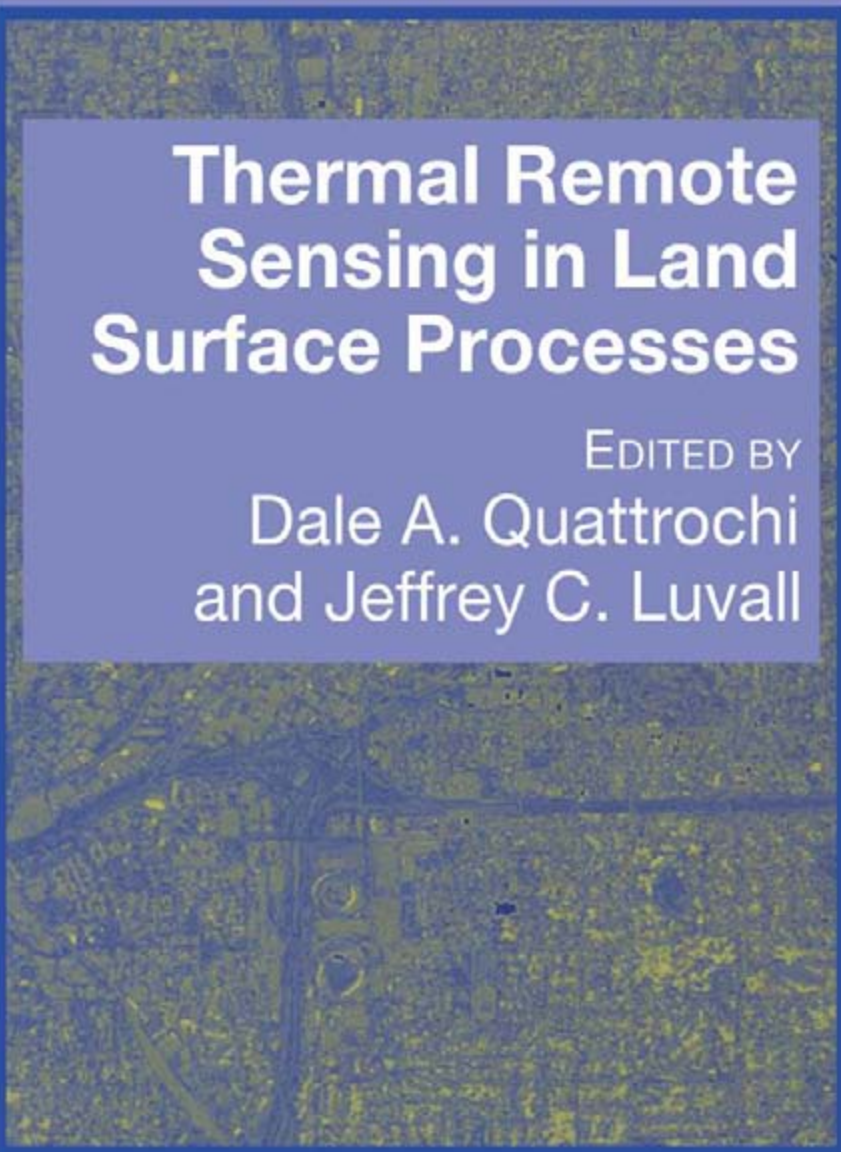
Thermal Remote Sensing in Land Surface Processes

EDITED BY

Dale A. Quattrochi
and Jeffrey C. Luvall



CRC PRESS

Also available as a printed book
see title verso for ISBN details

Thermal Remote Sensing in Land Surface Processes

Thermal Remote Sensing in Land Surface Processes

EDITED BY
Dale A.Quattrochi
and Jeffrey C.Luvall

CRC PRESS
Boca Raton London New York Washington, D.C.

This edition published in the Taylor & Francis e-Library, 2005.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

Library of Congress Cataloging-in-Publication Data

Catalog record is available from the Library of Congress

This is a work of the United States Government under the provisions of Title 17, Section 105 of the U.S. Code and, therefore, U.S. copyright protection is not available. Under U.S. law, no U.S. copyright may be assigned. U.S. Government works are in the public domain and may be used by members of the U.S. public without copyright restrictions.

The work shall not be used in any manner that would suggest or imply endorsement by NASA or any NASA employees. NASA provides no warranties of any kind, including but not limited to any warranty of mechantability or fitness for a particular purpose, and any warranty with respect to infringement of copyright or other rights of others; and nothing herein shall be construed to constitute the grant of any license or permission under any NASA patent, patent application, or other rights.

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

Visit the CRC Press Web site at www.crcpress.com

ISBN 0-203-50217-5 Master e-book ISBN

ISBN 0-203-57286-6 (Adobe eReader Format)

International Standard Book Number 0-415-30224-2

Contents

Biographical sketch of Dale A.Quattrochi

Dale A.Quattrochi is a Senior Research Scientist with the NASA Marshall Space Flight Center in Huntsville, Alabama, and has over 23 years of experience in the field of Earth science remote sensing research and applications. He received his PhD degree from the University of Utah, his MS degree from the University of Tennessee, and his BS degree from Ohio University, all in Geography. Dr Quattrochi's research interests focus on the application of thermal remote sensing data for analysis of heating and cooling patterns across the diverse urban landscape, which form the dome of elevated air temperatures over cities known as the urban heat island effect. He is also conducting research on the applications of geospatial statistical techniques, such as fractal analysis, to multiscale remote sensing data.

Dr Quattrochi is the recipient of numerous awards including the NASA Exceptional Scientific Achievement Medal, NASA's highest science award, which he received in 2001 for his research on urban heat islands and remote sensing. He is also a 2002 recipient of the Ohio University College of Arts and Science, Distinguished Alumni Award. Dr Quattrochi is the co-editor of *Scale in Remote Sensing and GIS* (with Michael Goodchild), published in 1997 by CRC/Lewis Publishers.

Dr Quattrochi is an adjunct faculty member in the Department of Geography and Anthropology at the Louisiana State University. He is also an adjunct professor in the Department of Plant and Soil Science and the Center for Hydrology, Soil Climatology and Remote Sensing at Alabama A&M University, and is an adjunct associate professor in the Department of Atmospheric Science at the University of Alabama in Huntsville.

Biographical sketch of Jeffrey C.Luvall

Jeffrey C.Luvall is currently employed by NASA as a senior research scientist at Marshall Space Flight Center. He holds a BS (1974, Forestry) and an MS (1976, Forest Ecology) from Southern Illinois University, Carbondale, IL, and a PhD (1984, Tropical Forest Ecology) from the University of Georgia, Athens, GA.

His current research involves the modeling of forest canopy thermal response using airborne thermal scanners on a landscape scale. He is also investigating the relationships of forest canopy temperatures and the evapotranspiration process. He has used remotely sensed surface temperatures to develop evapotranspiration estimates for eastern deciduous and tropical rain forests. These investigations have resulted in the development of a Thermal Response Number (TRN) which quantifies land surface's energy response in terms of kJ m^{-2} C^{0-1}, which can be used to classify land surfaces in regional surface budget modeling by their energy use. A logical outgrowth of characterizing surface energy budgets of forests is the application of thermal remote sensing to quantify the urban heat island effect. One important breakthrough is the ability to quantify the importance of trees in keeping the city cool. His current research involves alternate mitigation strategies to reduce ozone production through the use of high albedo surfaces for roofs and pavements and increasing tree cover in urban areas to cool cities. His recent work on urban heat islands has been the focus of several CNN, CBN, CBS Evening News, NBC, and ABC Discovery News programs during 1998. It was also featured in a November 23, 1998, *Newsweek* article "Blue Skies Ahead: Hot Ways to Cool Down Our Cities." He is also working closely with the Salt Lake City's 2002 Olympic Organizing Committee in revitalizing the city by planting greenways and high albedo surface materials. Invited by the *USSR Academy of Sciences* and the United Nations Environment Program to speak at the International Symposium on the State of the Art of Remote Sensing Technology for Biosphere Studies in Moscow (September 1989). Invited participant and co-authored a paper at the *Space Conference of the*

Americas in San Jose, Costa Rica, March 12–16,1990, and an invited delegate in August 1991. *Steering Committee for organizing a symposium* "Thermal Remote Sensing of the Energy and Water Balance over Vegetation in Conjunction with Other Sensors," La Londe Les Maures, France, September 1993. *Organized symposium* at the Intecol 1994 meeting at Manchester, England, "A Thermodynamic Perspective of Ecosystem Development" (with J.Kay and E.Schneider). Appointed to serve a three-year term (1994–1997) on the *La Selva Advisory Committee* by the Organization for Tropical Studies (OTS). La Selva is a biological research field station in Costa Rica, funded by the National Science Foundation. OTS is a consortium of 50 US and international universities that manage several field stations and courses in Costa Rica.

Selected awards

1999–2000: Walter Bean/Canada Trust Visiting Professor of the Environment, University of Waterloo, Faculty of Engineering. Sigma Xi, The Scientific Research Society of America Gulf Coast Chapter's Kaminski Award 1990. Given for the best scientific research paper published in a peer review journal during 1989. NASA's Marshall Space Flight Center's Director's Award for outstanding CDDF project, 1996.

Contributors' biographies

Martha C.Anderson is an Assistant Scientist in the Department of Soils at the University of Wisconsin, Madison, Wisconsin. She received her PhD in Astrophysics from the University of Minnesota in 1993, and then shifted her research focus towards the use of remote sensing data in deducing the land-surface energy balance. She has collaborated in developing a suite of related soil-plant-atmosphere models designed for practical application in agricultural and hydrologic monitoring applications, which utilize a combination of readily available satellite and surface synoptic data. Other interests include studying land use impacts on groundwater quality and local climate.

Julia A.Barsi is a Calibration Analyst in the Landsat Project Science Office at NASA/Goddard Space Flight Center, Greenbelt, Maryland. She received her BS and MS in Imaging Science from the Rochester Institute of Technology. After completing her Master's thesis on vicarious thermal calibration of Landsat ETM+, she joined the Landsat Project Science Office.

Scott D.Brown is on the Research Staff at the Digital Imaging and Remote Sensing (DIRS) Laboratory in the Center for Imaging Science at the Rochester Institute of Technology, Rochester, New York.

Toby Carlson is a Professor of Meteorology in the Department of Meteorology at The Pennsylvania State University in University Park, Pennsylvania. He has taught courses in remote sensing, synoptic meteorology, hydrology, oceanography, and boundary layer and land surface processes. He has done extensive research on the measurement and mathematical modeling of land surface properties.

Vicente Caselles is a Professor of Applied Physics and Head of the Thermal Remote Sensing Group at the Universitat de Valencia, Valencia, Spain. He has 20 years of expertise in the physical processes involved in the temperature measurement using remote sensing techniques that has been documented through 20 books, 15 doctoral theses, 80 papers in

international journals, 60 conference papers, and 30 reports. He collaborated with the European Space Agency (ESA) as member of the Advisory Group for the Land-Surface Processes and Interactions Mission. At present, he is the Chairman of the Spanish Remote Sensing Society and Chairman of the Spanish Atmosphere and Climate Programme.

Kevin P.Czajkowski is an Associate Professor in the Department of Geography and Planning at the University of Toledo, Toledo, Ohio. His research interests include remote sensing and land/atmosphere interactions. In addition, he works with local decision makers to develop remote sensing applications.

Ralph O.Dubayah is an Associate Professor in the Department of Geography at the University of Maryland, College Park, Maryland. He has developed applications for Lidar remote sensing and has additional research interests in the hydrological modeling and applications using remotely sensed data.

Philippe Duthil was initially a Space Systems Engineer (Ecole Nationale Supérieure de l'Aéronautique et de l'Espace Toulouse France, 1979). He is currently responsible for the development of new applications and Earth Observation products in the area of Environment and Water Management for Astrium in Toulouse, France. He rapidly specialized in the conception of Earth Observation systems and progressively evolved from space system engineering to remote sensing applications developments. His experience includes the analysis of overall space system requirements and performance assessment (SPOT series, HELIOS series, various projects). He also developed a simulator of space borne imagers, which can be used to simulate earth observation data and products and hence serve as a tool for products definition. He has been conducting several research and development projects under the 4th and 5th European Commission Framework program, in the areas of environment and agriculture, such as the MUST study of an infrared mission for water resources and risk management.

Roydon A.Fraser is currently a Professor on the Mechanical Engineering faculty at the University of Waterloo, Waterloo, Ontario, Canada. He obtained his undergraduate degree in Engineering Physics at Queen's University, Kingston, Ontario, in 1983 and his Master's and PhD in the Department of Mechanical and Aerospace Studies at Princeton University, Princeton, New Jersey, in 1985 and 1989, respectively. His research efforts include studies of ecosystem thermodynamics, the exergy analyses of complex systems, turbulent combustion, and non-intrusive combustion diagnostics as applied to internal combustion engines, alternative fuel vehicle development with particular emphasis on natural gas and ethanol,

methanol fuel cells, glazing system heat transfer studies, and energy utilization and conversion in general.

Robert R.Gillies is an Associate Professor in the Departments of Aquatic, Watershed and Earth Resources, and Plants, Soils and Biometeorology at Utah State University in Logan, Utah. He obtained his PhD at the University of Newcastle-upon-Tyne, England, in 1995. He publishes in the areas of remote sensing and meteorology and recently has been f focusing on interdisciplinary research in the application of remote sensing in ecological assessments of urbanization and epidemiological issues, such as vectorborne disease detection. His recent publications include "Determination of scaling characteristics of AVHRR data with wavelets: application to SGP97" with Nathaniel Brunsell, published in the *International Journal of Remote Sensing,* "The simulation of canopy transpiration under doubled CO2: The evidence and impact of negative feedbacks on transpiration in two 1-D soil-vegetation-atmosphere-transfer (SVAT) models" with Jonathon Gottschalk and Toby Carlson in *Agricultural and Forest Meteorology,* and "An application of remotely derived climatological fields for risk assessment of vector-borne diseases: A spatial study of filariasis prevalence in the Nile delta, Egypt" with Kate Crombie, Ray Arvidson, Paul Brookmeyer, and Gary Weil, in *Photogrammetric Engineering and Remote Sensing.*

Scott J.Goetz is an Assistant Research Scientist in the Department of Geography at the University of Maryland, College Park, Maryland. His research interests include ecological remote sensing and application projects in collaboration with decision makers.

Samuel N.Goward is a Professor of Geography at the University of Maryland, College Park, Maryland. He is the Landsat-7 Science Team Leader. He additionally has research interests in thermal infrared remote sensing including the derivation of environmental variables from satellite data.

Ray Hardy was an MA student in the Department of Geography at the University of Oklahoma in Norman, Oklahoma, at the time of authorship. He now resides in Oklahoma City.

Karen Humes is an Associate Professor in the Department of Geography at the University of Idaho in Moscow, Idaho. Humes obtained her PhD in Hydrology from the University of Arizona in 1992. She has worked at the NASA Jet Propulsion Laboratory and the USDA Agricultural Research Hydrology Laboratory. She joined the faculty of the Department of Geography at the University of Oklahoma in 1995 and moved to the University of Idaho in 1999. Her research interests include the estimation of spatially distributed surface energy fluxes with a combination of ground

and remotely sensed data, quantifying the spatial variability in land surface characteristics that control land/atmosphere interactions, improvements of measurements from *in situ* soil moisture sensors, and, most recently, the use of remotely sensed data to monitor forest health.

James J.Kay is an Associate Professor of Environment and Resource Studies at the University of Waterloo (with cross-appointments in Systems Design Engineering, Geography, Management Sciences, the School of Planning, and the School of Rural Planning and Development, University of Guelph), Waterloo, Ontario, Canada. He did his undergraduate work in Physics at McGill University and received his PhD in 1984 in Systems Design Engineering at the University of Waterloo. His PhD thesis was entitled Self-Organization in Living Systems. His research over the last 25 years has focused on complexity and systems theory and their application to the development of an ecosystem approach as a way of understanding and managing our role in the biosphere. His research activities span the full spectrum from the theoretical and epistemological basis for an ecosystem approach, to the formulation of ecosystem based environmental policy, the development of ecosystem monitoring programs, to on the ground ecosystem planning both in the context of urban, industrial, and natural ecosystems and the greening of institutions.

Yann H.Kerr is a Research Scientist at the Centre d'Etudes Spatiales de la BIOsphère (CESBIO) in Toulouse, France. He received an engineering degree from Ecole Nationale Supérieure de l'Aéronautique et de l'Espace (Radar and telecommunications), an MSc in optoelectronics from Glasgow University in E&EE, and PhD from Université Paul Sabatier in Physics and remote sensing. From 1980 to 1985 he was employed by CNES. In 1985 he joined LERTS. He spent 19 months at JPL, Pasadena, in 1987–88. He has been working at CESBIO since 1995. Kerr's fields of interest are in the theory and techniques for microwave and thermal infrared remote sensing of the Earth, with emphasis on hydrology and vegetation monitoring from space. He was involved in the organization of the HAPEX Sahel Experiment in 1992, and in the SALSA experiment in Mexico. He was also an EOS principal investigator (interdisciplinary investigations), and he was the science lead on the MIRAS project for ESA, and is the Lead-Investigator of the SMOS mission.

William P.Kustas is a Research Hydrologist with the United States Department of Agriculture (USDA)-Agricultural Research Service (ARS), Hydrology and Remote Sensing Laboratory in Beltsville, Maryland. He has been a Research Hydrologist with the USDA-ARS since receiving his PhD from Cornell University in 1986. His research encompasses all phases of hydrology, with a main emphasis on understanding the processes involved in the transfer of energy from the earth's land surface to the lower

atmosphere at both micro- and macro-scales. This has led to his research focusing on the application of remote sensing data for modeling spatially distributed water and energy fluxes from local to regional scales, requiring the integration of a broad range of disciplines. The interdisciplinary nature of this work and the need for large-scale spatial data led to the participation and coordination of large-scale multidisciplinary field studies covering a wide range of ecosystems and climates. These data have provided information critical in the development and testing of energy balance models and scaling methodologies recently published in the *Journal of Hydrometeorology* and *Water Resources Research*.

Jean Pierre Lagouarde is a Scientist in the Department "Environnement et Agronomie" at INRA (Institut National de la Recherche Agronomique) in Bordeaux, France. He graduated as an engineer from the Ecole Centrale de Lyon in 1975, and obtained a thesis in 1979. He joined INRA, the French Institute of Agronomic Research, in 1981. His field of research deals with the monitoring of surface fluxes using remote sensing, with a special emphasis on forest areas. He is also involved in the problem of spatial integration, and the development of new methodologies (scintillometry) for validating area-averaged fluxes.

Charles A.Laymon is currently a Research Scientist with the Universities Space Research Association (USRA) at the Global Hydrology and Climate Center within the National Space Science and Technology Center in Huntsville, Alabama. His research interests include hydrologic modeling and scaling of hydrologic processes, remote sensing of land surface properties and processes, such as soil moisture, vegetation parameters, surface temperature, and energy fluxes, and in the assimilation of these data in hydrologic and climate models for a wide variety of applications. He received the BS degree in Geology with honors in 1982 from St Lawrence University, Canton, New York, and the PhD degree in Geological Sciences in 1988 from the University of Colorado, Boulder, Colorado.

M.Susan Moran is a Research Hydrologist and Research Leader at the United States Department of Agriculture (USDA)-Agricultural Research Service (ARS), Southwest Watershed Research Center in Tucson, Arizona. She is also an Adjunct Professor in the Department of Soil, Water and Environmental Science at the University of Arizona. Her area of research is the development of theory, principles, and methods for estimation of soil moisture and evapotranspiration, detection of physical and biological stress in plants, and evaluation of energy balance and water balance at local and regional scales utilizing a combination of models and remote sensing techniques. She has made significant contributions to the three main types of remote sensing: visible-infrared, thermal and radar, and is also doing practical research to bring this technology to the average farmer or rancher.

Theresa Mulhern was a Research Scientist at the University of Toledo in Toledo, Ohio, at the time of her co-authorship. She currently resides in Denver, Colorado.

John Murtagh is a Business Development Manager at Infoterra Ltd. At the time of this research, John was a Remote Sensing Consultant at the National Remote Sensing Centre (NRSCL), Farnborough, United Kingdom.

Françoise Nerry is a Research Scientist at the Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection in Strasbourg, France. She received the Diploma from the Ecole Nationale de Physique de Strasbourg in 1984 and her PhD from Strasbourg University in 1988. She visited CARTEL (University of Sherbrooke Canada) in 1985 and NASA JPL (California, USA) in 1989–1990. Her field of interest focuses on thermal infrared data in laboratory and field experiments including spectro-radiometry. She is also working in the field of remote sensing for analysis of thermal infrared satellite data and emissivity and land surface temperature retrievals.

John M.Norman is the Rothemel Bascom Professor of Soil Science and also Professor of Atmospheric and Oceanic Science at the University of Wisconsin, Madison, Wisconsin. Following his PhD in 1971 from the University of Wisconsin-Madison, he was an Associate Professor of Meteorology at the Pennsylvania State University until 1978, and Professor of Agronomy at the University of Nebraska-Lincoln until 1988. He conducts biophysical research involving studies of the interaction between plants and their environment including measurements of soil, plant, and atmospheric characteristics and integrative modeling of the soil-plant-atmosphere system. Applications to ecology, agriculture, forestry, and meteorology have included plant productivity and water use efficiency, integrated pest management, irrigation water use, precision agriculture, agro-chemical leaching losses, remote sensing, and measurement and modeling of soil surface carbon dioxide fluxes. His recent research focuses on the sustainability of agricultural production and the importance of soil in the spatial and temporal distribution of crop production and environmental consequences. He is co-author of a text entitled *Introduction to Environmental Biophysics* published by Springer-Verlag in 1998.

Catherine Ottlé has been a Research Scientist at the Centre National d'étude des Environnements Terrestre et Planétaires (CETP) in Vélizy, France, since 1985. She received the PhD degree in Physics from the University of Paris, Paris, France, in 1983. Her research interests include the applications of remote sensing to the study of the land surface processes

for hydrology and vegetation monitoring at local and regional scales. In particular, she is working on the development of soil-vegetation-atmospheretransfer (SVAT) modeling for remote sensing data assimilation and on large-scale hydrology on the occasion of her participation to HapexMobilhy and Alpilles-ReSeDA programs and to ERS-ESA pilot projects. These activities are done in close collaboration with other Public Research Institutes or Organizations, aerospace companies and industry of the added value in remote sensing. She is also involved in thermal infrared remote sensing research activities, such as the modeling of radiative transfer in the atmosphere and development of atmospheric and emissivity correction methods, land and sea surface temperature determination techniques.

Stephen D.Prince is a Professor of Geography at the University of Maryland, College Park, Maryland. His main area of interest is ecological modeling using remotely sensed data.

John Prueger is a Research Scientist with United States Department of Agriculture (USDA)-Agricultural Research Service (ARS), National Soil Tilth Laboratory located in Ames, Iowa. He received his BS degree in 1981 and an MS degree in 1986 from California State University-Fresno, and his PhD in 1991 from Utah State University. His current research projects include atmospheric impacts of agricultural management practices, field scale evaluation of nitrogen soil water crop growth interactions, and carbon dioxide exchanges in Midwest cropping systems and nitrogen in corn production systems. He also works with satellite images and Model Simulations to study seasonal water balances of Ozark hill slopes.

David A.J.Ripley worked as a Research Associate with Professor Toby Carlson at The Pennsylvania State University and is now a computer systems engineer at Indiana State University, Terre Haute, Indiana.

John R.Schott is the Frederick and Anna B.Wiedman Professor of Imaging Science and Head of the Digital Imaging and Remote Sensing (DIRS) Laboratory in the Center for Imaging Science at the Rochester Institute of Technology, Rochester, New York. Schott received his PhD from the State University of New York College of Environmental Science and Forestry and his BS in Physics from Canisius College. Following 8 years at the Cornell Aeronautical Labs/CALSPAN, he joined the Rochester Institute of Technology in 1980. His career has focused on development of improved instrumentation and algorithms for extraction of information from remotely sensed data. He is the author of *Remote Sensing: The Image Chain Approach* published by Oxford University Press in 1997. He has served as principal investigator on numerous remote sensing programs for NASA and the Defense/Intelligence community.

Thomas J.Schmugge is a Research Physical Scientist with the United States Department of Agriculture (USDA)-Agricultural Research Service, Hydrology and Remote Sensing Laboratory in Beltsville, Maryland, working on the application of remote sensing techniques to the study of land surface hydrologic processes. He is a recognized expert in microwave measurements of soil moisture and other soil properties. He has a PhD in Physics from the University of California at Berkeley (1965). Prior positions include Assistant Professor of Physics at Trinity College, Hartford, Connecticut, and 15 years in the Hydrological Sciences Branch at NASA's Goddard Space Flight Center. His research interests include the use of microwave and thermal infrared remote sensing techniques to observe such parameters as soil moisture, surface temperature, and evapotranspiration. He is a member of the Joint US/Japan ASTER science team. He is a Fellow of the AGU and the IEEE.

David Shirey was a student at the University of Maryland at the time of authorship. He currently works for a computer consulting firm in the Washington DC area.

Stephen Stadler is a Professor of Geography at Oklahoma State University, Stillwater, Oklahoma. He received his PhD from Indiana State University in 1979, his MA in Geography from Miami University in 1976, and his BS in Educational Social Studies (cum laude) from Miami University in 1973. His research interests include both physical and social aspects of applied climatology. He has recently developed climatologies for wind generation in Oklahoma.

Patrick Starks is a Research Scientist with United States Department of Agriculture (USDA)-Agricultural Research Service (ARS), Grazinglands Research Laboratory located in El Reno, Oklahoma. His current research projects include the integration of climate forecasts into management and resources conservation tools, monitoring and evaluating runoff in several Oklahoma reservoirs and streams; and analysis of the integrated effects of management, land use, and climate on regional water resources.

Bekele Temesgen is an Associate Land and Water Use Analyst at the California Department of Water Resources, Sacramento, California. He obtained his PhD in Biometeorology from the Utah State University in 2001. His interest in remote sensing started while performing research on evapotranspiration as a graduate research assistant at the Utah State University. He studied the effects of terrain and biome on geophysical variables by coupling a boundary layer model with airborne data as part of his PhD dissertation. He is currently exploring the potential for coupling remotely sensed satellite data with a network of agricultural weather stations to map evapotranspiration for the State of California.

Alain Vidal is Head of the European and International Affairs Office of the Cemagref located in Montpellier, France. He graduated as an agricultural and environmental engineer from Engref, Paris, France, in 1985, and received his PhD in water sciences and remote sensing from the University of Montpellier, France, in 1989. He started his professional activity in Morocco where he worked as junior engineer for an irrigation agency (1986–88). He was project leader (1988–92) and then research leader at the Cemagref-Engref remote sensing laboratory (1992–96). From 1996 to 1998, he was senior scientist at the Cemagref Irrigation research unit, in charge of international networks and of remote sensing and GIS applications. His expertise covers bioclimatology, land surface fluxes, remote sensing applied to irrigation management and to forest fire risk assessment, and water conservation in agriculture. During his career, Vidal has been involved in research and development projects in Morocco, Pakistan, Mexico, and Ecuador, and via scientific cooperation with various universities and research institutes in Europe and in the United States. He has authored and co-authored more than 30 refereed papers or invited conferences, and has been editor of four scientific books or workshop proceedings. He is very active in ICID, the International Commission on Irrigation and Drainage, where he is Vice-Chairman of the Permanent Committee of Technical Activities.

Anita Walz was a student at the University of Maryland at the time of authorship. She currently resides in Rochester, New York.

Antonio Yagüe is the Founder and Chief Scientist of INFOCARTO, Spain.

Preface

The genesis of this book began on the sunny shores of southern France in September 1993. For five days in the delightful Mediterranean coastal town of La Londe Les Maures, a gathering occurred of a group of scientists interested in furthering both the understanding and use of thermal infrared (TIR) remote sensing data for analysis of land surface processes. Here the workshop on Thermal Remote Sensing of the Energy and Water Balance Over Vegetation in Conjunction with Other Sensors took place with the intent of assessing what the state-of-the-art of TIR remote sensing data was, and discussing how TIR data could be more widely used in research related to the analysis and modeling of land surface energy fluxes and land surface processes by the larger scientific community. Those in attendance at this workshop (including the editors of this book) were all of the same opinion that TIR data offered a tremendous amount of information on surface energy flux characteristics and dynamics, yet these data were vastly underutilized in land surface processes research. As noted in the Executive Summary and Overview of the La Londe workshop:

> The problem in demonstrating the value of thermal remote sensing lies in (a) the difficulty of calibration and correction of the measured radiance to consistent physical qualities, (b) the limited ability to estimate accurately the surface energy fluxes over complex terrain, which might consist of a mixture of vegetation (including forests), sloping surfaces, water bodies, bare soil and urban landscapes, and (c) the detection and removal of the effect of clouds. Nevertheless, despite reservations on the utility of thermal infrared measurements, many scientists think that a multispectral approach to remote sensing, including thermal infrared temperature measurements, will prove to be essential.
>
> (La Londe Workshop 1993)

Since the convening of the La Londe Workshop, there has been a substantial increase in both the amount and availability of TIR remote sensing data,

particularly from satellites. This is particularly true with the advent of the National Aeronautics and Space Administration's (NASA) Terra Earth Observing System that has a number of TIR sensors associated with it, such as the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Enhanced Thematic Mapper Plus (ETM+), and Moderate-Resolution Imaging Spectroradiometer (MODIS) specifically designed for Earth observation and analysis (see the NASA Terra website at http://terra.nasa.gov and the Landsat 7 website at http://landsat.gsfc.nasa.gov for more information on these sensors). Despite the increased availability of TIR, however, we see where the wide application of these data to land surface processes has been limited for five fundamental reasons:

1 What examples exist of the application of TIR data for analysis of land surface processes are fragmented across the literature (e.g. forestry, geology, geography, meteorology, climatology) and, thus, there is an absence of a strong or concerted focus for using TIR data specifically in land surface processes research.
2 Because of this fragmentation of references, TIR data are little understood from a theoretical and applications perspective across the Earth science research community.
3 The theory of TIR remote sensing is perceived as being recondite and difficult to understand, which severely limits the application of these data to only those who have the desire, background, and need to work through the basics of thermal theory.
4 The perceived difficulties in calibration and correction of TIR data to obtain consistent physical measurements of land surface properties.
5 Despite the increased availability of TIR data from satellites, there is still the perception that TIR data are inaccessible or difficult to obtain.

It is our purpose here to assist in overcoming these misconceptions on the uses and applications of TIR data for land surface processes research. By doing so, we hope to promote wider use of TIR data for analysis of land surface processes for more robustly examining landscape and landatmosphere dynamics in Earth system science studies. It is our intent through the material presented in this volume to:

1 Present studies where TIR data have been applied to deriving quantitative measurements of the fluxes and redistribution of surface thermal energy balance characteristics for developing a better understanding of land surface process and land-atmosphere interactions.
2 To promote the wider usage of TIR data in research and modeling to further our understanding of the role of thermal energy balance and surface energy fluxes in driving land processes.

3 To elucidate both the prospects and problems of using TIR data in land processes research that will be useful to those wishing to employ these data as a major component in Earth system science research.
4 To illustrate the virtues and importance of TIR data in remote sensing research of the land surface to facilitate the development of new and improved satellite and airborne TIR remote sensing systems in the future.

Thus, it is our overall intent in preparing this book to fill a significant void in the remote sensing literature and also to develop a more well defined niche for furthering the use of TIR data in future research on land surface processes. Above all, it was our purpose to make this a "how to" book as much as possible—one that illustrates how TIR data have been used in, or applied to, land surface processes research and to assess the utility of new TIR sensors for analysis of surface energy flux parameters and characteristics—rather than being a volume that just discusses the prospects and problems of using TIR data. We trust with the chapters included in this volume that we have "hit our mark" and that readers will find this book an informative and useful reference in exploring the utility of TIR data in their own research applications and initiatives.

We wish to thank the NASA George C.Marshall Space Flight Center (NASA/MSFC) in Huntsville, Alabama, for the support given to us throughout the development of this book. NASA/MSFC has both permitted, and encouraged us, to pursue the editing of this volume as part of our day-today work activities—which we sincerely appreciate. Additionally, we are indebted to the NASA Earth Science Enterprise as a whole for providing us with project funding for various TIR research endeavors that have spurred us on to produce a book of this type as a resource for the entire Earth science community. Moreover, we are most grateful for the diligence, patience, and contributions provided by the authors of the chapters that are included in this book. Without their interest and support, this book would never have come to fruition. We must, too, extend our heart-felt thanks to our families for what they have had to endure throughout the organization and compilation of this book. Our absences away from them for activities related to the development of this book, such as for meetings or conferences, has been definitely noted—but accepted—by our respective families. We cannot adequately express our appreciation to them for their forbearance and for their continued love and support throughout this endeavor. For this, we wish to dedicate this book to them.

We must also give our humblest and deepest thanks to the "unsung heros" of this book—the reviewers—for their thoughtful and insightful comments on each of the chapters. Their review comments and suggestions on content, theory, and overall structure of each chapter are extremely appreciated by both ourselves and the chapter authors, in helping to make this a technically

and scientifically sound volume. We wish to acknowledge the reviewers of the chapters for this book as listed below:

Reviewers

Marvin Bauer
Toby Carlson
William Capehart
Richard Crago
William Crosson
Kevin Czajkowski
George Diak
Mark Fiedl
Narendra Goel
Sam Goward
J.L.Hatfield
C.Ross Hinkle
Fred Huemmrich
And three anonymous reviewers
Karen Humes
Rob Kaiser
William Kustas
Charles Laymon
Massimo Menenti
John Norman
Howard Odum
Thomas Schmugge
James Smith
Francesco Tubiello
Robert Ulanowicz
Craig Wiegand
Stephen Yool

References

La Londe Workshop (1993) Workshop on thermal remote sensing of the energy and water balance over vegetation in conjunction with other sensors. Organized by The Pennsylvania State University, College of Earth and Mineral Sciences, Earth System Science Center, University Park, Pennsylvania, the Centre d'etude des Environnements Terrestre et Planétaires (CETP), Centre Universitaire Technique, Vélizy, France, and the CEMAGREF-ENGREF, Remote Sensing Lab, Montpellier, France. La Londe Les Maures, France, September 20–23, CEMAGREF-ENGREF, Montpellier Cedex 5, France, 330 pages.

Introduction

The Earth science research community and even the general public are widely familiar with thermal infrared (TIR) remote sensing data, but the virtue and applicability of these data still remain in many ways an enigma. The research community is aware of TIR data from scientific publications and presentations that have described the analyses and results from working with TIR data for Earth science-related research; for example, in providing measurements of surface thermal properties of geologic materials or of land surface thermal energy fluxes for forest, agricultural, or other landscape attributes. The public is aware of TIR data from television broadcasts that show these data as part of weather forecasts and even from night-time thermal images showing heat loss from their homes. Still, the utility of TIR remote sensing data has not been fully realized by the larger scientific community because of a number of perceptions—or misconceptions—that have prevented the broader usage of TIR data for research on Earth processes, particularly those related to land surface processes. With the launch of the NASA Terra suite of Earth remote sensing instruments in 1999, a number of which have thermal IR sensors, TIR data are becoming much more readily available than in the past. With the increasing availability of these data, it appears that TIR data are poised to become a major source of quantitative and qualitative information on land surface processes and for their characterization, analysis, and modeling.

Land surfaces processes may be loosely defined as those attributes, exchanges, and relationships that contribute to the overall functioning of the concomitant physical, biophysical, and hydrologic interactions that come together to form "the landscape." These include the processes that occur across (or even just below) the land surface as well as between the land surface and the atmosphere. Using this definition as a foundation or a baseline, there are two fundamental reasons why TIR data contribute to an improved understanding of land surfaces processes: (a) through measurement of surface temperatures as related to specific landscape and biophysical components; and (b) through relating surface temperatures with energy fluxes for specific landscape phenomena or processes (Quattrochi and Luvall 1999). The

magnitude and distribution of longwave thermal energy emitted from surfaces across the landscape are primary components that can be measured using TIR data. Measurement of surface temperature or thermal energy provides quantitative (and qualitative or descriptive) information on one of the basic inputs to the overall energy budget of land surface processes. Given the role of thermal energy responses as part of the overall input and output of land surface and land-atmosphere interactions, understanding how thermal energy is partitioned across a landscape and determining the magnitude or variability in surface temperatures emanating from various landscape elements (e.g. forest, crops, water, pasture, urban) is essential to defining the mechanisms that govern land surface processes (Quattrochi and Luvall 1999). The energetic dynamics that drive these land surface and landatmosphere interactions may be defined as: (a) the coupling of extant energy balances with the environment; (b) the level of energy inputs (and subsequent outputs); (c) the kinds of energy transformations that occur, especially those that are biologically controlled; and (d) the mix of energy outputs that can be regarded as yields from an ecosystems perspective (Quattrochi and Luvall 1999). Variability in the magnitude of surface thermal energy for specific land surfaces or processes often affect the density, dynamics, and importance of other energy fluxes linked to specific landscape characteristics (e.g. evapotranspiration, nutrient cycling) (Miller 1981; Quattrochi and Luvall 1999).

We have provided in an earlier reference (see Quattrochi and Luvall 1999) the results from past research that has either directly or indirectly demonstrated the potential application of TIR remote sensing data to land surface processes research, particularly as related to the discipline of landscape ecology. Here we identified six general thematic areas where TIR remote sensing data have been applied to the analysis of landscape attributes or land surface processes: (a) landscape characterization; (b) thermal inertia and landscape analysis; (c) estimation of energy fluxes; (d) evaporation/evapotranspiration/soil moisture; (e) quantification of energy balance or energy flux; and (f) forest energy exchange. In this book, we wish to build upon this earlier reference and provide explicit examples of how TIR remote sensing data can further elucidate the analysis of land surface processes within the general purview of these six thematic areas. By doing so, we wish to provide further quantitative and descriptive evidence on how TIR data have and will continue to be of fundamental importance to the analysis and modeling of land surface processes. Additionally, we wish to provide more credence for the development and launching of satellite TIR remote sensing systems that are better calibrated, collect data at higher spatial resolutions (e.g. $\leq$20m), and have the capability to measure temporal dynamics of surface energy fluxes (e.g. diurnal) than what is presently available from satellite platforms, to enable more robust quantitative analysis and modeling of land surface processes. Although current TIR satellite systems onboard or

associated with the NASA Terra platform, such as the Landsat ETM+, ASTER, and MODIS, provide better spectral, and to some extent spatial, resolutions than could be obtained in the past, there is still a dire need to have satellite-based TIR systems that combine multispectral thermal radiometric characteristics with landscape-scale spatial resolutions; that is, spatial resolutions that are able to discern and quantify land processes that operate at multiple spatial scales from very fine (e.g. ≤1 m) to local (e.g. ~10–20m) and regional (~30–100m) (Quattrochi and Goel 1994). Moreover, the TIR data available from satellites today only capture "snapshots" of the surface thermal energy regime at a particular instant in time. What is critically needed are satellite-based TIR sensors that are able to collect data on surface thermal energy dynamics on a short-term (e.g. diurnal) or even a continuous basis (e.g. hourly) to provide inputs for modeling the temporal variability of thermal energy dynamics. This then is the premise for this book: to illustrate the critical importance of TIR remote sensing data in the analysis and modeling of land surface processes, and to stimulate the development of new TIR satellite and airborne remote sensing systems that are needed to advance research in examining the inter-relationships of thermal energy responses within the overall foci of Earth system science research.

A brief sojourn though the contents

The contributors to this book have done an excellent job in building upon, and advancing, the basic tenets defined at the La Londe workshop as noted in the Preface (La Londe 1993) by providing updated overviews of the prospects—and challenges—of utilizing TIR data for analysis of land surface processes. More importantly, the contributors have presented case studies that directly show the great applicability of TIR data for the quantification and analysis of surface energy balance characteristics that exist as the driving force in land surface and land-atmosphere energy exchanges. Within this perspective, the contributions to this volume have been organized into three overarching categories: (a) TIR Data for Assessment and Quantification of Surface Energy Fluxes and Soil Moisture; (b) TIR Data for Assessment of Ecosystem Health; and (c) TIR Instruments and Calibration.

Kevin Czajkowski and his colleagues (Chapter 1) lead off the first group of seven chapters embodied within Part I on Thermal-Infrared Data for Assessment and Quantification of Surface Energy Fluxes and Soil Moisture. Here, they report on the advances in using satellite TIR remote sensing data to derive environmental variables, specifically surface temperature, near-surface air temperature, and near-surface water vapor. Thermal data from the US National Oceanic and Atmospheric Administration's (NOAA), Advanced Very High Resolution Radiometer (AVHRR) series of satellites constitute the basis for their overview and analyses. They discuss the rudiments

of interpreting TIR signals, derivation of radiometric surface temperature (T_s), air temperature (T_a), and near-surface water vapor from AVHRR data, as well as illuminating the challenges associated with deriving these variables from AVHRR data. Additionally, Czajkowski *et al.* discuss the applications of TIR methods to sensors onboard the NASA Terra system of Earth remote sensing instruments, such as the ASTER and MODIS sensors.

Chapter 2 in Part I is a comprehensive assessment by Yann Kerr and his other co-authors from France on land surface temperature retrieval techniques and applications using data from the NOAA AVHRR satellite. They provide an excellent overview of the theoretical background associated with temperature retrieval and the problems encountered with using these methods. They then present a review of existing land surface temperature retrieval algorithms and discuss the theoretical and practical considerations that must be considered in applying each of these algorithms. These authors go on to examine one of the real "challenges" in analyzing TIR data—that of deriving emissivity values. Here, they review a number of methods that have been used for emissivity retrieval from AVHRR data and discuss the uncertainties or errors that are associated with these methods. Atmospheric water vapor retrieval is an important component in algorithms that are used to derive temperatures from AVHRR data, as well as being an important component in correcting for the effects of the atmosphere in accurately measuring land surface temperatures from these and other TIR satellite remote sensing data. Kerr *et al.* present an overview of the existing methods used for assessing water vapor retrieval and the subsequent errors associated with these algorithms. These authors also present the results from an inter-comparison of the algorithms discussed in their chapter and examine the errors and related issues in using these techniques. Finally, Kerr and his colleagues discuss the potential applications of land surface temperature retrieval methods from AVHRR and other satellite TIR data.

Chapter 3 in Part I by Karen Humes, Ray Hardy, William Kustas, John Prueger, and Patrick Starks illustrates the utility of TIR data for mapping spatially distributed information on a number of key land surface characteristics and state variables that control the surface energy balance. They provide results from an investigation that used data from the Landsat Thematic Mapper in conjunction with a relatively simple "snapshot" model, to compute spatially distributed values of net radiation, soil heat flux, sensible heat flux, and latent heat flux over the Little Washita Experimental Watershed located in south-central Oklahoma, USA. The key information provided by remotely sensed data in the model included surface temperature, land cover type, and estimates of vegetation density. Thus, remotely sensed data were important for quantifying the components of the surface energy balance at a landscape-scale to better understand the exchange of heat and moisture between the land surface and the lower atmosphere.

Charles Laymon and Dale Quattrochi (Chapter 4) demonstrate a method for using Landsat Thematic Mapper data to estimate instantaneous regionalscale energy fluxes over an arid valley in eastern Nevada, USA. Here, pointbased models of surface energy and water balance fluxes were applied to individual pixels of a Landsat Thematic Mapper scene over the study area. Although the method used to estimate these instantaneous fluxes requires certain assumptions be made about the spatial distribution of several physical parameters, the results from this analysis and modeling suggest that it is possible to scale from point measurements of environmental state variables (i.e. net radiation flux, surface heat flux, sensible heat flux, and latent heat flux) to regional estimates of energy exchange to obtain an understanding of the spatial relationship between these fluxes and landscape variables.

Chapter 5 by Robert Gillies and Bekele Temesgen also focuses on the derivation of biophysical variables from thermal and multispectral remote sensing data. This chapter focuses on the use of Soil-VegetationAtmosphere-Transfer (SVAT) models as coupled with remote sensing data, so that biophysical variables are derived via an inverse method. Gillies and Temesgen focus on the application of an SVAT technique that employs the "triangle method" to derive estimates of vegetation cover and surface radiant temperature. The triangle method is by definition a multispectral remote sensing method that combines measurements of surface radiant temperature (T_o) and reflectance in the red and near-infrared portions of the electromagnetic spectrum. The reflectance measurements are used to calculate the Normalized Difference Vegetation Index (NDVI). NDVI is then plotted as a function of T_o to evaluate the relationship between these two variables, as well as providing an overlaying index of moisture availability to establish a "warm edge" and "cold edge" index. The triangle method within the overall scheme of an SVAT model, therefore, presents itself as a technique for inversely deriving biophysical variables relating to fractional vegetation cover, moisture availability, evapotranspiration, and sensible heat flux. Gillies and Temesgen offer a unique application of this technique to the mapping of environmental risk factors related to the prevalence of filariasis (a disfiguring and disabling disease caused by a parasite that is carried by mosquitoes) in the Nile Delta of Egypt.

In Chapter 6, Toby Carlson, David Ripley, and Thomas Schmugge examine rapid soil drying and its implications for derivation of quantitative estimates of soil moisture and surface energy fluxes from remote sensing data. The problems related with associating a correct level or depth for a soil moisture estimate are particularly acute when comparing estimates made via *in situ* measurements and those derived through analysis of remote sensing data. Carlson and his co-authors discuss methods for quantifying soil water content and surface radiant temperature using both *in situ* and remote sensing data, as well as the inter-relationships between vegetation (as an additional source of uncertainty for determining soil water content) and surface energy fluxes.

To elucidate the salient points of their chapter, they present the significant results from a soil experiment that was designed to study the soil drying process in relation to surface radiant temperature.

Chapter 7 by William Kustas, John Norman, Thomas Schmugge, and Martha Anderson concludes the overall theme of Part I of surface temperature and energy flux derivation from remote sensing data. Here the authors describe the use of a detailed Plant-Environment (PE) model (referred to as Cupid) to investigate the aerodynamic-radiometric temperature relationship for high and low wind speeds, stressed and unstressed vegetation, and for wet and dry soil moisture over a range of fractional vegetation cover conditions. The predicted radiometric temperature from Cupid is used as input to a Simplified Two-Source (STS) model that can be used operationally with either single or multiple directional radiometric surface temperature observations. Using the Cupid simulations, the capability of the STS model to predict energy flux partitioning between soil and vegetation under more extreme conditions is evaluated. Kustas *et al.* provide an example of energy flux maps over a heterogeneous landscape as generated by the STS model, and discuss the robustness of the STS model for predicting surface heat flux estimations at regional scales using satellite remote sensing data.

Chapter 8 by Susan Moran is one of two chapters in Part II dealing with the use of thermal IR data for assessment of ecosystem health. The term "ecosystem health" has generally been used to indicate the proper functioning of a complex ecosystem. A healthy system is one where the biophysical processes are operating adequately to maintain the ecosystem's structure, organization, and optimal activity over time. Moran focuses her discussion on the determination of plant transpiration and plant photosynthetic rates, which are key indicators of cropland and rangeland health. Transpiration has such an important role in plant health that even a slight reduction in plant water content can impact both growth and other physiological functions such as photosynthesis and respiration. It is the importance of these key indicators that has led to extensive efforts in using TIR measurements to evaluate the spatial and temporal distribution of plant transpiration. The direct link between the process of transpiration and plant thermal response offers significant potential for the use of TIR data for monitoring and managing plant ecosystem health. Moran provides a history of the physics and technology that have led to development of TIR spectral indices of plant ecosystem health. She then presents descriptions of several TIR indices and offers algorithms that can be implemented to measure plant ecosystem health using these indices, along with a discussion of the robustness of these algorithms in assessing the health of these ecosystems.

Chapter 9 in Part II by Roydon Fraser and James Kay greatly expands upon Chapter 8 by providing a detailed description of ecosystem thermodynamics as an indicator of ecosystem health. In particular, this chapter examines the theoretical foundations of energy quality, which measures the

capacity of energy, in its various forms, to do useful work. As Fraser and Kay note, the study of energy quality has the potential to provide a quantitative method for characterizing the status, maturity, or stage of development of ecosystems, and to provide fundamental physical explanations, at least in part, on survival strategies and structures employed within ecosystems as they evolve. Inherent to the measurement of ecosystem health as described in this chapter is the concept of *exergy,* which recognizes energy quality—not energy magnitude—as the appropriate criteria for assessing the most effective use of an energy source. Fraser and Kay consider this concept of *exergy* within the overall role of surface temperature and thermal remote sensing as an indicator of ecosystem maturity and health.

