohammed L, Ryan JG and Hargreaves JNG (1991) Assessing the value of seasonal rainfall predictors to agronomic decisions: The case of response farming in Kenya. In: Muchow RC and Bellamy JA (eds) *Climate Risk in Crop Production Models and Management for the Semiarid Tropics and Subtropics*, CABI, Wallingford, pp. 383–410. **[13]**

McCreary JP (1985) Modeling equatorial ocean circulation. *Ann. Rev. Fluid Mechs*., **17**, 359–409. **[3, 4]**

McCullagh P and Nelder JA (1989) *Generalized Linear Models*, Chapman-Hall, Boca Raton, FL 511 pp. **[7]**

McDonnell KA and Holbrook NJ (2004) A Poisson regression model of tropical cyclogenesis for the Australian–Southwest Pacific ocean region. *Weather Forecasting*, **19**, 440–455. **[7]**

Mielke PW, Berry KJ, Landsea CW and Gray WM (1996) Artificial skill and validation in meteorological forecasting. *Weather Forecasting*, **11**, 153–169. **[7]**

Mo KC and Wang XL (1995) Sensitivity of the systematic-error of extended–range forecasts to sea surface temperature anomalies. *J. Clim.*, **8**, 1533–1543. **[8]**

Montgomery DC and Peck EA (1992) *Introduction to Linear Regression Analysis*, Wiley, New York, 527 pp. **[7]**

Moum JN, Hebert D, Paulson CA and Caldwell DR (1992) Turbulence and internal waves at the equator. Part I: Statistics from towed thermistor and a microstructure profiler. *J. Phys. Oceanogr.*, **22**, 1330–1345. **[6]**

Ni-Meister W, Houser PR and Walker JP (2006) Soil moisture initialization for climate prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model. *J. Geophys. Res. Atmos.*, **111**, D20. **[5]**

O'Brien K and Vogel C (eds) (2003) *Coping with Climate Variability: The Use of Seasonal Climate Forecasts in Southern Africa*, K. Ashgate Studies in Environmental Policy and Practice. **[12]**

Oliver RM and Smith JQ (eds) (1990) *Influence Diagrams, Belief Nets and Decision Analysis*. Proceedings of the Conference entitled 'Influence Diagrams for Decision Analysis, Inference and Prediction', held at the Engineering Systems Research Center, University of California at Berkeley, USA, 9–11 May 1988, Wiley, New York, 465 pp. **[2]**

Ostrom E, Dietz T, Dolsak N, et al. (eds) (2002) *The Drama of the Commons*, National Academy Press, Washington, D C. **[11]**

Penland C (1989) Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.*, **117**, 2165–2185. **[7]**

Penland C and Magorian T (1993) Prediction of Niño3 sea surface temperatures using linear inverse modeling. *J. Clim.*, **6**, 1067–1076. **[7]**

Reichle RH, Koster RD, Dong JR and Berg AA (2004) Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *J. Hydrometeor.*, **5**, 430–442. **[5]**

Rice JA (2006) *Mathematical Statistics and Data Analysis*, Duxbury Press, Belmont, CA, 672 pp. **[9]**

Rutherford S, Mann ME, Delworth TL and Stouffer RJ (2003) Climate field reconstruction under stationary and nonstationary forcing. *J. Clim.*, **16**, 462–479. **[7]**

Saha S, Nadiga S, Thiaw C, et al. (2006) The NCEP Climate Forecast System. *J. Clim.*, **19**, 3483–3517. **[8]**

Salinger MJ, Sivakumar MVK and Motha RP (2005) Reducing the vulnerability of agriculture and forestry to climate variability and change: Workshop summary and recommendations. *Clim. Change*, **70**, 341–362. **[12]**

Sarewitz D, Pielke Jr RA and Byerly Jr R (eds) (2000) *Prediction: Science, Decision Making and the Future of Nature*, Island Press, Washington, DC, 400 pp. **[2, 14]**

Schneider EK, DeWitt D, Rosati A, et al. (2003) Retrospective ENSO forecasts: Sensitivity to atmospheric model and ocean resolution. *Mon. Wea. Rev.*, **131**, 3038–3060. **[5]**

Schneider T (2001) Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Clim.*, **14**, 853–871. **[7]**

Silbiger SA (2005) *The Ten-Day MBA*, 3rd edn., HarperCollins, New York, 420 pp. **[2]**

Sivakumar MVK, Motha RP and Das HP (eds) (2005) *Natural Disasters and Extreme Events in Agriculture: Impacts and Mitigation*. Proceeding from a CAgM Expert Team Meeting, 16–20 February 2004, Beijing, China. Springer, New York, 367 pp. **[12]**

Skyllingstad ED, Smith WD, Moum JN and Wijesekera H (1999) Upper-ocean turbulence during a westerly wind burst: A comparison of large-eddy simulation results and microstructure measurements. *J. Phys. Oceanogr.*, **29**, 5–28. **[6]**