Part III focuses on TIR instruments and calibration. To achieve accurate measurements of surface temperature using TIR data, it is essential that TIR sensors be properly calibrated. Calibration of TIR sensors, therefore, is critical to both the acquisition and analyses of TIR data to avoid spurious or erroneous results in interpreting and quantifying these data. Additionally, one of the key objectives of this book is to hopefully provide impetus to spur the development and implementation of new sensors to advance research in TIR data utilization and analysis. The two chapters in Part III focus on these issues of sensor calibration and new instrumentation.

In Chapter 10, John Schott, Scott Brown, and Julia Barsi examine radiometric calibration of TIR sensors from an end-to-end systems perspective. This chapter provides information as a basis for calibration of laboratory, field, and flight TIR instruments and to describe the radiometric and spectral standards that need to be achieved to provide accurate temperature measurements from TIR data. Additionally, they present an in-depth discussion of the errors that can result from incorrectly calibrating TIR data.

Chapter 11 focuses on a discussion of the information derived from a demonstration mission of a proposed satellite TIR imaging system. Here, Alain Vidal, Philippe Duthil, Catherine Ottlé, Vicente Caselles, Antonio Yagüe, and John Murtagh present the results from a study using a prototype of the MediUm Scale surface Temperature (MUST) sensor. The objective of the MUST study was to define and demonstrate the utility and applicability of a large-swath, medium-resolution thermal IR sensor dedicated to the assessment of the environment and agriculture. As they relate, the specific objectives of this investigation were: (a) to demonstrate the relevance and efficiency of prospective MUST mission products to their relevant areas of application and to evaluate the economical benefits of such a mission; (b) to further develop methodologies for retrieving thermal and water related surface parameters from the sensor data; (c) to produce the design of a medium-resolution, large-swath thermal imager that is compact and affordable; and (d) to analyze the operational implementation of the ground segment needed to support such a mission. Vidal and his colleagues conclude that from the

results of the study presented in their chapter, the MUST sensor appears to have wide potential applicability for obtaining new and useful TIR data to support the improved analyses of a number of environmental and agricultural surface energy characteristics.

Penultimate comments

We hope this introduction has given some insight on both the breadth of the topics covered in this book, as well as providing a "roadmap" of the book's contents. We encourage readers to map out their own "itinerary" for navigating through the chapters and to choose their "sojourn" in whichever way will best help them better understand the capabilities and utility that TIR remote sensing can provide as related to their own interests. The Epilogue that appears at the end of this volume provides more reflection on the overall impetus for this book project, along with some final thoughts on the need for continued and robust research using TIR remote sensing data for analysis of land surface processes.

References

La Londe Workshop (1993) Workshop on thermal remote sensing of the energy and water balance over vegetation in conjunction with other sensors. Organized by The Pennsylvania State University, College of Earth and Mineral Sciences, Earth System Science Center, University Park, Pennsylvania, the Centre d'etude des Environnements Terrestre et Planétaires (CETP), Centre Universitaire Technique, Vélizy, France, and the CEMAGREF-ENGREF, Remote Sensing Lab, Montpellier, France. La Londe Les Maures, France, September 20–23, CEMAGREF-ENGREF, Montpellier Cedex 5, France, 330 pages.

Miller, D.H. (1981) *Energy at the Surface of the Earth*. Academic Press, New York.

Quattrochi, D.A. and J.C.Luvall (1999) Thermal infrared remote sensing data for analysis of landscape ecological processes: methods and applications. *Landscape Ecol* 14 (6): 577–98.

Quattrochi, D.A. and N.S.Goel (1995) Spatial and temporal scaling of thermal infrared remote sensing data. *Remote Sens. Rev.* 12:255–86.

Part I

Thermal infrared data for assessment and quantification of surface energy fluxes and soil moisture

Chapter 1

Estimating environmental variables using thermal remote sensing

Kevin P.Czajkowski, Samuel N.Goward, Theresa Mulhern, Scott J.Goetz, Anita Walz, David Shirey, Stephen Stadler, Stephen D.Prince and Ralph O.Dubayah

1.1 Introduction

There have been considerable advances in the estimation of land surface environmental conditions from satellite observations, particularly from thermal infrared remote sensing data (Running and Nemani 1988; Carlson *et al.* 1994; Norman *et al.* 1995; Prince and Goward 1995; Sun and Mahrt 1995; Andersen 1996; Susskind *et al.* 1997). Near-surface temperature and water vapor are of critical importance to the study of terrestrial hydrology (Dubayah *et al.* 2000), biospheric processes (Prince and Goward 1995), and other Earth System Science processes (Ehrlich *et al.* 1994).

Traditionally, ground-based meteorological observations have been used in biospheric and hydrologic modeling. Satellites provide higher spatial resolution data over the entire Earth and is especially important over isolated locations where meteorological observations are sparse. Goetz *et al.* (2000) incorporated thermal remote sensing of surface temperature, air temperature, and atmospheric water vapor into the Global Production Efficiency Model (Glo-PEM) to estimate global net primary production (NPP). They used Advanced Very High Resolution Radiometer (AVHRR) data from 1982 to 1990 to monitor interannual variability plant growth and carbon uptake worldwide. Their results showed a global decrease in NPP with an increase in Northern Hemisphere, high-latitude regions.

Hydrologic modeling can also benefit from satellite-derived surface and lower atmosphere conditions. Determining the energy and water balances for hydrologic modeling is dependent upon both the difference in temperature between the surface and some level in the atmosphere and the amount of water vapor in the atmosphere. Dubayah *et al.* (2000) used AVHRR estimates of air temperature and water vapor to drive the Land Surface Process Model, VIC-2L, for the Mississippi River Watershed. O'Donnell *et al.* (2000) applied similar techniques to the Ohio River watershed and found runoff estimates

to be very similar between satellite-derived and ground-based input data. Lakshmi and Susskind (2001) used surface temperature derived from TIROS Operational Vertical Sounder (TOVS) to adjust soil moisture in a Land Surface Process Model.

In this chapter, we will report on advances in using thermal infrared remotely sensed satellite observations to derive environmental variables, specifically surface temperature, air temperature, and water vapor. Throughout this chapter, we will use the AVHRR as an example. We will discuss the limits of the data and the pitfalls that need to be avoided. Finally, we will discuss the way to use other thermal sensors such as Landsat 7, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and the Moderate Resolution Imaging Spectrometer (MODIS).

1.2 Interpreting thermal infrared signals

The radiant energy detected by thermal sensors is a composite of energy emitted by the land surface that is transmitted through the atmosphere (not absorbed) and energy that is emitted by the atmosphere. This landatmosphere coupling complicates interpretation of the remotely sensed signal. However, this complication allows the estimation of a number of environmental variables of interest in Earth System Science modeling (Czajkowski et al. 2000).

Thermal bands on remote sensing instruments observe the wavelengths in the atmospheric window region of the electromagnetic spectrum, approximately between 8 and 14μm. Figure 1.1 shows the relationship between the two thermal bands on AVHRR and the thermal window region

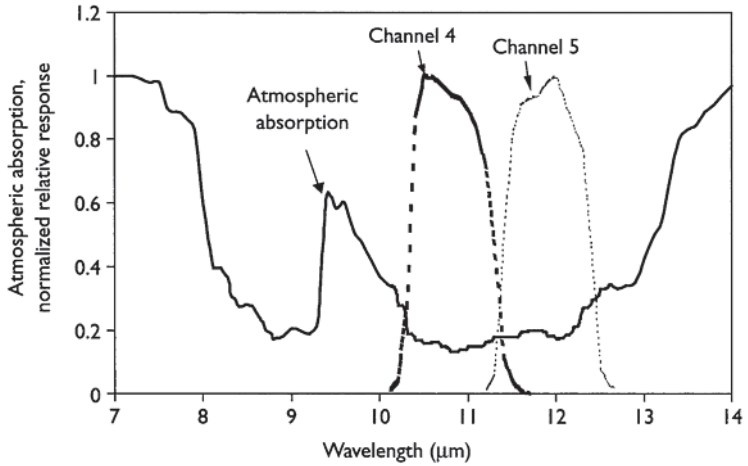


Figure 1.1 Relationship between AVHRR channels 4 and 5 (dashed lines) and the atmospheric window in the thermal infrared part of the spectrum (solid line).

in the atmosphere. The radiance (L) observed by the thermal channels can be expressed by

$$L = \int \varepsilon_\lambda B_\lambda(T_s)\tau_\lambda \, d\lambda + \int (1 - \tau_\lambda) B_\lambda(T_a) \, d\lambda \tag{1.1}$$

where B_λ is the wavelength-dependent black body radiation that would be emitted from the surface at temperature T_s, while ε_λ and τ_λ represent the wavelength-dependent emissivity of the surface and transmission of energy through the atmosphere. The first term in equation (1.1) represents the portion of the observed radiance that is not attenuated by the atmosphere before it reaches the satellite. The second term represents emission by the atmosphere at its effective temperature, T_a. Atmospheric water vapor content and the effective temperature of the atmospheric layer that contains the water vapor are the two primary atmospheric factors contributing to the thermal signals. The majority of electromagnetic radiance observed by AVHRR channels 4 and 5 originates from the surface, while the remainder of the signal originates from the atmosphere below 2 km. The water vapor profile and the air temperature profile both influence the observed thermal radiance (equation (1.1). Therefore, it should be possible to estimate surface temperature, atmospheric water vapor content, and the effective temperature of the atmospheric layer that contains the water vapor from the two thermal infrared AVHRR observations.

1.3 Radiometric surface temperature (T_s)

1.3.1 T_s algorithms

Approximately 80% of the energy thermal sensors receive in the 10.5–12.5 μm wavelength region is emitted by the land surface, making surface temperature the easiest variable to extract from the thermal infrared signal. Extensive work has gone into the development of algorithms to estimate land surface temperature from AVHRR channels 4 and 5 (Price 1984; Becker and Li 1990). The primary approach is the so-called "split window" technique that uses the difference in brightness temperature between AVHRR channels 4 and 5 to correct for atmospheric effects on sea surface and land surface temperatures. The split window technique works independent of other data sources and takes advantage of the differential effect of the atmosphere on the radiometric signal across the atmospheric window region. The basic form of the split window equation for AVHRR channels 4 (T_4) and 5 (T_5) is

$$T_s = a + T_4 + b(T_4 - T_5) \tag{1.2}$$

where a and b are constants that can be estimated from model simulations (Becker and Li 1990) or correlation with ground observations (Prata 1993).

Table 1.1 Split window equations used to estimate land surface temperature

Author	*Equation with emissivity assumed to be 1.0*
McClain *et al.* (1983)	$T_s = 1.035T_4 + 3.046(T_4 - T_5) - 10.934$
Price (1984)	$T_s = T_4 + 3.33(T_4 - T_5)$
Becker and Li (1990)	$T_s = 1.274 + T_4 + 2.63(T_4 - T_5)$
Prata and Platt (1991)	$T_s = T_4 + 2.45(T_4 - T_5)$
Sobrino *et al.* (1993)	$T_s = 1.06(T_4 - T_5) + 0.46(T_4 - T_5)^2$
Ulivieri *et al.* (1994)	$T_s = T_4 + 1.8(T_4 - T_5)$

Channel 5 is more sensitive to atmospheric water vapor than channel 4, so the difference between T_4 and T_5 is larger for humid atmospheric conditions than for dry conditions.

Five split window equations including McClain *et al.* (1983), Price (1984), Prata and Platt (1991), Sobrino *et al.* (1993), and Ulivieri *et al.* (1994) are given in Table 1.1. Each of these equations takes on the form of equation (1.2) if emissivity in all wavelengths is assumed to be 1.0. The equations are much more complicated if emissivity is allowed to vary. We have found that assuming an emissivity of unity is realistic in heavily vegetated areas that we are most concerned with. Differences in emissivity need to be addressed to estimate T_s in areas with sparse vegetation such as semiarid or desert regions. Sobrino *et al.* (1993) is different from the others in that it is a second-order equation that tries to account for the non-linear changes in the difference between T_4 and T_5. It is best to use a second-order equation whenever possible because it increases the accuracy of T_s somewhat.

One main difference between each of the split window equations in Table 1.1 that is often forgotten is that they were derived for different AVHRR sensors. For instance, the Price (1984) algorithm was derived from AVHRR data from the NOAA-7 satellite, Becker and Li (1990) equation was derived using NOAA-9 AVHRR data, and Sobrino *et al.* (1993) equation was derived using NOAA-11 AVHRR data. Prata (1993) developed regression coefficients for both NOAA-11 and -12. It is important to derive a new T_s equation for each new sensor that is used because of variations in filter functions (relative spectral response) between sensors (Czajkowski *et al.* 1998). No two satellite sensors are made exactly alike (Figure 1.2). These changes in spectral band width and sensitivity create different T_4 and T_5 values for different sensors. For example, if we use the Becker and Li (1990) technique, which was derived for use with NOAA-9 data, with data from another NOAA satellite errors larger than 2 K are possible (Figure 1.3). Czajkowski *et al.* (1998) derived split window equations for AVHRR data from NOAA satellites 7, 9, 11, 12,

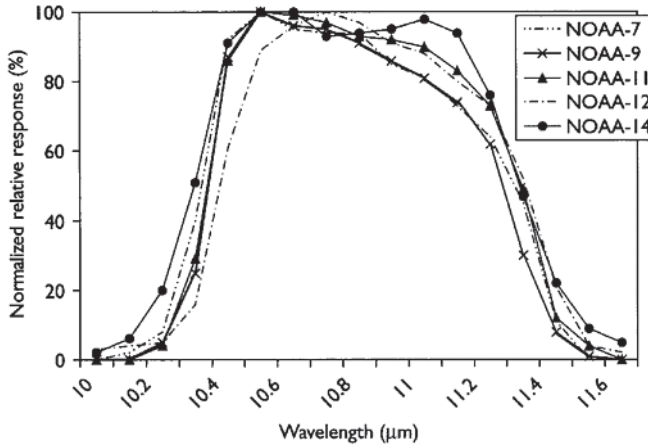


Figure 1.2 AVHRR normalized relative response functions for AVHRR channel 4 from NOAA satellites 7, 9, 11, 12, and 14 (data from Kidwell 1997).

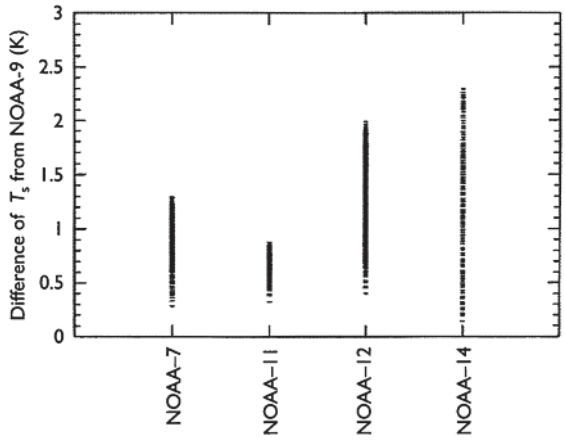


Figure 1.3 Difference in T_s due to use of the Becker and Li (1990) split window equation with data from an AVHRR instrument not on NOAA-9.

and 14 using MODTRAN3 (Berk *et al.* 1989) simulations for a variety of atmospheric air temperature and water vapor profiles and surface temperature conditions (Table 1.2).

1.3.2 T_s validation

Testing of the split window concept has produced mixed results. Figure 1.4 shows comparison of ground infrared radiative thermometer (IRT)

Table 1.2 Sensor-specific coefficients for the split window equation: $T_s = a + T_4 + b(T_4 - T_5)$ (Czajkowski et al. 1998)

Sensor	*a*	*b*
AVHRR-7	2.67	2.24
AVHRR-9	1.95	2.56
AVHRR-11	2.86	2.40
AVHRR-12	7.86	2.34
AVHRR-14	5.54	2.08

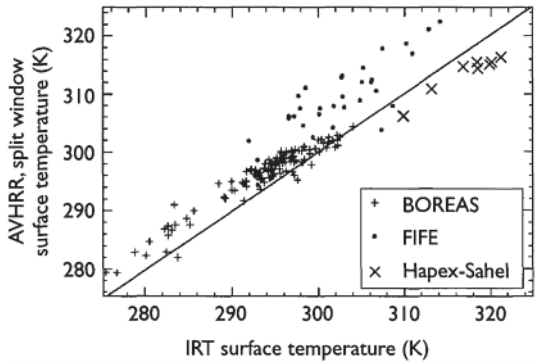


Figure 1.4 Comparison of ground-based surface temperature measurements from IRTs with surface temperature estimated using the split window technique of Becker and Li (1990).

observations with T_s derived from AVHRR observations from FIFE, HapexSahel, and BOREAS (Goetz *et al.* 1995; Czajkowski *et al.* 1997; Prince *et al.* 1998). For all three of these cases, the Becker and Li (1990) split window equation was employed even though some of the data were from NOAA11 and the Becker and Li (1990) equation was derived for NOAA-9. The results from BOREAS produced errors of about ±3 K with some errors as large as ±8 K. The satellite algorithm had a warm bias for FIFE of 6.75 K and BOREAS of 2.34 K and a cool bias for Hapex-Sahel of –3.67K. The source of estimation errors stems from (a) assumptions used to derive the split window equations, (b) unknown variations in spectral emissivity, changes in spectral response functions between satellites, and (c) mismatches between ground observations and the satellite field of view (Table 1.3).

The assumption of linearity on which many split window equations are based limits T_s estimations because the relationship between the brightness temperatures of AVHRR channels 4 and 5 is non-linear. Figure 1.5 shows

Table 1.3 Application of temperature/vegetation index (Tvx) to EOS satellite sensors

Sensor	*NDVI*	*Surface temperature*
AVHRR	Use band 1 (0.55–0.90 μm) and 2 (0.72–1.10 μm)	Use a split window equation derived for the specific sensor used
Landsat 7	Use band 3 (0.63–0.69 μm) and 4 (0.75–0.90 μm)	Adjust band 6 with an atmospheric model or profile
ASTER	Use band 2 (0.63–0.69 μm) and 3 (0.76–0.86 μm)	Use the temperature/emissivity separation (TES) algorithm (Kahle and Alley 1992)
MODIS	Use band 1 (0.55–0.90 μm) and 2 (0.72–1.10 μm)	Use the surface temperature product derived from the MODIS Product (MOD11) (Wan and Dozier 1996)

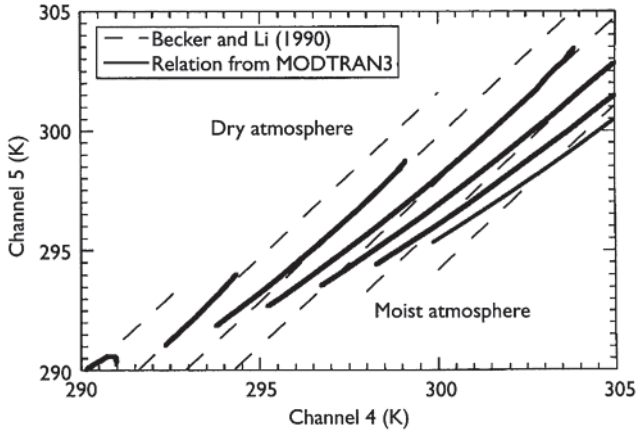


Figure 1.5 Relation of surface temperature within channel 4 and channel 5 space. The dashed lines indicate the surface temperature that would be calculated using the linear relationship of the Becker and Li (1990) split window equation while the sold lines indicate the surface temperature found using MODTRAN3 simulations. Each line represents a different surface temperature.

the relationship between AVHRR channels 4 and 5 derived from MODTRAN3 simulations for a range of surface temperature and atmospheric water vapor conditions. These simulations were compared to the relationship of linear channels 4 and 5 used in the Becker and Li (1990) split window equation. The Becker and Li (1990) equation performs well for dry atmospheric conditions. Errors introduced by non-linearity increase, however, with increasing water vapor in the atmosphere. The second-order algorithm proposed by Sobrino *et al.* (1993) is better at accounting for these non-linearities. We conclude that the non-linear relationship between channels 4 and 5 is an area that needs more research.

Errors on the order of 1–2 K may result in the estimation of T_s due to emissivity (Coll *et al.* 1994; Becker and Li 1995; Caselles *et al.* 1995; Wan and Dozier 1996). Emissivity varies significantly, particularly when silicate minerals are present (Bowman 1996). In contrast, emissivity in the 10.5–12.5 μm region is generally much less variable.

The errors discussed above may account of some for the error in our validation, however, the comparison in Figure 1.4 and results from other studies suggest that the validation of T_s with ground observations from large field studies is difficult (Prince *et al.* 1998; Schmugge and Schmidt 1998). A significant portion of the comparison errors, as indicated by the clustering of each field studies' points in Figure 1.4, appears to be due to fundamental differences in resolution scales between the satellite and ground-based IRTs; the satellite field of view is typically much larger than the area observed by the tower mounted IRTs. The ground measurements are, therefore, not representative of the complex landscape that is observed by the satellite sensor. This was especially apparent for the FIFE site where the IRTs viewed an ungrazed, fenced-in area surrounding each tower while the majority of the satellite footprint included grazed land (Goetz *et al.* 1995; Steyn-Ross *et al.* 1997; Schmugge and Schmidt 1998). Therefore, the Split Window Techniques may be more accurate than reported.

1.4 Air temperature (T_a)

Estimation of air temperature from the thermal signal is very complex due to atmospheric emission. Specifically, interpretation of AVHRR thermal observations is an under-determined problem because there are just two AVHRR thermal channels with many unknowns including T_s, ε, and T_a and water vapor profiles. There are more unknowns than thermal channels, thus an exact solution cannot be found.

1.4.1 Temperature/vegetation index

To overcome the indeterminacy of thermal remote sensing, a contextual array of contiguous pixels can be used to increase the number of observations available for each calculation (Goward *et al.* 1994). To estimate nearsurface air temperature, the normalized difference vegetation index (NDVI) is merged with thermal infrared measurements of surface temperature over a 9×9 contextual array (Tucker 1979). This TVX has been used successfully to estimate near-surface air temperature from AVHRR data (Goward *et al.* 1994; Czajkowski *et al.* 1997; Prihodko and Goward 1997; Prince *et al.* 1998). The TVX approach was developed from the empirical observation that when T_s and NDVI measurements are compared, they generally display a linear and mostly negative correlation (Figure 1.6) (Hope *et al.* 1987; Nemani and Running 1989; Carlson *et al.* 1990; Goward *et al.* 1994).

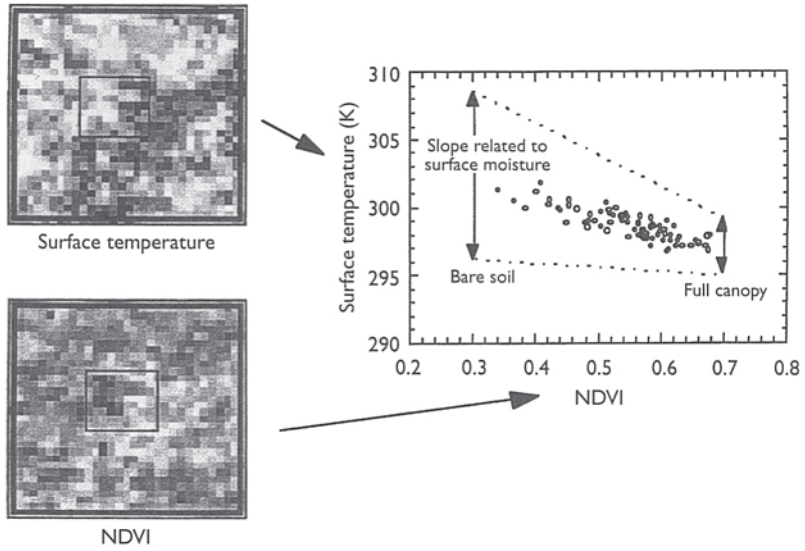


Figure 1.6 The "classic" TVX plot. Bare surfaces have lower NDVI and tend to be warmer than vegetated surfaces.

Where NDVI measurements are high, indicating extensive vegetation cover, the surface temperature is relatively low, and appears to closely estimate the near-surface air temperature. The temperature of vegetation canopies rarely deviates from ambient air temperature by more that ±2.0K, whether or not the vegetation is actively transpiring (Geiger 1965). In contrast, specific studies have found that individual leaves (or needles) of vegetation can be as much as ±5–10 K different from air temperature (Hatfield 1979; Jackson *et al.* 1981; Williams 1989). Assemblages of canopies with a multitude of leaves, some shadowed, in 1 km or larger AVHRR pixels are expected to exhibit surface temperatures comparable to the surrounding air temperature.

The TVX technique is generally applied to images using a 9×9 moving window although the appropriate window size to use may depend on the topography and ecosystem heterogeneity. By fitting a linear regression to the T_s/NDVI relationship for each 9×9 array and extending the regression to a full canopy, the surface temperature at the NDVI of the full canopy can be used to approximate air temperature:

$$T_a = \mathrm{NDVI}_{\mathrm{full}} S + I \tag{1.3}$$

where S and I are the slope and intercept from the least-squares regression of the 9×9 array of pixels, and $\mathrm{NDVI}_{\mathrm{full}}$ is the NDVI of a full canopy, 0.9 for

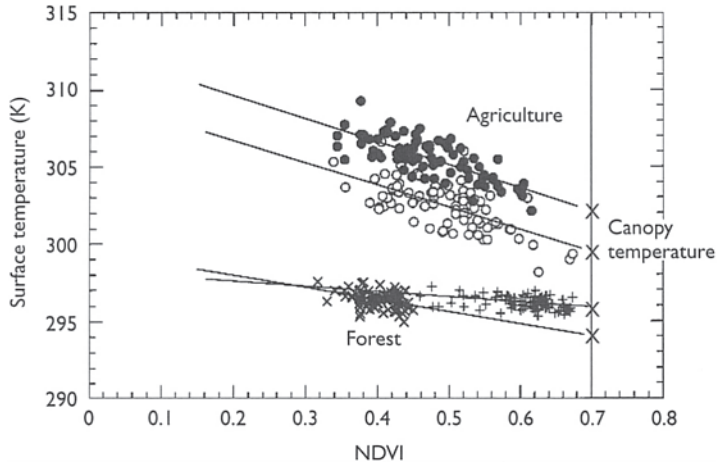


Figure 1.7 TVX relationships for agricultural sites, Rosetown and Saskatoon, and forested sites, Southern Study Area-Old Aspen (SSA-OA) and Northern Study Area-Old Jack Pine (NSA-OJP), for July 24, 1994 during BOREAS. Lines represent the least-squares regression through each station's 9×9 contextual array.

atmospherically corrected images and 0.7 for uncorrected images (Czajkowski *et al.* 1997).

Figure 1.7 shows the TVX slope for four separate 9×9 arrays of pixels from a single AVHRR image from BOREAS that includes agriculture and the nearby boreal forest. Note that for a given NDVI, the T_s of agriculture is warmer than the nearby forest. Agriculture then appears to produce a more negative slope than the forest. This is most likely due to differences in canopy structure, evapotranspiration, and energy exchange (Nemani and Running 1989; Carlson *et al.* 1990; Goetz 1997).

1.4.2 T_a validation

Air temperature recovery with the TVX technique was tested and showed consistent accuracy across a range of field studies (Figure 1.8): FIFE (Prihodko and Goward 1997), Hapex-Sahel, BOREAS (Czajkowski *et al.* 1997; Prince *et al.* 1998), and the Oklahoma Mesonet (Czajkowski *et al.* 2000). Results show that air temperature can be estimated with an rmse of 3.9 K. Errors in T_s may contribute significantly to the errors in the TVX technique. For example, the satellite estimation of T_s showed a warm bias of 2.34 K for the BOREAS study and the satellite estimation of T_a had a warm bias of 3.2K.

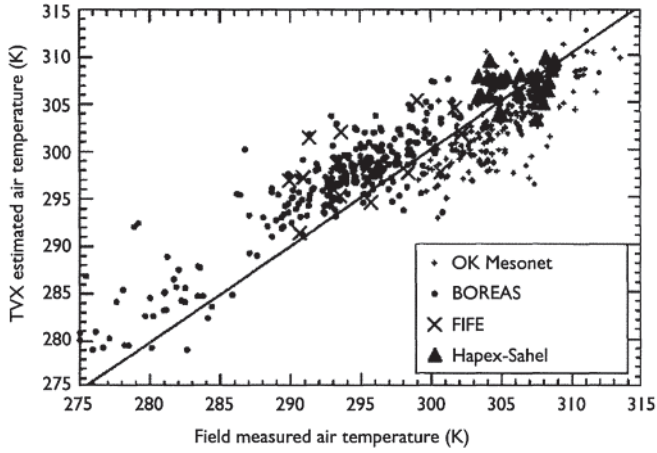


Figure 1.8 Comparison of TVX estimated air temperature with shelter height air temperature for FIFE, Hapex-Sahel, and BOREAS and from the Oklahoma Mesonet.

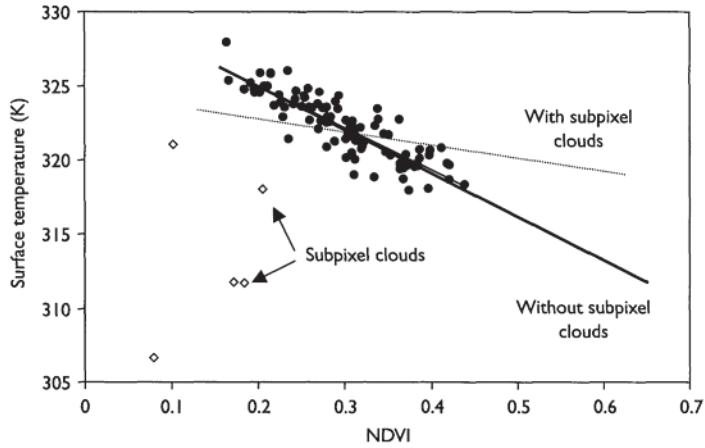


Figure 1.9 An example of cloud contamination within the 9×9 TVX contextual array from a site in Oklahoma. The open diamonds represent pixels with subpixel clouds that were not masked using typical cloud detection techniques. The dotted line is the TVX regression with cloud contaminated pixels included, and the solid line is the regression after those clouds have been removed.

Cloud contamination is a significant problem with the TVX air temperature technique. Figure 1.9 shows the TVX T_s/NDVI relation for a site in Oklahoma. The positive slope, shown by the dashed line, is caused by subpixel cloud contamination that was not detected during cloud screening. Pixels with

subpixel clouds generally have lower NDVI and lower T_s than the land surface if an atmospheric inversion is not present. Czajkowski *et al.* (1997) used iterative TVX slope calculations to perform additional cloud screening to remove subpixel clouds. Basically, once a TVX slope has been identified as having pixels with subpixel cloud contamination (i.e. a positive slope with a large scatter around the regression line), pixels in which the NDVI or T_s fall below the first standard deviation are masked and the slope is computed again. This is done iteratively until the slope of the TVX relation shows no sign of cloud contamination or the number of available pixels becomes small, that is, 40 pixels. For Figure 1.9, correcting for subpixel cloud contamination, only 5 pixels will lower the TVX air temperature estimate by 7K.

1.5 Near-surface water vapor

Water vapor has the least direct impact on the thermal infrared signal observed by satellites. It is thus the hardest of the three environmental variables (T_s, T_a, and water vapor) to estimate. In the literature, several techniques have been proposed to estimate atmospheric water vapor (Jedlovec 1990; Kleespies and McMillin 1990; Justice *et al.* 1991; Eck and Holben 1994; Goward *et al.* 1994; Prince and Goward 1995; Andersen 1996). in particular total column water vapor, the integration of water vapor from the surface to the top of the atmosphere. For hydrologic and biospheric modeling applications, we are most concerned with near-surface water vapor, that is, within the boundary layer. We have developed a technique to estimate nearsurface water vapor because most modeling applications need near-surface water vapor. Radiosonde profiles from BOREAS under clear-sky conditions shows that water vapor falls off rapidly above the boundary layer, 2 km on average at BOREAS (Figure 1.2). Chesters *et al.* (1983) showed similar results, that is, 80% of atmospheric water vapor is within 2 km of the ground under clear-sky conditions. Given that the AVHRR thermal channels are positioned in an atmospheric window region of the water vapor absorption continuum, they are well suited to estimate the amount of water vapor in the lower layer of the atmosphere (Prabhakara *et al.* 1974). In contrast, channels from satellite sounders such as VISSR Atmospheric Sounder (VAS) and TOVS are sensitive to water vapor mostly in the upper part of the troposphere (Chesters *et al.* 1983; Smith 1991).

1.5.1 Look-up table approach

Dalu (1986) suggested that an alternative approach to estimate water vapor would use radiative transfer model output to produce a look-up table of radiometric signals observed in the two thermal channels under a variety of atmospheric and surface conditions. A radiative transfer model is able to

capture the non-linear response of channels 4 and 5 to atmospheric water vapor. Prince and Goward (1995) and Andersen (1996) used output from the radiative transfer model, Lowtran7, to derive regression equations for an arbitrary set of atmospheric and surface conditions. Andersen (1996) recognized that knowledge of the actual surface temperature is needed to estimate water vapor accurately. For this study, we used MODTRAN3 simulation output to develop a look-up table that uses channels 4 and 5, surface temperature, and near-surface air temperature to derive near-surface water vapor (Figure 1.10).

Czajkowski *et al*. (2002) generated a look-up table for a large number of atmospheric conditions similar to those used by Prince and Goward (1995) and Andersen (1996). MODTRAN3 calculates the surface emitted radiance and radiometric transmittance and emitted flux of energy through and by the atmosphere at specific wave numbers using the specified surface and atmospheric conditions. An average mid-latitude summer atmospheric profile was used as a base for the simulations. The profile was varied to simulate a variety of atmospheric temperature and moisture conditions. The atmospheric temperature profile was varied such that the shelter height air temperature ranged from 270 to 320 K by an increment of 2 K. Vapor pressure was varied only in the lowest 2 km of the atmosphere, approximately the boundary

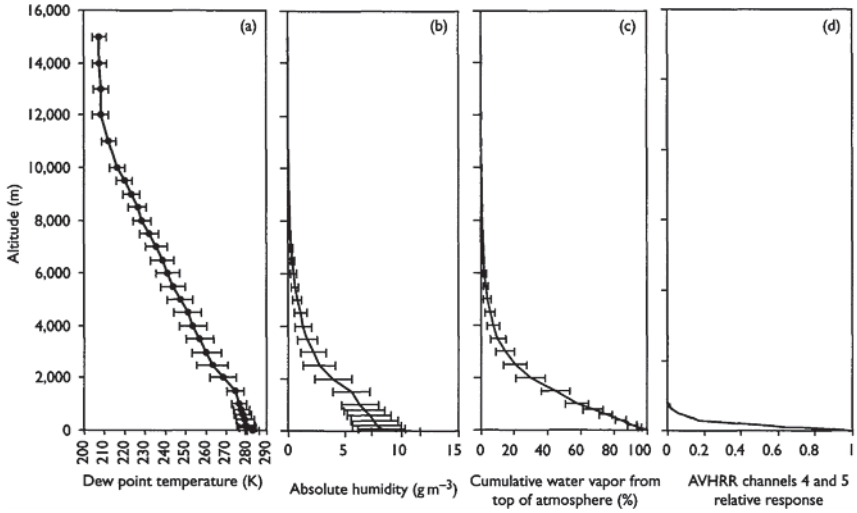


Figure 1.10 Average atmospheric profiles with standard deviations produced from 63 radiosondes from Saskatoon and The Pas, Saskatchewan from the BOREAS study 1994: (a) dew point temperature; (b) near-surface absolute humidity; (c) the cumulative percentage of water vapor; and (d) the influence of the water vapor profile on AVHRR channels 4 and 5 derived from MODTRAN3 simulations.

layer, from dry, 0.25 mb, to nearly saturated conditions at increments of 0.25 mb. Each layer of the profile was checked to ensure that saturation did not occur anywhere within the profile. Surface temperature was varied from 10 K less than the air temperature to 20 K more than the air temperature at increments of 0.1 K accounting for a total of several hundred thousand simulations. For each surface temperature, air temperature, and water vapor profile, top of the atmosphere radiances in AVHRR channels 4 and 5 were calculated. The radiances were converted to brightness temperature by combining the monochromatic radiance output from MODTRAN3 with the spectral response function of AVHRR channels 4 and 5. A table with five columns containing the brightness temperatures in AVHRR channels 4 and 5, input surface temperature, input near-surface air temperature, and input near-surface water vapor was constructed. Each MODTRAN3 simulation represents one line in the look-up table. We have found that knowledge of T_s and T_a are needed to successfully search the look-up table and find a unique solution. If only satellite imagery is available to search for the near-surface water vapor value, the Split Window Technique can be used to estimate surface temperature and the TVX approach can be used to estimate air temperature.

1.5.2 Near-surface water vapor validation

Czajkowski *et al.* (2002) used ground observations from BOREAS and the Oklahoma Mesonet to validate the look-up table approach. Some 359 BOREAS and 222 Oklahoma Mesonet comparisons were made between satellite estimations and ground observations (Figure 1.11). The comparisons produced an r^2 of 0.36 for the Oklahoma Mesonet and 0.26 for the BOREAS study site. The atmospheric profiles used in the MODTRAN3 simulations to construct the look-up table were all based on an average profile for the mid-latitude summer. Use of an actual profile or a more realistic one would, therefore, be an improvement.

Errors in both surface temperature and air temperature will influence the water vapor estimation. Errors in T_s are compounded because they propagate through both the TVX T_a estimation and water vapor estimation. If we substitute ground station T_a into the look-up table approach, we are able to improve the results to r^2 of 0.48 for the Oklahoma Mesonet and 0.5 for BOREAS.

1.6 Combined retrieval

It is through the solution of both the surface and atmospheric contributions to the satellite observations that the complete thermal signal can be understood. We have shown above that all of the biospheric variables that

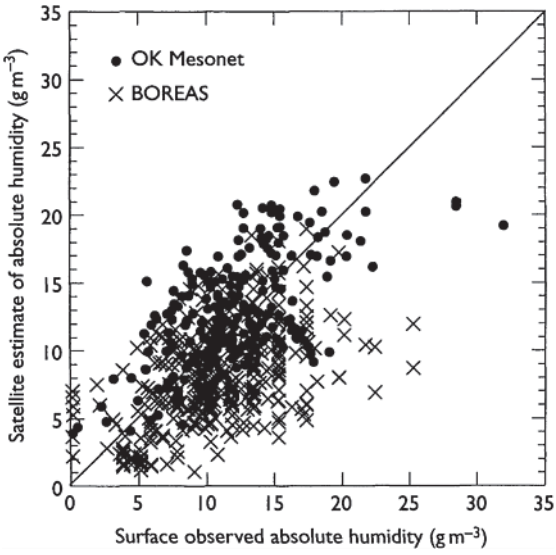


Figure 1.11 Comparison of atmospheric water vapor (absolute humidity) estimated from AVHRR data using the look-up table approach with ground measurements from the Oklahoma Mesonet and BOREAS meteorological stations.

contribute to the thermal signal can be derived, although sufficient accuracy with surface water vapor retrievals has yet to be realized. The various processing routines needed to acquire environmental variables from the thermal signal are summarized in Figure 1.12. Cloud and water screening play a very important part in deriving accurate T_s, T_a, and water vapor estimations. AVHRR channels 1 and 2 are used to derive NDVI while channels 4 and 5 are used in a sensor-specific Split Window model to derive T_s. Finally, atmospheric water vapor is solved using T_4, T_5, T_s, and T_a. Figure 1.13 shows how the variables relate to one another for a clear day, August 16, 1994, in Oklahoma. T_s and T_a are highly correlated, however, T_s varies over a larger range of temperatures. Satellite estimates of both T_a and water vapor match the spatial patterns of the Mesonet station data. They show a typical summertime pattern in Oklahoma where the eastern part of the state is warm and moist while the western part of the state is hot and dry. On this day, the synoptic conditions showed a high pressure area centered over the northeast corner of Oklahoma (NOAA 1994). Under such conditions, the pressure gradient was small in the eastern half of the state producing light or no winds. On the western part of the state, both the satellite estimates and ground data show a dry area.

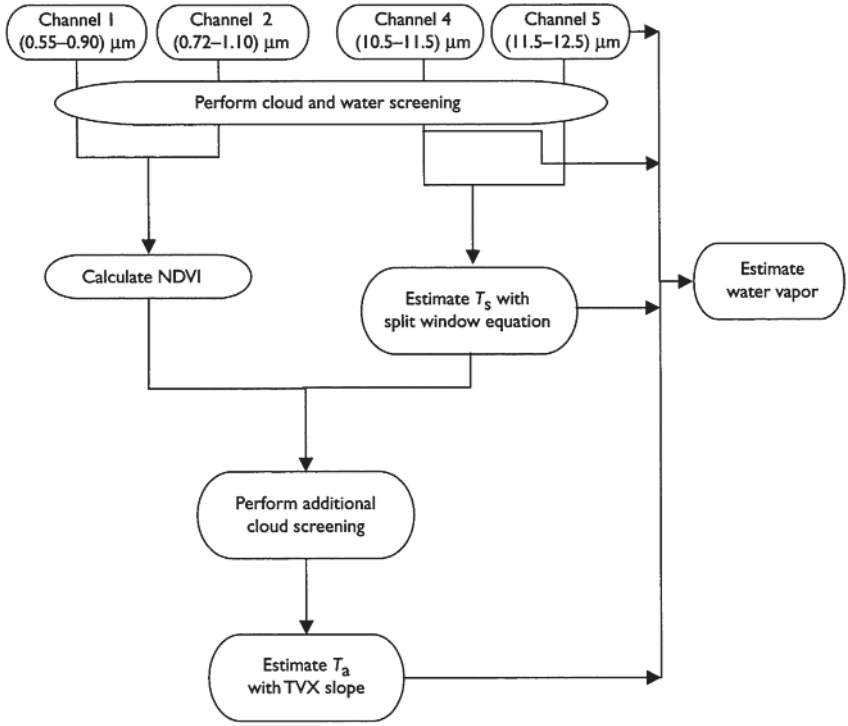


Figure 1.12 Schematic of steps taken to recover surface temperature (T_s), air temperature (T_a), and water vapor from the TIR signal.

1.7 Applications of TIR methods to other space-based sensors

Within NASA's Earth Science Enterprise there is a major initiative to observe the Earth on a global scale. This initiative, the Earth Observing System, is implementing a 15-year effort to study the physical, chemical, biological, and social processes that influence global climate (NASA 1996). Among the first new sensors to be launched as part of this effort were Landsat 7 and ASTER and MODIS on the Terra satellite. Each satellite has channels with wavebands similar to those of the AVHRR instrument. The differences in spatial resolution between the sensors holds the potential to improve new land surface studies.

For MODIS, NDVI can be derived using channels 1 and 2 (King *et al.* 1992) and surface temperature can be derived with a split window equation from channels 31 and 32 (Table 1.3). Channel 29 in the 8.5-μm region has the potential to be used to detect ice crystals and may also be used in surface temperature estimation (Wan and Dozier 1996). The atmospheric profile

products available from channels 27, 28, 30, and 33–36 may help determine the temperature and moisture characteristics of the atmosphere (Menzel and Gumley 1996).

ASTER's five channels in the thermal region of the spectrum should provide accurate surface temperature estimations using the (TES) technique (Kahle and Alley 1992). ASTER will also have the capability to estimate T_a and

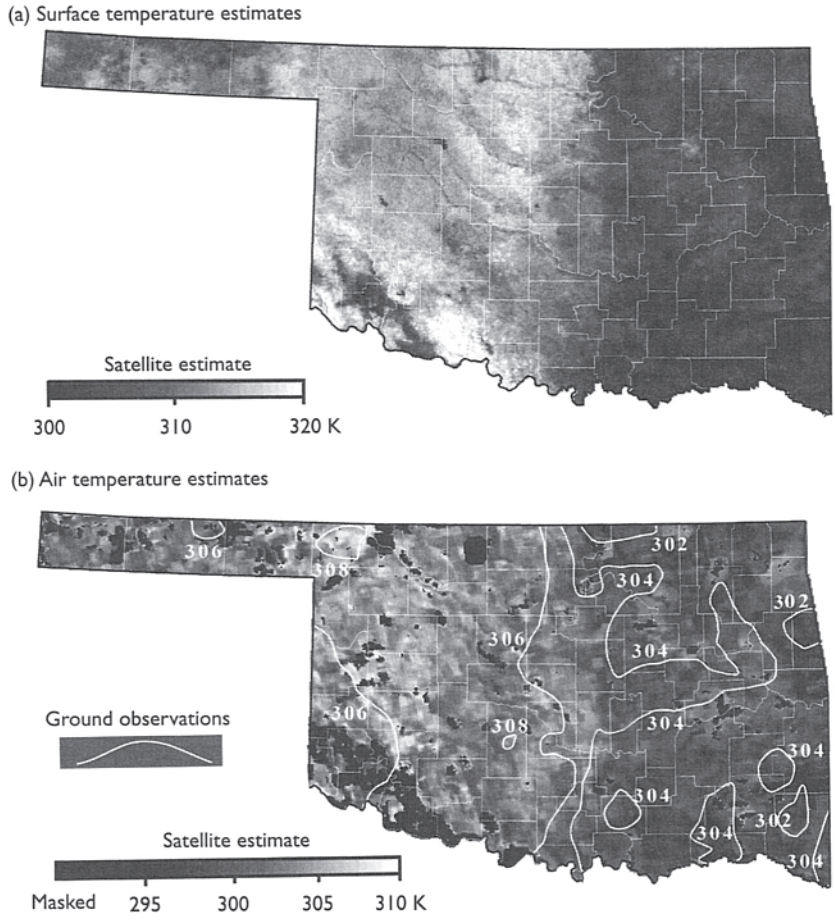


Figure 1.13a Images over Oklahoma for August 16, 1994 of (a) surface temperature, (b) air temperature comparison, (c) water vapor comparison, and (d) Mesonet station sites. The Mesonet values for air temperature and water vapor have been contoured with an inverse distance interpolation. Dark blotchy areas in the air temperature and water vapor images are cloud or water contaminated.

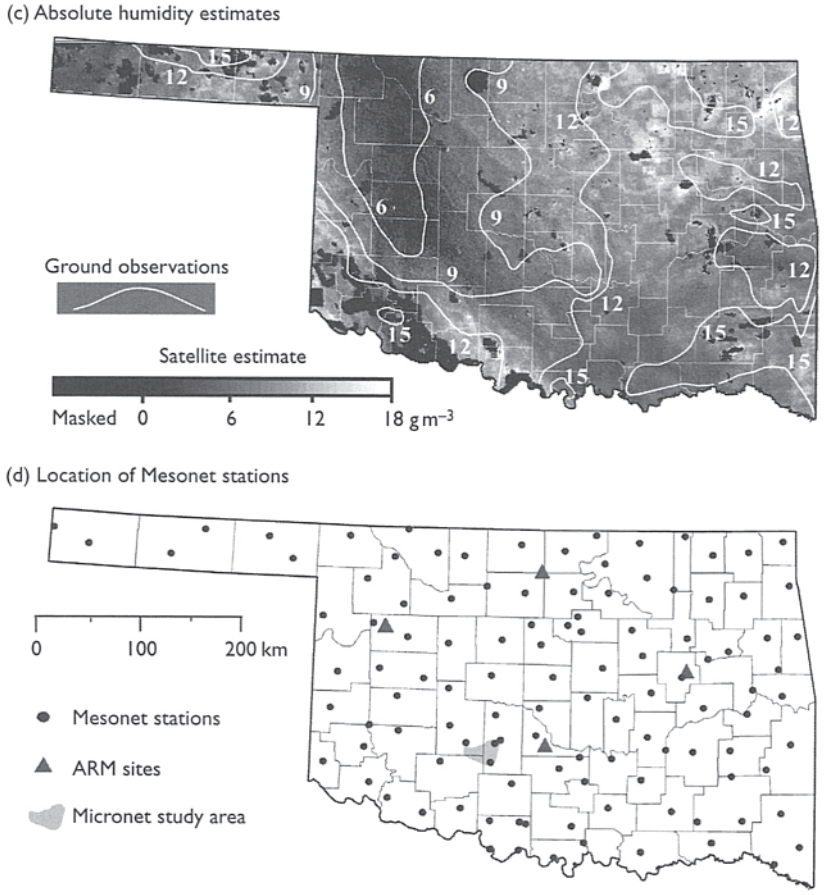


Figure 1.13b (Continued).

water vapor, but its 16-day revisit interval and an "on demand" collection mode suggest it may not be useful for routine environmental monitoring over large areas. The benefit of ASTER is the 90-m thermal resolution, and the fact that it shares the platform with MODIS. Observations from ASTER can be used to address the subgrid scale heterogeneity in the MODIS signal, thus bridging the gap between the MODIS scale and ground observations. It will be possible to test whether vegetation processes acting at spatial scales less than 1 km affect the performance of the TVX technique. This has not been possible in the past because the high-resolution Thematic Mapper and coarse resolution AVHRR sensors have been on different satellites and had different overpass times.

The Landsat-7 Enhanced Thematic Mapper Plus (ETM+) has capabilities comparable to ASTER due to its high spectral resolution visible and

nearinfrared channels (30 m) and one thermal channel (60 m). Spectral vegetation indices can be derived from channels 3 and 4; however, Landsat-7 is limited in thermal remote sensing because it has only one thermal infrared channel. Also, its one thermal channel has lower precision, about 0.2 K, than the other sensors. Atmospheric correction typically with radiative transfer codes and radiosonde soundings is needed to get an accurate estimate of the surface temperature.

As with air temperature, the techniques to estimate near surface water vapor can be applied to MODIS and ASTER data. The five thermal channels of ASTER can be used to develop more accurate water vapor retrievals. In addition, the contextual techniques may prove more fruitful with ASTER's higher quality measurements and resolution.

The potential of thermal remote sensing in Earth System Science is increased number of thermal sensors available. It is apparent that there is a great deal of research to be done, and a great deal of critical information on land surface processes to be gained.

Acknowledgments

This work was supported by NASA grants NAG 53851 and NAG 58052 from the Terrestrial Ecology Program, Oklahoma's NASA EPSCOR Project and the State of Oklahoma. The Oklahoma State University portion of the work was funded by NASA Cooperative Agreement NCC5–171 and the Oklahoma State Regents for Higher Education. We would like to thank Ms Pat Hurlburt of the Manitoba Remote Sensing Centre, Stanley Shewchuk and the observers of the Saskatchewan Research Council, Barry Goodison of Environment Canada, and Jeff Newcomer and many others responsible for organizing the data on the BORIS data system.

References

Andersen, H.S. (1996) Estimation of precipitable water vapour from NOAAAVHRR during the Hapex Sahel experiment. *Int. J. Remote Sens.* 17:2783–801.

Becker, F. and Z.-L.Li (1990) Towards a local split window method over land surfaces. *Int. J. Remote Sens.* 11:369–93.

Becker, F. and Z.-L.Li (1995) Surface temperature and emissivity at various scales: definition, measurement and related problems. *Remote Sens. Environ.* 12: 225–53.

Berk, A., L.S.Bernstein, and D.C.Robertson (1989) MODTRAN: a moderate resolution model for LOWTRAN 7, PL-TR-89–0122. US Air Force Phillips Laboratory, Hanscom Air Force Base, MA.

Bowman, A. (1996) Empirical solutions for land surface temperature estimation from thermal remote sensing: the emissivity factor. Thesis, Masters of Science, The University of Maryland.

Carlson, T.N., E.M.Perry, and T.J.Schmugge (1990) Remote estimation of soil moisture

availability and fractional vegetation cover for agricultural fields. *Agric. For. Meteorol.* 52:45–69.
Carlson, T.N., R.R.Gilles, and E.M.Perry (1994) A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. *Remote Sens. Environ.* 9:161–73.
Caselles, V., C.Coll, E.Valor, and E.Rubio (1995) Mapping land surface emissivity using AVHRR data application to La Mancha, Spain. *Remote Sens. Environ.* 12: 311–33.
Chesters, D., L.W.Uccelini, and W.D.Robinson (1983) Low level water vapour fields from the VISSR Atmospheric Sounder (VAS) "split-window" channels. *J. Clim. Appl. Meteor. 22:*725–43.
Coll, C., V.Caselles, and T.J.Schmugge (1994) Estimation of land surface emissivity differences in the split-window channels of AVHRR. *Remote Sens. Environ.* 48: 127–34.
Czajkowski, K.P., T.Mulhern, S.N.Goward, J.Cihlar, R.O.Dubayah, and S.D.Prince (1997) Biospheric environmental monitoring at BOREAS with AVHRR observations. *J. Geophy. Res.* 102:29651–63.
Czajkowski, K.P., S.N.Goward, and H.Ouaidrari (1998) Impact of AVHRR filter functions on surface temperature estimation from the split window approach. *Int. J. Remote Sens.* 19:2007–12.
Czajkowski, K.P., S.N.Goward, J.S.Stadler, and A.Walz (2000) Thermal remote sensing of near surface environmental variables: application over the Oklahoma Mesonet. *Prof. Geogr.* 52:345–57.
Czajkowski, K.P., S.N.Goward, D.Shirey, and A.Walz (2002) Thermal remote sensing of near surface water vapor. *Remote Sens. Environ.* 79:253–65.
Dalu, G. (1986) Satellite remote sensing of atmospheric water vapour. *Int. J. Remote Sens.* 7:1089–97.
Dubayah, R.O., E.F.Wood, T.E.Engman, K.P.Czajkowski, and M.Zion (2000) A remote sensing approach to macroscale hydrological modeling. In G.Schultz and E.Engman (eds), *Remote Sensing in Hydrology and Water Management*. Springer-Verlag, Berlin, pp. 85–102.
Eck, T.F. and B.N.Holben (1994) AVHRR split window temperature differences and total precipitable water over land surfaces. *Int. J. Remote Sens.* 15:567–82.
Ehrlich, D., J.E.Estes, and A.Sing (1994) Applications of NOAA-AVHRR 1km data for environmental monitoring. *Int. J. Remote Sens.* 15:145–61.
Gieger, R. (1965) *Climate Near the Ground*. Rev. Ed. Sci. Tech., Inc., Harvard University Press, Cambridge, MA, 611 pp.
Goetz, S.J. (1997) Multi-sensor analysis of NDVI, surface temperature, and biophysical variables at a mixed grassland site. *Int. J. Remote Sens.* 18:71–94.
Goetz, S.J., R.N.Haltore, F.G.Hall, and B.L.Markham (1995) Surface temperature retrieval in a temperate grassland with multiresolution sensors. *J. Geophys. Res.* 25: 25397–410.
Goetz, S.J., S.D.Prince, J.Small, and A.C.R.Gleason (2000) Interannual variability of global terrestrial primary production: results of a model driven with satellite observations, *J. Geophys. Res.* 105:20077–91.
Goward, S.N., R.H.Waring, D.G.Dye, and J.Yang (1994) Ecological remote sensing at OTTER: Macroscale satellite observations. *Ecol Appl.* 4:322–43.
Hatfield, J.L. (1979) Canopy temperatures: The usefulness and reliability of remote measurements. *Agron. J.* 71:889–92.
Hope, A.S., D.E.Petzold, S.N.Goward, and R.M.Ragan (1987) Simulating canopy reflectance and thermal infrared emissions for estimating evapotranspiration. *Water Resour. Bull.* 22:1011–19.

Jackson, R.D., S.B.Idso, R.J.Reginato, and P.J.Pinter (1981) Canopy temperature as a crop stress indicator. *Water Resour. Res.* 4:1133–8.
Jedlovec, G.J. (1990) Precipitable water estimation from high-resolution split window radiance measurement. *J. Appl. Meteorol.* 29:863–77.
Justice, C.O., T.F.Eck, D.Tanre, and B.N.Holben (1991) The effect of water vapour on the normalized difference vegetation index derived for the Sahelian region from NOAA AVHRR data. *Int. J. Remote Sens.* 12:1165–87.
Kahle, A.B. and R.E.Alley (1992) Separation of temperature and emittance in remotely sensed radiance measurements. *Remote Sens. Environ.* 42:107–11.
Kidwell, K.B. (1997) *Polar Orbiter Data User's Guide.* Department of Commerce, NOAA/NCDC, Washington, DC, 128 pp.
King, M.D., Y.J.Kaufman, W.P.Menzel, and D.Tanre (1992) Remote sensing of cloud, aerosol, and water vapor properties from the Moderate Resolution Imaging Spectrometer (MODIS). *IEEE Trans. Geosci. Remote Sens.* 30:2–26.
Kleespies, T.J. and L.M.McMillin (1990) Retrieval of precipitable water from observations in split window over various surface temperatures. *J. Appl. Meteorol.* 29:851–62.
Lakshmi, V. and J.Susskind (2001) Utilization of satellite data in land surface hydrology: Sensitivity and assimilation. *Hydrol. Process.* 15:877–92.
McClain, E.P., W.G.Pichel, and C.C.Walton (1983) Comparative performance of AVHRR-based multichannel sea surface temperatures. *J. Geophys. Res.* 90: 11587–601.
Menzel, W.P. and L.E.Gumley (1996) MODIS atmospheric profile retrieval: algorithm theoretical basis document. NASA-MTPE.
NASA (1996) NASA Strategic Plan, Office of Mission to Planet Earth. National Aeronautics and Space Administration.
Nemani, R.R. and S.W.Running (1989) Estimation of surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. *J. Clim. Appl. Meteorol.* 28:276–94.
NOAA, National Meteorological Operation Division and Climate Analysis Center (1994) *Daily Weather Maps: Weekly Series.* NOAA, NWS.
Norman, J.M., M.Divakarla, and N.S.Goel (1995) Algorithms for extracting information from remote thermal-IR observations of the Earth's surface. *Remote Sens. Environ.* 51:157–68.
O'Donnell, G.M., K.P.Czajkowski, R.O.Dubayah, and D.Lettenmaier (2000) Macroscale hydrological modeling using remotely sensed inputs: Application to the Ohio River basin. *J. Geophys. Res.* 105:12499–516.
Prabhakara, C., G.Dalu, and V.G.Kunde (1974) Estimation of sea surface temperature from remote sensing in the 11 and 13 μm window region. *J. Geophys. Res.* 79:5039–44.
Prata, A.J. (1993) Land surface temperature derived from the Advanced Very High Resolution Radiometer and the Along-Track Scanning Radiometer 1. theory. *J. Geophys. Res.* 98:16689–702.
Prata, A.J. and C.M.R.Platt (1991) Land surface temperatures measurements from AVHRR. In *5th AVHRR Data Users Conference,* June 25–28, Tromso, Norway, pp. 433–8.
Price, J.C. (1984) Land surface temperature measurements from the split window channels of the NOAA 7 advanced very high resolution radiometer. *J. Geophys. Res.* 89:7231–7.
Prihodko, L. and S.N.Goward (1997) Estimation of air temperature from remotely sensed observations. *Remote Sens. Environ.* 60:335–46.
Prince, S.D. and S.N.Goward (1995) Global net primary production: a remote sensing approach. *J. Biogeogr.* 22: 2929–49.

Prince, S.D., S.J.Goetz, R.Dubayah, K.Czajkowski, and M.Thawley (1998) Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using AVHRR satellite observations: validation of algorithms. *J.Hydrol* 213:230–49.

Running, S.W. and R.R.Nemani (1988) Relating seasonal patterns of the AVHRR vegetation index to simulated photosynthesis and transpiration of forests in different climates. *Remote Sens. Environ.* 24: 347–67.

Schmugge, T.J. and G.M.Schmidt (1998) Surface temperature observations from AVHRR in FIFE. *J. Atmos. Sci.* 55: 1239–46.

Smith, W.L. (1991) Atmospheric soundings from satellites-false expectation or the key to improved weather prediction? *Q.J.R.Meteorol. Soc.* 117: 267–97.

Sobrino, J.A., V.Caselles, and C.Coll (1993) Theoretical split window algorithms for determining the actual surface temperature. *Il Nuovo Cimento.* 16:219–36.

Steyn-Ross, M.L., D.A.Steyn-Ross, and W.J.Emery (1997) A dynamic water vapor correction method for the retrieval of land surface temperatures from AVHRR. *J. Geophys. Res.* 102:19629–43.

Sun, J. and L.Mahrt (1995) Relationship of surface heat flux to microscale temperature variations: applications to BOREAS. *J. Bound. Layer Meteorol.* 76: 291–302.

Susskind, J., P.Piraino, L.Rokke, L.Iredell, and A.Mehta (1997) Characteristics of the TOVS Pathfinder Path A data set. *Bull. Am. Meteorol. Soc.* 78:1449–72.

Tucker, C.J. (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8:127–50.

Ulivieri, C., M.M.Castronuovo, R.Francioni, and A.Cardillo (1994) A split window algorithm for estimating land surface temperature from satellites. *Adv. Space Res.* 14:59–65.

Wan, Z. and J.Dozier (1996) A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.* 34: 892–905.

Williams, D.L. (1989) The radiative transfer characteristics of spruce (Picea spp.): implications relative to the canopy microclimate. PhD thesis, University of Maryland.

Chapter 2

Land surface temperature retrieval techniques and applications

Case of the AVHRR

Yann H.Kerr, Jean Pierre Lagouarde, Françoise Nerry and Catherine Ottlé

2.1 Introduction

Except for solar irradiance components, most of the fluxes at the surface/atmosphere interface can only be parameterized through the use of surface temperature. Land surface temperature (LST) can play either a direct role, such as when estimating long wave fluxes, or indirectly as when estimating latent and sensible heat fluxes. Moreover, many other applications rely on the knowledge of LST (geology, hydrology, vegetation monitoring, global circulation models—GCM). Consequently, for many studies, it is crucial to have access to reliable estimates of surface temperature over large spatial and temporal scales.

As it is practically impossible to obtain such information from groundbased measurements, the use of satellite measurements in the thermal infrared appears to be very attractive since they can give access to global and uniform (i.e. with the same sensor and measurement characteristics) estimates of surface temperature. As a matter of fact, satellite thermal infrared sensors measure a radiance, which can be translated into top-ofthe-atmosphere brightness temperature. If the sensor is designed to work in a part of the spectrum where the atmosphere is almost transparent (e.g. 10.5–12.5 μm), access to surface temperature would seem to be an easy task. It is not generally the case however, due to the fact that the atmosphere, even though almost transparent, still has a non-negligible effect. Moreover, the surface emissivity is almost always unknown when land surfaces are not black or even grey bodies (i.e. the emissivity is not unity and may also be frequency dependent).

In summary, with satellites, we have a means of deriving spatial and temporal values of surface temperature, provided we can perform accurate atmospheric corrections and account for the surface emissivity.

Since thermal infrared data have been available, several approaches have been developed to infer surface temperature. The first problem to be solved

is to translate the satellite radiance into surface brightness temperature. After calibration and conversion of radiance into temperature using inverse Planck's law, it is necessary to account for the atmospheric contribution. It is then necessary to transform surface brightness temperature into surface temperature, and thus to take into account emissivity, and directional effects. Actually, the problem is slightly more complicated as atmospheric, emissivity, and directional effects are coupled and these modulating factors cannot be approached independently. The rationale here is to establish which are the most relevant factors.

The goal of this chapter is to give an overview of existing methods to retrieve surface temperature. Based on the existing space system we will assume that we have access to two thermal infrared channels around 11 and 12 μm. The practical aspects will be done with data from the Advanced Very High Resolution Radiometer (AVHRR) on board the National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellites. The different issues and possible solutions will then be presented. Finally, several examples of uses of surface temperatures will be briefly delineated.

In the following, we will not consider data calibration issues and assume that we have access to accurately calibrated top-of-the-atmosphere brightness temperatures.

In the second part, we will consider potential and/or proven applications of LST with associated problems.

2.1.1 Theoretical background

Without unnecessary details, we will now give the very basic concepts necessary to define the problem.

Role of the atmosphere

The energy going through an elementary solid angle per unit time and unit wavelength can be written as (Chandrasekhar 1960):

$$\mathrm{d}I_\lambda/\mathrm{d}s = (I_\lambda + B_\lambda)\tau_\lambda \tag{2.1}$$

where I_λ is the intensity of radiation at wavelength λ passing through an absorbing and emitting layer, s is the path length, B_λ is the blackbody emission of the layer given by the Planck function, and τ_λ is the optical depth.

After integrating equation (2.1) along the complete path between the surface and the top of the atmosphere, we have:

$$I_\lambda(\theta) = \varepsilon_\lambda(\theta)B_\lambda(T_s)\tau_\lambda(\theta) + R_{\mathrm{atm}\uparrow}(\theta) + (1-\varepsilon_\lambda(\theta))R_{\mathrm{atm}\downarrow}\tau_\lambda(\theta)$$

with

$$R_{\text{atm}\uparrow} = \int_0^h B_\lambda\,(T(z))\,\frac{\partial\,\tau_\lambda\,(\theta,z)}{\partial z}\,\mathrm{d}z$$
$$R_{\text{atm}\downarrow} = \frac{1}{\pi}\int_0^{2\pi}\int_0^{\pi/2}\int_\infty^0 B_\lambda\,(T(z))\,\frac{\partial\,\tau_\lambda\,(\theta,z)}{\partial z}\mathrm{d}z\,\sin(\theta_i)\cos(\theta_i)\,\mathrm{d}\theta_i\,\mathrm{d}\phi_i \quad (2.2)$$

where τ_λ, (θ) is the total directional transmission defined by

$$\tau_\lambda\,(\theta) = \exp\left[\int_0^h \frac{\alpha\,(\lambda,z)}{\cos\,(\theta)}e(z)\,\mathrm{d}z\right]$$

and *a* is the absorption coefficient for water vapor and *e* the water vapor concentration.

The first term of equation (2.2) is related to the surface contribution, the second to the atmospheric contribution along the upward path, and the third to the atmospheric contribution along the downward path, reflected by the surface and attenuated along the upward path.

Equation (2.2) indicates that it is necessary to take into account the atmospheric effects and advantage to make measurements in a spectral region where the atmospheric contribution is as small as possible. In most cases the satellite-borne sensors are designed to work in one of the thermal region "atmospheric windows" (10.5–12.5 μm in this chapter). When this condition is met, the first term of equation (2.2) will be least affected, while the relative importance of the second term will be very variable depending upon meteorological conditions (thin cirrus will have a significant influence, for instance). The role of the third term is related to the surface characteristics: the larger the emissivity, the smaller the contribution.

In this section, we neglect scattering in the atmosphere, this effect being small when visibility is higher than 5 km (McClatchey *et al.* 1971). We also neglect the influences of carbon dioxide (CO_2) and ozone (O_3), as they are much smaller than the effect of water vapor. However, the simulations take these effects into account (MODTRAN). Finally, we assume that the characteristics of the sensor (normalized response function, calibration) are well known and perfectly taken into account during the data calibration. It is worthwhile to note, however, that usually the modulation transfer function (MTF) of the sensor is not perfectly taken into account and that future systems would greatly benefit from an improved MTF.

Radiance temperature relationship

This section is again basic, but allows defining the terms given in some of the presented algorithms. Digital counts recorded by the radiometer are first converted into radiances and subsequently into brightness temperature values. For this, calibration is performed, giving the radiance in channel *i*: I_i.

The radiance I_i is related to the brightness temperature $T_{B,i}$ through the integration over the channel bandwidth $[\lambda_1, \lambda_2]$ of Planck's black-body function for the temperature $T_{B,i}$ weighted by the sensor's normalized response:

$$I_i(T_{B,i}) = \int_{\lambda_1}^{\lambda_2} f_{\lambda,i} \frac{2hc^2}{\lambda^5 \exp\left[(hc/\lambda\kappa T_{B,i}) - 1\right]} d\lambda \tag{2.3}$$

where h is Planck's constant, c is the velocity of light, and κ is Boltzman's constant.

Consequently, the brightness temperature $T_{B,i}$ can be retrieved from equation (2.3) either through a look-up table relating radiances to brightness temperatures, or by defining a central wavelength λ_i for each channel:

$$I_i(T_{B,i}) = \frac{2hc^2}{\lambda_i^5 \exp\left[(hc/\lambda_i\kappa T_{B,i}) - 1\right]}$$

It should be noted that the central wavelength λ_i is temperature dependent and is usually defined by the temperature range.

This section presented the main concepts necessary to compute a surface brightness temperature provided the sensor characteristics, the atmosphere, and the surface emissivity are all perfectly known. We subsequently will detail the evaluation of atmospheric and emissivity effects.

2.1.2 The AVHRR data

The NOAA meteorological polar orbitors are sun synchronous satellites whose altitude is nominally 825 km. They carry a scanning radiometer: the AVHRR. We will consider hereafter the case of the AVHRR/2 onboard the NOAA satellites, since only this version of the AVHRR has two different thermal infrared channels. The AVHRR/2 has five channels in the short wave (red and near-infrared), mid-infrared, and thermal infrared. The AVHRR/2 field of view is of$\pm 55^{\circ}$, which enables the system to view almost any point of the Earth's surface twice a day (ascending and descending orbits). Nominally (i.e. without considering the drift of the satellite), the overpass time is around 2 pm local solar time. Even though a given point of the surface is viewed every day, it must be noted that it will be viewed at different viewing angles on subsequent days, with the viewing conditions being approximately repeated only every 9 days. It is worth mentioning that a second satellite operates simultaneously, but on a different orbit (overpass time around 7 am local solar time).

The AVHRR/2 spectral bands are: 0.58–0.68, 0.725–1.1, 3.55-3.93, 10.30–11.30, and 11.50–12.50 μm. Nadir resolution is of the order of 1.1km. Algorithms using the 3.7-μm channel will not be discussed here, since they

can only be efficiently used at night due to the reflected solar signal (the directional reflectance in channel 3 is not well known) and since this channel can be saturated during daytime over some areas.

The first two bands are in the short-wave part of the spectrum and are widely used to derive the Normalized Difference Vegetation Index (NDVI), which is the ratio of the difference to the sum of the reflectances $\rho 1$ and $\rho 2$:

$$\mathrm{NDVI} = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1} \tag{2.4}$$

It has been shown (e.g. Tucker and Sellers 1986) that this ratio can be used to monitor biophysical properties of vegetation such as the Leaf Area Index (LAI)—which is the total area of the leaves per unit area—and photosynthetic capacity. However, the NDVI can only be used to quantify the vegetation LAI when the LAI does not exceed 3–5 due to a saturation effect. Another use of the short-wave channels is in estimating the vegetation fractional cover by using another index, the Modified Soil Adjusted Vegetation Index (MSAVI) (Chehbouni *et al.* 1994; Qi*et al.* 1994), which is insensitive to soil reflectances, but has to be computed from surface reflectances (hence requiring atmospheric corrections):

$$\mathrm{MSAVI} = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1 + L} \tag{2.5}$$

with

$$L = 1 - 2\gamma\,\mathrm{NDVI}\,(\rho_2 - \gamma\rho_1)$$

and where λ is the bare soil slope (λ=1.06).

We will tentatively use either the NDVI or the MSAVI to quantify the vegetation cover (i.e. the ratio between bare soil and vegetation).

Finally, another vegetation index the Global Environment Monitoring Index (GEMI) (Pinty and Verstraete 1992) is:

$$\mathrm{GEMI} = \eta(1 - 0.25\eta) - (\rho_1 - 0.125)/(1 - \rho_1) \tag{2.6}$$

with

$$\eta = [2(\rho_2 - \rho_1) + 1.5\rho_2 + 0.5\rho_1)/(\rho_1 + \rho_2 + 0.5)]$$

This index is rather insensitive to the atmosphere (Leprieur *et al.* 1996), but very sensitive to surface reflectances. Hence, we propose to use it as a surrogate for finer cloud discrimination. Actually, partial cloud cover might not be readily visible using conventional methods (especially due to partial cloud cover and to cirrus clouds) when, according to our experience gained from HAPEX-SAHEL (Prince *et al.* 1995), the GEMI has a tendency to show

partially covered pixels. However the method will require some sort of thresholding which is delicate to implement on an operational scheme. Other indices do exist and it might be interesting to check whether some of them would not prove more interesting.

Thermal band calibration is rather straightforward. The sensor views thermistances and deep space that gives the calibration curve (Kidwell 1986). Non-linearities can be taken into account (Brown 1985). The procedure is simple and reliable even though some questions were recently raised concerning the "hot target" blackbodies. Consequently, deriving brightness temperatures at the top of the atmosphere is relatively simple and reliable. The problem we will study now is the atmospheric correction procedure. The errors induced by the atmospheric contribution will be especially large for hot surfaces with humid atmospheres.

2.1.3 Practical satellite-based methods

Problem I: atmospheric profile method

A primary method to perform atmospheric corrections is to use a radiative transfer model coupled with a characterization of atmospheric structure. The characterization can be made from "standard values" such as climatological means, but this characterization is bound to introduce large errors due to the spatial and temporal variability of the atmosphere. It has also been suggested to use indirect methods, such as the use of a reference target (typically a large water body) of known and uniform temperatures (which is another challenge), to assess atmospheric contribution, assuming that the atmosphere characteristics will not change spatially, which, obviously, is not the general case. Moreover, this method relies on only one measurement, which is a "cold" reference in the case of a water body, when at least two are necessary (hot and cold as the lower levels of the atmosphere are affected by surface temperature).

It is thus necessary to use more accurate characterizations of the atmosphere. Several methods have been used to assess the pressure, temperature, and humidity (PTU) profiles of the atmosphere. The most evident being to use radiosoundings. The PTU profile can then be used as an input to a radiative transfer code such as the "4A" (Scott and Chedin 1981), LOWTRAN (Kneizys *et al.* 1983 and subsequent updates), MODTRAN, or even WINDOW (Price 1983). This approach can give very satisfactory results, provided the radiosoundings are synchronous and collocated with the satellite measurement. Otherwise, large errors can be introduced (up to 10 K as shown by Cooper and Asrar 1989). Moreover, the use of radiosoundings is hampered by the insufficient density of the network in some areas (3, for example, for the whole Sudanian Sahel), by the timing (usually 12:00 UT and, in some

cases, for 00:00 UT), which is not the satellite overpass time, by the poor representativity in some cases (e.g. near the coast in arid areas), and by the difficulty to access the data in a timely fashion or in digital form. Generally, ground-based radiosoundings do not really fit our needs. In this study, we, nevertheless, relied heavily on atmospheric profiles and RT codes for assessing the different methods.

An alternative to radiosoundings is to use atmospheric profiles derived from satellite measurements (Susskind *et al.* 1984; Chedin *et al.* 1985), but in this case the inversion algorithms are time consuming and very complex. Moreover, the existing sounders do not have the capacity to accurately retrieve profiles near the surface face, where most of the atmospheric water vapor is located. Large errors may result from the resolution (30 km) and related surface emissivity variability within the pixel as shown by Ottlé and Stoll (1993). Such methods are thus not yet relevant, but they will need to be investigated further when we enter the Earth Observing System (EOS) era since the NASA and ESA polar platform will carry more sophisticated sounding instruments (AIRS, Atmospheric InfraRed Sounder; IASI, Interféromètre Atmosphérique de Sondage dans l'Infra-rouge). The simultaneous use of such a sensor coupled with (MODIS) MODerate resolution Imaging Spectroradiometer or (MERIS) MEdium Resolution Imaging Spectrometer should allow us to derive accurate surface brightness temperatures.

Another possibility is to use the output of meteorological forecasting models. Actually this is the most appropriate method for the time being. The reanalyses are global and available one a roughly 1×1° grid every 6 h. Crude interpolation might be sufficient to derive accurate enough estimates of the integrated water content to be used with differential absorption methods. For a radiative code correction however, the reanalyses will not be accurate enough, and more importantly, they are available only at UT times (usually 0, 6, 12, 18) posing temporal interpolation issues.

Consequently, based on existing systems and ancillary data, we will focus here on alternative methods, which, even though less accurate theoretically, have the advantage of being suitable for global applications and can be run "operationally," without sophisticated ancillary data. We have mainly investigated the differential absorption method (the so called Split Window Techniques, SWT).

The differential absorption method: background

When two channels, or more, corresponding to different atmospheric transmissions, are available, it is possible to use the differential absorption to estimate the atmospheric contribution to the signal. This method was first suggested by Anding and Kauth (1970) and put in its now "classical" form

by Prabhakara *et al.* (1974). It has been since adapted and tested successfully with AVHRR data, mainly over sea surfaces (Njoku 1985). Its general name is the SWT. The SWT has been tested mainly for Sea Surface Temperature (SST) retrievals, Some comparisons over land surfaces have also been done, but with varying degrees of success (Price 1984; Lagouarde and Kerr 1985; Cooper and Asrar 1989).

The SWT relies on the different absorption characteristics of the atmosphere within two different but close wavelengths. The algorithm consists simply of a linear combination of the thermal channels, which gives a surface temperature pseudo-corrected for the atmospheric contribution. For the AVHRR/2 the equation is of the type:

$$T_s = a_0 + a_1 T_{10.8} + a_2 T_{11.9} \tag{2.7}$$

with

$$a_1 + a_2 = 1 \tag{2.8}$$

where $T_{10.8}$ and $T_{11.9}$ are the brightness temperatures at the top of the atmosphere in the two infrared bands. The a_i coefficients are estimated using various methods depending on the authors. The SWT is now used operationally over oceans with a claimed accuracy of 0.7K (McClain *et al.* 1985: AVHRR data).

We note that an alternative method, consisting in using different view angles can be used. It can rely on measurements made by two different satellites (Becker 1982; Chedin *et al.* 1982), or the same satellite provided it can view along track with two different angles. The Along Track Scanning Radiometer (ATSR) on board ERS-1 satisfies this dual viewing angle and differential absorption technique simultaneously (Eccles *et al.* 1989; ESA 1989; Prata *et al.* 1990).

Even though the SWT works satisfactorily over sea surfaces, when used directly as developed for SST over land surfaces, the errors can reach 6K (Lagouarde and Kerr 1985). This is mainly due to the fact that the assumptions made for the SWT over sea surfaces are not applicable for land surfaces. We are now going to study the SWT assumptions and describe why this is so. We will then describe the main methods currently proposed for land surface temperature estimation.

PROBLEM 2: ROLE OF THE EMISSIVITY

The SWT has been developed for sea surfaces. It is a simplified way to take into account atmospheric effects, and thus relies on a number of assumptions such as (see Becker 1987 for a more detailed analysis):

1 the surface is lambertian;
2 the surface temperature is close to the temperature in the lower layers of the atmosphere, the latter varying slowly (Planck's law linearization);

3 the surface temperature does not exceed 305 K;
4 absorption in the atmosphere is small and occurs essentially in the lower layers;
5 the surface emissivity is very stable spatially and close to unity;
6 the emissivities $\varepsilon_{10.8}$ and $\varepsilon_{11.9}$ are almost identical and $\varepsilon_{10.8}>\varepsilon_{11.9}$.

It is obvious that these conditions are not usually met over land surfaces, hence the problems encountered when using the SST-SWT over land surfaces. Nevertheless, provided we accept a somewhat reduced accuracy, the SWT could be adapted to land surface temperature retrieval.

In the specific case of the AVHRR, several limitations linked to the instrument itself are to be considered:

1 the sensor saturates for temperatures higher than 320K;
2 the ascending node time may drift;
3 due to its large scanning angle, the sensor views simultaneously points whose local solar time are quite different (almost 2 h from one end of the scan to the other);
4 for two successive overpasses, the sensor views a given point at different angles and at a different solar time. Thus, over heterogeneous areas, angular effects are bound to exist between subsequent acquisitions.

The perturbing effects on the SWT when used over land surfaces are mainly the following:

1 the surface spectral emissivity is a priori unknown and different from unity;
2 spatial variability of the emissivity can be high;
3 surface temperature may have high spatial variability at scales smaller than the resolution of the AVHRR;
4 a strong difference between air and surface temperature may exist.

We are now going to analyze the influence of the emissivity on the split window algorithms.

EMISSIVITY INDUCED ERRORS

Land surface emissivity has two characteristics whose effects are negative in terms of retrieval accuracy:

1 The spectral emissivity in the band 10.3–12.5 μm is not equal to 0.99 but presents such a spectral variability (Buettner and Kern 1965; Fuchs and Tanner 1966; Fuchs *et al.* 1967; Salisbury and D'Aria 1992, 1994) that integrated values over the AVHRR thermal channels might range from 0.92 to 1.

2 The spectral emissivity is generally constant over the two AVHRR channels. It has been shown by Becker (1987) that if we assumed that the spectral emissivities $\varepsilon_{10.8}$ and $\varepsilon_{11.9}$ were equal to 1 when they are actually different from one another and different from one, the error ΔT induced by such an assumption on the retrieved surface temperature using the SWT could be written (Becker 1987):

$$\Delta T = 50\,\frac{1-\varepsilon}{\varepsilon} - 300\,\frac{\Delta\varepsilon}{\varepsilon} \tag{2.9}$$

where

$$\varepsilon = \frac{\varepsilon_{10.8} + \varepsilon_{11.9}}{2} \quad \text{and} \quad \Delta\varepsilon = \varepsilon_{10.8} - \varepsilon_{11.9} \tag{2.10}$$

The difference $\Delta\varepsilon$, when positive, reduces the errors since the second term compensates the first in equation (2.9). This case occurs for water and vegetation. Moreover, the closer to 1 is ε the smaller will be the errors on the retrieved surface temperature T_s. In conclusion, the classical SWT will give good results over water, slightly less over fully vegetated areas, and poor results on dry bare soil.

It is thus necessary to know the two spectral emissivities to accurately derive surface temperature, which gives us a system of two measurements for three unknowns. The problem is thus a priori not solvable. Several authors thus proposed local SWT with coefficients a_i being functions of the surface, atmosphere and view angles, and derived from either exact knowledge of the emissivity or empirically.

2.2 Review of existing algorithms

2.2.1 Algorithms that do not satisfy either accuracy or global applicability requirements

We assume here that the surface temperature algorithm has to be applicable nearly globally and limit ourselves to AVHRR-type data. We have not considered algorithms requiring both night and day data, since it not practical in many areas (the probability of having regularly successive night and day cloud free acquisitions proving to be very small) and algorithms using variance/covariance methods to infer atmospheric variations (Ottlé *et al.* 1998) as there is still some controversy on the global efficiency of such methods. The accuracy goal is 1.5 K.

Empirical methods

The first way to approach atmospheric corrections is to use methods qualified here as "empirical." Using bodies of known temperature (oceans, water

bodies, ice caps, etc.) a relation is established between the surface temperature and the top of atmosphere temperature. This relationship is then applied to the neighboring pixels over land. This method is, of course, not applicable globally and is highly subject to errors. The errors are linked: (a) to the accuracy with which the "known" temperatures are established and, more important, (b) to the variations of emissivity between the "calibrating" targets and neighboring pixels, and finally (c) to the variations in atmospheric characteristics over water or ice and land surfaces.

Another method uses estimates of low-level atmospheric temperature and humidity to establish the atmospheric corrections. Here again the problem is the validity of the relationships and the availability of global fields of air temperature or humidity.

Similarly, empirical methods using the difference between temperatures at the top of the atmosphere in bands 4 and 5 of the AVHRR have been developed to infer atmospheric water content. The accuracy of such approaches is highly linked to having minimal emissivity differences between channels 4 and 5 (Choudhury *et al.* 1995). As such, they are bound to fail when applied globally.

All the empirical methods, even though they might in some cases deliver very good results over a specific test site and at a given time, are, a priori, not applicable globally and over long periods of time. Consequently, they will not be considered as applicable to our problem.

Radiative transfer approaches

The methods using radiative transfer approaches to infer atmospheric contributions are by essence exact, provided the radiative transfer model is correct (and, thus, all the line contributions well described) and the surface and atmospheric characteristics are very well known. Several such radiative transfer codes exist, of which the most used are LOWTRAN7 and MODTRAN. A simplified method developed by Price (WINDOW 1983) also gives satisfactory results. The main problems in using such codes are the following:

1 They are rather complex and require significant computing resources. This problem is, however, not drastic, as a look-up table approach can be developed.
2 They require a good knowledge of surface spectral emissivity that is not globally available.
3 They require an accurate description of the atmosphere. To have this description, several approaches have been considered namely: use of radiosoundings, satellite sounders, or reanalyses. As described in section "Problem 1: atmospheric profile methods" none of these are fully satisfactory.

Consequently, the methods relying on radiative transfer approaches do not seem to be applicable to our present problem, even though they can be used quite efficiently for validating other methods or establishing the coefficients for a parameterized approach.

2.2.2 Split window techniques

For the problem under consideration, we are thus left with the differential absorption methods, which we will now review, by categories.

It has been noted that between channels 4 and 5 of the AVHRR, the main difference between the top of atmosphere temperatures was directly linked to the difference in water vapor absorption, assuming the surface emissivities in the two channels are really identical. If this assumption is correct, extracting surface temperature is rather straightforward. The problem is usually much more complicated as land surface characteristics are often not the same in the two bands.

The SWT can be summarized with the following formula, similar to equation (2.7) but now employing three channels:

$$T_s = a_0 + a_3 T_3 + a_1 T_4 + a_2 T_5 \tag{2.11}$$

where the subscript in T denotes the AVHRR channel. The use of band 3 is subject to some care. At nighttime it can be used as is but for daytime data, the reflective component must be taken into account. Moreover, bidirectional reflectances in channels around 3.7 μm are not known which can significantly increase errors.

The SWT has been used for some time now over ocean surfaces with very good results (see section on "The differential absorption method: background"). Numerous authors developed new schemes to retrieve surface temperature over land using "improved" or "adapted" SWT for land surfaces. The general conclusion is that even though the accuracies achieved for ocean surfaces will not be feasible with such approaches, with our current knowledge of surface spectral emissivity, accuracies of 1–2 K are attainable. Consequently, we tested mainly SWT methods. They can be sorted out in the categories:

1 purely empirical;
2 depending on the spectral emissivities;
3 depending upon water vapor content;
4 depending upon view angle;
5 depending on any combination of the above methods.

We will now briefly describe these categories.

Empirical and semi-empirical SWT

From existing data sets an empirical set of coefficients is established. The set of coefficients is not globally applicable. Sometimes the empirical coefficients are coupled with physically based expressions. The global validity, however, is not assured.

Empirical approaches are not global in nature, we have tested only one such algorithm for the sake of comparison. It is from May *et al.* (1992), from NESDIS (National Environmental Satellites, Data, and Information Service) and is only dependent on the view angle. Several similar algorithms of the type were developed in the 1980s and early 1990s. We investigated only this one because of its NESDIS background and because it is representative of similar algorithms with almost constant coefficients but having an "angle correction" adjustment in the constant term (a_0).

The May *et al.* algorithm gives a surface temperatures T_m given by

$$T_m = \left[1.0162\ T_4 + 2.657(T_4 - T_5) - 0.5265\ (S_v - 1)(T_4 - T_5) - 4.58\right] \tag{2.12}$$

Where S_v is the secant of the view angle, T_4 and T_5 are, respectively, the brightness temperatures measured in channels 4 and 5 of the AVHRR.

Emissivity-dependent algorithms

In these algorithms, the coefficients are functions of spectral emissivity. The Kerr *et al.* (1992) algorithm was the first attempt to take the emissivity into account (very coarsely) when no fields of surface emissivity are currently available. Strictly speaking, this algorithm should classified in the section on "Empirical and semi-empirical SWT." To account for the varying emissivities of vegetation and soil, two temperatures are computed, one for typical bare soil emissivities T_{bs}, the other for vegetation T_{ve}, and the two are added with a weighting proportional to the vegetation cover derived from the NDVI or MSAVI (see Section 2.1.2) resulting in T_k given by

$$\begin{aligned} T_{ve} &= (-2.4 + 3.6T_4 - 2.6T_5) \\ T_{bs} &= (3.1 + 3.1T_4 - 2.1T_5) \\ T_k &= \left[CT_{ve} + (1 - C)T_{bs}\right] \end{aligned} \tag{2.13}$$

C being the canopy fraction.

However, if the spectral emissivities are known, it is possible to infer "exact" coefficients as derived by Becker and Li (1990b).

THEORETICAL DEVELOPMENT

Becker and Li (1990b) propose, after a linearization of the radiative transfer equations, a method to derive the coefficients a_i of the SWT. This method is only valid locally, since emissivities have a high spatial variability. Consequently, as was suggested by Kerr and Lagouarde (1989), for land surfaces, an SWT can only work locally and has to be tailored to local surface characteristics. Becker and Li (1990b) give an expression for the SWT, where the coefficients a_i are derived from the actual emissivity values. For the specific case of the AVHRR/2 on board the NOAA-9 satellite, they obtain the analytical expression of the coefficients a_i given by

$$\begin{aligned} a_0 &= 1.274 \\ a_1 &= 3.63 + 2.068\frac{(1-\varepsilon)}{\varepsilon} + 18.924\,\frac{\Delta\varepsilon}{\varepsilon^2} \\ a_2 &= -2.63 - 1.912\frac{(1-\varepsilon)}{\varepsilon} - 19.406\,\frac{\Delta\varepsilon}{\varepsilon^2} \end{aligned} \tag{2.14}$$

where $\varepsilon=(\varepsilon_{10.8}+\varepsilon_{11.9})/2$ and $\Delta\varepsilon=\varepsilon_{10.8}-\varepsilon_{11.9}$.

This method has been tested in the framework of HAPEX MOBILHY (André *et al.* 1988) over one location in the South West of France, where the spectral emissivities had been estimated (Becker and Li 1990b). The results were reasonable, but the method is yet to be validated over other types of surfaces. The main results from that study were that the coefficients a_i are independent of the atmosphere and can be derived directly from the knowledge of the spectral emissivities. Actually, this did not prove to be accurate, as we will see later, and the impact of atmospheric effects on the a_i also needed to be ascertained. Values of surface emissivities can be found in the literature, even though they are scarce (Buettner and Kern 1965; Fuchs and Tanner 1966; Fuchs *et al.* 1967; Elvidge 1988; Takashima and Matsuda 1988; Eastes 1989), but we do not have access yet to global data sets of emissivity at sufficient spatial resolution. A large database is nevertheless now available and was used here establish the ranges of emissivity values and variations. To apply such a method routinely, global atlases of emissivities are required, or a method to infer emissivities needs to be developed. This last point will also been investigated here.

OTHER METHODS USING EMISSIVITY

A number of algorithms have been developed and proposed in the literature. They are somewhat similar in formulation and several are directly inspired by Becker and Li's formulation. Otherwise they are empirically derived from ground data. A good example of the empirical approach is given in Prata and Platt (1991). Several authors have developed modified algorithms,

notably Sobrino *et al.* (1994). In summary, we have a large number of algorithms that are described below:

With the following definitions:

$$\Delta\varepsilon = \varepsilon_4 - \varepsilon_5, \quad \varepsilon = \frac{\varepsilon_4 + \varepsilon_5}{2} \tag{2.15}$$

$$\varepsilon_{1\mathrm{bl}} = \frac{1-\varepsilon}{\varepsilon} \tag{2.16}$$

$$\varepsilon_{2\mathrm{bl}} = \frac{\Delta\varepsilon}{(\varepsilon)^2} \tag{2.17}$$

$$\varepsilon_{3\mathrm{v}} = \frac{\Delta\varepsilon}{\varepsilon} \tag{2.18}$$

we have the following formulations, each retrieve temperature T having a subscript recalling the author's initials:

Becker and Li(1990):

$$T_{\mathrm{bl}} = \left[1.274 + \left[\frac{T_4 + T_5}{2}(1 + 0.15616\,\varepsilon_{1\mathrm{bl}} - 0.482\,\varepsilon_{2\mathrm{bl}}) + \frac{T_4 - T_5}{2}(6.26 + 3.98\,\varepsilon_{1\mathrm{bl}} + 38.33\,\varepsilon_{2\mathrm{bl}})\right]\right] \tag{2.19}$$

Becker and Li 1990 (modified by Sobrino *et al.* 1994):

$$T_{\mathrm{bls}} = \left[1.737 + \left[\frac{T_4 + T_5}{2}(1 + 0.00305\,\varepsilon_{1\mathrm{bl}} - 0.376\,\varepsilon_{2\mathrm{bl}}) + \frac{T_4 - T_5}{2}(5.17 + 21.44\,\varepsilon_{1\mathrm{bl}} + 30.67\,\varepsilon_{2\mathrm{bl}})\right]\right] \tag{2.20}$$

Prata and Platt (1991):

$$T_{\mathrm{pp}} = \left(3.45\frac{T_4 - T_0}{\varepsilon_4} - 2.45\frac{T_5 - T_0}{\varepsilon_5} + 40\frac{1-\varepsilon_4}{\varepsilon_4} + T_0\right) \tag{2.21}$$

Prata and Platt (1991) (modified by Caselles *et al.* 1997):

$$T_{\mathrm{ppc}} = \left(\frac{3.46}{\varepsilon}T_4 - \frac{2.46}{\varepsilon}T_5 + 40\frac{1-\varepsilon}{\varepsilon}\right) \tag{2.22}$$

Prata and Platt (1991) (modified by Sobrino *et al.* 1994):

$$T_{\mathrm{pps}} = \left(3.56\frac{T_4 - T_0}{\varepsilon_4} - 2.61\frac{T_5 - T_0}{\varepsilon_5} + 30.7\frac{1-\varepsilon_4}{\varepsilon_4} + T_0\right) \tag{2.23}$$

Price (1984):

$$T_p = \left[\left[T_4 + 3.33\,(T_4 - T_5) \right] \frac{5.5 - \varepsilon_4}{4.5} + 0.75\,T_5\,\Delta\varepsilon \right] \tag{2.24}$$

Price (1984) (modified by Sobrino *et al.* 1994):

$$T_{ps} = \left[\left[T_4 + 2.79\,(T_4 - T_5) \right] \frac{7.6 - \varepsilon_4}{6.6} + 0.26\,T_5\,\Delta\varepsilon \right] \tag{2.25}$$

Ulivieri and Cannizzaro (1985):

$$T_{u0} = \left[T_4 + 3\,(T_4 - T_5) + 51.57 - 52.45\,\varepsilon \right] \tag{2.26}$$

Ulivieri *et al.* (1992):

$$T_u = \left[T_4 + 1.8\,(T_4 - T_5) + 48(1 - \varepsilon) - 75\,\Delta\varepsilon \right] \tag{2.27}$$

Ulivieri *et al.* (1992) (modified by Sobrino *et al.* 1994):

$$T_{us} = \left[T_4 + 2.76\,(T_4 - T_5) + 38.6(1 - \varepsilon) - 96.0\,\Delta\varepsilon \right] \tag{2.28}$$

Vidal (1991):

$$T_{av} = T_4 + 2.78\,(T_4 - T_5) + 50\,\varepsilon_{1bl} - 300\,\varepsilon_{3v} \tag{2.29}$$

Coll *et al.* (1997):

$$T_c = \left[T_4 + 2.13\,(T_4 - T_5) + 0.18 + 50\,(1 - \varepsilon_4) - 200\,\Delta\varepsilon \right] \tag{2.30}$$

Sobrino *et al.* (1993):

$$T_{so} = \left[T_4 + 1.06\,(T_4 - T_5) + 0.46\left[(T_4 - T_5)\,(T_4 - T_5) \right] + 53\,(1 - \varepsilon_4) - 53\,(\varepsilon_4 - \varepsilon_5) \right] \tag{2.31}$$

Methods using both emissivity and water vapor

We have identified only one such method, developed by Sobrino *et al.* (1991). This method poses some problems as it uses the opacity of the atmosphere in AVHRR bands 4 and 5, and thus requires a good knowledge of the atmosphere and a radiative transfer code. It is a bit too complex to be detailed in this work.

Methods using both emissivity and view angle

Again, in this section, we have identified only one method, developed by François and Ottlé (1996) and called QUAD (noted T_{quad}). This method was developed for the Along Track Scanning Radiometer (ATSR) and thus only established for the two ATSR view angles. We have tested it as is, but strictly speaking, one should recompute the coefficients for the AVHRR and for various angles, as angular interpolation is very risky with such approaches:

$$T_{quad} = \left[T_4 + c_{21}\,(T_4 - T_5) + c_{22}\,(T_4 - T_5)^2 + c_{23}\right] \tag{2.32}$$

The coefficients C_{ij} for equation (2.32) are tabulated (LUT) and available in their paper. Of note is that the expression is not linear in T.

Methods using emissivity, atmosphere, and view angle

These are the most sophisticated methods and were all developed fairly recently. We have tested four of them.

1 The first one developed was by Ottlé and Vidal Madjar in 1992 (T_{ovm}). It was developed with the use of Radiative Transfer (RT) code for a variety of conditions. Three coefficients are extracted according to water vapor content, emissivity values and view angle. We have compiled a set of values for different surface conditions [emissivity and angles (LUT)], and different types of atmosphere (tropical, temperate, and polar).
2 Sobrino *et al.* (1994) proposed a method based on the use of the atmospheric opacities in AVHRR bands 4 and 5 and their ratio (T_{sob2}). This requires knowledge of the atmosphere and use of the RT code. For this study, we have used coarse values for the water vapor derived t, as it will be more representative of real-world application where accurate knowledge of water vapor is missing.
3 Becker and Li (1995) proposed a method inspired by their 1990 formulation (T_{grtr}), but including angle and water vapor contribution in the coefficients.
4 Finally, François and Ottlé (1996) proposed an algorithm (T_{wvd}) with coefficients that are quadratic functions of water vapor.