Stone RC, Hammer GL and Marcussen T (1996) Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. *Nature*, **384**, 252–255. **[7]**

Swinbank R, Shutyaev V and Lahoz WA (eds) (2003) *Data Assimilation for the Earth System*, NATO Science Series, Kluwer, Dordrecht, The Netherlands, 377 pp. **[5]**

Tang BY, Hsieh WW, Monahan AH and Tangang FT (2000) Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures. *J. Clim.*, **13**, 287–293. **[7]**

Tangang FT, Tang BY, Monahan AH and Hsieh WW (1998) Forecasting ENSO events: A neural network extended EOF approach. *J. Clim.*, **11**, 29–41. **[7]**

Tarhule A and Lamb PJ (2003) Climate research and seasonal forecasting for West Africans: Perceptions, dissemination, and use? *Bull. Amer. Meteor. Soc.*, **84**, 1741–1759. **[2, 14]**

Thompson PD (1961) A dynamical method of analyzing meteorological data. *Tellus*, **13**, 334–349. **[5]**

Timmermann A and Jin F-F (2006) Predictability of coupled processes. In: Palmer TN and Hagedorn R (eds) *Predictability of Weather and Climate*, Cambridge University Press, Cambridge, pp. 251–274. **[3]**

Trenberth KE (1997) Short-term climate variations: Recent accomplishment ands issues for future progress. *Bull. Amer. Meteor. Soc.*, **78**, 1081–1096. **[3]**

van den Dool HM (1994) Searching for analogs, how long must we wait? *Tellus*, **46A**, 314–324. **[7]**

Various authors (2006) Advances in applying climate prediction to agriculture. *Clim. Res.*, **33**(1), Special Issue 16. **[12]**

Vitart F (2006) Seasonal forecasting of tropical storm frequency using a multi-model ensemble. *Quart. J. Roy. Meteorol. Soc.*, **132**, 647–666. **[3]**

von Storch H, Bürger G, Schnur R and von Storch J-S (1995) Principal oscillation patterns: A review. *J. Clim.*, **8**, 377–400. **[7]**

Wang G, Alves O, Zhong A, et al. (2004) POAMA an Australian ocean-atmosphere model for climate prediction, American Meteorological Society Symposium on Global Change and Climate Variations, Vol. 15. **[5]**

Weare BC and Nasstrom JS (1982) Examples of extended empirical orthogonal function analysis. *Mon. Wea. Rev.*, **110**, 481–485. **[7]**

Weber EU, Baron J and Loomes G (eds) (2000) *Conflict and Tradeoffs in Decision Making*, Cambridge University Press, Cambridge, 347 pp. **[2, 11]**

Vargish T and Mook DE (1999) *Inside Modernism: Relativity Theory, Cubism, Narrative*, Yale University Press, New Haven, CT, 185 pp. **[2]**

Wheeler M and Kiladis GN (1999) Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *J. Atmos. Sci.*, **56**, 374–399. **[6]**

WHO (2003a) Climate change and human health: Risks and responses. Geneva, World Health Organization (ISBN 92 4 156248 X, http://www.who.int/globalchange/publications/cchhsummary/en/). **[13]**

WHO (2003b) Methods of assessing human health vulnerability and public health adaptation to climate change. WHO Regional Office for Europe (ISBN 92 890 1090 8). **[13]**

WHO (2004) Public health responses to extreme weather and climate events, in Fourth Ministerial Conference on Environment and Health 2004: "The future for our children", WHO Regional Office for Europe, 2004 (EUR/04/5046269/15) **[13]**

WHO-UNICEF (2005) *World Malaria Report*. Roll Back Malaria. 294 pp. http://www.rbm.who.int/wmr2005/index.html **[13]**

WMO (2004) WMO statement on the status of the global climate in 2003. World Meteorological Organization, Geneva (WMO No. 966). **[13]**

Wright G and Goodwin P (eds) (1998) *Forecasting with Judgement*, Wiley, Chichester, 297 pp. **[2]**

Xie PP and Arkin PA (1998) Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. *J. Clim.*, **11**, 137–164. **[3, 6]**

Xu J-S and von Storch H (1990) Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. *J. Clim.*, **3**, 1316–1329. **[7]**

Yu Z-P, Chu P-S and Schroeder T (1997) Predictive skills of seasonal to annual rainfall variations in the US Affiliated Pacific Islands: Canonical correlation analysis and multivariate principal component regression approaches. *J. Clim.*, **10**, 2586–2599. **[7]**

Zhou YH, McLaughlin D and Entekhabi D (2006) Assessing the performance of the ensemble Kalman filter for land surface data assimilation. *Mon. Wea. Rev.*, **134**, 2128–2142. **[5]**

Index

Q

R

S