2.2.3 Conclusion

An overview of the existing algorithms in the literature shows that there are many different algorithms, even though the general approaches can be classified in a few categories. The methods proposed vary considerably in complexity and ancillary data requirements. A primary question is whether accuracy is related to increased complexity, and thus when increasing

complexity is of no additional benefits. The following sections are devoted to analyzing the main perturbing factors and then evaluating the most efficient algorithm.

2.3 Emissivity retrieval

2.3.1 Introduction

Current emissivity databases (Salisbury and D'Aria 1992, 1994) have only consisted of laboratory measurements and do not represent the emissivity at the scale of the NOAA/AVHRR pixel (1×1 km^2). The emissivity parameters, necessary to tune or compute Split Window coefficients, need to be retrieved directly from satellite data and preferably simultaneously with temperature determination to take into account changes of the state of the surface due to humidity, vegetation growth and others temporal changes. Methods have been developed using multispectral data (Schmugge *et al.* 1998) and maybe applicable to newer multispectral systems such as ASTER (Advanced Spaceborne Thermal Emission and Reflection radiometer). They are not thus considered here, but are worth mentioning, as they should enable the derivation of global spectral surface emissivities.

2.3.2 Review of existing methods

Two main methods exist to assess emissivity from space:

1 The methods that establish an empirical relationship between shortwave (0.58–0.68 and 0.725–1.1 μm) channel measurements and the emissivity (Van de Griend and Owe 1993; Valor and Caselles 1996).
2 The second method aims at deriving the emissivity by solving the radiometric equation at the surface in the thermal infrared (Becker and Li 1990b; Kealy and Hook 1993; Li and Becker 1993).

Method 1: Van de Griend and Owe (1993)

This method links emissivity to NDVI using an empirical relationship established from ground measurements.

$$\varepsilon = a + b\log(\mathrm{NDVI}) \tag{2.33}$$

This method has the drawback that the coefficients a and b have been determined only for the 8–14 μm band and are surface dependant.

Method 2: "vegetation cover" (Valor and Caselles 1996)

Emissivity is written as a function of bare soil and vegetation emissivities weighted with the vegetation cover fraction. A term due to cavity effect

is added:

$$\varepsilon = \varepsilon_v P_v + \varepsilon_s(1 - P_v) + d\varepsilon \tag{2.34}$$

with

$$d\varepsilon = (1 - \varepsilon_s)\varepsilon_v F(1 - P_v) + [(1 - \varepsilon_v)\varepsilon_s G + (1 - \varepsilon_v)\varepsilon_v F']P_s \tag{2.35}$$

or, on a simplified form:

$$d\varepsilon = 4\langle d\varepsilon \rangle P_v(1 - P_v) \tag{2.36}$$

where P_v is the vegetation cover fraction, P_s is the soil cover fraction, ε_v is the vegetation canopy emissivity, ε_s is the bare soil emissivity, and F, G, F' are shape factors.

The vegetation cover fraction is estimated through the NDVI:

$$NDVI = NDVI_v P_v + NDVI_s(1 - P_v) + di \tag{2.37}$$

where NDVI is the pixel's NDVI, $NDVI_v$ is the vegetation NDVI, $NDVI_s$ is the soil NDVI, and di is the error related to the approximation, which gives for the emissivity:

$$\varepsilon = \frac{\varepsilon_v - \varepsilon_s}{NDVI_v - NDVI_s} NDVI + \frac{\varepsilon_s(NDVI_v + di) - \varepsilon_v(NDVI_s + di)}{NDVI_v - NDVI_s} + d\varepsilon \tag{2.38}$$

This approach requires a preliminary knowledge of (a) the vegetation and bare soil NDVI; (b) the vegetation and bare soil emissivities; (c) the shape factors. Also, expression (2.34) is only valid for an isothermal media, that is, when soil and vegetation temperatures are close to each other, which is not necessarily the case especially around noon over a dry and hot area.

Method 3: "alpha residuals" (Kealy and Hook 1993)

This method utilizes Wien's approximation of the Planck function:

$$R_j = \varepsilon_i \frac{C_1}{\lambda_j^5 [\exp(C_2/\lambda_j T)]} \tag{2.39}$$

Emissivity in channel j is then written as:

$$\varepsilon_j = \exp \frac{\alpha_j + (1/N)\sum_{i=1}^{N} \lambda_i \ln \varepsilon_i}{\lambda_j} = \exp \frac{\alpha_j + X}{\lambda_j} \tag{2.40}$$

with

$$\alpha_j = \lambda_j \ln R_j - \frac{1}{N}\sum_{i=1}^{N} \lambda_i \ln R_i + K_j, \quad K_j \text{ being a constant}$$

The X parameter is estimated from the alpha variance: $X=c\left(\sigma_\alpha^2\right)^{1/M}$, with the coefficients c and M retrieved from laboratory measurements.

This approach is applicable to ground radiances. Consequently, atmospheric corrections need first to be performed; the X parameter is dependent on the spectral response of the surface. More importantly, this method depends on the number of spectral channels used and their central wavelengths. The method is thus essentially multispectral and not really applicable to AVHRR-type sensors with only two channels in the thermal infrared. This approach also relies on homogeneous pixels (ASTER type), which is not generally the case for the AVHRR resolution.

Method 4: "TISI" (Becker and Li 1990a; Li and Becker 1993)

In AVHRR channel 3, the daytime radiance is a combination of the emitted radiance by the surface and a reflected radiance due to sun illumination. Day-night image pairs are used to estimate emitted radiance and thus the reflected contribution. Reflectivity is retrieved and emissivity is deduced using a Lambertian assumption. The emissivities in the three AVHRR channels are given by the following formulae:

$$\varepsilon_3 = 1 - \pi \frac{R_3^{\mathrm{d}} - R_3^{\mathrm{n}}\left[\left(R_4^{\mathrm{d}}/R_4^{\mathrm{n}}\right)\right]^{\mathrm{n}_3}}{R_{\mathrm{sun}}} \tag{2.41}$$

$$\varepsilon_4 = \varepsilon_3^{\mathrm{n}_4/\mathrm{n}_3}\left[\frac{T_4^{\mathrm{n}}}{T_3^{\mathrm{n}}}\right]^{\mathrm{n}_4} \tag{2.42}$$

$$\varepsilon_5 = \varepsilon_3^{\mathrm{n}_5/\mathrm{n}_3}\left[\frac{T_5^{\mathrm{n}}}{T_3^{\mathrm{n}}}\right]^{\mathrm{n}_5} \tag{2.43}$$

where R_i^{d} is the daytime radiance in channel i, the nighttime radiance in channel i, R_{sun} is the solar irradiance in channel 3, and $\mathrm{n_i}$ is defined by $R_{\mathrm{i}} = a_{\mathrm{i}}.R_i^{\mathrm{n}}$

This methodology is also applicable only to surface radiances, so atmospheric corrections need to be performed first. It is also necessary to have access to day and night acquisitions sufficiently close in time so that the surface conditions have not changed (moisture/vegetation). There are some concerns about the lambertian behavior expected in the 3.7-μm channel, which is probably not generally the case.

The "TISI" method is given here for information as it is an interesting approach. However, the issues linked with the possibility to acquire contiguous day-night acquisitions and the uncertainties associated with nonlambertian behavior of the 3.7-μm channel makes it non-applicable in a routine scheme.

2.3.3 Choice of a suitable method and related uncertainties (errors)

Method 1 is valid only locally where the coefficients a and b have been determined, and so can only be applied locally. Method 3 appears to be adequate for multispectral sensors such as ASTER. The fourth method is not applicable for sensors not having a 3.7-μm channel and when day-night pairs are not always available, significantly reducing the potential uses. It must be stressed that the 3.7-μm channel has to be well designed, as problems linked to noise or saturation effects (as encountered with the AVHRR) are prohibitive. Method 2 requires the knowledge of the emissivity and the NDVI of the pixels. This simplistic method has the advantage of being easy to implement, and thus will be analyzed further below and related errors estimated.

The error related to the determination of the emissivity using the "vegetation cover" depends on how accurate is the a priori knowledge of the emissivity of the pure (i.e. homogeneous) pixels ($\Delta\varepsilon_v \Delta\varepsilon_s$), the estimation of the vegetation cover (ΔP_v), and the knowledge of the shape factors A ($d\varepsilon$).

The error of $\varepsilon(\Delta\varepsilon)$ is given by

$$\Delta\varepsilon = P_v \Delta\varepsilon_v + (1 - P_v)\Delta\varepsilon_s + \left[(\varepsilon_v - \varepsilon_g) + 4\langle d\varepsilon\rangle - 8\langle d\varepsilon\rangle P_v\right]\Delta P_v + 4P_v(1 - P_v)\Delta\langle d\varepsilon\rangle \tag{2.44}$$

We note the following points:

1 $\Delta\varepsilon_v$: the emissivity of vegetation can be assumed to be well known in the thermal infrared, with a flat spectral response and a high value. ε_v=0.985±0.005 may be considered as a good approximation.
2 $\Delta\varepsilon_s$: the bare soil emissivity is more variable with a spectral variability depending on the wavelength. An emissivity database can give information on the range of emissivity values. Using the database described by Salisbury and D'Aria (1992) and integrating over the spectral responses of channels 4 and 5 of the AVHRR, mean values with associated statistical errors can be determined. We obtained ε_s^4 =0.95±0.02 and ε_s^5 =0.97±0.01.
3 ΔP_v: the vegetation cover fraction is determined using NDVI. The accuracy obtainable for this parameter depends on the error of NDVI, and the type of soil and on atmospheric conditions. Leprieur *et al.* (1996) show that the signal-to-noise ratio can vary from 2 to 50. For the purpose of emissivity error estimation, we will vary $\Delta P_v/P_v$ from 0% to 20%.

4 $\Delta(d\varepsilon)$: $(d\varepsilon)$ depends on the vegetation structure. This term is not measurable from space and only knowledge of the local vegetation structure allows its determination. After Valor *et al.* (1996), this term ranges from 0–0.025. We will assume here $(d\varepsilon)=0.01\pm0.005$.

Figure 2.1 (a) shows emissivities in channels 4 and 5 of AVHRR as a function of the vegetation cover fraction. Emissivities are estimated using equation (2.34) and the configuration described above (i.e $.\varepsilon_s^4$ =0.95 and ε_s^5=0.97, ε_v=0.985, and $(d\varepsilon)$=0.01).

The relative error on channel 4 emissivity (respectively, channel 5) is presented as a function of the vegetation cover fraction in Figure 2.1(b) (respectively (2.1(c)). $\Delta P_v/P_v$ takes values of 0%, 10%, and 20%.

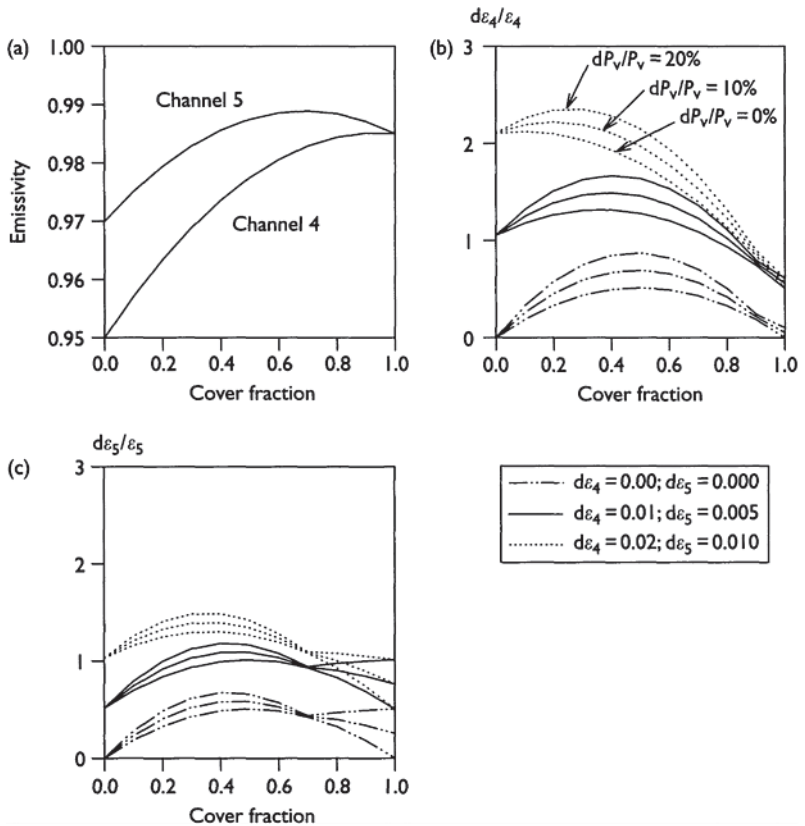


Figure 2.1 Emissivity in channels 4 and 5 and relative errors associated with the vegetation cover method.

Three error configurations are considered:

1 $d\varepsilon_s^4$.= 0.00 and = 0.00: soil emissivity is known exactly; it is the ideal case;
2 $d\varepsilon_s^4$= 0.01 and $d\varepsilon_s^5$= 0.005: we have a rather good *a priori* knowledge of soil emissivity (e.g. when field measurements of emissivity are available); it is the most favorable case;
3 $d\varepsilon_s^4$= 0.02 and $d\varepsilon_s^5$= 0.010: no a priori knowledge is available; it is the general case.

Curves 2.1(b) and (c) show that the error on emissivity decreases with the vegetation cover fraction. The error is driven mostly by the uncertainty on the soil emissivity. For non-vegetated areas, errors will be a maximum, whereas for vegetated areas pixel errors will be a minimum. The error induced by uncertainty on the cover fraction is negligible compared to the error on soil emissivity. When the emissivities are unknown, errors in channel 4 emissivities range from 1% to 2% and in channel 5 it is around 1%.

2.3.4 Special cases

Until now we have only considered average landscapes apart from the obvious case of ocean surfaces. However, the data collected may correspond to a snow-covered area, especially in the high-latitude or high-elevation area. In this case, the emissivity will be significantly different from unity and must be taken into account. Unfortunately, snow and ice emissivity values found in the literature (Table 2.1) are not reliable as the emissivity of snow and ice varies significantly over time (fresh or old; wet or dry).

To illustrate the importance of taking into account snow/ice in a retrieval algorithm (if only to mask the area), Figure 2.2 shows the results obtained with a set of AVHRR temperatures in bands 4 and 5, using Ulivieri's algorithm with emissivity values corresponding to snow, ice (rough and smooth) from Table 2.1, and typical soil.

Figure 2.2 depicts the difference of behavior for the different algorithm outputs. T_u, T_{uir}, T_{uis}, T_{us} standing for Uliveri's algorithm output for values of

Table 2.1 Emissivity values retained (from Salisbury and D'Aria 94)

Surface type	ε_4	ε_5
Ice		
Smooth	0.977	0.973
Rough	0.984	0.970
Snow	0.997	0.996

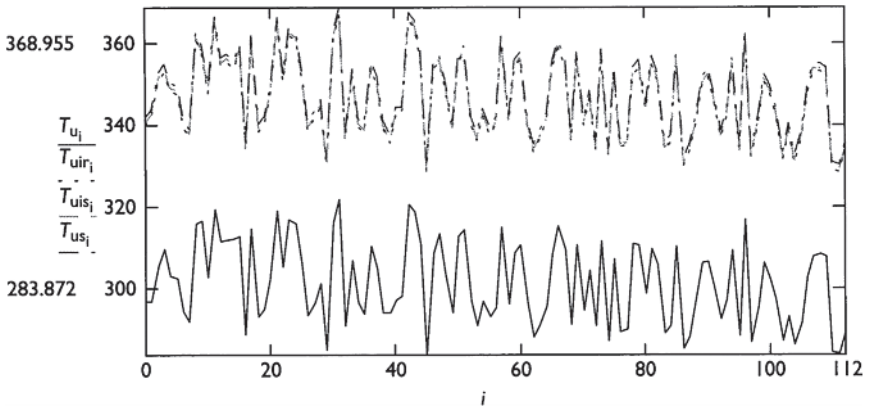


Figure 2.2 Different results obtained with emissivities of bare soil (red), or emissivities given in Table 2.1 (dotted lines). The x-axis represents the ground data number (arbitrary), the y-axis is the temperature in K (see Colour Plate 1).

emissivity corresponding, respectively, to standard soil, rough ice, smooth ice, and snow. A difference of 45–47 K is encountered. Such large errors justify taking great care in identifying snow/ice covered areas in global surface temperature retrieval schemes.

2.3.5 Conclusion

Various methods exist to retrieve emissivity from satellite data. Their usefulness depends on the *a priori* knowledge of surface spectral emissivities in the spectral channels of the sensor, as well as the nature of the data available (temporal composites, day night couples, atmospheric corrections, etc).

The vegetation cover method, due to its simple formulation, is easily applicable and seems rather robust. The drawback of this simplicity is that the method relies on an a priori knowledge of emissivity. Thus, good a priori knowledge is necessary to obtain a correct accuracy. Actually, the error budget analysis shows that a relative uncertainty in the soil emissivity of $n\%$ will induce, in the case of low vegetated area, an error of $n\%$ on the emissivity estimation of the pixel.

If a possibility exists in the near future to access a global data set of emissivity values, representative of the pixel (cf. ASTER or MODIS LST projects), the vegetation cover method should allow a viable estimation of the emissivity. Otherwise, the method is inefficient and taking an average value for the emissivity could prove just as reliable, provided snow- and ice-covered regions can be identified and proper emissivity values estimated.

2.4 Water vapor retrieval

2.4.1 Introduction: existing sources

In the 10–12 μm spectral region, water vapor is the most important atmospheric variable influencing atmospheric corrections. Since it is explicit in some algorithms, it is necessary to review methods for its estimation and their corresponding accuracy.

The methods may be separated in two classes: those based on satellite measurements and those relying on external data (forecast models or meteorological networks). Only methods that may be applied over land are presented here.

2.4.2 Existing methods, applicability, and errors

Satellite estimates

The only method that can be applied to data of AVHRR channels 4 and 5 and does not require ancillary data is called the differential absorption technique (introduced in the section on "The differential absorption method: background").

As explained above, the brightness temperature difference measured in channels 4 and 5 is related to total atmospheric water vapor amount (W_v). Different authors have tried to directly correlate this difference to W_v, fitting radiosounding water vapor amounts to satellite data, but obtained very poor correlation. The reason is that the radiance difference is also dependent on surface emissivity and thus any relationship found is surface dependent, and thus *local* (as shown by Choudhury *et al.* 1995).

Other methods to retrieve W_v have been proposed, which are based on statistical analysis of the images. If we consider the case of a cloudless region where the atmosphere is homogeneous, when the radiative transfer equation is written for two neighboring pixels whose surface temperatures are different, Kleespies and McMillin (1990) have shown that the ratio of the brightness temperature difference at the two wavelengths $R_{11,12}$ is equal to the ratio of the transmittances balanced by the respective surface emissivities in the two channels, that is:

$$R_{11,12} = \frac{(\Delta T_{ij})_{12}}{(\Delta T_{ij})_{11}} = \frac{(T_i - T_j)_{12}}{(T_i - T_j)_{11}} = \frac{\varepsilon_{12}\tau_{12}}{\varepsilon_{11}\tau_{11}} \tag{2.45}$$

The subscripts 11 and 12 refer to the split window channels, and i and j refer to two neighboring pixels for which the surface temperature changes measurably and have the same surface emissivity. Over land, these conditions are rarely satisfied and the varying surface emissivities must be accounted

for. Sobrino *et al.* (1993) obtained the following equation for this evaluation:

$$R_{12,11}\,(\Delta T_{ij})_{11} = (\Delta T_{ij})_{12} + \tau_{12}\left(\Delta\varepsilon_{11}\frac{\varepsilon_{12}}{\varepsilon_{11}}F_{11} - \Delta\varepsilon_{12}F_{12}\right) \tag{2.46}$$

with

$$R_{12,11} = \frac{\tau_{12}\varepsilon_{12}}{\tau_{11}\varepsilon_{11}} \tag{2.47}$$

and

$$F_i = \frac{B_\lambda\,(T_s) - R_{\mathrm{atm}_i}\downarrow}{\partial B_\lambda\,(T_s)\,/\partial T} \tag{2.48}$$

where $B_\lambda(T_s)$ is the Planck function,$\delta B_\lambda\,(T_s)/\,\delta T$ its derivative, and $R_{\mathrm{atm}_i}\downarrow$ is the downwelling atmospheric radiation.

In the case where the surface emissivity differences between pixels may be neglected, equation (2.45) is recovered. Sobrino *et al.* (1993) discussed the order of magnitude of the $\Delta\varepsilon$ term in the case of adjacent surfaces with different emissivities and the error introduced when the difference is neglected. Their results show that the mean error increases with the difference $\Delta\varepsilon$ and that for extreme cases (sand and sea over adjacent pixels), it varies between 3% and 13% (depending on the surface temperature difference), and being maximum for low-temperature differences between pixels. Then, the mean value of $R_{12,11}$ may be calculated over a zone where the atmosphere is assumed to be homogeneous. By least-squares analysis, the ratio may be written as

$$R_{12,11} = \frac{\sum_{k=1}^{N}\left(T_{11k} - T_{11j}\right)\left(T_{12k} - T_{12j}\right)}{\sum_{k=1}^{N}\left(T_{12k} - T_{12j}\right)^2} \tag{2.49}$$

where the numerator and the denominator represent, respectively, the covariance and variance of the brightness temperatures measured in the split window channels and where T_{11j} and T_{12j} are two reference temperatures that can be the mean temperature of the pixels considered.

As shown by Jedlovec (1990), Kleespies and McMillin (1990), Harris and Mason (1992), and Sobrino *et al.* (1993), $R_{12,11}$ (which is the ratio of the transmittances in these two channels) is related to the water vapor (W_v) in the atmospheric column by an inverse relationship, provided the total transmittance due to the other atmospheric gases is assumed constant:

$$R_{12,11} = K_1 W_v + K_2 \tag{2.50}$$

where K_1 and K_2 are constants related to the absorption coefficients of the atmospheric constituents. These authors have assessed this relationship through regression between rawinsonde measurements and computed transmittances ratios using radiative transfer models. Their results, however, do not show a good agreement with rawisonde values, with differences on the estimated water vapor greater than 50%.

Ottlé *et al.* (1998) also investigated this methodology, taking advantage of two different AVHRR and ATSR databases containing coincident water vapor measurements. They show that this method may be applied over land with some precautions, in particular, the application of an accurate cloud mask to eliminate all the cloud-contaminated pixels. In this case, the accuracy obtained is less than 0.4g/cm^{-2}. Since this technique is based on the calculation of radiance variances around the pixel under study, it cannot be applied on composite images.

Atmospheric water vapor may be also estimated from atmospheric sounders like TOVS (HIRS-MSU-SSU) on NOAA satellites but because of the poor accuracy of these retrievals near the surface and the poor spatial resolution, the water profiles derived with these instruments are not accurate enough and cannot be used to correct AVHRR channels (Ottlé and Stoll 1993).

In situ estimations

The atmospheric water vapor content is measured in the stations of the worldwide meteorological network by radiosoundings, two times per day at 0 and 12UT. The accuracy generally attributed to relative humidity as given by the radiosoundings is±2%. But in specific conditions, close to the saturation (RH>90%), or for very dry atmospheres (RH<20%), this uncertainty may locally reach±5% to±10% (Westwater *et al.* 1990). Such measurements are local and, generally, are not simultaneous with the satellite pass. Consequently, they cannot be used to correct most satellite images.

Meteorological models can also provide users with global maps of all the atmospheric fields. The European Centre for Medium-range Weather Forecasts (ECMWF) have archived global analysis and model-predicted fields for about 20years at the grid scale of their model, which is about 1°×1° at the equator. The nearest (in time and space) atmospheric water vapor profile corresponding to the satellite measurement may be obtained from these data and used for the atmospheric correction. The accuracy of such analysis/model fields has been investigated by different authors. Phalippou (1996) noted that the typical standard deviation of the first-guess relative humidity profiles (comparison between forecasts and observations) is around 7%. Nerry *et al.* (1998), using AVISO data archive provided by MeteoFrance, found slightly larger errors of about 0.3 g cm^{-2}, when comparing total atmospheric water

content predicted to radiosounding measurements for seven Spanish meteorological stations and ten different days (personal communication). Since the analyses are available at 6-h intervals and at UT times, the time lag with the satellite measurement may introduce an error. Nevertheless, we conclude that meteorological models can provide the atmospheric water vapor content needed with an accuracy better than 10%.

2.4.3 Conclusions

In conclusion, in most cases, the only possibility to obtain the total atmospheric water vapor content necessary to apply the water vapor dependent Split Window algorithms is to use the global fields provided by forecasting models like those archived at ECMWF.

2.5 Intercomparison of algorithms' accuracy and efficiency

Based on material presented in Chapters 2, 3, and 4, an intercomparison of the 21 algorithms was performed. This was done using two approaches. The first one consisted in using RT code simulations with realistic and exhaustive atmospheric profiles and surface conditions. The second method was performed using ground and satellite actual data. This part was performed using AVHRR data, as we wanted to show a practical example.

2.5.1 Simulations

In order to compare and test the different Split Window algorithms published in the literature, a large database containing different atmospheric, surface and viewing conditions associated with Brightness Temperatures (BT) at the top of the atmosphere has been compiled. Here, only 20 algorithms were compared as the 21st (Sobrino *et al.* 1994) is probably too complicated to implement for this study.

Since it is not possible to find coincident LST and BT measurements for a large range of situations, this database has been built by simulations of the radiative transfer of the emitted surface and atmospheric radiances along the path (surface-satellite) using a large set of radiosonde profiles: the TOVS Initial Guess Retrieval (TIGR) database (Chédin *et al.* 1985). This data set was compiled by the Laboratoire de Météorologie Dynamique (LMD) and represents a worldwide set of atmospheric situations (1,761 radiosoundings) from polar to tropical atmospheres, with varying water vapor amounts ranging from 0.1 to 8 g cm^{-2}.

The TIGR database was used together with the MODTRAN3 radiative transfer model to simulate the BT measured in the spectral channels 4 and 5 of the AVHRR radiometer on board NOAA-11 satellite for different conditions. The varying conditions are:

1 different values of LST ranging around the air temperature measured at the first level of the radiosounding, between $T_{low\text{-}layer}$–5 K and $T_{low\text{-}layer}$+10 K in steps of 5 K;
2 different values of AVHRR viewing angles: 0°, 20°, 40°, and 55°;
3 different values of the spectral surface emissivity: between 0.95 and 1 for channel 4. As a first step, a spectral difference of 1 % between channels 4 and 5 was implemented.

Comparison of the different available algorithms

The 20 algorithms selected have been compared using this database, called TIGRMOD henceforth. The statistical results (biases, standard deviations, correlation coefficients, linear regression slopes) on the whole database are summarized in Table 2.2 and illustrated on Figures 2.3 and 2.4 They show that three algorithms give better results than the others: they are Ulivieri *et al.* 1992 (U94), Becker and Li 1995 (GRTR), and Sobrino *et al.* 1993 (So93), and give the lowest standard deviations together with the lowest biases. They all account for spectral emissivities and only one of them, the GRTR, accounts for water vapor, as well as viewing angle. (For color schemes for Figures 2.3, 2.4, 2.6–2.12, see Appendix.)

A detailed analysis of the results by class of emissivity and class of viewing angle yields the following observations:

1 The performance of all the algorithms degrades with increasing water vapor amount.
2 The mean errors between LST retrievals and TIGR are generally greater at high viewing angles than at the nadir, especially for high water vapor amounts associated with increasing scatter.
3 A maximum error is observed between 0 and 1 g cm^{-2} for U94 and So93 is probably related to the method for deriving the empirical coefficients (the influence of minor gases like CO_2, where contributions is not negligible for very low water vapor amounts, may not have been accounted for in the radiative transfer simulations).
4 The performances of U94 degrades when the water vapor amount exceeds 4 g cm^{-2}, leading to errors sometimes greater than 5 K.
5 So93 does not have a stable behavior and can give large errors greater than 5 K in some atmospheric situations, especially for low surface emissivities.

6 In contrast, GRTR seems to have a very stable behavior, less accurate than U94 at high viewing angles and high water content with maximum errors never exceeding 5K whatever the value of the surface emissivity. It also provides the best overall agreement with the TIGR data set for the whole range of water vapor amounts and emissivity values.

The sensitivity of these algorithms to atmospheric water vapor content and to surface emissivity has been investigated for estimating the order of magnitude of the errors that may result from poor knowledge of these two variables, when they are needed for the LST retrieval.

Table 2.2 LST comparisons (TIGR/retrievals)–statistics

No	*algo param.*		*Mean (K), LST − TS*	*Std. dev. (K), LST − TS*	*Correl. coeff.*	*Reg. slope, LST = aTS + b*
		GRTR				
1	W_v, θ, ε	Becker and Li (1995)	−0.149	0.637	0.9997	1.0270
2	ε	Becker and Li (1990b)	−1.522	0.941	0.9993	1.0358
3	ε	Prata and Platt (1991)	1.189	1.618	0.9993	1.0814
4	ε	Price (1984)	2.360	1.550	0.9981	1.0592
5	ε	Ulivieri and Cannizzaro (1985)	1.730	1.259	0.9987	1.0467
6	ε	Ulivieri *et al.* (1992)	0.305	0.520	0.9996	1.0023
7	ε	Sobrino *et al.* (1993)	−0.172	0.720	0.9992	1.0041
8	W_v, ε	Sobrino *et al.* (1991)	1.940	1.288	0.9978	1.0250
9	θ	NESDIS May *et al.* (1992)	2.385	1.468	0.9983	1.0553
10	ε	Becker and Li, acc. to Sobrino (1990)	−0.591	0.936	0.9989	1.0203
11	ε	Prata and Platt, acc. to Caselles	−5.786	4.099	0.9778	1.0532
12	ε	Prata and Platt, acc. to Sobrino	1.269	1.108	0.9990	1.0408
13	ε	Price, acc. to Sobrino	1.654	1.123	0.9989	1.0413
14	ε	Ulivieri *et al.*, acc. to Sobrino	0.255	1.074	0.9990	1.0378
15	ε	Kerr *et al.* (1992)	1.862	1.015	0.9987	1.0227
16	ε	Vidal (1991)	−2.191	1.089	0.9989	1.0386
17	ε	Coll *et al.* (1997)	−1.450	0.638	0.9995	1.0145
18	(W_v, θ, ε)	Ottle and Vidal-Madjar (1992)	0.081	0.883	0.9988	1.0041
19	W_v, (θ, ε)	François and Ottle, WVD (1996)	0.582	1.052	0.9984	1.0145
20	(e, θ, ε)	François and Ottle, QUAD (1996)	0.626	1.745	0.9960	1.0349

Notes
Data points: 140,880.
LST: algorithm retrieval.
TS: LST from TIGR.
1,761 atmospheres (TIGR).
$T_{low\text{-}layer}$-5°<TS<$T_{low\text{-}layer}$+10°.
0°<θ<55°.
0.955<e<0.995 [e=(e_4+e_5)/2].

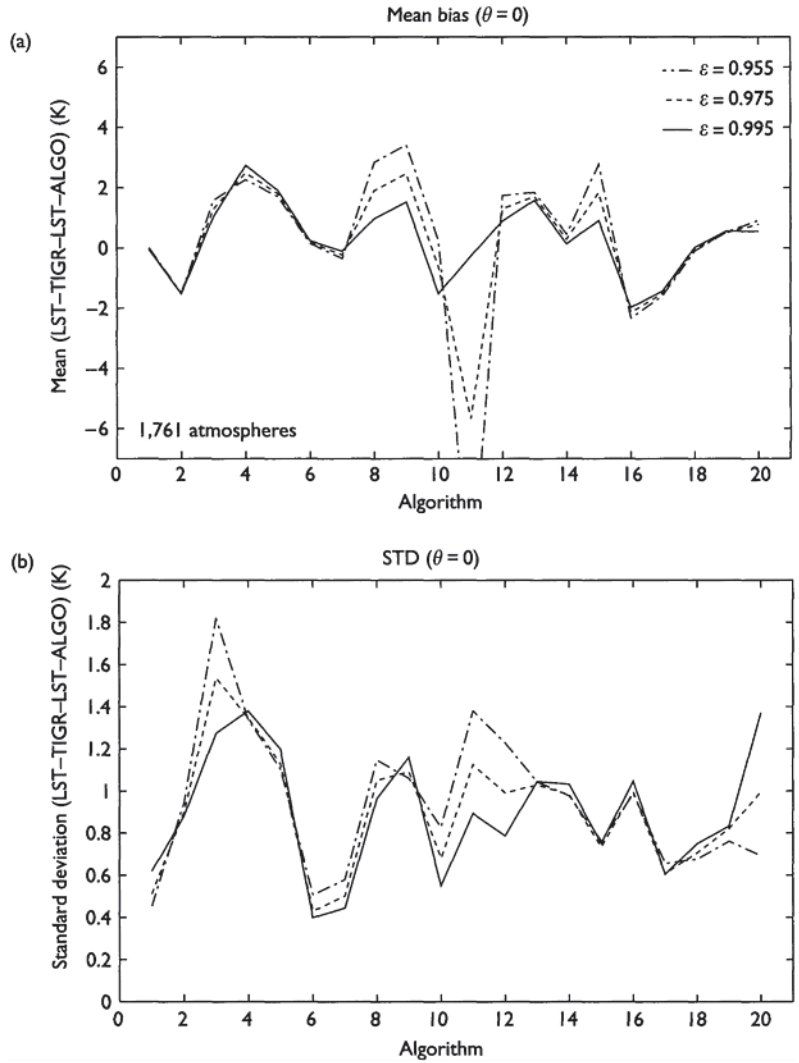


Figure 2.3 LST retrievals comparison with TIGR global statistics (see Colour Plate II).

Sensitivity analysis to surface emissivity

We investigated first the sensitivity of the different models to surface emissivity, which is the most influential parameter. For this purpose, we have introduced in our simulations a linear dependence of the emissivity spectral difference between channels 4 and 5, with channel 4 emissivity. According to analyses

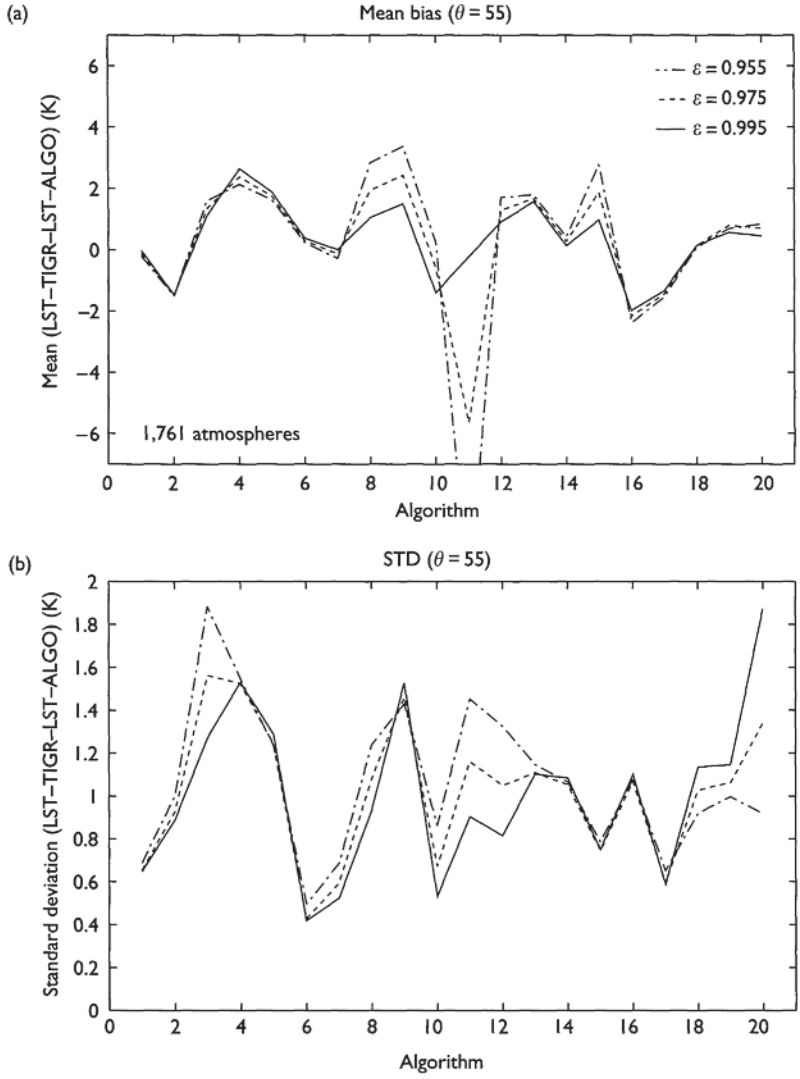


Figure 2.4 LST retrievals comparison with TIGR global statistics (see Colour Plate III).

of Salisbury's surface emissivity databases, we assumed a difference ranging from 3% (for ε_4=0.95) to 0 when ε_4=0.99 (Figure 2.5), as given by the following equation:

$$\varepsilon_5 = 0.25(\varepsilon_4 - 0.95) + 0.98 \tag{2.51}$$

Lower values of emissivities have not been used because too few few observations were available (three samples) and no relation appeared between ε_4 and ε_5.

The uncertainty on the estimation of surface emissivity has been first evaluated. The GRTR experience allowed us to assume that the maximum error on the emissivity is of the order of±2.5% for bare soils (which means for emissivities lower than 0.97).

According to the spectral dependence applied to the surface emissivities in channels 4 and 5, this maximum error finally ranges from±2.5% (for 0.95= ε=0.97) to±1.7% for ε=0.99, with the condition that the emissivity never exceeds unity. This error may be understood as the uncertainty on the surface emissivity (input value in the algorithm) but also as the variability of this parameter inside the pixel.

The 20 algorithms have been tested and compared in terms of absolute errors on the LST retrievals resulting from the given uncertainties on surface emissivity. As expected, the error due to a poor knowledge of surface emissivity is large, around 2 K, for dry atmospheres but does not vary according to the humidity of the atmosphere in the same way for all the models. The error is generally larger for the algorithms that account for the spectral value of the emissivity ($\varepsilon_4 \neq \varepsilon_5$). For these simulations, we took a constant value for the vegetation fractional cover (equal to 0.5) but, ideally, a variation of the vegetation cover amount should have been performed.

The results indicate that the GRTR and WVD model behave as expected theoretically, with an error decreasing when atmospheric water vapor content increases. It can be noted that WVD model gives lower errors than GRTR. The other algorithms show an opposite behavior for large water vapor amounts, which has not yet been explained.

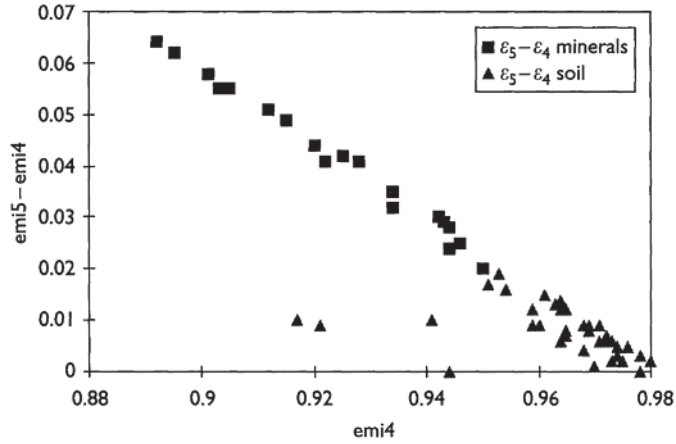


Figure 2.5 Emissivity difference between AVHRR channels 5 and 4 against channel 4 emissivity calculated from Salisbury's emissivity database.

If we compare GRTR, So93, and U94, we can note that GRTR is more sensitive to an error of the surface emissivity, that So93 has a large dispersion, and that U94 gives the lowest errors for water vapor amounts lower than 4 g cm^{-2} which is the range of variation of most atmospheres.

Sensitivity analysis to water vapor content

Finally, we analyzed the sensitivity of the algorithms to this parameter. The sensitivity to water vapor amount is the sum of two terms: the error on the input value of this variable for water vapor dependent algorithms (GRTR, WVD, and Sobrino *et al.* 1991 noted So91), plus the error due to the variability of this parameter inside the pixel. According to the previous analysis (part 4), this error is at least equal to±5%.

Consequently, we calculated the absolute error on LST retrieval resulting from a±5% error on the total water vapor amount W_v. This was achieved by building a second database of simulated MODTRAN's brightness temperatures associated with LSTs.

As expected, the absolute error on LST retrieval related to water vapor uncertainties is lower than the one linked to surface emissivity errors. The errors are about twice as low on average. The best results are obtained for the WVD algorithm, with errors lower than 0.5 K for water vapor lower than 5g cm^{-2}, but up to 1.5 K for very high water vapor amount, nadir viewing, and low surface emissivity. This algorithm has a high standard deviation of error when used on the whole TIGR database. Also, the WVD algorithm has been fitted for the ATSR radiometer, which may explain part of the deviation.

For the GRTR algorithm, the errors are always lower than 1.1 K whatever the emissivity. The last water vapor dependent algorithm (Sobrino *et al.* 1991) also shows low errors, smaller than 0.5 K, but with singular values around 4 g cm^{-2}(up to 3 K).

Consequently, these water vapor dependent algorithms are sensitive to the accuracy of the input value of W_v, and, in particular, the GRTR algorithm.

Conclusions

The total absolute error on LST, due to both water vapor amount and surface emissivity errors (5% on W_v; varying on ε, depending on the emissivity value, as explained in the section on "Sensitivity analysis to surface emissivity"), has been estimated for the 20 algorithms. This error is defined by the following equation:

$$\Delta T_s = \left|\frac{\delta T_s}{\delta \varepsilon}\right| \Delta\varepsilon + \left|\frac{\delta T_s}{\delta w}\right| \Delta w \qquad (2.52)$$

The results are presented in Figures 2.6–2.9. These figures show results for each algorithm referred by a number (see Table 2.2):

1 The mean absolute error on LST by class of water vapor amount in 1 g cm-2 steps for a surface emissivity (average of channels 4 and 5) of 0.9650 and 0.9837, and for two selected viewing angles: 0° and 55°.
2 The mean standard deviations associated with these errors.

These results allow us to draw a conclusion and identify the most suitable algorithm for retrieving surface temperature. They show that the maximum errors for U94 and So93 are obtained for high water vapor content and low emissivities and are around 2 K (at the nadir) and 2.5 K (at 55°). They are a little lower for GRTR. For dry atmospheres (W_v lower than 1 g cm^{-2}) the error is larger for GRTR (about 1.5 K in average) and decrease with W_v becoming lower than 0.5 K for W_v greater than 3 g cm^{-2}. On the contrary, U94 and So93 show lower errors for W_v lower than 3 g cm^{-2}(between 0.5 and 1.2 K for U94 and between 0.5 and 1.5 K for So93) and larger errors for high values of W_v (even with a larger scatter for So93) as mentioned earlier.

In conclusion, the GRTR algorithm seems to give the best performances but it is very sensitive to surface emissivity and to water vapor. Since these parameters are known with uncertainties that are minimally of the order of the values assumed in this study, algorithms independent of surface emissivity and water vapor will perform better (smaller errors) than GRTR when these parameters are not well known (usual case). Consequently, if the accuracy on ε and W_v cannot be improved, it is better to use U94 or So93. Considering the unstable behavior of So93, and the fact that it is more sensitive to radiometric noise because of its quadratic term, we have a preference for U94. The good performances of simple algorithms (Kerr *et al.* 1992) can also be noted.

2.5.2 Intercomparison with actual data

In order to perform the intercomparison, we relied on existing data sets. As far as ground data are concerned we have used all the Australian data sets, and the Hapex-Sahel data set. We have collected a few others, but in most cases they lack some of the inputs necessary to test all the algorithms. Even the Prata data set was not complete for our specific purpose.

The quality of the results is much poorer that that given in Section 2.5.1. This is, however, not a surprise as we are now dealing with actual data. The pixels are not pure and the ground data representativity may be questioned. This led us to scrutinize a bit more the ground data and we found some noticeable caveats. First of all, as seen in Figure 2.10, some of the results are systematically off whatever the algorithm, indicating a problem more linked with the inputs than the method. It could be attributed

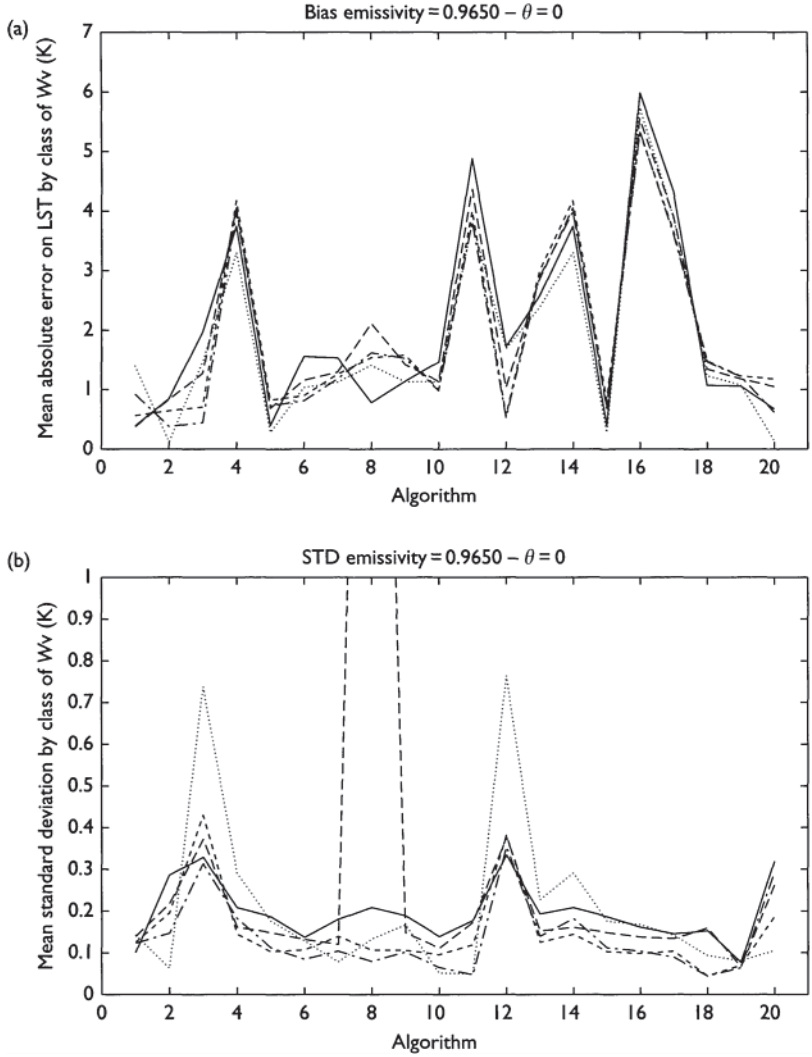


Figure 2.6 For the 20 algorithms, for θ=0 and an emissivity of 0.9650 absolute errors on LST retrievals (mean values per water vapor category) resulting from an error on water vapor and emissivity (see text) (see Colour Plate IV).

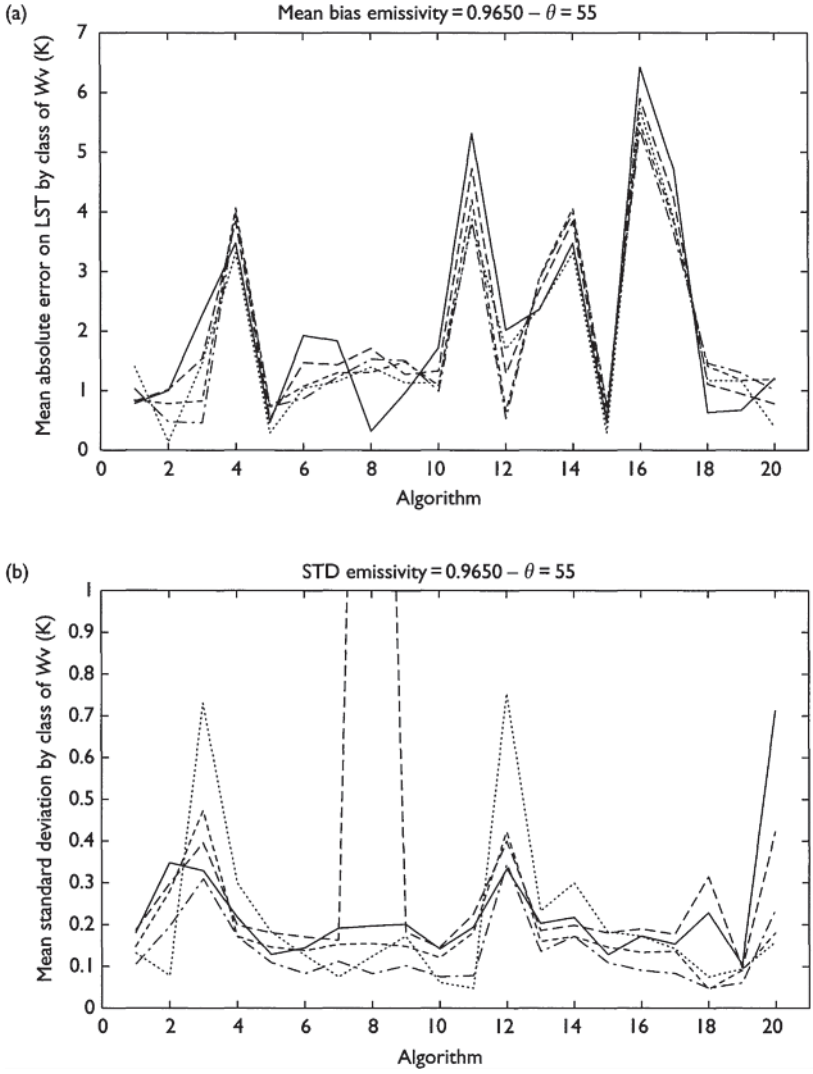


Figure 2.7 For the 20 algorithms, for θ=55 and an emissivity of 0.9650 absolute errors on LST retrievals (mean values per water vapor category) resulting from an error on water vapor and emissivity (see text) (see Colour Plate V).

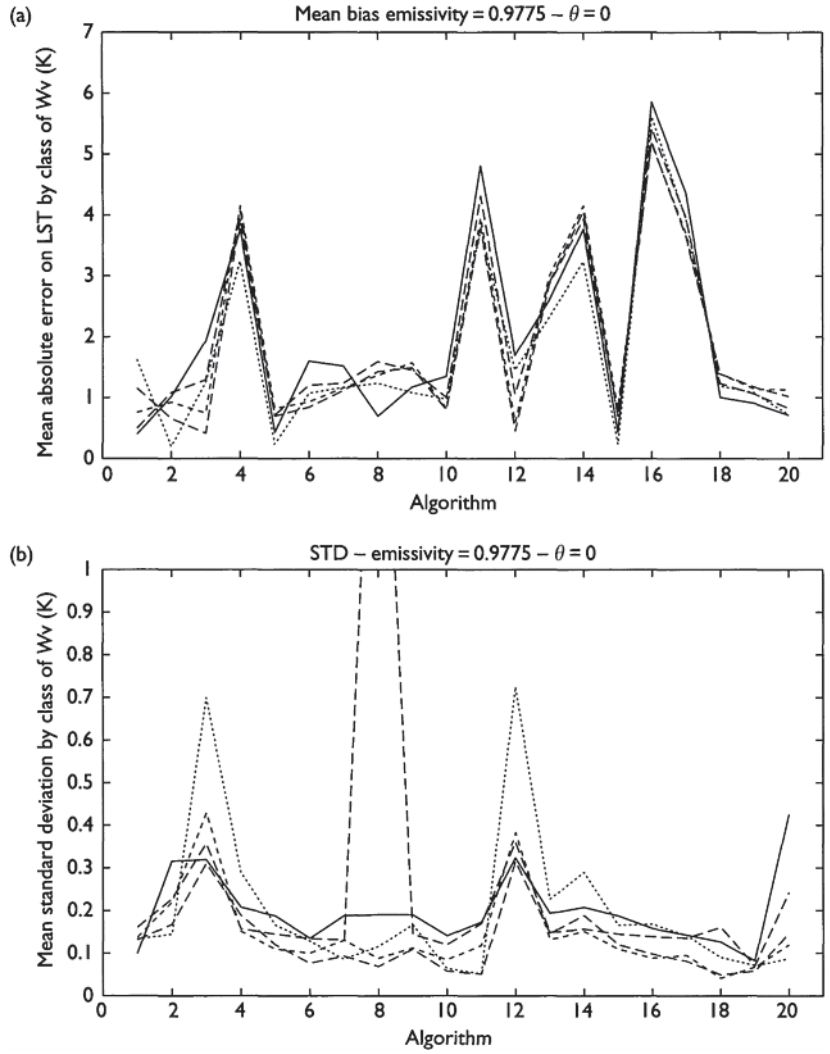


Figure 2.8 For the 20 algorithms, for θ=0 and an emissivity of 0.9775 absolute errors on LST retrievals (mean values per water vapor category) resulting from an error on water vapor and emissivity (see text) (see Colour Plate VI).

to either representativity of the measurements, a flaw in the measurements or in the AVHRR data. The latter point was sorted out when validating the algorithm. We found then that most of the errors were due to time differences between ground and satellite data acquisitions. When editing the data set by rejecting such points, the results improved slightly.

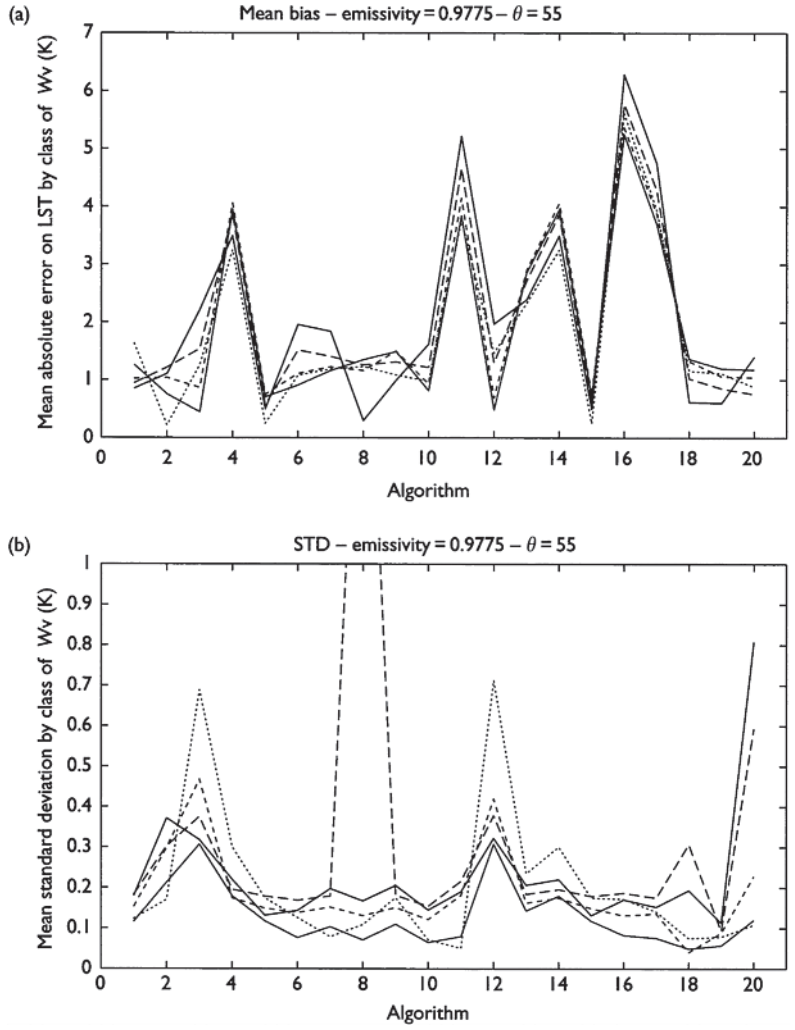


Figure 2.9 For the 20 algorithms, for θ=55 and an emissivity of 0.9775 absolute errors on LST retrievals (mean values per water vapor category) resulting from an error on water vapor and emissivity (see text) (see Colour Plate VII).

The two previous analyses were performed using all the data available. If we limit ourselves to data taken by the noon overpass we do not notice drastic changes. This seems to show the robustness of the algorithms.

One of the big problems with such intercomparison is the actual time of overpass, as with the high view angles, the time could be different.

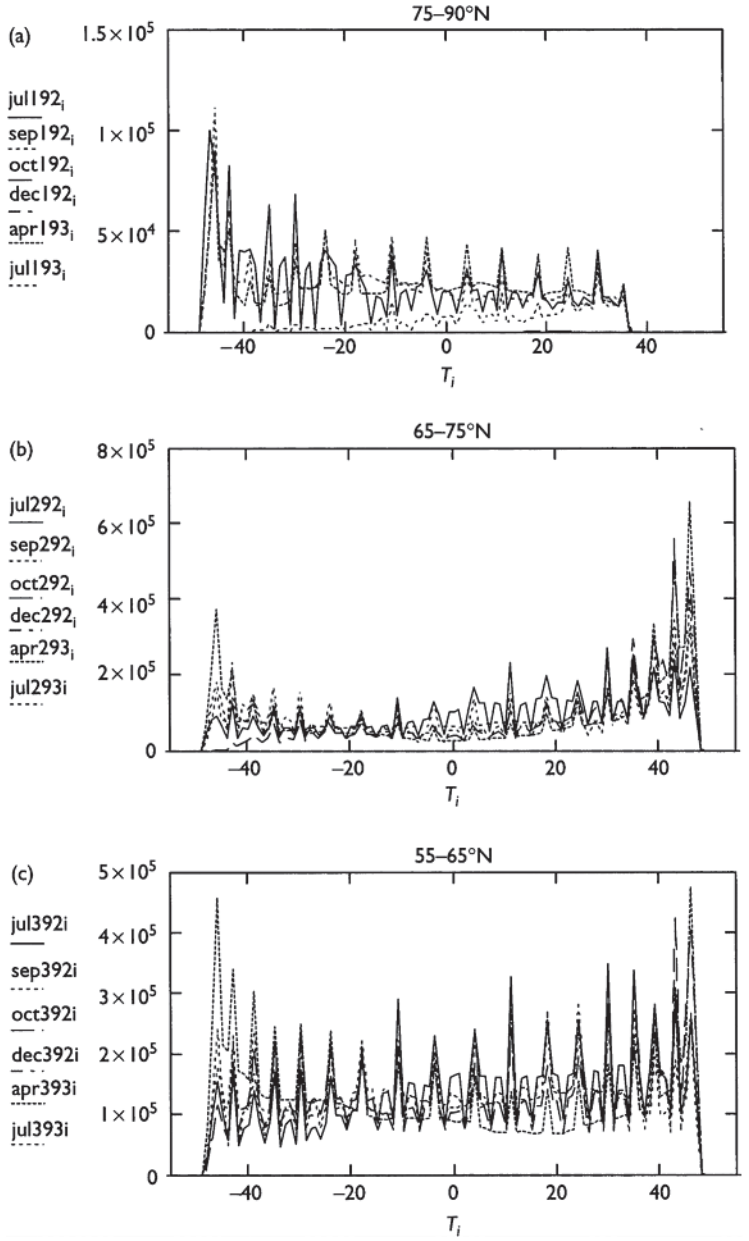


Figure 2.10 Histograms (10-day composites) of view angles for different decades: (a) latitudinal band 75–90ºN; (b) 65–75ºN; (c) 55–65ºN (see Colour Plate VIIIa-c).

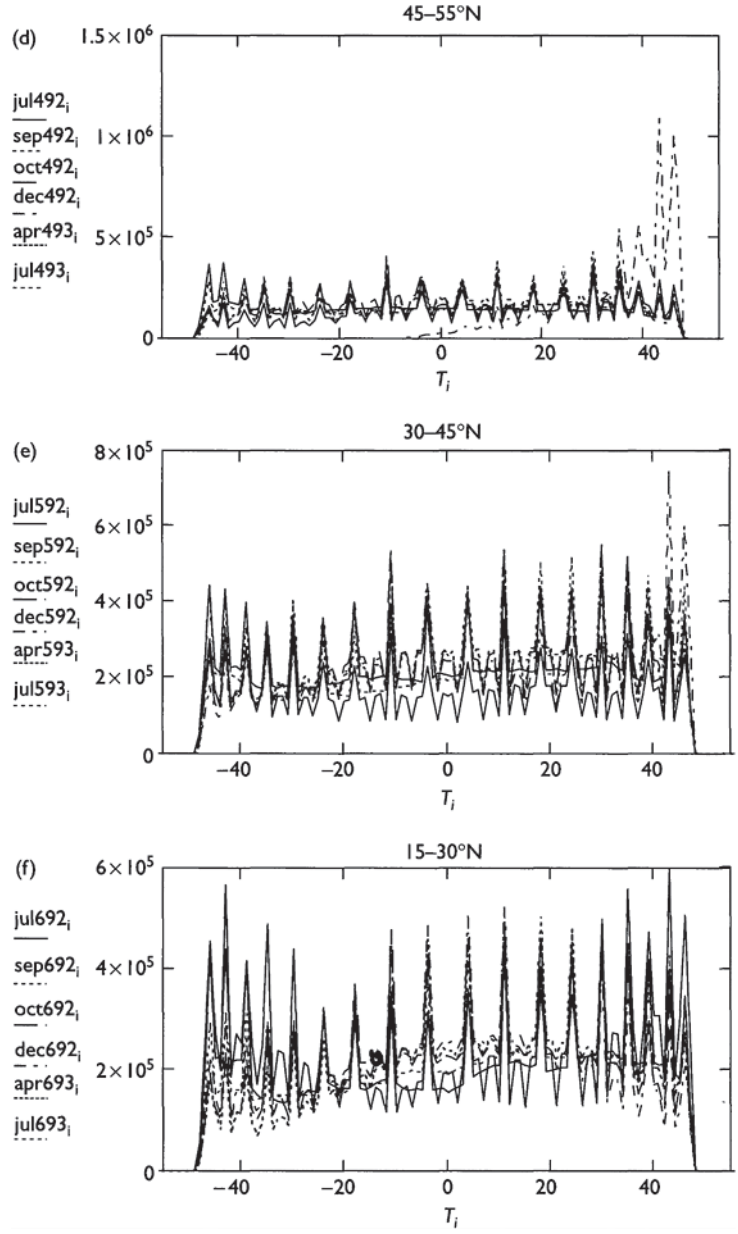


Figure 2.10 (d) Latitudinal band 45–55ºN; (e) 3–45ºN; (f) 15–30ºN (see Colour Plate VIIId-f).

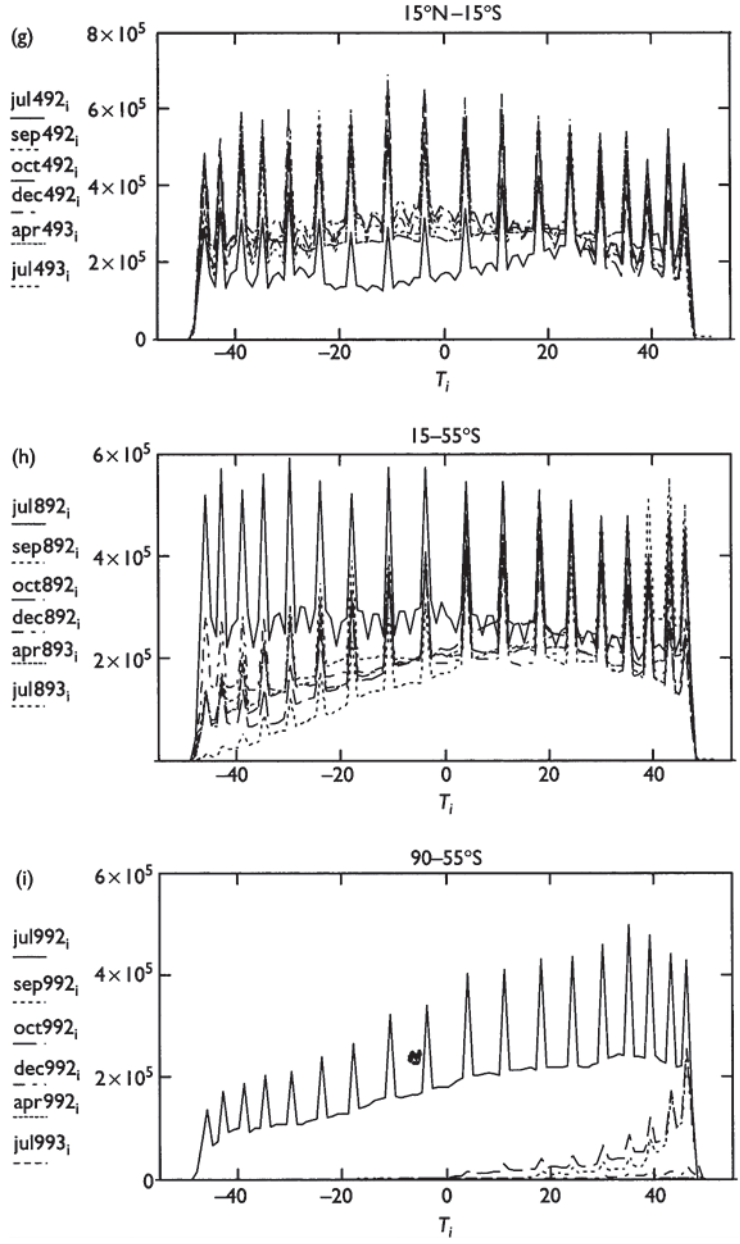


Figure 2.10 (g) Latitudinal band 15ºS-15ºN; (h) 55-l5ºS; (i) 90–55ºS (see Colour Plate VIIIg-i).

Table 2.3 Results obtained with the whole data set and the 21 tested algorithms

Algorithm, All data (371)	*Mean* (K)	STD (K)	*Max. positive error* (K)	*Max. negative error* (K)
Becker and Li (1995)*	0.136	2.327	14.657	−8.428
Becker and Li (1990a,b)	−1.034	2.56	11.54	−9.841
Prata and Platt (1991)*	0.06	2.495	12.828	−8.521
Price (1984)	−0.984	3.25	12.411	−12.892
Ulivieri and Cannizzaro (1985)	0.602	2.687	13.704	−9.329
Ulivieri *et al.* (1992)	0.698	2.127	15.58	−7.078
Sobrino *et al.* (1993)	0.333	2.333	14.119	−9.702
Sobrino *et al.* (1991)	1.603	3.00	23.651	−6.719
May *et al.* (1992)	1.008	2.726	16.753	−9.347
Becker and Li (Sobrino)*	−0.285	2.333	14.518	−8.221
Prata and Plat (Caselles)	−6.696	3.704	1.244	−30.574
Prata and Platt (Sobrino)*	0.775	2.289	13.892	−6.796
Price (Sobrino)*	0.007	2.304	14.413	−9.791
Ulivieri (Sobrino)*	0.321	2.526	13.764	−8.542
Kerr *et al.* (1992)	0.742	2.241	16.986	−6.462
Vidal *et al.* (1991)	0.243	2.979	10.613	−9.576
Coll *et al.* (1997)	0.431	2.427	13.215	−7.188
Ottlé and Vidal Madjar (1992)	−0.732	2.351	17.136	−9.262
François and Ottlé WVD (1996)	0.865	1.935	13.728	−7.295
François and Ottlé QUAD (1996)	0.057	3.025	19.085	−16.897
Sobrino *et al.* (1994)*	−0.208	2.645	12.536	−8.229

Globally all the algorithms give satisfactory results except maybe T_{quad}, T_{ppc}, and T_{sob2}. The results are summarised in Table 2.3.

These results require some comments. The quality of the retrievals is not necessarily commensurate with the complexity of the approach. However, some results have to be compared with care, as they are not necessarily representative:

1 Several of the algorithms were "tuned" using the Australian data set: thus, they should give better results in these tests than globally (they are marked with an *).
2 Some algorithms were not tested fully due to a lack of necessary data and/or due to their design. This applies especially to QUAD and WVD (ATSR), and Sobrino 91 (τ was not always available). For Kerr *et al.* (1992) we did not have the reflective channels, and thus the canopy fraction was kept constant for each sub-dataset, probably inducing errors.
3 There are problems with the time difference between acquisition and actual local solar time (not the case in Prata's data set).

2.5.3 Identified problems

Angular effects

Not including time of acquisition problem, two other types of angular effects can affect the measurements:

1 Angular effects linked to the atmosphere (see Section 2.1.1).
2 Angular effects linked to the surface. For vegetation, angular effects due to shadowing can be quite significant (Kimes and Kirchner 1983; Lagouarde *et al.* 1995). As for the AVHRR, between acquisitions, the viewing and solar angles can be significantly different; it may thus be necessary to consider this effect. Without any tractable analytical model for the angular effects available, several authors have used empirical or simplified models (see Prata and Platt 1991).

Angular effects/models are generally established using ground measurements. The validity of these ground-based models for heterogeneous surfaces and AVHRR-type pixels is as yet unclear. Most probably, due to averaging effects of the heterogeneous signals over the pixel, angular effects will be significantly reduced, making such corrections more or less necessary, depending on the surface type.

In order to show the amplitude of the angular variations; Figure 2.11 shows the range of view angles of AVHRR data. The data set used there is the 1 km global AVHRR data set produced by the Eros Data Center (Sioux Falls N Dakota) in the framework of the International Geosphere Biosphere Program (IGBP) initiative. The data consist of 10-day maximum value composites.

Figure 2.11 shows, for different latitudinal bands, the histogram of the view angle. Each curve corresponds to a different decade and month. The graphs are in the order of the latitudinal bands, that is 90–75°N; 75–65°N; 65–55°N; 55–45°N; 45–30°N; 30–15°N; 15°N–15°S; 15–55°S; 55–90°S.

Several viewing directions tend to be dominant, varying with latitude and season. For instance, for the first band (90–75°N) the view angles are mainly over –40° in July 92, while they are almost the opposite in September (+30°). This implies that the actual solar time on the imaged pixel varies by about 150min in relation to the almost 70° change in viewing angle. Another point worth mentioning is the regular spikes on many of the histograms. This implies that there are features most probably linked to the compositing scheme, done on a 10-day basis, close to the satellite repeat cycle. We finally note that there are systematic differences between July 1992 and 1993.

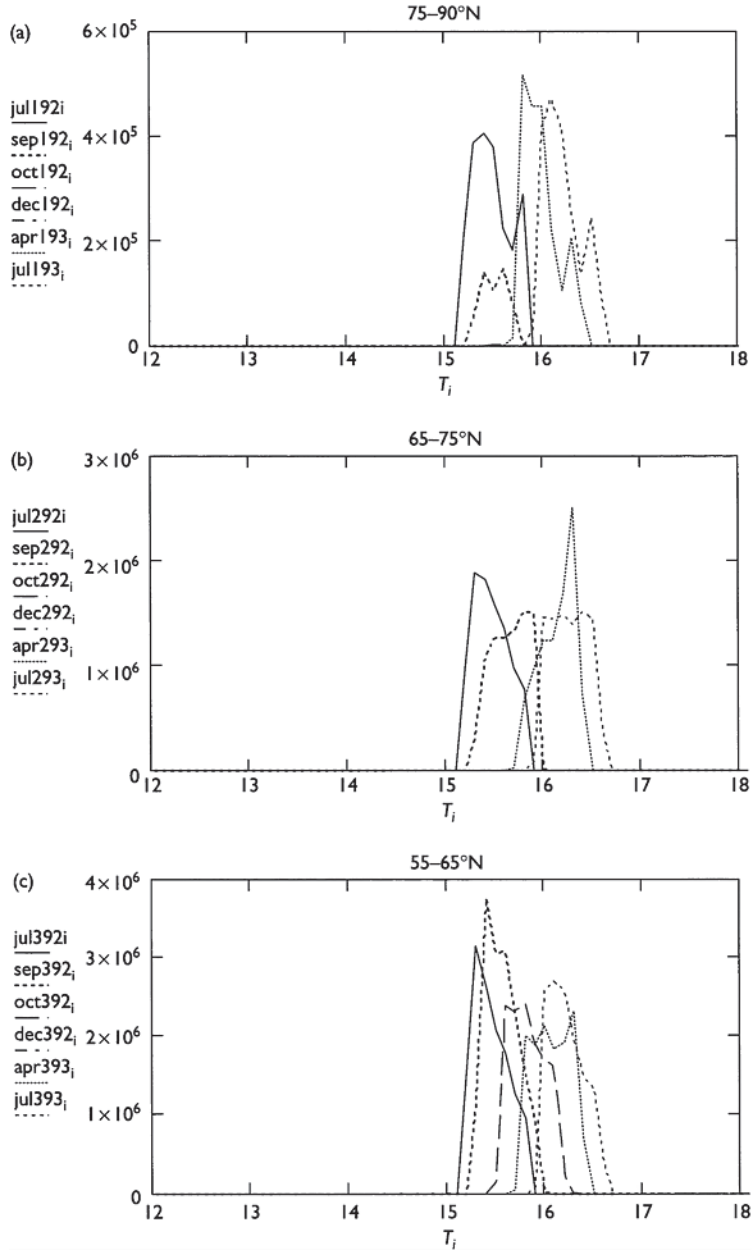


Figure 2.11 Histograms (10-day composites) of the local solar time for different decades. *x*-Axis is Local Solar Time: (a) latitudinal band 75–90ºN; (b) 65–75ºN; (c) 55–65ºN (see Colour Plate IXa-c).

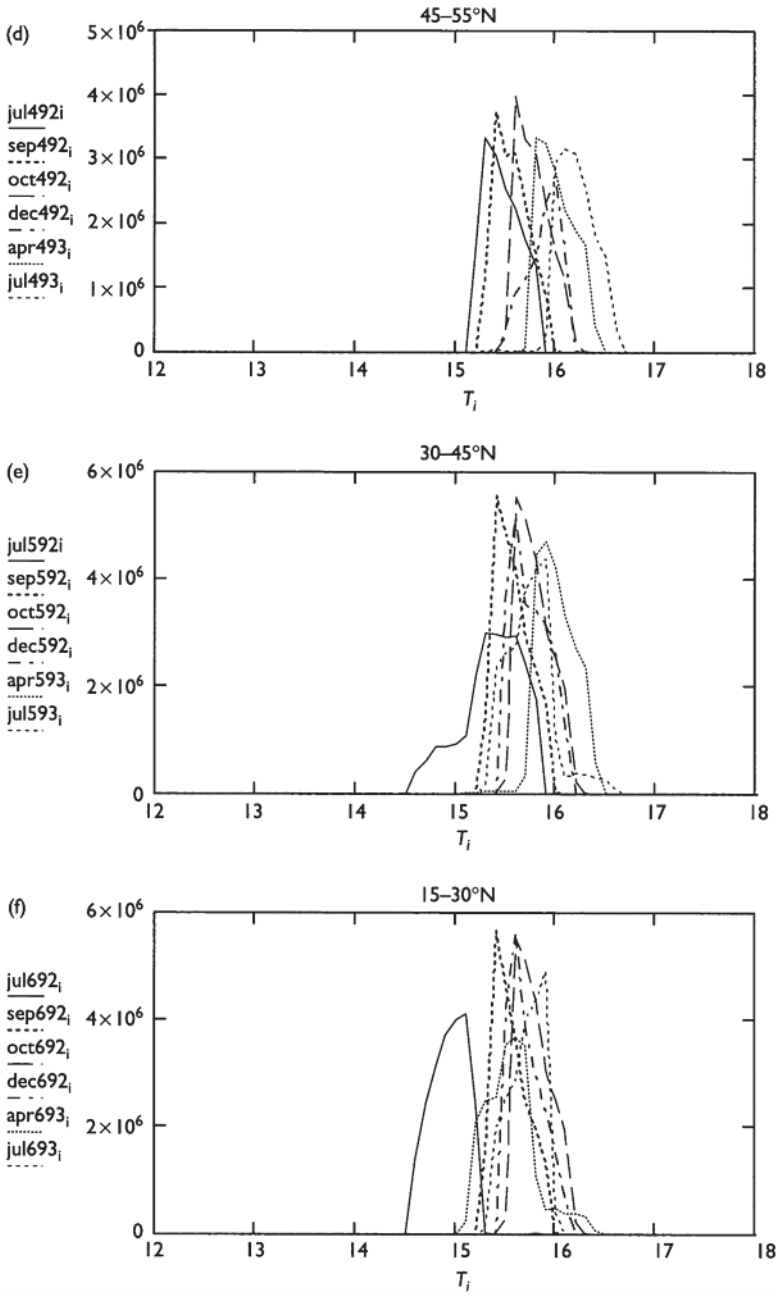


Figure 2.11 (d) Latitudinal band 45–55ºN; (e) 30–45ºN; (f) 15–30ºN (see Colour Plate IXd-f).

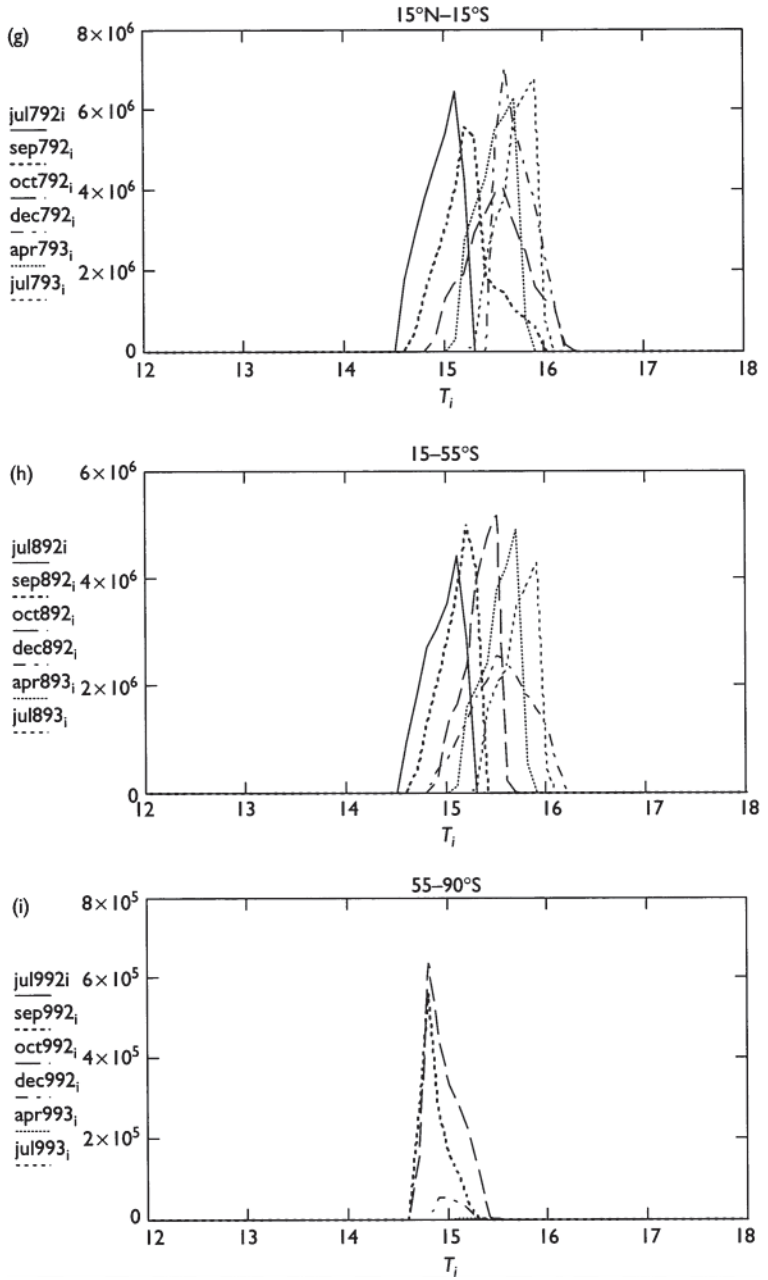


Figure 2.11 (g) Latitudinal band 15ºS-15ºN; (h) 55–15ºS; (i) 90–55ºS (see Colour Plate IXg-i).

To depict the local solar time effect, Figure 2.12 shows for the same data set the histogram of the local solar time of acquisition in identical format to Figure 2.11. The most striking feature of this set of figures is the evidence of the drift of the orbit, which amounts to about 1 h over 1 year. Other factors might contribute to this apparent effect and exaggerate it. The actual time of acquisition varies according to the orbit and view angle selected by the compositing algorithm. In the worst case, one can acquire pixels with a local solar time difference of up to 100 min.

Fractional vegetation cover

Vegetation cover is usually defined as the proportion of surface completely covered by vegetation. In the infrared, the important considerations are first the proportion of bare soil seen by the sensor and, within this, the proportion of shaded and sunlit surfaces. There are consequently angular effects that are also linked to the vegetation height and structure and depend on the

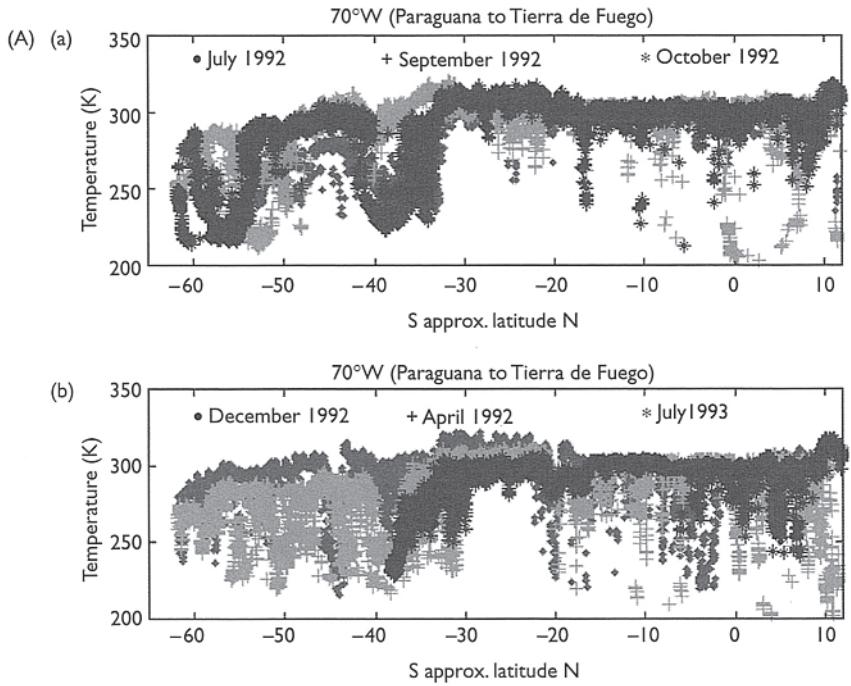


Figure 2.12 (A) Transects at longitude 70ºW for different decades: (a) July 1992 (red), September 1992 (green), October 1992 (blue); (b) December 1992 (red), April 1993 (green), July 1993 (blue) (see Colour Plate XA).

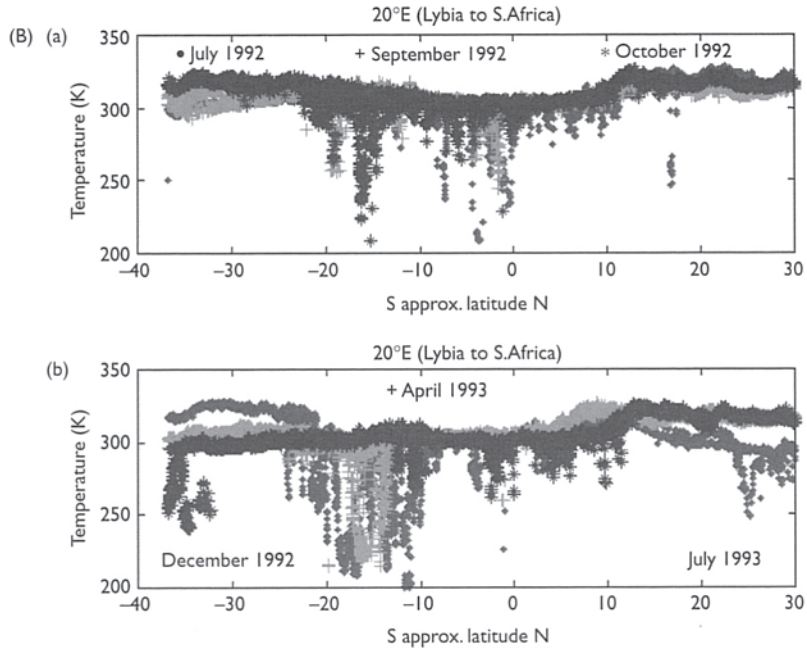


Figure 2.12 (B) Transects at longitude 20ºE for different decades: (a) July 1992(red), September 1992 (green), October 1992 (blue); (b) December 1992 (red), April 1993 (green) July 1993 (blue) (see Colour Plate XB).

spatial distribution of vegetation, as the same fractional cover can give way to very different behaviors.

Validation of satellite-retrieved surface temperatures

Validation of satellite sensors having a coarse resolution in thermal bands presents several challenging problems. The problem of validation can be summarized as follows:

1 The pixel size, even at nadir view, covers an area of about several square kilometres (1.2 km^2, for the AVHRR). It is thus almost impossible to make measurements over such an area accounting for all the spatial variability of the landscape. To overcome this problem, several initiatives have been taken. The most obvious is to choose a validation site that is several kilometers wide (at least 3×3 km), very uniform and perfectly flat. Such an area does not exist, and additionally one would want to have at least two such sites, one bare soil and one fully vegetated. However, several approximations of such targets have been used, one in

Australia (see Prata 1993 and 1994), one in Southeast France (see Kerr *et al.* 1992), and one in the United States (FIFE 1987, 1989). Another approach, less reliable, consists in measuring the temperature over the different elements constituting the observed area and weighting them by the fractional area that they cover to have an estimate of a pixel's temperature. This approach was performed by Prince *et al.* (1995) and Kerr *et al.* (1996) for HAPEX SAHEL. Such validation sites, to be effective, should provide data not only on surface temperature but also on emissivities and atmospheric composition, and thus require coincident ground data acquisition.

2 The second and most difficult validation problem is related to temporal sampling. LST is characterized by a very high temporal frequency. In seconds, the surface skin temperature may change by several degrees (wind, shadow, etc.). To overcome this problem ground radiometers often integrate the temperature over several minutes. The satellite acquisition, however, is done in a fraction of a second. In some cases the spatial aggregation compensates the temporal variability, but this is not always the case. Moreover, due to the integration times, which often vary between 10 and 60 minutes during field experiments, the ground measurements cannot always be related to a single satellite overpass.

3 Finally, one could think of large targets of uniform characteristics to perform validation. But they generally consist in either water bodies, snow packs, or ice sheets, which display the emissivity problems and very often specific atmospheric conditions.

In summary, an accurate validation is very unlikely for large-pixel sensors. An improvement will be available once high-resolution sensors such as ASTER provide high-resolution measurements compatible with ground data. The next issue being then to upscale these high-resolution measurements.

Other issues

Some sensors (such as the AVHRR) may have saturation effects for high radiances. This is particularly true for the AVHRR in the 3.7-μm band, as well as in channels 4 and 5. We will not deal specifically with this problem, as the new generation of sensors overcomes this limitation. Worth mentioning are problems linked to the modulation transfer function of both the sensor and the atmosphere. This problem, leading to somewhat "blurred images" is not always fully assessed, especially as far as the atmospheric contribution is concerned.

The last problem linked to the sensor characteristics is due to the scanning acquisition mode. Apart from the local solar time problem mentioned above, it also has an obvious effect on spatial resolution (hence the size and thus

components of an observed target varies with time over heterogeneous areas), as well as the emissivity (which is prone to show directional effects). These two problems are not yet fully quantified and in many cases may not have a significant effect on the retrieval.

Finally, with medium-resolution sensors, the problems of mixed pixels complicate further interpretation of measured brightness temperatures. The problem worsens with the cloud cover issue, and in many cases, the pixel will be partially filled with clouds, leading to temperature retrieval colder than that at ground level and not always easily detected. This is particularly true for some cirrus-type clouds. With current sensors, the solution is to apply sophisticated filtering methods using all the sensor channels. The next generation of sensors should have spectral channels designed to account for partially cloud-filled pixels.

To illustrate this point LST are plotted for two long transects in Figure 2.13. Even with the use of composited data (10-day composites), cloud cover is not completely filtered out, requiring the use of more sophisticated cloud identification algorithms.

These different points induce problems not only with compositing and potential uses of the LST, but also with validation of retrieval algorithms.

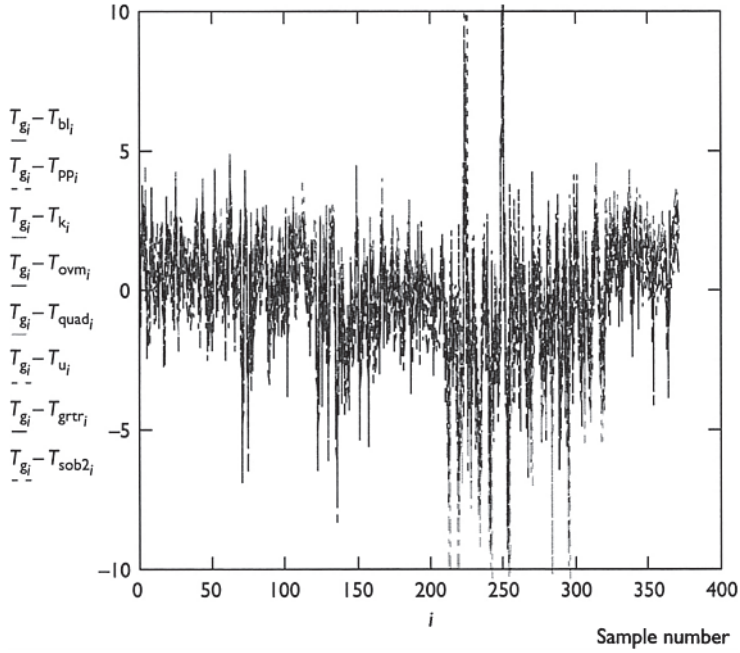


Figure 2.13 Difference between algorithm output and ground measurements for the complete data set used (see Colour Plate XI).

2.5.4 Conclusion

Figures 2.14 and 2.15 depict the global results that support the conclusions of Sections 2.5.1 and 2.5.2. The analysis performed with ground data does not give, however, a clear-cut answer. So, to isolate the most efficient retrieval algorithms, a two-step approach was thus performed. The first step consisted in rejecting all the algorithms that did not satisfy a set of arbitrary criteria. This step is designed to reject any algorithm having a flaw even if they provide good overall statistics. The second step consisted in "classifying" the remaining algorithms according to their accuracy.

The set of arbitrary criteria to reject algorithms is as follows:

1 The mean temperature difference should be less than 1 K. As some of the algorithms tuned using Prata's data set and ground measurements are not always fully representative, this criteria seemed appropriate.
2 The standard deviation of error should be less than 1.75K (same arguments).
3 The correlation coefficient should be better than 0.99.
4 The range (i.e. the maximum difference between ground measurements and retrievals) should be less than 12 K. This last item was selected rather arbitrarily to reduce the number of algorithms to five in the eight left from the three previous criteria.

We are then left with five algorithms (numbers 1, 6, 10, 14, 15) that can be sorted as follows by decreasing order of merit:

1	Mean Error	10, 1, 14, 6, 15
2	STD of error	6, 10, 15, 1, 14
3	Correlation coefficient	6, all the others equivalent
4	Range	10, 6, 1, 14, 15

We thus isolate Becker & Li (1990b) according to Sobrino, Kerr (1992), Ulivieri (1992) and GRTR (1995). Taking into account the results from Section 2.5.1, it is straightforward to recommend Ulivieri's algorithm. A

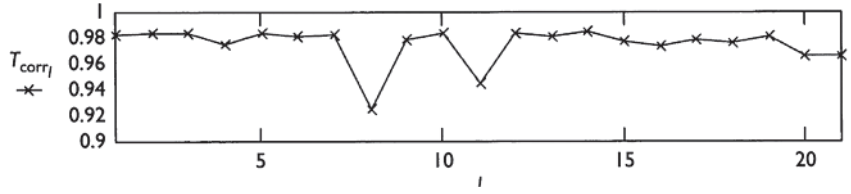


Figure 2.14 Correlation coefficient between ground and AVHRR-derived LST for the 21 tested algorithms. The statistics correspond to the data set taken between noon and 16 h and after removing dubious data points.

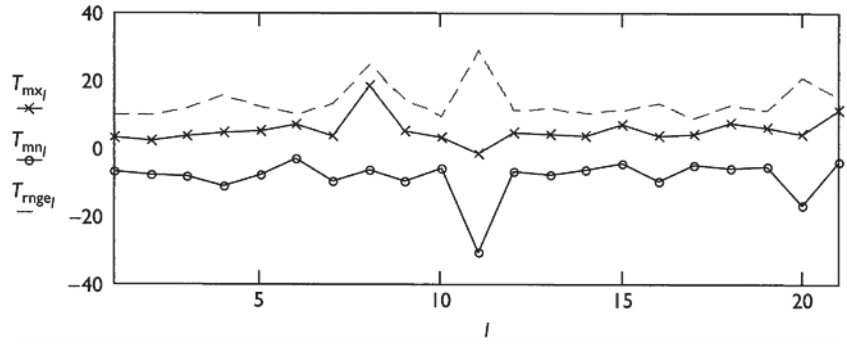


Figure 2.15 Comparison between the different algorithms in terms of maximum differences encountered and range of differences between ground and AVHRR retrieved with all the algorithms tested and for data taken between 12 and 16 h.

study by Vasquez *et al.* (1997) using a Spanish data set, and comparing five algorithms, gives similar results, supporting our analysis. They found, however, an interesting fact, algorithms 6 and 15 are statistically indistinguishable. As they have different approaches, and 15 does not require emissivity data, algorithm 15 is probably the most robust available.

2.6 Applications of land surface temperature

2.6.1 Introduction

It is thus possible to obtain daily values of land surface temperatures with an accuracy of 2 K or better, depending on the knowledge of surface emissivity and eventually atmospheric contribution. This statement applies to large field-of-view/low-resolution sensors having suitable channels. With higher spatial resolution sensors, having more thermal bands (such as the ASTER instrument), one may expect to get a much improved retrieval accuracy. However, spatial resolution is currently to the detriment of temporal sampling. For many applications, low temporal sampling reduces the number of observations and thus limits the application to the monitoring of phenomena slowly changing or unchanging with time. Consequently, for many applications daily coverage implies low-resolution sensors. We note that several satellite missions consider the possibility of merging high spatial resolution and temporal sampling, at the cost of a non-global coverage unless a constellation of satellites is used. One such example is SEXTET (Lagouarde 1997). Another approach proposes satellites with a low inclination orbit (i.e. Bird or Focus for instance).

Considering the currently available data sets, we will concentrate our discussion on AVHRR-type sensors.

Despite the feasibility of retrieving surface temperature, not many applications are operational at this time due to several reasons, which we will enumerate below.

2.6.2 Issues

Frequency of acquisition

Land surface temperature is a very highly variable quantity. Ideally, for many applications (fluxes, hydrology, etc.) it is necessary to capture the diurnal cycle or at least a sufficient number of measurements to characterize it. Geostationary satellites (GOES, METEOSAT) allow such monitoring, but until fairly recently the thermal channels were not adapted to an accurate temperature retrieval (spectral bands and calibration). Moreover, the spatial resolution was of the order of 5 km. The new generation of meteorological satellites is much better suited and should stem the development of applications. Work by Anderson *et al.* (1997) and Mecikalski *et al.* (1999) show how GOES data can be efficiently put to use to take advantage of the high frequency of acquisition merged with AVHRR data.

Polar orbiting satellites (NOAA/AVHRR, ERS/ATSR) do not allow to track well enough the diurnal cycle as their acquisition time are not necessarily optimal.

Cloud cover

When the sky is overcast, surface measurements are impossible. Consequently, the user will only access cloud-free measurements, implying that the data set will be biased. This is a serious issue, which can be worsened by the use of composited data (as depicted in 2.7.1). To address this problem several approaches have been identified:

1 The simplest remedy is to apply an empirical correction factor. This is obviously bound to fail in many instances and cannot be considered as a proper approach.
2 A more realistic approach consists in merging geostationary satellite data to be able to monitor the diurnal cycle and thus interpolate missing data. This method, however, requires large amount of data processing, and a good knowledge of the cloud radiative characteristics so as to be able to compute the energy budget at the surface (Anderson *et al.* 1997; Mecikalski *et al.* 1999).
3 A third method relies on the use of passive microwave data such as the 19-GHz SSM/I (Special scanning microwave/imager) channel. As passive

microwaves are much less sensitive to atmospheric conditions, including cloud cover, they allow all-weather monitoring of the surface. However, the retrieval accuracy in the microwave is generally inferior to the thermal infrared. This is due to the effect of surface characteristics (roughness, moisture), which have a large influence on the emissivity. Moreover, the microwave signal probes the soil over a thickness linked to the wavelength used and vegetation is not totally opaque. Consequently, the temperatures retrieved in the microwaves and in the infrared are not directly comparable. To overcome these problems a "calibration" of the passive microwave signal is performed when thermal infrared is available to adjust for changes in microwave surface emissivity and vegetation cover.

4 The most promising approach is based on the so-called "data assimilation techniques" similar to what is done in meteorological models. A model is run continuously, with the surface temperature as an output, and whenever satellite data are available, the retrieved and computed values are compared and any discrepancy corrected by modifying the related variables/parameters. Such an approach, if based on sound models, enables a continuous and reliable monitoring of surface temperature, thus addressing frequency of acquisition in spite of cloud cover (Cayrol *et al.* 2000).

What is surface temperature

The most important issue when dealing with satellite surface temperature is to know what the radiometric temperature measurements correspond to the aerodynamic temperature used in flux computation for instance. In other words, how the two temperatures relate. Several authors have addressed the problem of the definition of the surface temperature (Norman *et al.* 1994). The "aerodynamic" surface temperature is the one appearing in the bulk transfer equations and is imposed by the equilibrium of the different energy exchanges at the surface at any moment. It may differ by several degrees from the "radiative" temperature as measured in the thermal infrared by satellite (Chehbouni *et al.* 1997b). This largely depends on the structure of the canopy, and on the vertical distribution of temperature inside it. Consequently, when using thermal infrared data for assessing fluxes, one has to take into account the extra resistance linked to heat transfer in aerodynamically rough surfaces (Norman and Becker 1995). A correction factor the so-called kB-1 has thus to be applied (see the workshop on thermal remote sensing proceedings—1993, for more details) but the distinction should be made between the real aerodynamic kB-1 (Norman and Becker 1995) and the empirical one used to correct for view angle and other situation dependent effects.

Scale effects

Finally, surface temperature and related fluxes are not linearly related. When comparing ground data and satellite data, or when using satellite data in models, this factor has to be taken into account (Chehbouni *et al.* 1995). This problem has been addressed (see Stewart *et al.* 1996), but clear-cut answers are not yet available and more work is still required.

2.6.3 Potential applications

In this section, we will give a quick outlook of several applications, either proven or in the definition stages as the main ones are covered in the other chapters.

Geology

Several authors have investigated the potential use of thermal infrared for applications in geology and in particular looking at the thermal inertia of surfaces. They showed that in geology several applications are possible, provided multispectral data are acquired. This led to the ASTER concept.

Several other applications emerged, such as the monitoring of volcanoes (Bonneville *et al.* 1985, 1987). In this case, the thermal anomalies on the slopes of a volcano are detected as forerunners of eruptions. The approach is certainly interesting but requires frequent monitoring with a high spatial resolution.

Evapotranspiration, sensible heat fluxes

Several authors concentrated on the derivation of evapotranspiration and or rainfall monitoring using a simplified approach (Kerr *et al.* 1989; Seguin *et al.* 1989). It was found that these methods had several inherent limitations due to the influence of surface roughness and wind speed, as well as problems linked to the cloud cover. More sophisticated methods are now being investigated (Seguin *et al.* 1994; Chehbouni *et al.* 1997a).

Spring frosts

Spring frosts occur when radiative cooling is high (i.e. no clouds). Consequently, there are no clouds and provided the nocturnal temperature decrease can be estimated (one or two measurements between sunset and before sunrise) the areas prone to freezing can be identified and the information relayed to the field owners so that they can take the necessary steps to protect their crops in a timely manner and only as necessary (Kerr *et al.* 1983; Kerdiles *et al.* 1996). Similar methods have been used to estimate the damage during abnormally cold events (Choisnel *et al.* 1987). Recently,

François *et al.* (1999) have proposed a methodology to link the AVHRR radiative temperature measured at 2 am to minimum air temperature and frost risks over the Bolivian Altiplano. This method is based on empirical relationships between series of AVHRR data and meteorological station data coupled wirth a classification of the area.

Fires

It is well known that when a 3.7-μm band is available, active fires can be detected. This capability is now widely used (see the ATSR fire atlas for instance). Authors have also investigated the possibility of early warning of fire risks, essentially in the Mediterranean area. The concept here is to build a water stress indicator based on the difference between air and surface temperature and couple it to meteorological forecasts (wind mainly). From this information, it is possible to derive maps of areas where risks are high. It is also possible to analyze the surroundings of burning area to estimate in which direction the fire is most likely to propagate and estimate its speed.

Regional hydrology and "drought indices"

It was found that by cumulating the difference between surface temperature and air temperature a climatic index could be retrieved. This information is useful to estimate drought extent or to analyze climatic trends at a regional level (Courault *et al.* 1994; Savane *et al.* 1994).

Novel approaches

Recently, with the availability of the directional measurements of the ATSR, several authors investigated several new applications. One of them relies on the use of the directional measurements to assess both surface and canopy temperatures (François *et al.* 1997). The theoretical aspects have been demonstrated but practical use is still under investigation. Applications in crop yield prediction are also currently being investigated.

2.6.4 Conclusion

Research work carried out in approximately 15 years have shown that many applications of thermal infrared data are possible. Due to the problems associated with LST retrievals, very few are in operational use now. However, with the emergence of a new generation of satellite sensors and of dedicated space missions, and advances in the understanding of physical phenomena, in a very near future many new applications may emerge on an operational basis.

2.7 Impact of compositing on relevance of surface temperature

2.7.1 Simulation of LST composites

Most work performed in the domain of the thermal infrared has been done using instantaneous values of surface temperature measured in the early afternoon, around 14.00 UT. So as to overcome cloudy conditions, the maximum value composite method has been developed. It consists of selecting over a given period of time (typically 10-days), on a pixel-per-pixel basis, the data corresponding to the highest vegetation index value. However, the use of a composite LST first requires evaluating the physical meaning of such data for various applications (estimation of actual evapotranspiration, monitoring of water stress, etc.). In the following, we will analyze the impact of using such composites for typical applications of thermal infrared.

We based a study on simulations performed with an agrometeorological model, MAGRET, a soil-plant-atmosphere transfer model. A description of the model can be found in Brunet *et al.* (1996). It runs with standard input meteorological data (either at daily scale with a statistical reconstruction of hourly data, or directly with hourly data provided by automated stations). It predicts surface fluxes and surface temperatures on an hourly time step. This model couples the classical equations of the turbulent transfers in the atmosphere with a simplified soil water budget based on a two-reservoir system. The surface temperature is computed as the solution of the non-linear energy budget equation.

We present results obtained from simulations performed with INRA meteorological hourly data recorded at Bordeaux for the years 1993–6. The surface considered is a "standard" grass (as in meteorological stations) with constant roughness (2cm) and LAI (3).

The criteria for extracting the composite LST here consists of the selection of the T_s value corresponding to the maximum of sunshine fraction along the 10-day period considered. The sunshine fraction is defined as the ratio between the global solar radiation measured on the weather station and the theoretical value corresponding to a clear day (the latter being estimated statistically with a sunshine fraction equal to unity). This criterion obviously differs from the one based on the maximum NDVI value. Nevertheless, it seems reasonable to assume they are equivalent, the maximum NDVI generally being obtained for the most clear atmospheric conditions.

For some particular meteorological conditions or for some regions (northern countries, equatorial zone, etc.) significant cloud cover may result in missing composite data. The number of missing data will obviously depend on the nature of the criteria imposed for selecting the composite LST. As an example, we have computed for Bordeaux the percentage of days displaying a sunshine fraction greater than a given threshold at different times.

Table 2.4 Frequency of days with a sunshine fraction higher than a given value for the different months for a 14.00 UT acquisition time (Bordeaux data, 1980–6)

Sun. fract.	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
0.00	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.0	100.0
0.55	100.0	96.1	100.0	98.0	100.0	100.0	100.0	100.0	100.0	98.0	94.1	94.1
0.60	98.0	94.1	96.1	98.0	100.0	100.0	100.0	100.0	98.0	96.1	94.1	92.2
0.65	84.3	92.2	96.1	98.0	98.0	100.0	100.0	100.0	96.1	96.1	82.4	80.4
0.70	66.7	88.2	94.1	96.1	96.1	96.1	100.0	98.0	96.1	90.2	76.5	64.7
0.75	56.9	78.4	88.2	92.2	90.2	90.2	96.1	96.1	94.1	82.4	58.8	43.1
0.80	33.3	54.9	72.5	84.3	82.4	82.4	90.2	90.2	90.2	54.9	33.3	25.5
0.85	19.6	29.4	41.2	51.0	58.8	58.8	72.5	68.6	58.8	37.3	21.6	17.6
0.90	15.7	15.7	9.8	29.4	31.4	31.4	29.4	3.9	29.4	13.7	11.8	9.8
0.95	2.0	5.9	2.0	5.9	0.0	3.9	0.0	0.0	3.9	5.9	0.0	3.9

The results at 14.00 UT are given in Table 2.4. They have been established using 17 years of meteorological data, only 4 years being hourly data (for the other years, hourly data have been statistically rebuilt). They must, therefore, be considered with care. Nevertheless, it can be seen that summer months offer the greatest probability of clear conditions (90% of days with a sunshine fraction higher than 0.80 in July and August, for instance), whilst in winter data are sometimes missing.

Another composite product easy to derive consists in the maximum LST over 10-day periods. Its potential will also be very briefly analyzed here.

2.7.2 Temporal evolution of composite LST

In order to describe the significance of the different "temperatures" with regards to the maximum vegetation index 10-day composite, Figure 2.16(a) displays the evolution throughout the years 1993–6 of:

1 The composite LST retrieved at 14.00 UT ($T_{s_{c14}}$), which is a good approximation of the AVHRR composite retrieval.
2 The mean surface temperature over the 10-day period ($T_{s_{c14}}$), which could be related to the "average" temperature over the 10-day period.
3 The absolute maximum T_s over the same period ($T_{s_{xx}}$).

We notice that $T_{s_{c14}}$ most of the time strongly differs from the two other LSTs, the difference being approximately 3K (absolute value), with important fluctuations reaching up to 10 K (Figure 2.16(b)).

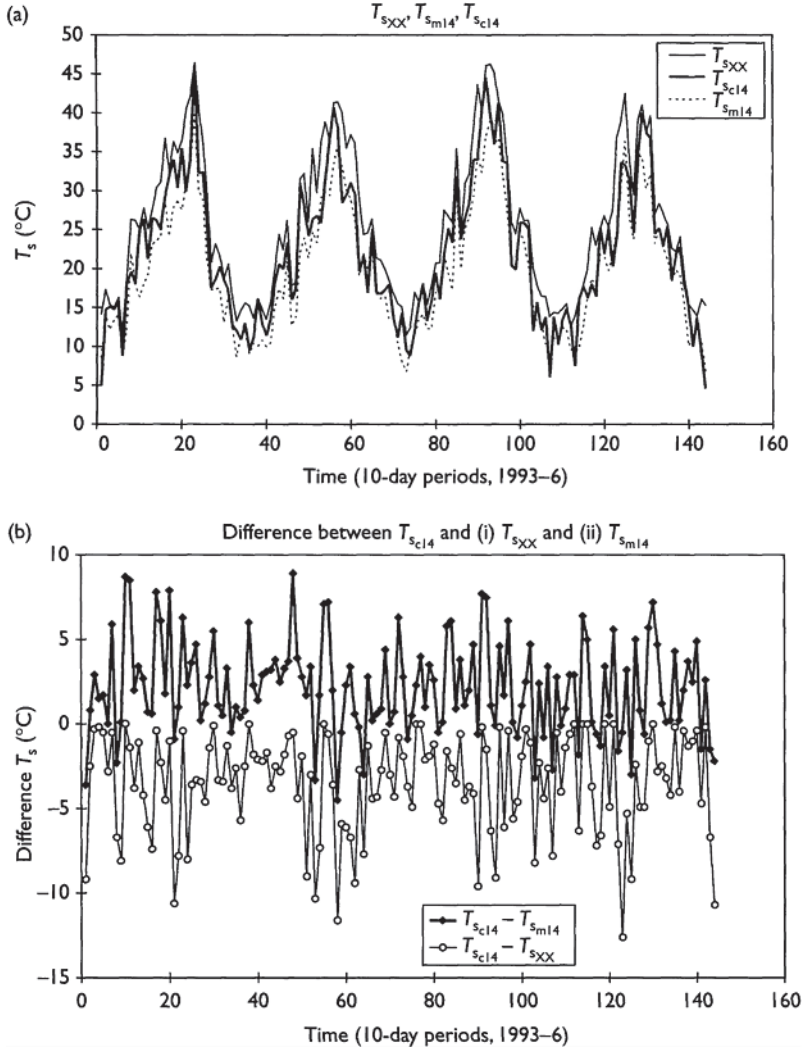


Figure 2.16 Evolution of T_{sc14}, T_{sm14}, and T_{sXX} (a) and of the differences $T_{sc14} - T_{sXX}$ and $T_{sc14} - T_{sm14}$ (b) for the years 1993–6 at Bordeaux.

2.7.3 Potential applications to the estimation of actual evapotranspiration using semi-empirical models

One of the applications of surface temperature is the derivation of actual daily evapotranspiration. The semi-empirical methods (Lagouarde 1991; Lagouarde and McAneney 1992; Lagouarde *et al.* 1996a,b) based on the

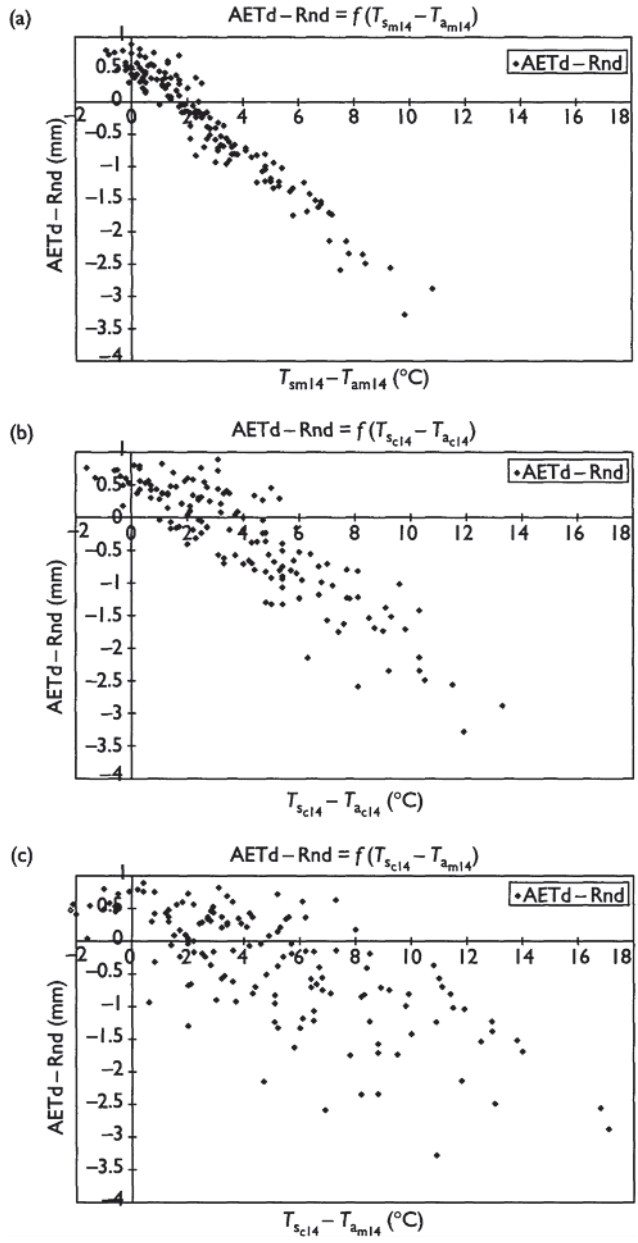


Figure 2.17 *Relationships between AETd—Rnd (in equivalent mm of water evaporated) and* $T_s—T_a$.

use of statistical relationships relating the difference AETd—Rnd (between daily actual evapotranspiration and net radiation) to the instantaneous value of T_s—T_a (difference between surface and air temperatures in the early afternoon, at 14.00 UT) provide a reasonable accuracy when used at a 10-day scale (or more) basis with mean values of LST and air temperature (Figure 2.17(a)). They require daily observations, their use however, is limited by the lack of LST satellite data for cloudy conditions. Despite the filtering of clouds it can perform; the composite LST does not seem suitable to flux estimation directly: as a matter of fact, it can be seen in Figure 2.17(b) and (c) that the scatter on the relationships strongly increases when $T_{s_{c14}}$ is combined with an air reference temperature that can be either the air temperature simultaneous with the composite $T_{a_{c14}}$ (but this practically requires using the date of the composite) or the mean air temperature at 14.00 UT over the 10-day period, $T_{a_{m14}}$. Thus, a quantitative assessment of actual evapotranspiration at a 10-day scale seems prone to be too large an error. The possibility of reducing the scatter by performing integration over longer periods (typically the month scale) is still to be explored.

2.7.4 Application to climatic conditions monitoring

Use of cumulated differences between air and surface temperature

The cumulated differences T_s—T_a over periods ranging from a few months to 1 year have been extensively used for a global characterization of water conditions at the regional scale (Seguin *et al.* 1991). Figure 2.18 displays classical results: the comparison for the four studied years between the cumulated evaporation (ΣAETd, Figure 2.18(a)) or sensible heat flux (ΣRnd—ΣAETd, Figure 2.18(b)) and the cumulated mean T_s—T_a at 14.00 UT, $T_{s_{m14}}$—$T_{a_{m14}}$ (Figure 2.18(c)) on a 7-month period (March-September) clearly illustrates the fact that the "driest" years correspond to both low ΣAET*d* and high $\Sigma(T_{s_{m14}}—T_{a_{m14}})$. This is not surprising, being in fact another representation of the plot in Figure 2.17(a).

The classification 1995–1993–1996–1994 of decreasing drought status does not appear as clear when using composite LST, the "hierarchy" between curves changing (Figures 2.19–2.22). For Figures 2.19–2.21, mean values over 10-day periods are considered for the reference air temperature. These have the advantage of being easily available from meteorological data. One may think that the advantage of the mean air temperature T_{a_m} (average value of 10 mean daily air temperatures) is to display less spatial variability than the mean air temperature at 14.00 UT ($T_{a_{m14}}$) or the mean maximum value ($T_{a_{Xm}}$). This is important for practical applications, as the air temperature is available only at the 1° square grid. In Figure 2.22, the instantaneous value of air temperature, $T_{s_{c14}}$, at the time of the composite T_s has been considered.

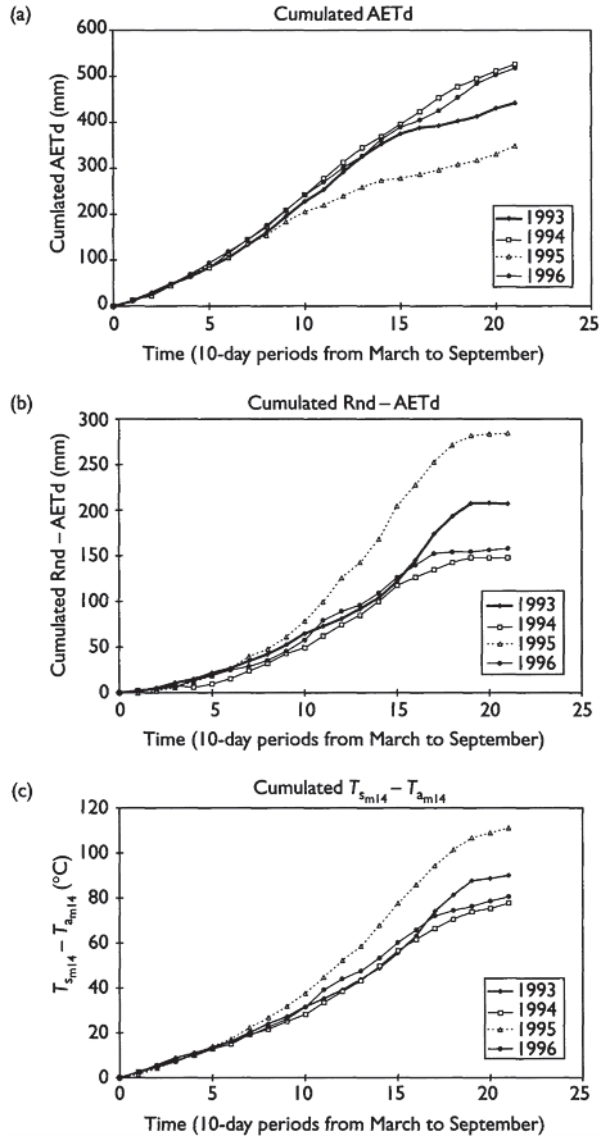


Figure 2.18 *Comparison of cumulated fluxes and* $\Sigma(T_{s_{m14}} - T_{a_{m14}})$ *for 1993–6.*

It appears that the difference between 1995 and 1996 is always clearly characterized. The use of T_{a_m} (Figure 2.21) seems best suited for characterizing years 1994 and 1996, which are quite comparable (see Figure 2.18). The difference appearing between these two years when using $T_{a_{m14}}$ and $T_{a_{Xm}}$

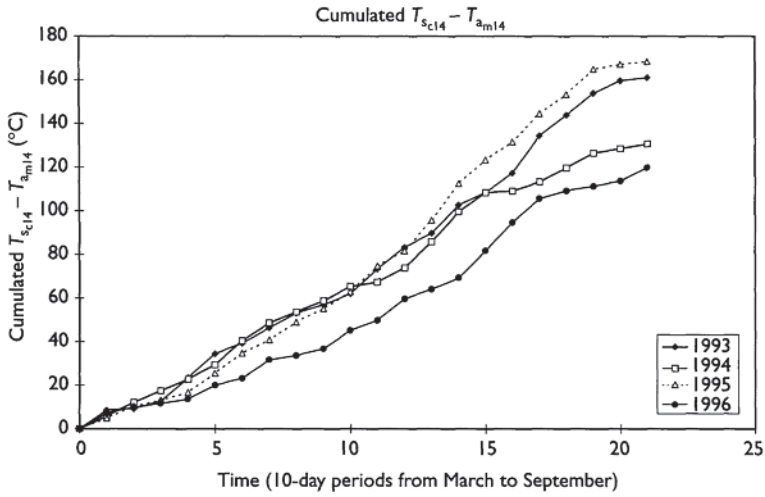


Figure 2.19 Evolution of $\Sigma(T_{s_{c14}} - T_{a_{m14}})$ over the period March-September for 1993–6.

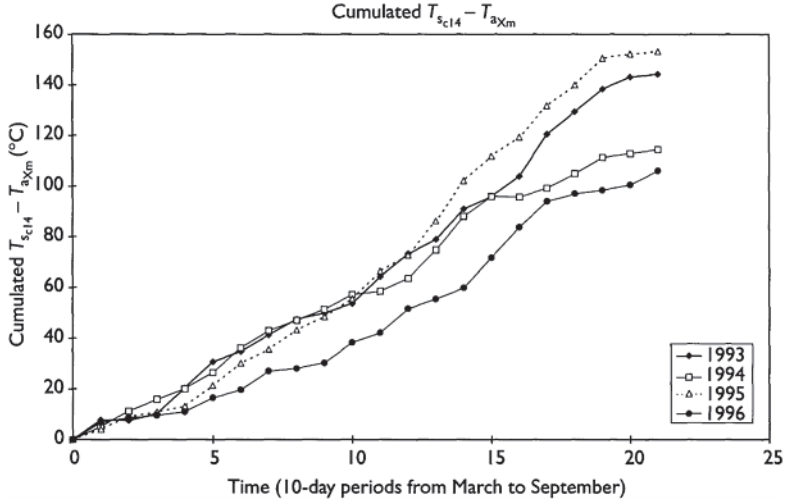


Figure2.20 Evolution of $\Sigma(T_{s_{c14}} - T_{a_m})$ over the period March-September for 1993–6.

and can be explained by inter-annual variations in meteorological conditions influencing the compositing process: Figure 2.23 displays the time-evolution of the sunshine fraction corresponding to the dates of extraction of $T_{s_{c14}}$, and illustrates the fact that in 1996 (characterized by lower fraction values), most of data are likely to be affected by atmospheric cloudiness or haze.

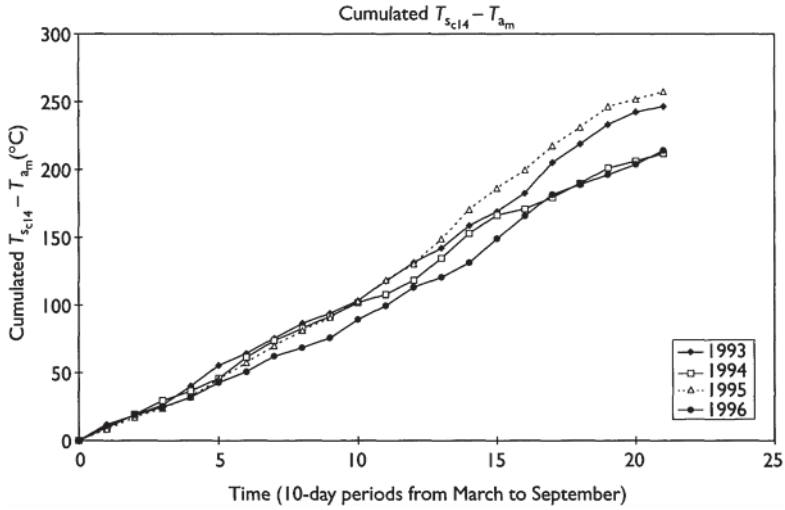


Figure 2.21 Evolution of $\Sigma(T_{s_{c14}} - T_{a_m})$ over the period March-September for 1993–6.

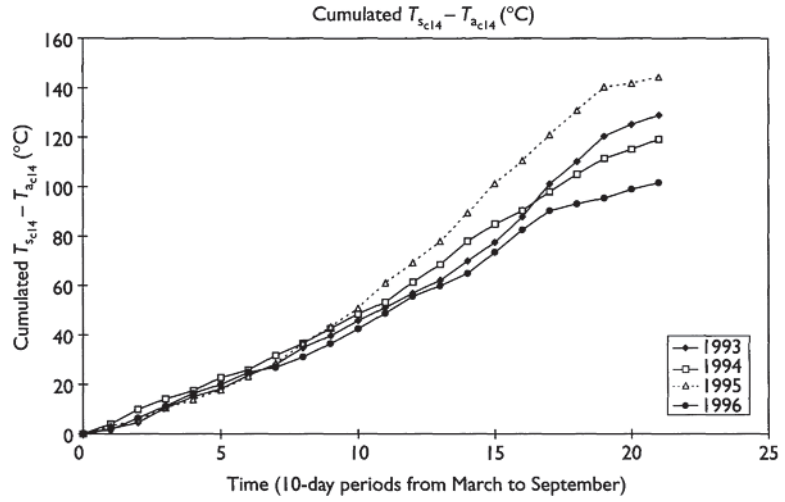


Figure 2.22 Evolution of $\Sigma(T_{s_{c14}} - T_{a_{mc14}})$ over the period March-September for 1993–6.

Another possibility for the differences could also lie in the relative dates of composite T_s and rainfall events: important differences on $T_{s_{c14}}$—for the same 10-day rainfall and evapotranspiration—might result from the fact that $T_{s_{c14}}$ would be acquired more often in the beginning (before rain) or at the end (after rain) of the period. The dates of composite T_s and of rains have been

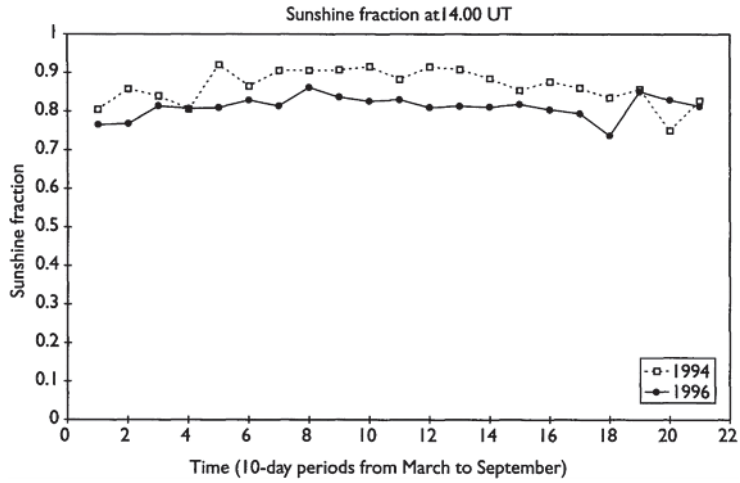


Figure 2.23 Comparison of sunshine fraction at 14.00 UT (time of LST composite) for 1994 and 1996.

compared in the present case and no systematic effect has been noted (Figures 2.24 and 2.25). The effect of wind speed, which could influence the transfer coefficient and consequently the difference T_s—T_a for the same values of surface fluxes, could also be invoked: Figure 2.26 displaying the temporal evolution of mean wind speed for years 1994 and 1996 led us to eliminate this explanation.

In the particular case studied, the use of the air temperature acquired simultaneously to the composite LST seems less interesting, the hierarchy between the four years 1993–6 being strongly affected (Figure 2.22).

The interest of the maximum LST, which is another composite product easy to build, has not yet been evaluated in detail. We only present here the evolution of the cumulated T_{sxx}—T_{a_m} (Figure 2.27): confusion appears between years 1993 and 1996, despite their different water budget status. On the other hand, the characterization of the three other years seems reasonable. Further work is necessary to evaluate the potential of T_{sxx}.

These observations are very preliminary. They have to be checked on other sites and for a large range of meteorological conditions and surface conditions (particularly roughness length). Moreover, possible methods for correcting and normalizing the cumulated values of T_s—T_a from errors (such as inter-annual variations of meteorological conditions) need to be defined.

Use of cumulated LST

Despite the differences, T_s—T_a are best suited for monitoring the water status of surfaces and it is interesting to analyze the potential of cumulated LST.

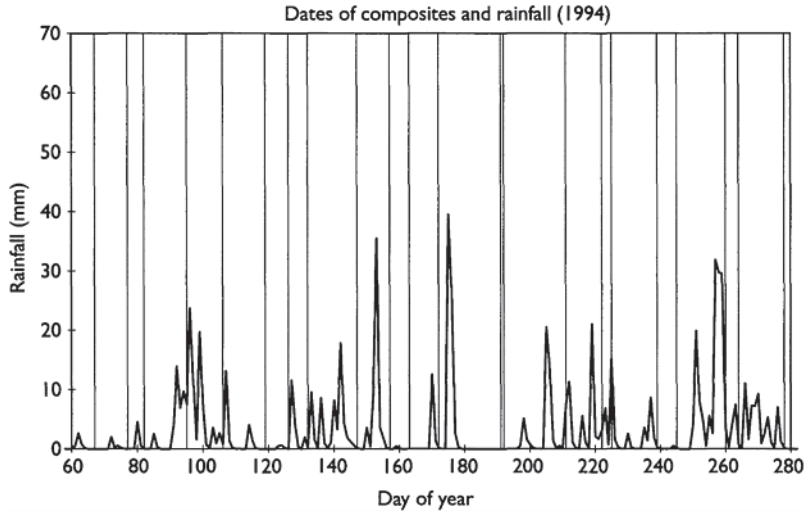


Figure 2.24 Dates of composite LST and rainfall for 1994.

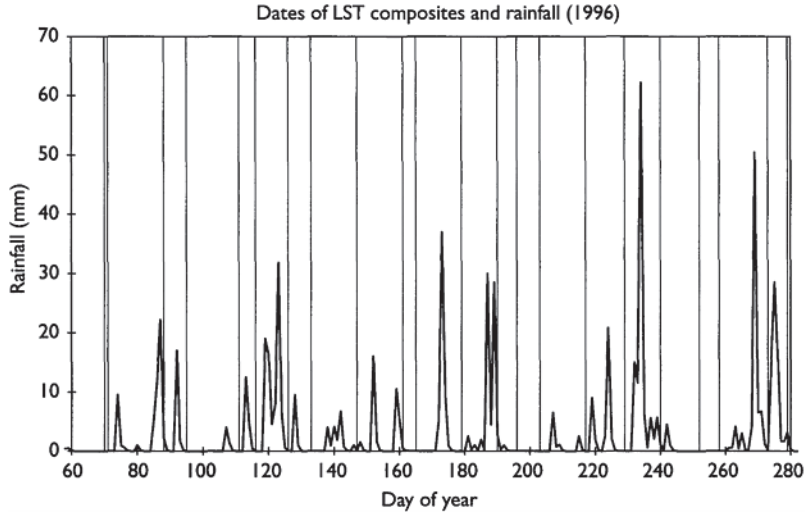


Figure 2.25 Dates of composite LST and rainfall for 1996.

Figures 2.28 and 2.29 display the cumulated $T_{s_{c14}}$ and $T_{s_{XX}}$, respectively. Here again some features of the evolution of surface fluxes can be found. In both cases, the "driest" year, 1995, displays high values of ST_s from the period 13 and can easily be distinguished from other years (the evolution of fluxes shows a drought beginning around period 8, see Figure 2.18). For 1993, drought

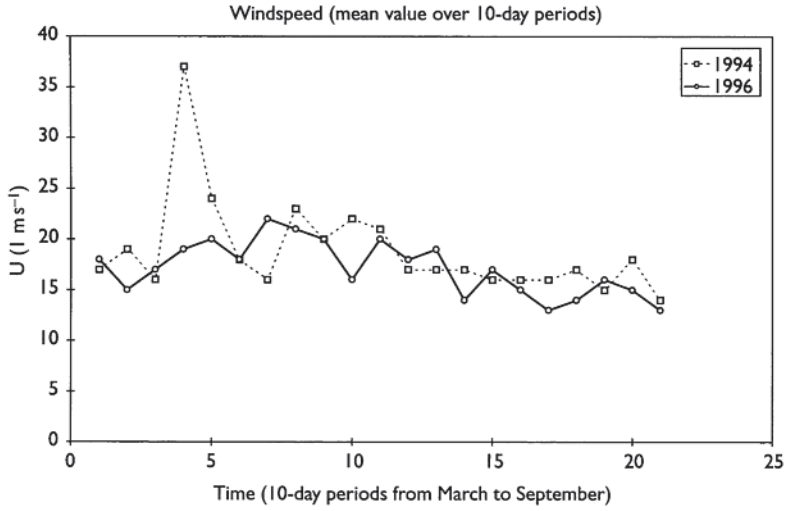


Figure 2.26 Comparison of windspeed from March to September for 1994 and 1996.

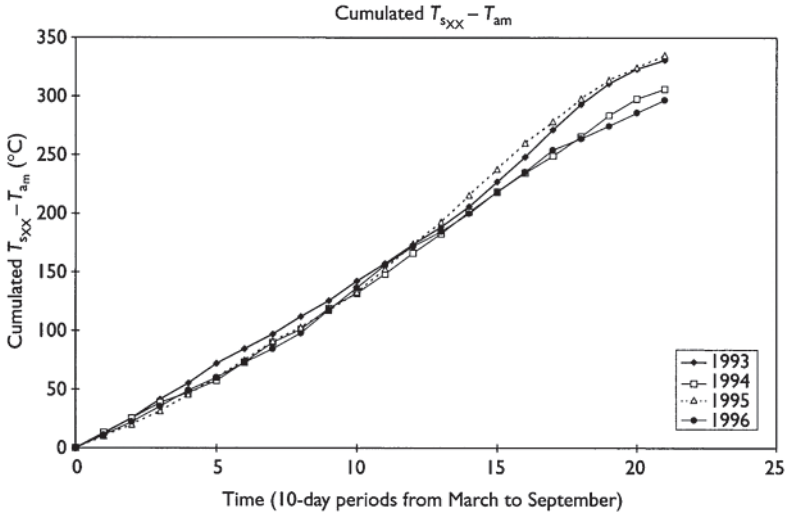


Figure 2.27 Evolution of $\Sigma(T_{sXX} - T_{am})$ over the period March to September for 1993–6.

occurs later around period 16, which can also be noticed on ΣT_{sc14} and ΣT_{sXX}. But some differences remain between ΣT_{sc14} and ΣT_{sXX} at the end of the season for 1994 and 1996, which does not correspond to differences of fluxes.

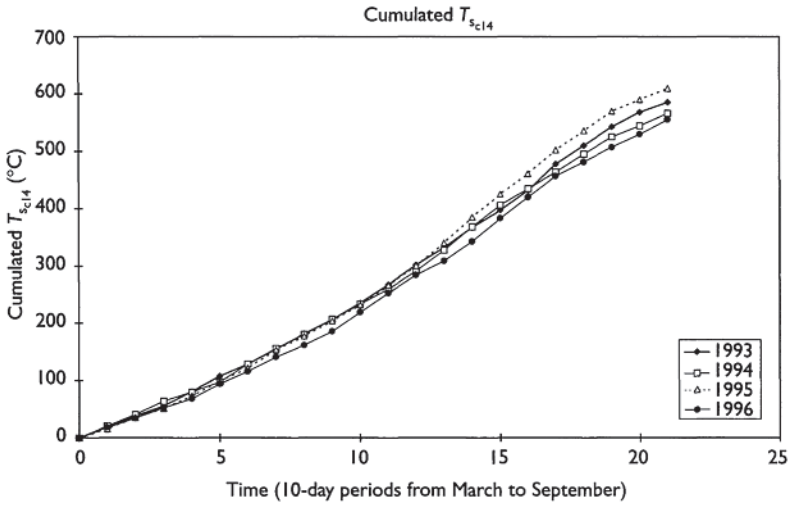


Figure 2.28 Evolution of $\Sigma T_{s_{c14}}$ over the period March to September for 1993–6.

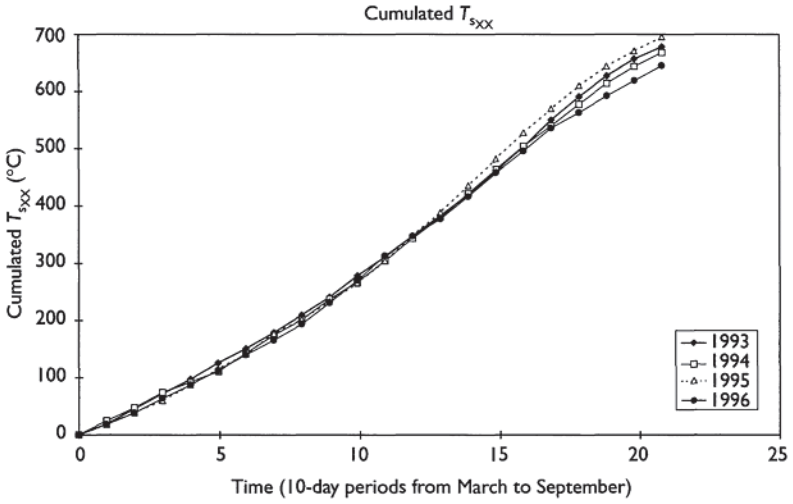


Figure 2.29 Evolution of $\Sigma T_{s_{XX}}$ over the period March to September for 1993–6.

2.7.5 Concluding remarks and further work

These preliminary results indicate that the composite LSTs cannot directly replace instantaneous values of surface temperature in the methods previously developed for monitoring water conditions. Nevertheless, the simulations performed (over Bordeaux) on a limited data set (1993–6) show that composite

T_s (derived on a criteria of maximum sunshine fraction, assumed to be comparable to the one based on NDVI maximum) combined with mean values of air temperature (mean T_a, or mean $T_{a_{14}}$, or mean T_{a_x} can provide information on the water status of the surface in some conditions. In some cases, however, the composite LST cannot be used directly and appears unable to characterize some years satisfactorily. Additional work is necessary to understand the reasons of such problems (cloud cover, temporal repartition of rains, effect of wind speed, etc.), and methods will have to be developed in the future to correct, if possible, these effects.

In the future, simulations should be generalized to other locations. They should seek to perform not only inter-annual comparisons as it has been done here for Bordeaux, but also to evaluate the potential of composite LST (maximum NDVI derived as well as maximum LST) for characterizing the spatial variability in surface water status.

2.8 Conclusions and perspectives

The results presented in this chapter show that retrieving surface temperature from space is not a straightforward operation, and issues and caveats have to be dealt with. It seems that we now have fairly good algorithms, and that in the near future, with the new generation of sensors, even better retrievals can be expected. The main issues are linked to a better knowledge of surface spectral emissivities and improved methods to take into account cloud cover and mixed pixels/aggregation issues.

Future improvements are thus closely linked with the availability of accurate and global fields of surface emissivity, a better modeling of scaling effects, and the use of assimilation models or improved and adapted compositing methods.

More challenging is the fact that apparently the MVC does not give necessarily cloud-free pixels and that cases of clouds are not always detected. Future improvement will rely on the use of:

1 a more sophisticated compositing method;
2 better identification of cirrus and scattered clouds;
3 additionally, the separation cloud/snow/ice with added dedicated wavelengths would greatly improve the products.

Validation of retrieved surface temperature may also be an issue for large pixels. The development of new ground techniques such as the scintillometry (Lagouarde *et al.* 1996) might provide the community with an improved tool.

Once these different steps are done it can be expected that thermal infrared will prove to be a most efficient tool for a wealth of applications in various domains. The added capabilities of the next generation of sensors (angular measurements, multispectral aspects, etc.) will provide added opportunities.

Appendix

Color scheme for figures: 2.3, 2.4, 2.6–2.12.

Algorithms #:

1 GRTR (1995)
2 Becker and Li (1990b)
3 Prata and Platt (1991)
4 Price (1984)
5 Ulivieri and Cannizzaro (1985)
6 Ulivieri *et al.* (1992)
7 *Sobrino et al.* (1993)
8 *Sobrino et al.* (1991)
9 NESDIS(1992)
10 Becker and Li, acc. to Sobrino (1990)
11 Prata and Platt, acc. to Caselles
12 Prata and Platt, acc. to Sobrino
13 Price, acc. to Sobrino
14 Ulivieri *et al.* acc. to Sobrino
15 *Kerr et al.* (1992)
16 Vidal (1991)
17 *Coll et al.* (1997)
18 Ottlé and Vidal-Madjar (1992)
19 WVD (1996)
20 QUAD (1996)
21 *Sobrmo et al.* (1994).

Color scheme

0 $<W_v<1$ pink
1 $<W_v<2$ blue
2 $<W_v<3$ green
3 $<W_v<4$ yellow
4 $<W_v<5$ red
5 $<W_v<6$ black (solid)
6 $<W_v<8$ black (dashed).

Acknowledgments

The authors of this work are thankful for the satellite data providers (EDC, ESA), the ground data providers (HAPEX-SAHEL, CSIRO, INRA), to Nelly Mognard, and to Fred Prata for informative discussions and data search, and to Susan Embrock at EDC who was most helpful in providing as soon as possible the new data files for computing AVHRR solar and view angles.

They are greatly indebted to the anonymous reviewers for their in depth comments and corrections.

References

Anderson, M.C., J.M.Norman, G.R.Diak, W.P.Kustas, and J.R.Mecikalski (1997) A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote Sens. Environ.* 60: 195–216.

Anding, D. and R.Kauth (1970) Estimation of sea surface temperature from space. *Remote Sens. Environ.* 1: 217–20.

Andre, J.C. *et al.* (1988) HAPEX-MOBILHY: first results from the Special Observing Period. *Ann. Geophys. 6:* 447–92.

Becker, F. (1982) Absolute sea surface temperature measurements by remote sensing and atmospheric corrections differential radiometry. In *Processing in Marine Remote Sensing,* F.J.Vernberg and F.P.Diemes, Eds, University of South Carolina Press, Columbia, pp. 151–74.

Becker, F. (1987) The impact of spectral emissivity on the measurement of land surface temperature from a satellite. *Int. J. Remote Sens.* 8 (10): 1509–22.

Becker, F. and Z.L.Li (1990a) Temperature-independent spectral indices in thermal infrared bands. *Remote Sens. Environ.* 32: 17–33.

Becker, F. and Z.L.Li (1990b) Towards a local split window method over land surface. *Int. J. Remote Sens.* 11: 369–93.

Becker, F. and Z.L.Li (1995) Surface temperature and emissivity at various scales: definition, measurements and related problems. *Remote Sens. Rev.* 12: 225–53.

Bonneville, A., G.Vasseur, and Y.H.Kerr (1985) Satellite thermal infrared observations of Mt Etna after the 17th March 1981 eruption. *J.Volcanol. Geotherm. Res.* 24 (3 & 4): 293–313.

Bonneville, A., G.Vasseur, and Y.H.Kerr (1987) A thermal forerunner of the March 28th 1983 Mt Etna eruption from satellite thermal infrared data. *J. Geodyn.* 7 (1): 1–31.

Brown, O.B. *et al.* (1985) Calibration of AVHRR infra-red observations. J. *Geophys. Res.* 90:11667.

Brunet, Y., J.P.Lagouarde, and V.Zouboff (1996) Estimating long-term microclimatic conditions for long-range sound propagation studies. In *7th Long Range Sound Propagation Symposium,* Ecole Centrale de Lyon, France, 24–26 July, 14pp.

Buettner, K.J.K. and C.D.Kern (1965) The determination of infra-red emissivities of terrestrial surfaces. *J. Geophys. Res.* 70:1329–37.

Caselles, V., C.Coll, and E.Valor (1997) Land surface temperature determination in the whole Hapex Sahel area from AVHRR data. *Int. J. Remote Sens.* 18 (5): 1009–27.

Cayrol, P., A.Chehbouni, L.Kergoat, G.Dedieu, P.Mordelet, and Y.Nouvellon (2000) Grassland modeling and monitoring with SPOT-4 VEGETATION instrument during the 1997–1999 Salsa experiment, Agricultural and Forest Meteorology, Volume 105, Issues 1–3, 91–115.

Chandrasekhar, S. (1960) *Radiative Transfer.* Dover, New York.

Chedin, A., N.A.Scott, and A.Berroir (1982) A single channel double viewing angle method for sea surface temperature determination from coincident METEOSAT and TIROS-N radiometric measurements. *J. Clim. Appl. Meteorol.* 21: 613–18.

Chedin, A., N.A.Scott, C.Wahiche, and P.Moulinier (1985) The Improved initialization inversion method: a high resolution for temperature retrievals from the TIROS-N series. *J. Clim. Appl. Meteorl.* 24:124–43.

Chehbouni, A., Y.H.Kerr, J.Qi, A.R.Huete, and S.Sorooshian (1994) Towards the development of multidirectional vegetation index. *Water Resour. Res.* 30 (50): 1281–6.

Chehbouni, A., E.G.Njoku, J.P.Lhomme, and Y.H.Kerr (1995) Approaches for averaging surface parameters and fluxes over heterogeneous terrain. *J. Clim.* 8 (5, II): 1386–93.

Chehbouni, A., D.Lo Seen, E.G.Njoku, J.P.Lhomme, B.Monteny, and Y.H.Kerr (1997a) Estimation of sensible heat flux over sparsely vegetated surfaces. *J. Hydrol.* 188–9 (1–4): 855–68.

Chehbouni, A., D.Nichols, E.G.Njoku, J.Qi, and Y.H.Kerr (1997b) On the use of radiative surface temperature to estimate sensible heat flux. *Remote Sens. Rev.* 15:99–112.

Choisnel, E., Y.H.Kerr, and P.Blanchet (1987) Etude du début de la vague de froid de Janvier 85 en France par l'analyse combinée de données satellitaires et des mesures au sol. *La Météorologie.* 19:11–28.

Choudhury, B.J., T.J.Dorman, and A.Y.Hsu (1995) Modeled and observed relations between the AVHRR Split-Window temperature difference and atmospheric precipitable water over land surfaces. *Remote Sen. Environ.* 51: 281–90.

Coll, C., E. Valor, T.J.Schmugge, and V.Caselles (1997) A procedure for estimating the land surface emissivity difference in the AVHRR channels 4 and 5. Application to the Valencian Area, Spain.

Cooper, D. and G.Asrar (1989) Evaluating atmospheric correction models for retrieving surface temperature from the AVHRR over a tall grass prairie. *Remote Sens. Environ.* 27: 93–102.

Courault, D., P.Clastes, J.P.Guinot, and B.Seguin (1994) Analyse des sécheresses de 1988 a 1990 en France a partir de l'analyse combinée de données satellitaires et d'un modèle agrométéorologique. *Agronomie* 14: 41–56.

Eastes, J.W. (1989) Spectral properties of halite rich mineral mixtures: implications for middle infra-red remote sensing of highly saline environments. *Remote Sens. Environ.* 27: 289–304.

Eccles, D., M.Gorman, D.Lee, P.Moffat, and A.Zavody (1989) Overview of the ATSR-IR off-line data processing system, RAL (internal note).

Elvidge, C.D. (1988) Thermal infra-red reflectance of dry plant materials: 2.5–20 μm. *Remote Sens. Environ. 26:* 265–85.

ESA (1989) *ERS-1 Reference Manual,* ESA, DC-MA-EOS-ED-0001.

François, C. and C.Ottlé (1996) Atmospheric corrections in the thermal infrared: global and water vapor dependent Split-Window algorithms. Application on ATSR data. *IEEE Trans. Geosci. Remote Sen.* 34: 457–70.

François, C., C.Ottlé, and L.Prévot (1997) Analytical parameterization of canopy emissivity and directional radiance in thermal infrared. Application on the retrieval of soil and foliage temperatures using two directional measurements, Part I, theory. *Int. J. Remote Sens.* 18: 2587–621.

François, C., R.Bosseno, J.J.Vacher, B.Seguin (1999) Frost risk mapping derived from satellite and surface data over the Bolivian Altiplano. *Agric. For. Meteorol.* 95(2): 113–37.

Fuchs, M. and C.B.Tanner (1966) Infra-red thermometry of vegetation. *Agron. J.* 58: 597–601.

Fuchs, M., E.T.Kanemasu, J.P.Kerr, and C.B.Tanner (1967) Effect of viewing angle on canopy temperature measurements with infrared thermometers. *Agron. J.* 59: 494–6.

Harris, A.R. and I.M.Mason (1992) An extension to the split-window technique

giving improved atmospheric correction and total water vapor. *Int. J. Remote Sens.* 5: 881–92.
Jedlovec, G.J. (1990) Precipitable water estimation from high-resolution SplitWindow radiance measurements. *J. Appl. Meteorol.* 29:863–77.
Kealy, P.S. and S.J.Hook (1993) Separating temperature and emissivity in thermal infrared muspectral scanner data: implications for recovering land surface temperatures. *IEEE Trans. Geosci. Remote Sens.* 31: 1155–64.
Kerdiles, H., M.Grondona, R.Rodriguez, and B.Seguin (1996) Frost mapping using NOAA/AVHRR in the Pampean region, Argentina. *Agric. For. Meteorol.* 79:157–82.
Kerr, Y.H. and J.P.Lagouarde (1989) On the derivation of land surface temperature from AVHRR data. EUM-P06, pp. 157–60.
Kerr, Y.H., J.Imbernon, G.Dedieu, O.Hautecoeur, J.P.Lagouarde, and B.Seguin (1989) NOAA/AVHRR and its uses for rainfall and evapotranspiration monitoring. *Int. J. Remote Sens.* 10(4 & 5): 847–54.
Kerr, Y.H., J.P.Lagouarde, and J.Imbernon (1992) Accurate land surface temperature retrieval from AVHRR data with use of an improved Split Window algorithm. *Remote Sens. Environ.* 41:197–209.
Kerr, Y.H., H.Makhmara, J.C.Meunier, and T.Valéro (1996). Ground Data v3 (08/95), (v3). H. CESBIO/CNES/ORSTOM, CD-ROM 3.
Kidwell, K.B. (1986) *AVHRR User's Guide.* US Department of Commerce, NOAA/NESDIS, Washington DC, 128 pp.
Kimes, D.S. and J.A.Kirchner (1983) Directional radiometric measurements of rowcrop temperatures. *Int. J. Remote Sens.* 4:299–311.
Kleespies, T.J. and L.M.McMillin (1990) Retrieval of precipitable water from observations in the Split-Window over varying surface temperatures. *J. Appl. Meteorol.* 29:851–62.
Kneizys F.X., E.P.Shettle, W.O.Gallery, J.H.Chetwynd Jr, L.W.Abreu, J.E.A.Selby, S.A.Clough, and R.W.Fenn (1983) Atmospheric transmittance/radiance: computer code Lowtran 6. Technical Report AFGL-TR-83–0187, Optical Physics Division, US AIR FORCE Geophysics Laboratory, Hanscom Air Force Base, MA, USA.
Lagouarde, J.P. (1991) Use of NOAA-AVHRR data combined with an agrometeorological model for evaporation mapping. *Int. J. Remote Sens.* 12 (9): 1853–64.
Lagouarde, J.P. (1997) SEXTET: a high resolution and high repetitivity infra red instrument on a mini satellite platform. Proposal to CNES call for ideas.
Lagouarde, J.P. and Y.H.Kerr (1985) On the calibration of the thermal NOAA/AVHRR sensor. ESA SP 248, p. 325.
Lagouarde, J.P. and K.J.McAneney (1992) Daily sensible heat flux estimation from a single measurement of surface temperature and maximum air temperature. *BoundLayer Meteorol.* 59: 341–62.
Lagouarde, J.P., Y.H.Kerr, and Y.Brunet (1995) An experimental study of angular effects on surface temperature for various plant canopies and bare soils. *Agric. For. Meteorol.* 77: 167–90.
Lagouarde, J.P., E.Pradel, and C.Sanz (1996) Estimation des flux de surface a l'échelle régionale. Actes de l'Ecole-Chercheurs INRA en Bioclimatologie. Départ. de Bioclimatologie/INRA Ed, pp. 361–82.
Lagouarde, J.P., K.J.McAneney, and A.E.Green (1996) Scintillometer measurements of sensible heat flux over heterogeneous surfaces. In Stewart, Engman, Feddes, and Kerr (eds) *Scaling up in Hydrology using Remote Sensing.* John Wiley, London.
Leprieur, C., Y.H.Kerr, and J.M.Pichon (1996) Critical assessment of vegetation indices

from AVHRR in a semi-arid environment. *Int. J. Remote Sens.* 17(13): 2549–63.

Li, Z.L. and F.Becker (1993) Feasability of land surface temperature and emissivity determination from AVHRR Data. *Remote Sens. Environ.* 43(1): 67–85.

May, D.A., L.L.Stowe, J.D.Hawkins, and E.P.McClain (1992) A correction for Saharan dust effects on Satellite sea surface temperature measurements. J. *Geophys. Res. 97* (C3): 3611–19.

McClain, E.P., W.G.Pichel, and C.C.Walton (1985) Comparative performance of AVHRR-based multichannel sea surface temperatures. *J. Geophys. Res.* 20: 11587–601.

McClatchey, R.A., R.W.Fenn, J.E.A.Selby, F.E.Volz, and J.S.Garing (1971) Optical properties of the atmosphere (Revised). Environmental Research Paper N° 311.910.8, Air Force Cambridge Research Laboratories, p. 24.

Mecikalski, J.R., G.R.Diak, M.C.Anderson, and J.M.Norman (1999) Estimating fluxes on a continental scale using remotely-sensed data in an atmospheric-land exchange model. *J. Appl. Meteorol.* 38: 1352–69.

Nerry, F., F.Petitcolin, and M.-P.Stoll (1998) Bidirectional reflectivity in AVHRR Channel 3: Application to a region in Northern Africa. *Remote Sens. Environ. 66* (3): 298–316.

Njoku, E.G. (1985) Satellite derived sea surface temperature workshop comparisons. *Bull. Am. Meteorol. Soc.* 66: 274–81.

Norman, J.M. and F.Becker (1995) Terminology in thermal infrared remote sensing of natural surfaces. *Agric. For. Meteorol. 77*: 153–66.

Norman, J.M., S.Castello, and L.K.Balick (1994) Directional infrared temperature and emissivity of vegetation: measurements and models. *Proc. VI, ISPRS Int. Symp. on Physical Measurements and Signatures in Remote Sensing.* CNES Ed., pp. 749–58.

Ottlé, C. and D.Vidal Madjar (1992) Estimation of land surface temperature with NOAA9 data. *Remote Sens. Environ.* 40:27–41.

Ottlé, C. and M.P.Stoll (1993) Effect of atmospheric absorption and surface emissivity on the determination of land surface temperature from infrared satellite data. *Int. J. Remote Sens.* 14:2025–37.

Ottlé, C., S.Outalha, C.François, and S.Le Maguer (1998) Estimation of total water vapour content from Split-Window radiance measurements. *Remote Sens. Environ.* 61:410–18.

Phalippou, L. (1996) Variational retrieval of humidity profile, wind speed and cloud liquid-water path with the SSM/I: Potential for numerical weather prediction. *Q. J. R. Meteorol. Soc.* 122:530.

Pinty, B. and M.M.Verstraete (1992) GEMI: a non linear index to monotor global vegetation from satellite. *Vegetatio* 101:15–20.

Prabhakara *et al.* (1974) Estimation of sea surface temperature from remote sensing in the 11 and 13mm window region. *J. Geophys. Res. 79*:5039–44.

Prata, A.J. (1993) Land surface temperatures derived from the AVHRR and the ATSR. Part 1 Theory. *J. Geophys. Res.* 98 (D9): 16689–702.

Prata, A.J. (1994) Land surface temperatures derived from the AVHRR and the ATSR. Part 2 Experimental results and validation of AVHRR algorithms. *J. Geophys. Res.* 99 (D6): 13025–58.

Prata, A.J., R.P.Cechet, I.J.Barton, and D.T.Llewellyn-Jones (1990) The Along Track Scanning Radiometer for ERS-1—Scan geometry and data simulation. *IEEE Trans. Geosci. Remote Sens.* 28:3–13.

Prata, A.J. and C.M.R.Platt (1991) Land surface temperature measurements from

the AVHRR. In *Proc. 5th AVHRR Data Users Conference,* June 25–28, Tromso, Norway, EUM P09, pp. 433–8.

Price, J.C. (1983) Estimating surface temperatures from satellite thermal infrared data. A simple formulation for the atmospheric effect. *Remote Sens. Environ.* 13: 353–61.

Price, J.C. (1984) Land surface temperature measurements from the Split Window channels on the NOAA 7 Advanced Very High Resolution Radiometer. *J. Geosphys. Res.* 89:7231–7.

Prince, S.D., Y.H.Kerr, J.P.Goutorbe, T.Lebel, A.Tinga, P.Bessemoulin, J.Brouwer, A.J.Dolman, E.T.Engman, J.H.C.Gash, M.Hoepffner, P.Kabat, B.Monteny, F.Said, P.Sellers, and J.Wallace (1995) Geographical, biological and remote sensing aspects of the hydrologic atmospheric pilot experiment in the Sahel (HAPEX-Sahel). *Remote Sens. Environ.* 51:215–34.

Qi, J., A.Chehbouni, A.R.Huete, Y.H.Kerr, and S.Sorooshian (1994) A Modified Soil Adjusted Vegetation Index: *Remote Sens. Environ.* 48 (2): 119–26.

Salisbury, J.W. and D.M.D'Aria (1992) Emissivity of terrestrial materials in the 8–14 mm atmospheric window. *Remote Sens. Environ.* 42: 83–106.

Salisbury, J.W. and D.M.D'Aria (1994) Emissivity of terrestrial materials in the 3–5 (μm atmospheric window. *Remote Sens. Environ.* 47: 345–61.

Savane, M., B.Seguin, and B.Guillot (1994) Suivi des conditions hydriques a l'échelle régionale a opartir de données satellitaires dans 1'infrarouge thermique. *Télédétection de l'environnement dans l'espace francophone,* F.Bonned, Presses de 1'université du quebec, pp. 345–58.

Schmugge, T.J., S.J.Hook, and C.Coll (1998) Recovering surface temperature and emissivity from thermal infra red multispectral data. *Remote Sens. Environ.* 65: 121–31.

Scott, N.A. and A.Chedin (1981) A fast line by line method foratmospheric absorption computations: the automatized atmospheric absorption atlas. *J. Clim. Appl. Meteorol.* 20: 801–12.

Seguin, B., E.Assad, J.P.Fréteaud, J.Imbernon, Y.H.Kerr, and J.P.Lagouarde (1989) Use of meteorological satellites for water balance monitoring in Sahelian regions. *Int. J. Remote Sens.* 10 (6): 1101–17.

Seguin, B., J.P.Lagouarde, and M.Savané (1991) The assessment of regional crop water conditions from meteorological satellite thermal infrared data. *Remote Sens. Environ.* 35:141–8.

Seguin, B., D.Courault, and M.Guerif (1994) Surface temperature and evapotranspiration: application of local scale methods to regional scales using satellite data. *Remote Sens. Environ.* 49: 287–95.

Sobrino, J.A., C.Coll, and V.Caselles (1991) Atmospheric correction for land surface temperature using NOAA-11 AVHRR Channels 4 and 5. *Remote Sens. Environ.* 38:19–34.

Sobrino, J.A., Z.L.Li, and M.P.Stoll (1993) Impact of the atmospheric transmittance and total water vapor content in the algorithms for estimating satellite sea surface temperatures. *IEEE Trans. Geosci. Remote Sens.* 31:946–52.

Sobrino, J.A., Z.L.Li, M.Ph. Stoll, and F.Becker (1994) Improvements in the Split-Window technique for land surface temperature determination. *IEEE Trans. Geosci. Remote Sens.* 32 (2): 243–53.

Stewart, J.B., E.T.Engman, R.A.Feddes, and Y.H.Kerr (1996) *Scaling up in Hydrology Using Remote Sensing.* John Wiley & Sons, Chichester, England.

Stewart, J.B., E.T.Engman, R.A.Feddes, and Y.H.Kerr (1998) Scaling up in Hydrology using remote sensing: summary of a workshop. *Int. J. Remote Sens.* 19 (1): 181–94.

Susskind, J., J.Rosenfield, D.Renter, and M.T.Chahine (1984) Remote sensing of weather and climate parameters from HIRS2/MSU on TIROS/N. *J. Geophys. Res.* 89:4677–97.

Takashima, T and K.Matsuda (1988) Averaged emissivities of Quartz and Sahara dust powders in the infra-red region. *Remote Sens. Environ.* 26:301–2.

Tucker, C.J. and P.J.Sellers (1986) Satellite remote sensing of primary production. *Int. J. Remote Sens.* 7:1395–416.

Ulivieri, C. and G.Cannizzaro (1985) Land surface temperature retrievals from satellite measurements. *Acta Astronautica* 12(12): 977–85.

Ulivieri, C., M.M.Castronouvo, R.Francioni, and A.Cardillo (1992) A SW algorithm for estimating land surface temperature from satellites. Presented at COSPAR, Washington DC, USA. [*Adv. Space Res.* 14 (3): 59–65.]

Van de Griend, A.A.and M.Owe (1993) On the relationship between thermal emissivity and the normalized difference vegetation index for natural surfaces. *Int. J. Remote Sens.* 14 (6): 1119–31.

Valor, E. and V.Caselles (1996) Mapping land surface emissivity from NDVI: application to European, African, and South American Areas. *Remote Sens. Environ.* 57:167–84.

Vasquez, D.P., F.J.Olmo Reyes, and L.Alados Arboledas (1997) A comparative study of algorithms for estimating Land surface temperature from AVHRR data. *Remote Sens. Environ.* 62:215–22.

Vidal, A (1991) Atmospheric and emissivity correction of land surface temperature measured from satellite using ground measurements or satellite data. *Int. J. Remote Sens.* 12:2449–60.

Westwater, E.R., J.B.Snider, and M.J.Falls (1990) Ground based radiometric observations of atmospheric emission and attenuation at 20.6, 31.65, and 90 GHz: A comparison of measurements and theory. *IEEE Trans. Antennas Prop.* AP-38: 1569–80.

Chapter 3

High spatial resolution mapping of surface energy balance components with remotely sensed data

Karen Humes, Ray Hardy, William P.Kustas, John Prueger and Patrick Starks

3.1 Introduction

3.1.1 Background

In order to better understand the exchange of heat and moisture between the land surface and lower atmosphere, it is important to quantify the components of the surface energy balance in a distributed fashion at the landscape scale. Remotely sensed data can provide spatially distributed information on a number of key land surface characteristics and state variables that control the surface energy balance. When combined with near-surface meteorological measurements and a relatively simple model, satellite and aircraft-based remotely sensed data can be used to create "maps" of spatially distributed surface energy balance components over a watershed. Assuming no advection of energy into an area, the simplest form of the surface energy balance is given by

$$R_{\text{net}} = G + H + LE \tag{3.1}$$

where R_{net} refers to the net radiation balance, G refers to the soil heat flux (i.e. the energy used to warm the near-surface soil layers), H refers to the sensible heat flux (the energy used to transfer heat from the surface to the atmosphere), and LE refers to the latent heat flux (the energy used to transfer water vapor from the surface to the atmosphere).

The influence of the land surface energy fluxes on regional and global atmospheric processes has become well recognized in the climate and meteorological modeling communities (e.g. Avissar and Pielke 1989; Chen and Avissar 1994; Betts *et al.* 1996). This has given rise to the development of quite a number of more sophisticated parameterizations for simulating land surface processes within mesoscale and global atmospheric models (Dickinson *et al.* 1986; Sellers *et al.* 1986; Entehkhabi and Eagleson 1989; Noilhan and Planton 1989; Avissar and Verstraete 1990; Xue *et al.* 1991).

The use of these schemes within atmospheric models has helped to improve the performance of both regional and mesoscale atmospheric models. However, most of these models, referred to as soil-vegetation-atmosphere transfer (SVAT) models, require a priori knowledge of a considerable number of surface parameters and detailed information for initialization. They also require pertinent ground data and substantial human effort for model calibration. Additionally, when complex point-scale models are run within the context of mesoscale or global atmospheric models, the grid cell resolution is generally on the order of hundreds to thousands of meters in size. Many of the key parameters and variables in the complex physically based models would be expected to vary considerably within grid cells of that size.

3.1.2 Objectives of this study

The primary objective of this study is to demonstrate the feasibility of using high spatial resolution remotely sensed data, combined with driving meteorological data from a ground network and a relatively simple model, to compute spatially distributed values of surface energy balance components. The model employed here is a relatively simple "snapshot" model. That is, it does not simulate any of the processes as a function of time; rather, it uses satellite and ground data to estimate the fluxes at the time of the satellite overpass. Almost all the model parameters and variables used by the model (such as surface temperature, land cover type, and vegetation density) are estimated from remotely sensed data. The meteorological inputs required by the model were derived from a ground network. This approach has the advantage of being very "data driven" and the model does not need to be calibrated or "tuned" for a particular site. Thus, the fluxes estimated from this approach can be useful for validation or assimilation into more complex simulation models.

The model was applied on a pixel-by-pixel basin across a watershed in a sub-humid climate zone. Although surface fluxes have been previously mapped using these types of approaches (Moran *et al.* 1990; Holwill and Stewart 1992; Humes *et al.* 1997), this study represents the application of a more complex (two-layer) model over more heterogeneous land cover types than these previous efforts. Additionally, the watershed studied here has a special instrumentation network that makes possible more detailed spatial analysis of the factors influencing the surface fluxes. The motivation for applying this model at relatively high spatial resolution over the watershed is twofold: (a) at higher spatial resolution the approach is more easily validated using ground-based point measurements and (b) mapping the fluxes at high spatial resolution allows an evaluation of the relative importance of various surface face and atmospheric variables in determining the surface face fluxes.

3.2 Study area

The USDA/Agricultural Research Service (ARS) Little Washita Watershed (LWW), operated by the ARS Grazinglands Research Station, is located in central Oklahoma. The land cover types present in the watershed include a mixture of cultivated areas (primarily winter wheat, soybeans, alfalfa, and corn), pastures with native grasslands and non-native species, varied management practices, and (depending considerably on climatological variables that vary considerably from east to west) wooded areas. The LWW has also been the site of several special experimental campaigns involving the simultaneous acquisition of ground and remotely sensed data. The watershed was a US Supersite for the SIR-C (Shuttle Imaging Radar) Experiments in 1992 and 1994. The SIR-C experiments became the focal point for one field campaign in 1992 and three field campaigns in 1994 which included many different ground measurements, as well the acquisition of many types of remotely sensed data from ground, aircraft, and satellite-based sensors (Jackson and Scheibe 1993; Starks and Humes 1996). Remotely sensed data sets included passive microwave, active microwave, and optical sensors.

Among the many special ground observations acquired during these campaigns were the measurement of surface energy fluxes by Bowen ratio and eddy correlation techniques (Prueger 1996; Kustas *et al.* 1999). These ground-based measurements were used for validation of the surface energy fluxes produced by this modeling effort. Observations also included groundbased radiometric measurements of surface reflectance and temperature. These were acquired with a backpack-type apparatus that facilitated the acquisition of ground data over a large, relatively uniform target area at the time of the Landsat satellite overpass. These data were used to validate the atmospherically corrected radiometric surface temperatures derived from satellite data. Additionally, the ARS operates the Micronet network in the LWW, which consists of 42 monitoring stations on a 5-km grid. These stations record meteorological variables such as incoming solar radiation and near-surface (1.8 m) air temperature and relative humidity. These measurements were used for meteorological input to the model.

The data sets used in this analysis were from the August 1994 field campaign on the CWW. A false color composite image from Landsat 5 Thematic Mapper (TM) data acquired on August 18, 1994, is shown in Figure 3.1. In this image, the data from the TM band 4 (near-infrared) are displayed as red, data from the TM band 3 (red) are displayed as green, and data from the TM band 2 (green) are displayed as blue. In August, the winter wheat fields are typically bare and thus appear bluish green on the false color composite image. It can be observed from the image that these areas are most extensive in the western portion of the watershed. The bright red areas of the image correspond to riparian vegetation along drainage areas, the relatively small watershed area corresponding to cultivated crops that

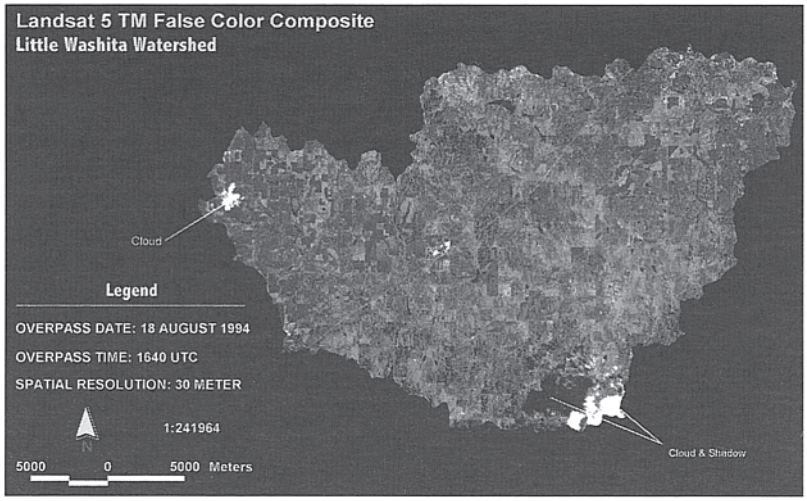


Figure 3.1 False color composite image from the Landsat TM sensor for the LWW from August 18, 1994 (see Colour Plate XII).

are green at this time of year (such as corn and alfalfa), and, to a lesser degree, the spatially extensive pastures of various density and species composition.

In the early morning hours of August 18, a relatively intense thunderstorm moved through the watershed. The cumulus clouds that can be seen in Figure 3.1, and the cirrus cloud contamination over a portion of the watershed evident in the thermal band, were remnants from that storm. The system moved out of the watershed region approximately 1 h before the image was acquired.

3.3 Model description and implementation

3.3.1 Model description

The model utilized here is described in detail in Norman *et al.* (1995). It is a two-source model, meaning that separate energy balance computations are done for the soil and vegetation layers of the surface. It was run on a pixelby-pixel basis to compute spatially distributed energy fluxes over the LWW during the time of the Landsat TM overpass during the August 1994 field campaign. A diagram of model inputs and outputs is shown in Figure 3.2. The conceptual model formulation is summarized here.

The four components of net radiation are quantified as follows: (a) incoming solar radiation is a model input typically provided by ground

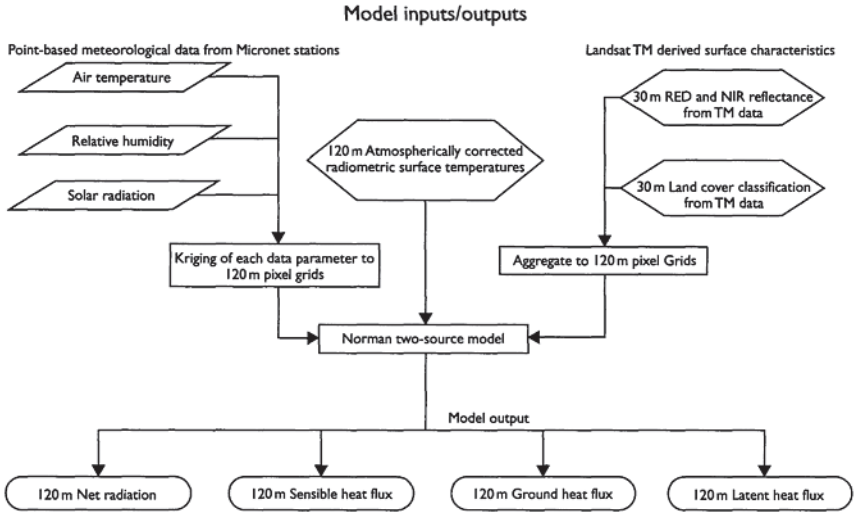


Figure 3.2 Conceptual diagram of the input and output quantities used for the application of the Norman *et al.* (1995) model to data from the August 18, 1994, Landsat scene over the LWW.

measurements; (b) outgoing solar radiation is computing using incoming solar radiation and assumed values of surface albedo f or dif ferent land cover types; (c) incoming longwave radiation is estimated using ground-based measurements of air temperature and relative humidity and an empirical expression for clear sky conditions (Idso 1981); (d) outgoing longwave radiation is computed using the surface temperature from the satellite data and an assumed emissivity of 0.98. It should be noted that for some "snapshot"type models for estimating fluxes, surface albedo is calculated using empirical functions that relate surface hemispherical albedo to reflectance in the finite wavebands of the Landsat TM sensor. This approach was not utilized in this application because of uncertainty in the atmospheric correction of the satellite data to absolute surface reflectance.

The net radiation at the surface is partitioned between the soil and vegetation layers using a typical "Beers law" formulation. The exponent in this relationship is controlled by an estimate of the fractional vegetation cover (which is estimated from remotely sensed data in the manner described in more detail below), and an assumption of spherical leaf inclination angle distribution. Soil heat flux is assumed to be a constant fraction (0.35) of the net radiation reaching the soil.

The total sensible and latent heat fluxes are simply taken to be the sum of the vegetation and soil contributions. Those contributions are determined

by doing a separate surface energy balance on the soil and vegetation layers and assuming that the flux of heat from the soil and vegetation layers act in parallel. This gives rise to a simpler resistance formulation than some multi-layer models (e.g. Shuttleworth and Wallace 1985; Xue *et al.* 1991), but several studies have shown that for low to moderate vegetation cover, the various levels of complexity in resistance networks are indistinguishable because air temperature gradients are small in the upper canopy (Norman and Campbell 1983). The key to estimating both contributions to the sensible heat flux is in the decomposition of the radiometric surface temperature (T_{rad}), derived from satellite observations, into soil and vegetation components using

$$T_{rad}(\theta) = [f(\theta)T_c^n + (1 - f(\theta))T_s^n]^{1/n} \tag{3.2}$$

where T_c is the vegetation canopy temperature, T_s is the soil surface temperature, n is the power of the temperature and approximates the blackbody function when considered over the entire wavelength of the sensor, θ is the view angle of the sensor, $f(\theta)$ is the fraction of the field of view of the radiometer occupied by canopy and is given by

$$f(\theta) = 1 - \exp\left(\frac{-0.5F}{\cos\theta}\right) \tag{3.3}$$

where F is the leaf area index.

At the value of θ=0:

$$f_c = 1 - \exp(-0.5F) \tag{3.4}$$

where f_c is the fractional vegetation cover.

The component surface temperatures and the turbulent flux components for the soil and vegetation layers are derived using a series of steps that sometimes require iteration. In the following equations, the symbols $R_{net,c}$, H_c, and LE_c refer to the canopy portion of the net radiation, sensible, and latent heat fluxes, respectively, and the symbols $R_{net,s}$, H_s, and LE_s refer to the soil contribution to the net radiation, sensible, and latent heat fluxes. First, an approximation for the canopy portion of the latent heat flux is estimated using a Priestly and Taylor (1972) type formulation with the canopy portion of the net radiation:

$$LE_c = 1.26 f_g \left(\frac{s}{s+\gamma}\right) R_{net,c} \tag{3.5}$$

where f_g is the fraction of the vegetation cover which is green, s is the slope of the saturation vapor pressure versus temperature curve, γ is the psychrometric constant (0.66 kPa C^{-1}).

The sensible heat flux of the canopy layer is then computed as the residual in the energy balance for the canopy layer:

$$H_c = R_{net,c} - LE_c \tag{3.6}$$

The canopy temperature is then estimated by inverting the equation for a simple resistance formulation for sensible heat flux from the canopy:

$$H_c = (T_c - T_{air})/r_{ah} \tag{3.7}$$

where T_c is the surface temperature of the canopy, T_{air} is the near-surface air temperature and r_{ah} is the aerodynamic resistance. The formulation for the aerodynamic resistance is derived from the diabatically corrected log temperature profile equations (Brutsaert 1982). The roughness lengths used in the calculation of aerodynamic resistance were set according to the land cover type as shown in Table 3.1.

Using this approximation for T_c and the satellite measurement of T_{rad}, equation (3.2) is used to solve for T_s, the soil temperature. This value of T_s is used to calculate the soil contribution to sensible heat flux using a bulk resistance formulation for the soil layer, given by

$$H_s = \rho C_p (T_s - T_{air})/(r_{ah} + r_s) \tag{3.8}$$

where r_s is the soil-surface resistance as derived in Norman *et al.* (1995). The soil component of latent heat flux is then computed as the residual in the soil energy balance:

$$LE_s = R_{net,s} - G - H_s \tag{3.9}$$

If the soil evaporation which results from this calculation is less than zero, then LE_s is set equal to zero and H_s is recomputed using equation (3.9), T_s is

Table 3.1 Roughness length (Z_0), canopy height (h), and albedo for each land cover type

Land cover	*Roughness length* (m)	*Canopy height* (m)	*Albedo*
Water	0.00035	—	0.10
Urban	0.25	—	0.25
Woodland	0.625	5	0.15
Mixed crops	0.0125	0.1	0.20
Pasture – dense	0.0625	0.50	0.15
Pasture – moderate density	0.0625	0.5	0.15
Pasture – less dense	0.0625	0.5	0.175
Sparse or senescent	0.0125	0.1	0.20
Bare soil and wheat stubble	0.01	0.1	0.15

recomputed using equation (3.8), new values of T_c and H_c are computed using equations (3.2) and (3.7), respectively, and a new value of LE_c is computed using equation (3.6).

One advantage of this model formulation over other resistance-based formulations and more complicated SVAT schemes is that it does not require an estimate of the canopy surface resistance to evaporation. Since this quantity is highly spatially variable, very dynamic in time, and not readily obtained from remotely sensed data, a formulation that can reliably estimate surface fluxes without the use of this parameter can be more readily applied to new areas and larger spatial scales.

3.3.2 Inputs derived from ground data

As discussed above, the meteorological inputs required for the data include: incoming solar radiation, near-surface air temperature, relative humidity, and windspeed. Spatially distributed values for the near-surface air temperature (1.8 m above the surface) and incoming solar radiation are shown in Figures 3.3 and 3.4, respectively. These maps were derived using point data from the USDA/ARS Micronet network of 42 stations located across the watershed, shown on the maps. The point data correspond to the data from the 5-min averaging interval closest in time to the satellite overpass

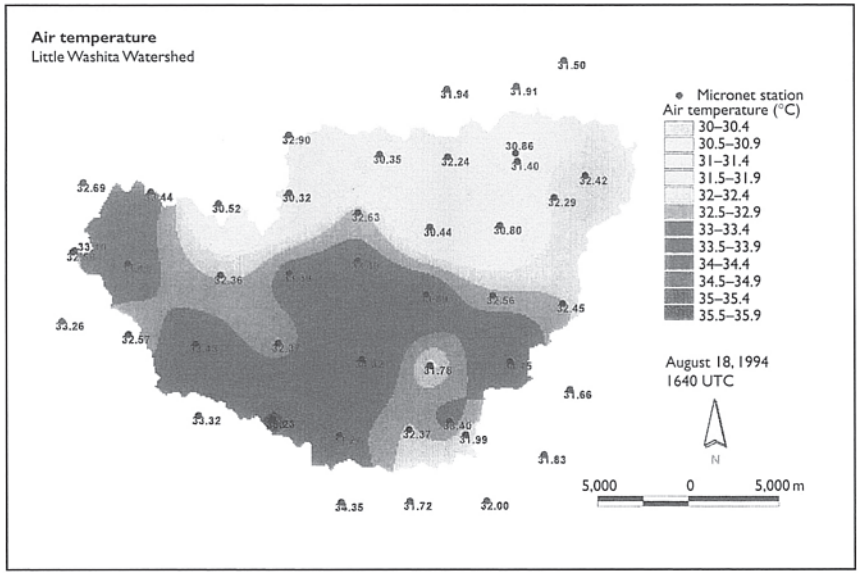


Figure 3.3 Gridded field of air temperature 2 m above the surface interpolated from measurements at Micronet stations (see Colour Plate XIII).

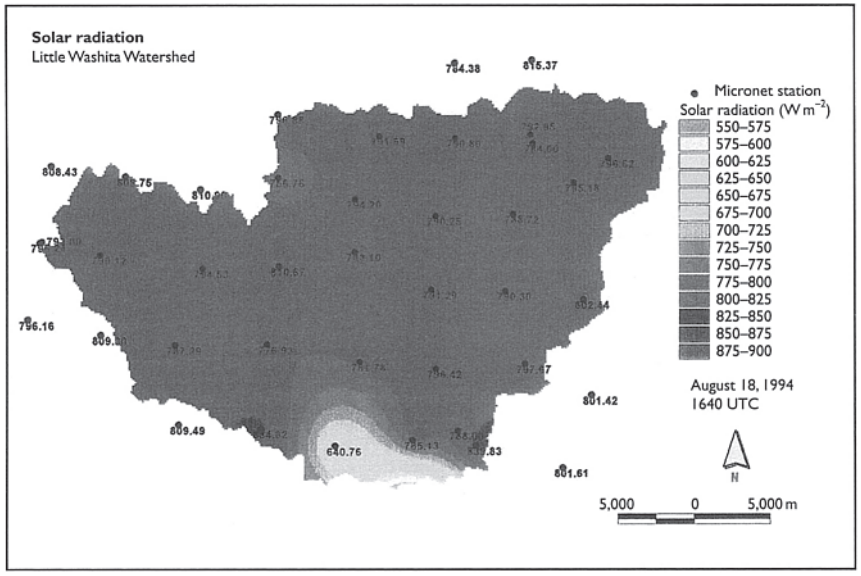


Figure 3.4 Gridded field of incoming solar radiation measurements interpolated from measurements at Micronet stations (see Colour Plate XIV).

time of approximately 1640 UTC. A universal kriging algorithm was used for spatial interpolation between the point measurements.

The Micronet does not include measurements of windspeed, which are required for model calculations of aerodynamic resistance. To obtain a reasonable watershed-wide average value of windspeed, data from four Oklahoma Mesonet stations were used. The Oklahoma Mesonetwork is a state-wide monitoring network consisting of 112 stations that provide measurements of meteorological and surface variables at 5-min intervals. Four of the Mesonet stations are located just outside the boundaries of the watershed. Values of the windspeed (at 9 m above the surface) and relative humidity from these four stations were averaged to compute a watershedwide average for these variables for the 5-min interval closest to the satellite overpass time.

3.3.3 Inputs derived from remotely sensed data

Radiometric surface temperature

One of the key inputs to the model is the radiometric surface temperature, in this case derived from TM Band 6 (bandpass 10.9–12.3 μm). Data from the Landsat thermal band were corrected for atmospheric effects by running the

radiative transfer code Lowtran 7 (Kniezys *et al.* 1988). Atmospheric temperature and water vapor profiles from on-site radiosonde data acquired by a team from the National Severe Storms Lab at the time of the satellite overpass was used as input to the radiative transfer model. The resulting corrections were applied on a pixel-by-pixel basis across the scene. The radiometric temperature of a large ground target area was measured at the time of the satellite overpass with instruments mounted on two backpack type apparatuses. The satellite pixels that most closely corresponded to this large target area were extracted from the scene and compared with the ground-based temperature measurement. The TM-derived temperature was slightly higher than the ground-based temperature (approximately 1.5°C). The ground radiometric measurements and radiosonde measurements were made just adjacent to one another at a site near the center of the watershed.

The map of surface temperature for the watershed is shown in Figure 3.5. The cool areas in the east-central portion of the image correspond to contamination by cirrus clouds, and the cool spots in the far southern and western portions of the watershed correspond to cumulus clouds and shadows.

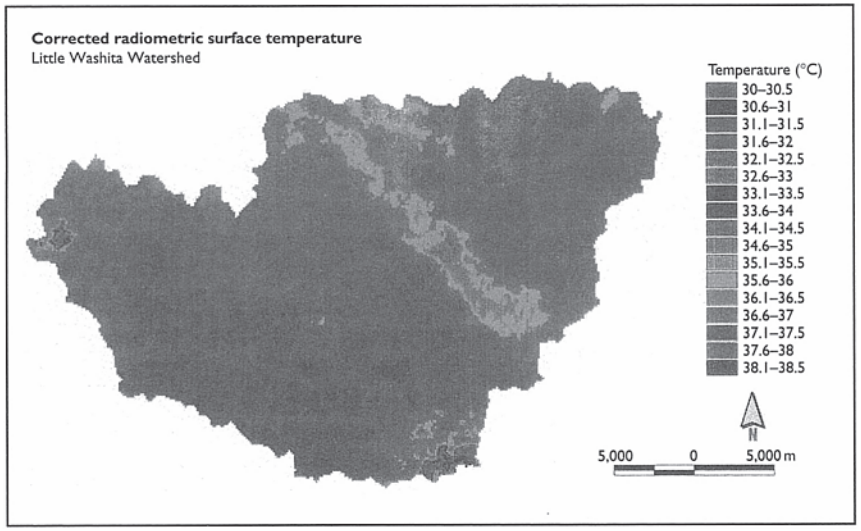


Figure 3.5 Atmospherically corrected radiometric surface temperature derived from a Landsat 5 TM scene acquired over the Little Washita Watershed on August 18, 1994. The dark areas in the east-central portion of the image corresponds to contamination by cirrus clouds.and the dark spots in the far southern and western edges of the watershed correspond to contamination by cumulus clouds (see Colour Plate XV).

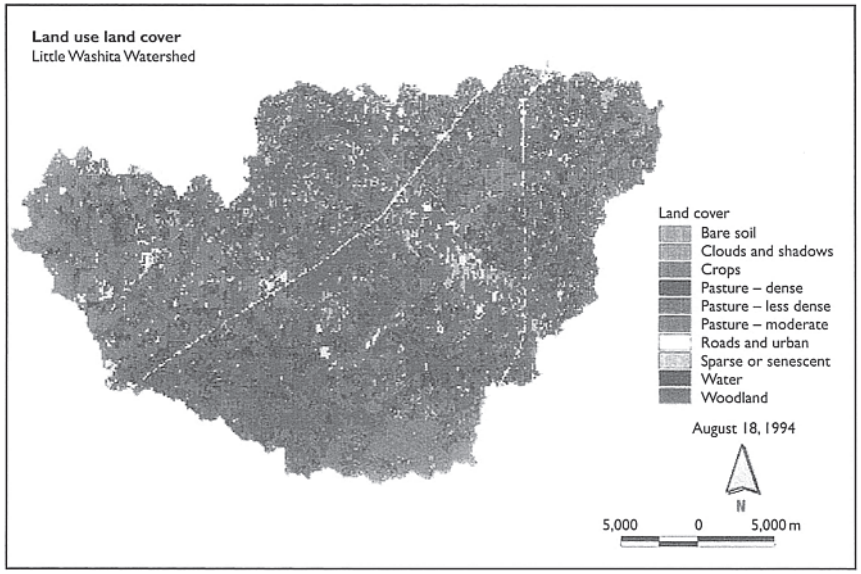


Figure 3.6 Land cover map derived from the unsupervised classification of data from six of the Landsat TM bands from the August 18, 1994, image (see Colour Plate XVI).

Land cover type

Information on land cover type across the watershed was derived from an unsupervised classification run on the six solar reflectance bands of the Landsat scene. The classes were identified and merged based on approximately 15 sites for which vegetation characteristics and density were known. A separate set of 197 points of known land cover type were used to assess the accuracy of this classification, and 81 % of these points were accurately classified. The land cover map was derived at the original 30-m pixel resolution for the TM reflective bands, then aggregated to 120-m resolution to match the resolution of the thermal band data. The aggregation procedure assigned the land cover type that occupied the majority of the area of the 120-m pixels. The resulting map is shown in Figure 3.6. This land cover map was used to assign a number of surface characteristics for individual pixels, namely the albedo, canopy height, and surface roughness. The values of these parameters used in this analysis are given in Table 3.1.

Vegetation cover

The data from Landsat TM Band 3 (0.63–0.69 μm) and Band 4 (0.76–0.90 μm) were converted to exoatmospheric reflectance using the calibration

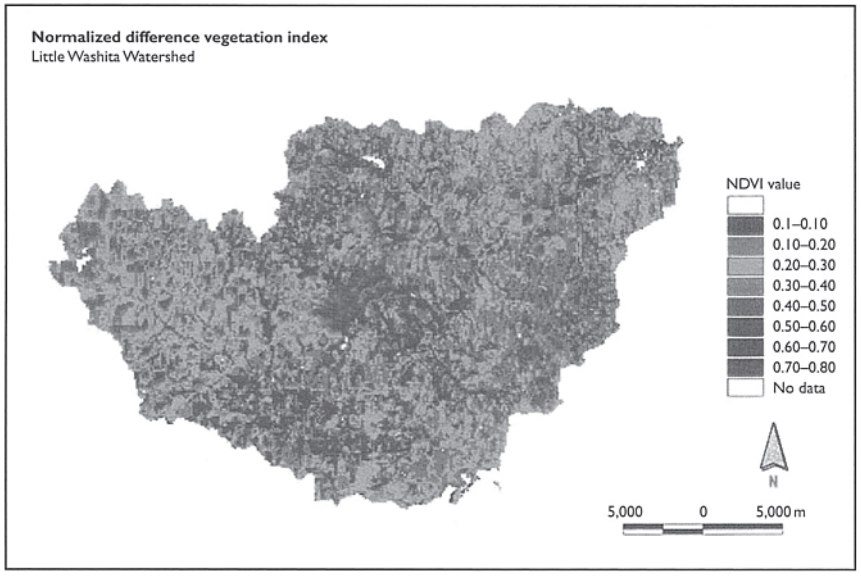


Figure 3.7 NDVI derived from the August 18, 1994, TM image (see Colour Plate XVII).

coefficients and solar irradiance data given by Markham and Barker (1986) and aggregated (via a simple average) to 120-m spatial resolution. These values were used to compute the Normalized Difference Vegetation Index (NDVI) for each pixel. The resulting map is shown in Figure 3.7. These NDVI values were used to estimate fractional cover for each pixel using the method of Carlson and Ripley (1997). They suggest using a "normalized" NDVI value, N*, given by

$$N^* = (\mathrm{NDVI}_i - \mathrm{NDVI}_{\min})/(\mathrm{NDVI}_{\max} - \mathrm{NDVI}_{\min}) \tag{3.10}$$

where NDVI_i is the value of NDVI for a given pixel, and $\mathrm{NVDI}_{\min}$ and $\mathrm{NDVI}_{\max}$ are the values of NDVI observed in the TM scene for pixels over bare soil and full vegetation cover, respectively. These values were selected by examination of the histogram of NDVI values for the watershed and set to be 0.04 and 0.60, respectively. The bands of TM data used to compute N^* were not corrected for atmospheric effects. However, Carlson and Ripley (1997) showed, through simulations with an atmospheric radiative transfer mode, that N^* is insensitive to atmospheric effects. The NDVI was used to compute f_c, the fractional vegetation cover parameter needed by the model, using the relationship obtained independently by both Choudhury *et al.* (1994) and Gillies and Carlson (1995):

$$f_c = (N^*)^2 \tag{3.11}$$

The value of leaf area index, F (also used by the model), was calculated using these values of f_c and equation (3.4).

3.4 Results and discussion

3.4.1 Validation of model flux estimates

As mentioned before, during the August 1994 Little Washita campaign, there were ground measurements of surface energy balance components at four sites across the watershed with Bowen ratio and eddy correlation instrumentation (Prueger 1996; Kustas *et al.* 1999). Three of the four sites were located on pastures of various species composition and vegetation density; one of the sites (#4) was located in a bare soil field. The model input quantities for these sites were extracted from the input data sets in order that the model estimates of surface energy balance components could be computed and compared with these measured values. The results of that comparison are shown in Table 3.2.

Data were acquired at the ground flux sites over an averaging interval of 30min, so the measured values shown in Table 3.2 represent the values of the energy balance components for the 30-min period bracketing the satellite overpass time. The model estimates are essentially an "instantaneous" estimate of the fluxes at the overpass time. Because of the storm system that moved through the area earlier in the morning, atmospheric conditions were rather dynamic in the time period just before the satellite overpass. These differences in integration time for the measured and modeled flux estimates should be taken into consideration in evaluating the results shown in Table 3.2. The site that has the poorest comparison between measured and modeled fluxes is the bare soil field (#4). It appears that the model overestimated the soil heat flux rather substantially, and since the soil component of LE (which is the only component of LE for this site) is computed as the residual in the

Table 3.2 The values of modeled surface fluxes and observed values at ground stations for the four components of the surface energy balance. The four components shown are R_{net} (net radiation), G (soil heat flux), H (sensible heat flux), and LE (latent heat flux). The observed values at ground stations are noted as (obs) and modeled values are noted as (mod). All values are in W m^{-2}

Site	R_{net} (obs)	R_{net} (mod)	G (obs)	G (mod)	H (obs)	H (mod)	LE (obs)	LE (mod)
1	582	566	46	80	132	83	404	402
2	528	541	82	93	100	49	346	399
3	527	550	77	91	90	44	360	415
4	603	591	74	148	14	34	515	409
RMS	16.57		41.68		43.35		65.33	

energy balance for the soil, all the errors in the other three components affect the model value of *LE*. Given that the measurements of the soil heat flux and turbulent fluxes would be expected to have an uncertainty of at least 30W m^{-2}, the agreement between the measured and modeled fluxes at the other sites appears to be very reasonable.

3.4.2 Spatially distributed fluxes

The maps of spatially distributed values for the net radiation (R_{net}), soil heat flux (G), sensible heat flux (H), and latent heat flux (LE) components of the energy balance are shown in Figures 3.8, 3.9, 3.10, and 3.11, respectively. The areas of contamination by cumulus and cirrus clouds are displayed in black and were not included in the calculations of median flux values and correlations described below.

In interpreting the spatial patterns observed in the fluxes, it is important to keep in mind the surface and meteorological conditions over the watershed at the time of overpass. Specifically, the image was acquired *very shortly* after a significant precipitation event. The precipitation totals over the watershed for approximately 9 h preceeding the satellite overpass are shown in Figure 3.12. There were no substantial precipitation events over the watershed for many days prior to this event. The isohyets shown in Figure 3.12 were derived from the precipitation measurements at the 42 Micronet

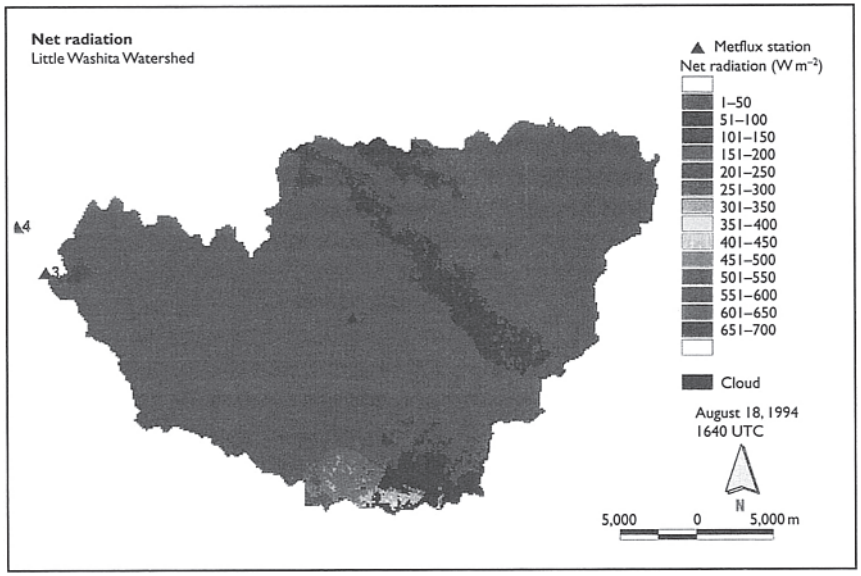


Figure 3.8 Map of net radiation (R_{net}) over the watershed computed with the Norman *et al.* (1995) model and Landsat TM data (see Colour Plate XVIII).

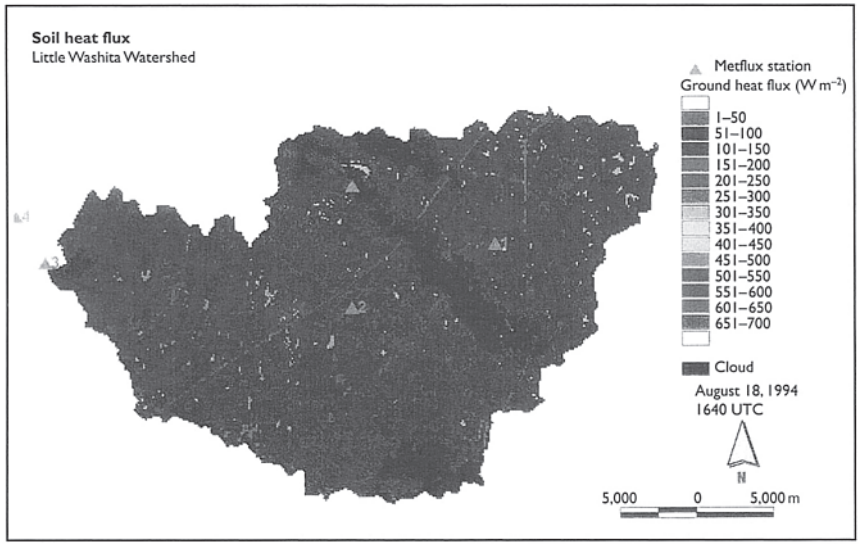


Figure 3.9 Map of soil heat flux (*G*) over the watershed computed with the Norman *et al.* (1995) model and Landsat TM data (see Colour Plate XIX).

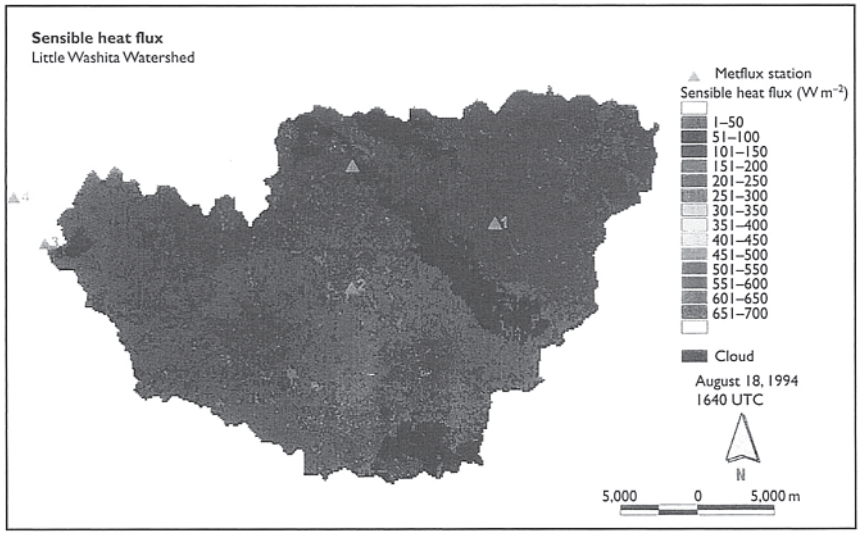


Figure 3.10 Map of sensible heat flux (*H*) over the watershed computed with the Norman *et al.* (1995) model and Landsat TM data (see Colour Plate XX).

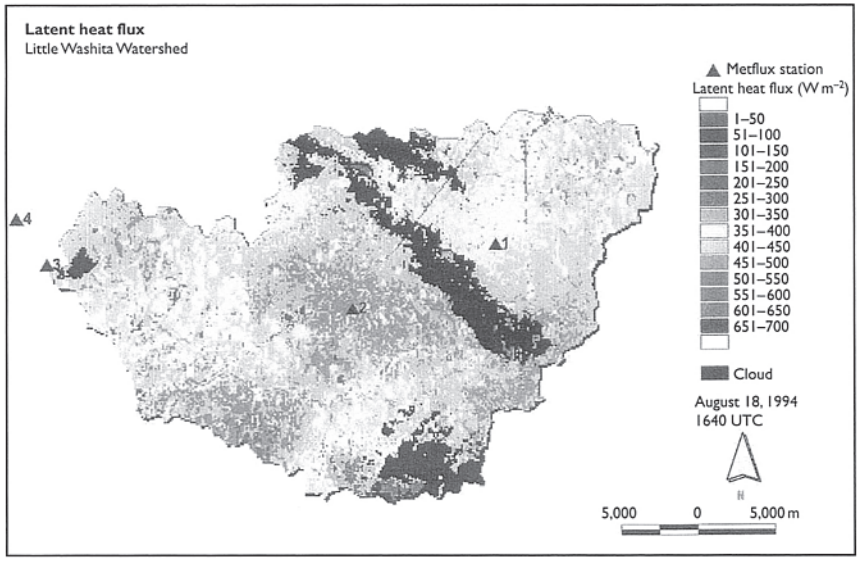


Figure 3.11 Map of latent heat flux (*LE*) over the watershed computed with the Norman *et al.* (1995) model and Landsat TM data (see Colour Plate XXI).

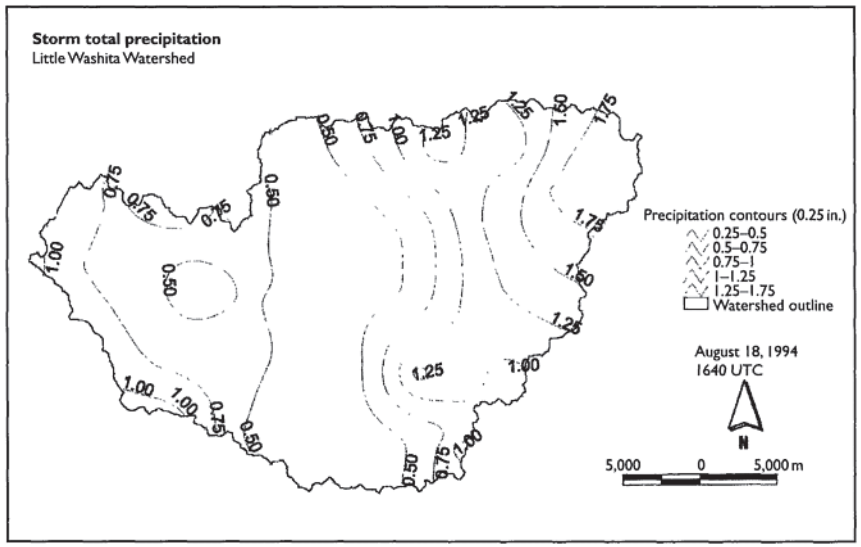


Figure 3.12 Isohyet map showing values of total precipitation from storm that occurred just prior to the acquisition of the Landsat TM image. Isohyets were derived from measurements at Micronet stations.

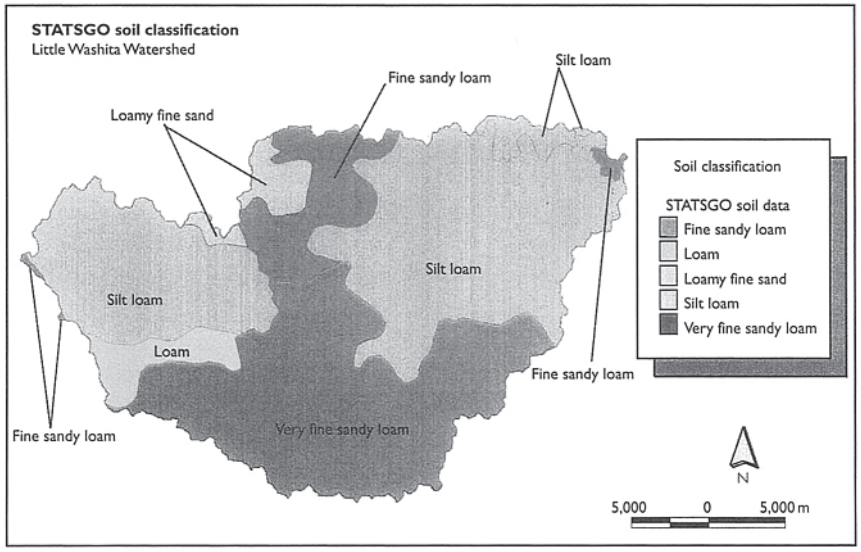


Figure 3.13 Map of surface soil texture from STATSGO database (see Colour Plate XXII).

stations across the watershed. The rainfall totals were highest in the eastern third and extreme western edge of the watershed.

The texture classes for the surface soils over the watershed, as derived from the STATSGO database are shown in Figure 3.13. The precipitation and soils data shown in Figures 3.12 and 3.13 were not used in the calculation of the fluxes; they are provided here to aid in the interpretation of the flux maps.

To further assist in the interpretation of the flux maps, the median flux values observed within each land cover type are shown in Figure 3.14. The numerical data corresponding to these plotted medians, together with the standard deviation among all the pixels belonging to a particluar land cover type, are summarized in Table 3.3.

Both the flux maps and the data shown in Figure 3.14 indicate that, overall, there were not major variations in the fluxes across the watershed during the time of the image acquisition. This is most likely due to the fact that the image was acquired immediately after a substantial rainfall event. The saturated conditions across the watershed, combined with minimal radiation loading that occurred the morning before the satellite data were acquired, would tend to minimize spatial variation among the factors controlling the energy balance. For example, the surface temperature map shown in Figure 3.5 does not show a discernible difference between the bare soil fields that dominate the western portions of the watershed and the vegetated surfaces

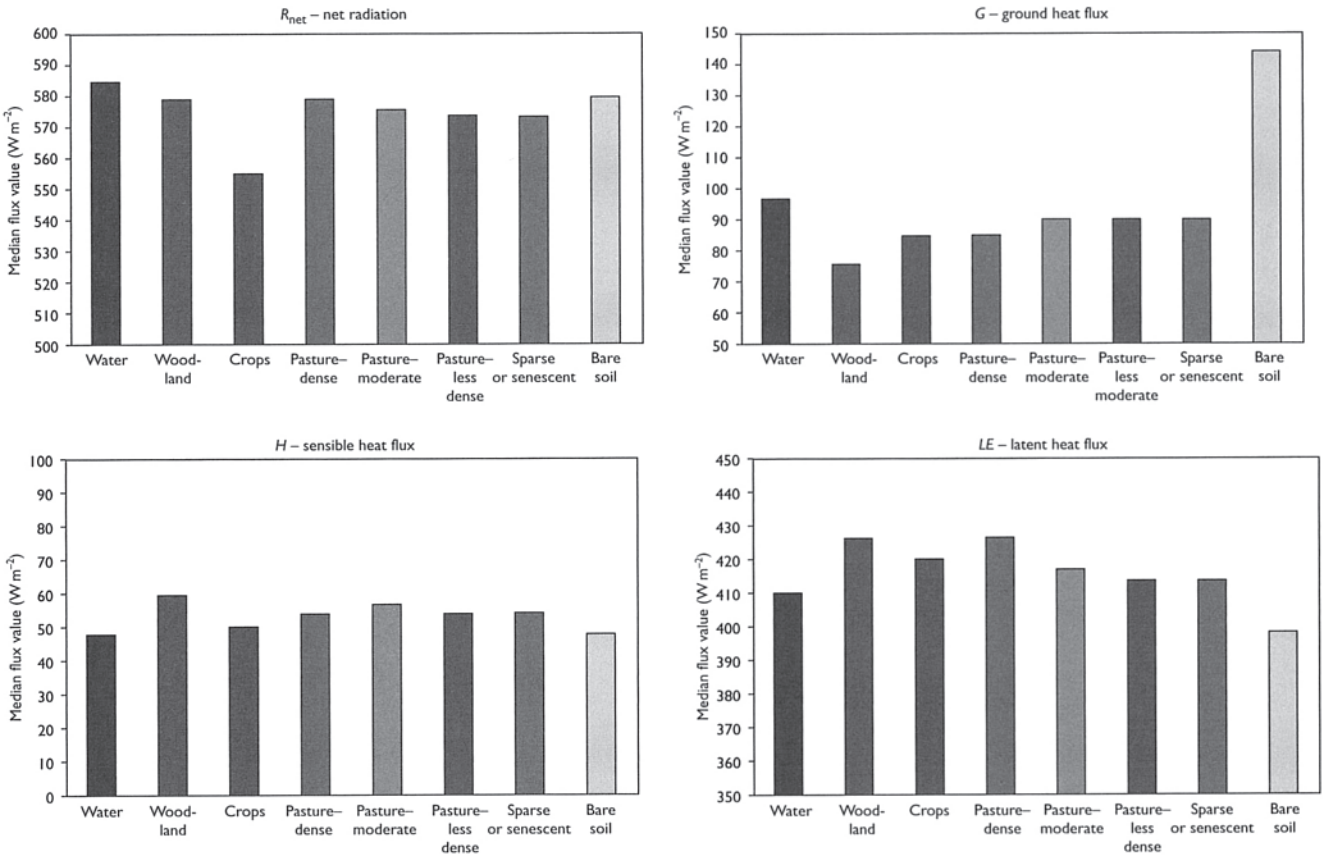


Figure 3.14 Median values of surface energy balance components for different land cover types.

Table 3.3 Standard deviation (STD) and median of flux values by land cover type observed within each land cover classification. The four components of the surface energy balance shown are R_{net} (net radiation), *G* (soil heat flux), *H* (sensible heat flux), and *LE* (latent heat flux). All values are in units of W m^{-2}

Land cover	*Pixel count*	*Median* R_{net}	*STD* R_{net}	*Median G*	*STD G*	*Median H*	*STD H*	*Median LE*	*STD LE*
Water	571	585	28.6	96	95.7	48	23.3	411	68.8
Woodland	8,590	579	29.9	75	44.9	60	41.6	426	51.5
Crops	6,814	555	27.2	84	42.3	51	22.4	420	47.3
Pasture – dense	3,705	579	25.7	84	38.3	54	25.7	426	44.0
Pasture – moderate	2,507	576	23.6	90	38.9	57	22.1	417	41.1
Pasture – less dense	3,893	573	28.1	90	39.0	54	20.9	414	42.7
Sparse or senescent	1,497	573	25.5	90	41.4	54	23.4	414	43.8
Bare soil	8,648	579	30.9	144	40.4	48	16.7	399	44.1

in other parts of the watershed, or major differences arising from the different precipitation totals observed across the watershed.

The map of sensible heat and latent heat flux appear to show a slight pattern of higher sensible heat flux (and therefore lower latent heat flux) in the portion of the watershed that received the greatest precipitation. This trend appears to be contrary to the expectation that higher precipitation values would yield lower sensible heat flux values as observed by Humes *et al.* (1997) for a more sparsely vegetated watershed. However, the turbulent heat fluxes are influenced by atmospheric factors as well as surface factors, particularly during the time period immediately after a precipitation event. The conditions under which this satellite image was acquired were very unusual, in that some of the factors controlling energy balance were more spatially variable in the near-surface atmosphere than they were on the land surface.

The dynamic and spatially variable atmospheric conditions that existed at the time of the overpass are manifested in the map of near-surface air temperature shown in Figure 3.3. The data shown in this figure indicate that there was a difference in air temperature over the watershed of more than 2°C, with the northeastern portion of the watershed (the portion that received more precipitation) being cooler than the southwestern portion of the watershed. Comparison of the air temperature map and the surface temperature map shows that the conditions at the time of the experiment were very unusual in that there was more variation in air temperature across the watershed than there was in surface temperture. The cooler air temperatures in the northeastern portion of the watershed account for one of the only clearly discernible spatial patterns in the sensible heat flux, which tended to be higher in the northeastern portion of the watershed (due to a

Table 3.4 Correlation coefficients for each input model grid and output flux map

Norman Model Input Grids (18 August 1994)	*Surface Energy Flux Map*			
	R_{net}	*G*	*H*	*LE*
Corrected radiometric surface temperature	−0.106	−0.006	−0.066	−0.020
Land cover	0.004	0.263	−0.157	−0.157
NDVI	−0.065	−0.394	0.222	0.206
Solar radiation	0.577	0.113	−0.078	0.284
Air temperature	0.034	−0.037	−0.316	0.238
Storm total precipitation	0.183	0.102	0.039	−0.009

slightly larger T_s - T_{air} values). The latent heat flux was conversely slightly lower in the northeastern portion of the watershed.

Correlation coefficients were computed between the four flux components and each of the data layers that served as input into the model, plus the precipitation totals (which were not used as input to the model). The results are shown in Table 3.4. None of these correlation coefficients are very large, which is likely due to the fact that there is not much spatial variation in the fluxes for this particular overpass time. However, the relative values of the coefficients bears out the qualitative observation above which air temperature values appear to have a dominant effect on the sensible heat flux, followed by a dependence on the NDVI, from which the fractional cover estimates are derived.

The consideration of the spatially variable near-surface meteorological conditions is important for accurate mapping of the flux components. A more preliminary version of these maps (Humes *et al.* 2000) were derived with the same model, but using areal average values of meteorological data (and several differences in the way the fractional vegetation cover parameter was estimated). The correlation of the turbulent fluxes with precipitation and land cover type appeared to be stronger in that case than is currently indicated.

The density of the Micronet observations makes it possible to observe this spatial variability in the meteorological data. Since it is rare to have this density in observations of meteorological quantities, these types of conditions may occur more frequently than is known, particularly in the period shortly after precipitation events. These results underscore the need for as much density as possible in the ground networks that provide near-surface meteorological inputs to these and other types of models.

3.5 Summary

A relatively simple, "snapshot"-type model for computing components of the surface energy balance data was run on a pixel-by-pixel basis for the LWW in central Oklahoma. The model uses ground and remotely sensed

data to calculate a separate energy balance for the vegetation and soil layers. The remotely sensed data, in this case Landsat TM data, were used to compute surface characteristics that affect the energy balance (such as surface temperature, land cover type (which was used to assign albedo and surface roughness parameters), and NDVI (which was used to estimate fractional vegetation cover). The model also requires near-surface meteorological inputs, which were derived from a very dense network of such observations at the LWW. Model estimates of surface fluxes were in good agreement with ground-based measurements, except that the model overestimated the soil heat flux rather substantially for the case of bare soil. This gave rise to a substantial error in the latent heat flux for bare soil.

For the particular date studied, the satellite data were acquired shortly (approximately 5 h) after a significant precipitation event occurred in the area. The saturated conditions across the watershed, combined with minimal radiation loading that occurred that morning before the satellite data were acquired, had the effect of minimizing the spatial variability in some of the surface state variables that control surface fluxes, such as surface temperature. It also gave rise to the rather unusual situation in which there was more variation in near-surface air temperature than in surface temperature. Thus, the spatial variability in the surface energy balance components was rather minimal, but did show some slight spatial patterns related to near-surface meteorologic conditions, precipitation totals in the hours prior to the satellite data acquisition, and land cover type.

Though the spatial variability in surface fluxes for the time period studied was relatively minimal, the work presented here demonstrates the utility of this type of modeling approach, which is primarily "data driven" and does not require special calibration for application to other areas. These results also underscore the need for as much density as possible in the ground networks that provide near-surface meteorological inputs to these and other types of models. Future work in this area will include the application of this type of model to satellite data with more coarse spatial resolution but finer temporal resolution.

Acknowledgments

The authors wish to thank Tom Jackson (USDA/ARS Hydrology Lab), Frank Scheibe (formerly USDA/ARS US Agricultural Water Quality Lab), and Ted Engman (NASA/GSFC) for their efforts in organizing the 1994 Little Washita field experiments; Conrad Ziegler and Les Shoal (National Severe Storms Laboratory), and Christa Peters-Lidard (Georgia Tech) for the radiosonde data used in this study; and especially Gary Heathman (USDA/ARS Grazinglands Research Center) and his staff at the Little Washita field office for their outstanding support during the experimental operations.

References

Avissar, R. and R.Pielke (1989) A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. *Mon. Weather Rev.* 117: 2113–36.

Avissar, R. and M.M.Verstraete (1990) The representation of continental surface processes in atmospheric models. *Rev. Geophys.* 28: 35–52.

Betts, A.K., J.H.Ball, A.C.M.Beljaars, M.J.Miller, and P.A.Viterbo (1996) The land surface-atmosphere interacton: a review based on observation and global modeling perspectives. *J. Geophys. Res.* 101 (D3): 7209–25.

Brutsaert, W. (1982) *Evaporation into the Atmosphere.* D.Reidel, Dordrecht.

Carlson, T.N. and D.A.Ripley (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ. 62:* 241–52.

Chen, F. and R.Avissar (1994) The impact of land-surface wetness on mesoscale heat fluxes. *J. Appl. Meteorol.* 33: 1324–40.

Choudhury, B.J., N.U.Ahmed, S.B.Idso, R.J.Reginato, and C.S.T.Daughtry (1994) Relations between evaporation coefficients and vegetation indices studied by model simulations. *Remote Sens. Environ.* 50: 1–17.

Dickinson, R.E., A.Henderson-Sellers, P.J.Kennedy, and M.F.Wilson (1986) Biosphere-atmosphere transfer scheme (BATS) for NCAR Community Climate Model. Technical Note, NCAR/TN-275+STR, National Center for Atmospheric Research, Boulder, Colorado.

Entekhabi, D. and P.S.Eagleson (1989) Land surface hydrology parameterization for atmospheric general circulation models including sub-grid scale variability. *J. Appl. Meteorol.* 2: 817–31.

Gillies, R.R. and T.N.Carlson (1995) Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. *J. Appl. Meteorol.* 34: 745–56.

Holwill, C.J. and J.B.Stewart (1992) Spatial variability of evaporation over FIFE area derived from aircraft and ground-based data. *J. Geop. Res. 97* (D17): 18673–80.

Humes, K.S., W.P.Kustas, and D.C.Goodrich (1997) Spatially distributed sen sible heat flux over a semiarid watershed. Part I: use of radiometric surface temperature & a spatially uniform resistance. *J. Appl. Meteorol.* 36:281–92.

Humes, K.S., R.Hardy, and W.P.Kustas (2000) Spatial patterns in surface energy balance components derived from remotely sensed data. *Prof. Geogr. 52* (2): 272–88.

Idso, S.B. (1981) A set of equations for the full spectrum and 8 to 14μm and 10.5–12.5 μm thermal radiation from cloudless skies. *Water Res. Res.* 17: 295–304.

Jackson, T.J. and F.Scheibe (eds) (1993) Data Report for Washita '92 Field Campaign. USDA/ARS National Agricultural Water Quality Laboratory, Durant, Oklahoma.

Kniezys, F.X., E.P.Shettle, L.W.Abreu, J.H.Chetwynd, and G.P.Anderson (1988) User Guide to Lowtran 7. Air Force Geophysical Laboratory Report No. AFGLTR-88–0177, Hanscom AFB, MA 01731.

Kustas, W.P., J.R.Prueger, K.S.Humes, and P.J.Starks (1999) Surface heat fluxes at field scale using surface layer versus mixed-layer atmospheric variables with radiometric temperature observations. *J. Appl. Meteorol.* 38:224–38.

Markham, B.L. and J.L.Barker (1986) EOSAT Landsat Technical Notes, #1, August, Eosat Corp., Lanham, MD.

Moran, M.S., R.D.Jackson, L.H.Raymond, L.W.Gay, and P.N.Slater (1990) Mapping surface energy balance components by combining Landsat Thematic mapper and ground-based meteorological data. *Remote Sens. Environ.* 30: 77–87.

Noilhan, J. and S.Planton (1989) A simple parameterization of land surface processes for meteorological models. *Mon. Weather Rev.* 117: 536–49.
Norman, J.M. and G.S.Campbell (1983) Application of a plant-environment model to problems in irrigation. In D.J.Hillel (ed.) *Advances in Irrigation.* Academic Press, New York, pp. 155–88.
Norman, J.M., W.P.Kustas, and K.S.Humes (1995) A two-source approach for estimating soil and vegetation energy fluxes from observations of directional radiometric surface temperature. *Agric. For. Meteorol. 77:* 263–93.
Priestley, C.H.B. and R.J.Taylor (1972) On the assessment of surface heat flux and evaporation using large scale parameters. *Mon. Weather Rev.* 100: 81–92.
Prueger, J. (1996) Surface flux measurements, Chapter XIV. In P.Starks and K.S.Humes (eds) *WASHITA 1994 Multidisciplinary Field Campaigns Data Report.* USDA/NAWQL Report.
Sellers, P.J., Y.Mintz, Y.C.Sud, and A.Dalcher (1986) A simple biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.* 43(6): 505–31.
Shuttleworth, W.J. and J.S.Wallace (1985) Evaporation from sparse canopies—an energy combination theory. *Q. J. R. Meteorol. Soc.* 111: 839–55.
Starks, P.J. and K.S.Humes (eds) (1996) Data report for 1994 remote sensing field campaigns over the Little Washita Experimental Watershed. USDA/ARS Grazing Lands Research Laboratory, El Reno, Oklahoma.
Xue, Y., P.J.Sellers, J.L.Kinter, and J.Shukla (1991) A simplified biosphere model for global climate studies. *J. Clim.* 4: 346–64.

Chapter 4

Estimating spatially distributed surface fluxes in a semi-arid Great Basin desert using Landsat TM thermal data

Charles A.Laymon and Dale A.Quattrochi

4.1 Introduction

Ground-based measurements of hydrologic and micrometeorologic processes are now available for many parts of the world, especially for the United States and Europe, on a nearly routine basis. These measurements, however, are only representative of a very small area around the sensors, and, therefore, provide little information about regional hydrology. The variability of the land surface precludes using these measurements to make inferences about processes that occur over an area of a hectare, much less the size of an entire valley. Recent developments have demonstrated an increasing capability to estimate the spatial distribution of hydrologic surface fluxes for very large areas with remote sensing techniques. A number of studies have focused on the use of remote sensing to measure surface water and energy variables in attempts to derive latent heat flux or evapotranspiration (ET) over semiarid regions (e.g. Kustas *et al.* 1989a,b, 1990, 1994a,b,d, 1995; Humes *et al.* 1994, 1995; Moran *et al.* 1994; Ottlé and Vidal-madjar 1994; Tueller 1994). All of these investigations have used aircraft-based instruments and were limited to small areas. In only a few investigations has satellite-based remote sensing data been used to estimate ET. The synoptic and real-time attributes of remote sensing data from satellites offer the potential for measuring landscape, hydrometeorological, and surface energy flux characteristics that can be used in both monitoring and modeling the state and dynamics of semi-arid regions. Choudhury (1991) reviewed the current state of progress in utilizing satellite-based remote sensing data to estimate various surface energy balance parameters. Kustas *et al.* (1994c) used Advanced Very High Resolution Radiometer (AVHRR) data to extrapolate ET estimates from one location containing near-surface meteorological data to other areas in a semi-arid basin in Arizona. Moran *et al.* (1989) and Moran and Jackson (1991) used Landsat Thematic Mapper (TM) data to estimate ET over a small agricultural area. In this paper, we present a method for scaling from point to spatial estimates of instantaneous surface fluxes for a Great Basin

desert valley using Landsat TM data and for characterizing the partitioning of fluxes among the different soil and landcover types found in the study area.

A field study was conducted from May 1993 through October 1994 to improve our understanding of the processes that govern the local energy and water fluxes in a Great Basin desert ecosystem. A survey of soils and vegetation was conducted for the study area. Six surface water and energy balance flux stations were deployed in major plant ecosystems. These stations operated nearly continuously throughout the study period, except for several of the winter months. Field work was conducted during special observing periods at the peak "green-up" in the early summer of 1993 and 1994 and at "dry-down" during late summer of 1993. These periods included deployment of several eddy correlation systems, soil moisture measurements using the neutron probe and time domain reflectometry techniques, and radiosonde observations of the lower atmosphere. This research program provided an infrastructure to further study the use of remote sensing to measure surface properties and processes.

4.2 Setting

The study was conducted in Goshute Valley of northeastern Nevada, a faulted graben valley of the Basin and Range Province of the western United States about 50 km west of the Great Salt Lake Desert (Figure 4.1). Although the entire valley is about 75km long and 16km wide, our study was restricted to

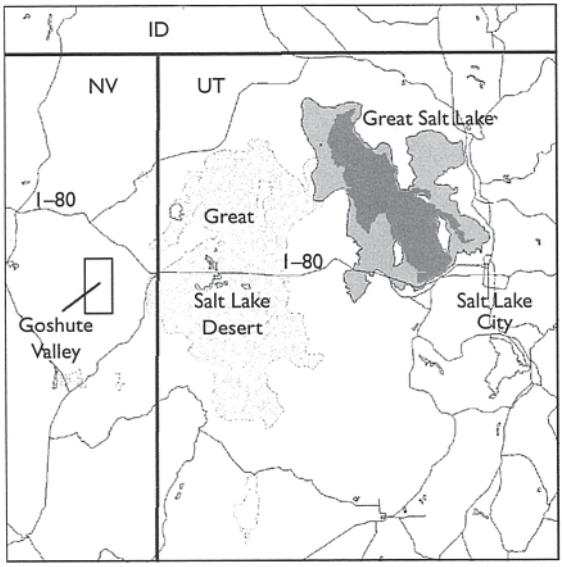


Figure 4.1 Map showing the location of Goshute Valley (40º44'N, 114º26' W) in northeastern Nevada in relation to state boundaries and Great Salt Lake Desert, Utah.

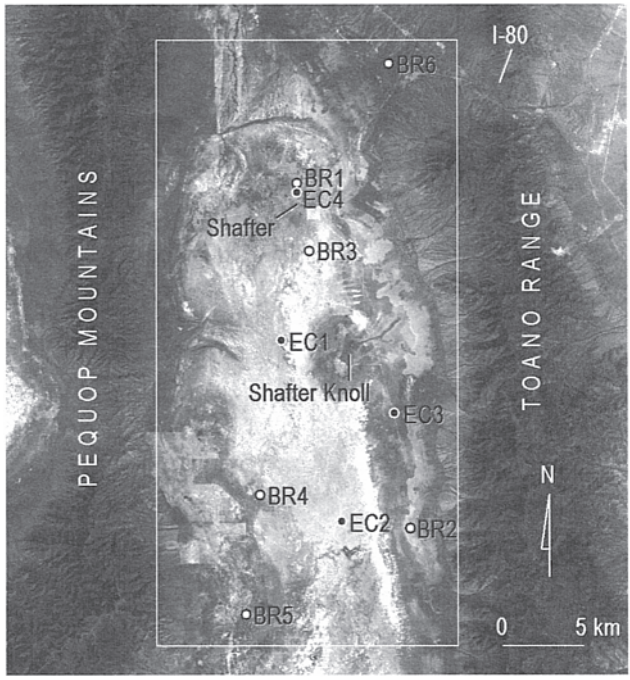


Figure 4.2 Landsat-5 TM image of the Goshute Valley, Nevada, study area showing the types and location of surface water and energy balance flux stations (BR=bowen ratio, EC=eddy correlation). The box defines the area over which energy balance components were derived and cooresponds to the area shown in Figure 4.9(a)–(d).

a 40 km long central section (Figure 4.2). The valley floor, with an elevation of about 1700m asl, is nearly flat with slopes of less than a few degrees. The valley is bordered by alluvial fans emanating from the mountains. A pluvial lake occupied the valley during the Late Pleistocene leaving strand lines and terraces on the alluvial fans and allowing for lacustrine silt and clay to accumulate in the valley. Because outflow drainage was limited, dissolved weathering products from the surrounding mountains became concentrated in the lake producing significant amounts of soluble salts and carbonates in the lacustrine sediments. As a result, salt content and pH of the lacustrine soils in the central reaches of Goshute Valley are high. Vegetation of the valley is dominated by shrubs with some understory forbs and grasses. Land within the valley has not been heavily grazed or developed, although small portions of the valley have been chained for grazing and are easily identified by the regular geometric patterns in Figure 4.2.

4.3 Methods

4.3.1 Approach

The general surface energy balance can be summarized as:

$$R_n = H + LE + G \tag{4.1}$$

where R_n is net radiation absorbed at the surface, G the flux of heat into the soil, and H and LE are the sensible and latent heat fluxes into the atmosphere. We use the sign convention that all the radiative fluxes directed toward the surface are positive, while other (non-radiative) energy fluxes directed away from the surface are positive and vice versa. LE, a product of the rate of evaporation E and the latent heat of vaporization L, is the rate of energy utilization in ET and is often treated as a proxy for ET. R_n, G, and H can be estimated from micrometeorological measurements, or in some cases, using remote sensing techniques exclusively (Jackson *et al.* 1985; Clothier *et al.* 1986). The remote sensing techniques, however, usually require assumptions about surface conditions that are best measured on the ground. Remote sensing reflectance and emittance data used in conjunction with surface meteorological data can be used to estimate parameters needed to characterize R_n, G, and H, leaving LE to be defined mathematically.

Our approach is to establish a one-to-one relationship between surface radiation and energy fluxes measured at points on the ground to corresponding reflectance and emittance values of a geolocated remote sensing image. The empirical relationships are then used to extrapolate from the point measurements to spatial estimates of surface fluxes. Our procedure is based on a Landsat-5 TM image of June 19, 1994. This date closely follows field observations that occurred between June 7 and June 14, 1994.

Five surface energy balance flux stations were installed in Goshute Valley in May, 1993, and a sixth station was added in June, 1994 (Figure 4.2). The most northerly and southerly stations were separated by 35 km. The stations were installed in different assemblages of dominant vegetation types present within the valley or in assemblages of vegetation with different plant density. Each station contained the same instrument configuration (Figure 4.3 and Table 4.1), with the exception that infrared thermometers were located at only four stations. Measurements were made every 5 s and then output as 20-min averages. Malek *et al.* (1997) and Malek and Bingham (1997) have discussed the annual radiation and energy balance from these stations.

The Bowen ratio method used to measure the surface energy balance in this experiment requires fetch. On the basis of instrument height and the wind speed measured during the hour that the TM scene was acquired, we assume that flux measurements are representative of an area within a 100 m radius of each Bowen ratio station. The image was geometrically corrected

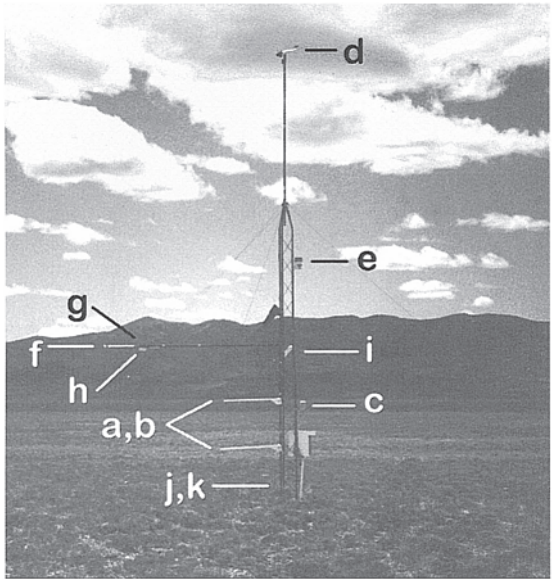


Figure 4.3 Photo of a surface water and energy balance flux station deployed in Goshute Valley during the experiment. The letters correspond to the instrument descriptions in Table 4.1.

Table 4.1 Instrument configuration at the surface energy balance stations

Variable[a]	*Instrument*	*Deployment*	*Vendor*
a. Air temperature	Thermocouple	1 and 2 m above sfc	Campbell Scientific
b. Dew point temperature	Cooled mirror hygrometer	1 and 2 m above sfc	General Eastern Corp.
c. Relative humidity	RH Sensor	2 m above sfc	Campbell Scientific
d. Wind speed/direction	Anemometer/Vane	10 m above sfc	RM Young
e. Rainfall	Tipping bucket	6 m above sfc	Texas Electronics
f. Net radiation	Net radiometer	4 m above sfc	REBS Fritschen
g. Downwelling solar radiation	Pyranometer	4 m above sfc	LI-COR
h. Reflected solar radiation	Pyranometer	4 m above sfc	Epply Lab
i. Surface temperature	IR Thermometer	4 m above sfc	Everest InterScience
j. Soil temperature	Temperature probe	2 and 6 cm below sfc at three locations	Campbell Scientific
k. Ground heat flux	Heat flux plate	2 and 8 cm below sfc at three locations	Campbell Scientific

Note
a Letters correspond to the letters in Figure 4.3.

to within one pixel of the true location. Thus, the station data were related to the mean remote sensing reflectivity values of an area corresponding to 7×7 pixels (~200 m×~200 m) centered over each station.

4.3.2 Geometric correction

A full Landsat-5 TM scene covering the Goshute Valley was obtained from the Earth Observation Satellite (EOSAT) Corp. for June 19, 1994 (09:39 h local standard time). The image was a system-corrected, orbit-oriented product (type "P" data). Using 16 ground control points defined with Global Positioning System (GPS) instruments, the image was more precisely rectified with a first-order Affine transformation with no resampling yielding a standard error of 4.6 m. Visual inspection of the control points in relation to the image revealed they were all within one pixel (<28.5 m) of the correct location. The data were not topographically corrected as the valley floor is essentially flat.

4.3.3 Radiometric correction of the reflected bands

Before remote sensors can measure components of the surface energy budget, the recorded digital values must be converted to measures of at-satellite radiance and then to surface reflectivity. Digital values in each band are converted to at-satellite spectral radiance (Chevez 1989) and then to apparent at-satellite reflectance after normalizing for the effects of variations in incident solar irradiation (Nicodemus *et al.* 1977; Markham and Barker 1986, 1987a,b; Hill and Sturm 1991; Markham *et al.* 1992; Gilabert *et al.* 1994). After accounting for viewing geometry, atmospheric scattering, and transmission losses, surface reflectance $\rho(\lambda)$ (unitless) is defined as

$$\rho(\lambda) = \frac{\pi(L_0(\lambda) - L_p(\lambda))d^2}{T(\lambda) \uparrow E_g(\lambda)\cos\theta_0} \tag{4.2}$$

where $L_0(\lambda)$ is the apparent at-satellite spectral radiance in band λ, and $L_p(\lambda)$ is the atmospheric path radiance resulting from scattering, d is the Earth-Sun distance (Sturm 1981), $T(\lambda) \uparrow$ is the direct beam transmittance of the atmosphere in the upward direction, $E_g(\lambda)$ is the global solar irradiance at the surface, and θ_0 is the solar zenith angle. Thus, surface reflectance can be determined with estimates of $L_p(\lambda)$, $T(\lambda) \uparrow$, and $E_g(\lambda)$.

Atmospheric path radiance is the sum of Rayleigh and aerosol (Mie) scattering (Gordon 1978):

$$L_p(\lambda) = L_r(\lambda) + L_a(\lambda) \tag{4.3}$$

The Rayleigh scattering contribution, $L_r(\lambda)$, is all but constant in the atmosphere, as it is based on the solar zenith and sensor view angles, and

thus, can be determined from image header information (Saunders 1990). Gilabert *et al.* (1994) developed a procedure that integrates the dark object subtraction and atmospheric transmission modeling techniques to estimate the aerosol scattering contribution to path radiance, $L_a(\lambda)$, on observed surface reflectances. The method consists of an inversion algorithm based on a simplified radiative transfer model in which characteristics of atmospheric aerosols are estimated from the observed radiance in TM bands 1 and 3. This is in contrast to many other procedures in which the characteristics of aerosols are measured or estimated a priori. The technique has the advantage over other methods in that it is based entirely on information derived from the image. The path radiance in TM bands 1 and 3 determined from dark objects in the image are used to define the aerosol spectral properties at the time the image was acquired. With this model, the parameters necessary to solve equation (4.2) can be determined from any Landsat-5 TM image that contains some dark pixels. The only information needed to apply this model is the mean elevation of the imaged terrain, the day of year the image was acquired, the solar zenith angle, and the dark object digital values for TM bands 1 and 3. The sun elevation reported in the header of each Landsat-5 TM image is used to determine the solar zenith angle at the time of image acquisition. The definition of digital values for dark pixels in the image is the most critical step in the entire procedure and should be done with great care. Dark object digital values were defined for spectral minima associated with water and shadows within the scene, but outside the study area.

4.3.4 Estimation of energy balance components

Net radiation

The net radiation flux in equation (4.1) can be written as

$$R_n = (1 - \alpha)R_s\downarrow + R_l\downarrow - \varepsilon_s \sigma T_s^4 \tag{4.4}$$

where α is the surface albedo, $R_s\downarrow$ is incoming shortwave radiation or irradiance, $R_1\downarrow$ is incoming longwave radiation, ε_s is surface emissivity, σ is the Stefan-Boltzmann constant, and T_s is the surface temperature. The actual amount of insolation received at the ground may be considerably smaller than at the top of the atmosphere because of scattering, absorption, and turbidity of the atmosphere. It is, therefore, usually measured in the field and assumed to be spatially invariant over the study domain. $R_1\downarrow$ emanates largely from the atmosphere and is spatially homogeneous relative to the land surface. Although $R_1\downarrow$ has been estimated using measurements of nearsurface air temperature and relative humidity (Brutsaert 1975; Humes *et al.* 1994), direct observations from the flux stations were used in this study. Thus, net radiation was determined with field measurements of the

downwelling radiation fluxes, $R_s\downarrow$, and $R_l\downarrow$, and remote sensing measurements of a, ε_s, and T_s.

Albedo

Albedo is the ratio of upwelling shortwave radiation to solar irradiance. For our purpose, solar irradiance at the land surface can be estimated satisfactorily using a radiative transfer model with parameters derived from atmospheric soundings. Solar irradiance at the surface in Goshute Valley was modeled for the day of the satellite overpass using the SPECTRL radiative transfer model (Justus and Paris 1985, 1987) and sounding data obtained from the National Weather Service at Ely, Nevada (0Z, June 20, 1994 = 17:00 h PST, June 19, 1994), about 140 km to the south-southwest.

Shortwave radiometers on today's satellites detect radiation in discrete bandwidths, not over the total solar spectrum (~0.3–4.0 μm). These narrow band samples of the solar spectrum must be extrapolated over the entire spectrum to estimate broadband albedo. The technique used here follows that of Brest and Goward (1987) and Starks *et al.* (1991) in which broadband albedo is the reflectance in multiple bands integrated over the total solar spectrum. Each band is weighted according to the ratio of radiance sampled to the total radiance for an extended bandwidth associated with each band. Thus, broadband albedo, aBB, is (Starks *et al.* 1991)

$$\alpha_{BB} = \pi \sum_{\lambda=1}^{6} (\rho(\lambda))(W(\lambda)) \tag{4.5}$$

where $\rho(\lambda)$ is the reflectance in TM band λ, and $W(\lambda)$, the weighting coefficient, is

$$W(\lambda) = \int_{L(\lambda)}^{U(\lambda)} E(\lambda)\, d\lambda \Big/ \int_{0.3}^{4.0} E(\lambda)\, d\lambda \tag{4.6}$$

where $E(\lambda)$ is the solar irradiance in band λ and $U(\lambda)$ and $L(\lambda)$ are the upper and lower wavelengths of each TM bandpass, respectively. An assumption that the surface responds as a Lambertian reflector is necessary because the remote sensing instrument is nadir viewing. Generalized reflectance curves were developed for vegetation, soil, bedrock, and water using data from the National Aeronautics and Space Administration (NASA). These curves were used in conjunction with the modeled solar irradiance curve to define the extended bandwidths for each reflected TM band based on inflection points. Thus, spatially distributed broadband albedo was computed for the study

area by

$$\alpha_{BB} = \pi[(0.111\rho(1)) + (0.119\rho(2)) + (0.078\rho(3)) + (0.124\rho(4)) + (0.041\rho(5)) + (0.019\rho(7))] \quad (4.7)$$

Emissivity

Emissivity was measured in the field at two sites (BR1, EC1; Figure 4.2) just before dawn using the procedure described by Hipps (1989). The measurements included apparent temperature with an infrared thermometer, temperature of the target covered by an aluminum cone, and apparent and actual temperatures of an aluminum plate of known emissivity. Emissivity was determined for bare soil (0.92–0.93), and the dominant vegetation species: greasewood (0.94–0.95), shadscale (0.95), and sagebrush (0.98). The percentage of ground surface covered by vegetation at each flux station and the proportion of total vegetation represented by different species was determined using the point quadrat method (Groeneveld 1997). Based on these data, the area-weighted mean emissivity was determined for each flux station. Because emissivity was generally higher for vegetation than bare soil, spatially distributed emissivity was estimated from the Normalized Difference Vegetation Index (NDVI). There was insufficient variability in emissivity among flux sites to define the nature of the relationship until emissivity at endpoint NDVI values of 0.1 and 1.0 for bare soil and complete vegetation coverage, respectively, were included. The resulting linear relationship is defined by

$$\varepsilon_s = 0.022\text{NDVI} + 0.928 \quad (4.8)$$

As this relationship is based on the observed emissivity for specific plant species, it is appropriate only for the study site and similar settings. More observations of emissivity over a broader range of NDVI values are required to more precisely define the ε_s to NDVI relationship (cf. Labed and Stoll 1991).

Surface temperature

Longwave radiation is emitted from the surface in proportion to its temperature as described by Planck's law. Using pre-launch calibration constants for Landsat-5 TM band 6, surface temperature $T_s(\lambda)$ is determined by (Markham and Barker 1986)

$$T_s(\lambda) = \frac{C_2}{\ln((C_1/L_s(\lambda)) + 1)} \quad (4.9)$$

where C_1 and C_2 are the calibration constants equal to 60.776 m W cm^{-2} ster^{-1} μm^{-1} and 1260.56 K, respectively (see also Goodin 1995). Surface radiation,

$L_s(\lambda)$, can be expressed in terms of the observed radiation, $L_0(\lambda)$, as (Schott and Volchok 1985)

$$L_s(\lambda) = \frac{L_0(\lambda) - \tau(1 - \varepsilon_s)L_d(\lambda) - L_p(\lambda)}{\tau\varepsilon_s} \tag{4.10}$$

where $L_0(\lambda)$ is the apparent at-satellite spectral radiance in band λ, $L_d(\lambda)$ is the downwelling longwave radiation reaching the surface, $L_p(\lambda)$ is the atmospheric path radiance, ε_s is the surface emissivity, and t is the atmospheric transmissivity.

For sensors with more than one thermal channel, various split-window algorithms have been developed for atmospheric correction. Only one thermal channel on the TM sensor prevents use of these algorithms. Instead, various alternative methods have been developed that use sounding data and radiative transfer models to characterize the atmosphere (cf. Vidal *et al.* 1994). Atmospheric transmissivity and downwelling and path radiance at the TM thermal waveband were calculated using the radiative transfer model SPECTRL and atmospheric sounding data from Ely, Nevada (described previously). The model was run for the TM-6 bandwidth with no surface reflectance to determine path radiance, and again, with surface reflectance (albedo) consistent with field measurements to determine downwelling radiance and atmospheric transmissivity. These values were assumed to be constant in space throughout the study area and were applied to calculate surface temperature for each image pixel.

Soil heat flux

The surface temperature at a given location is controlled by the surface energy balance, which, in turn, depends on the radiation balance and vegetation cover among other factors. Thus, the soil heat conduction flux can be estimated as a fraction of the net radiation (Clothier *et al.* 1986). Based on this theory, several investigations have attempted to define soil heat flux as a function of net radiation and reflectivity in the red and near-infrared wave bands (Reginato *et al.* 1985; Clothier *et al.* 1986; Jackson *et al.* 1987; Kustas and Daughtry 1990). Soil heat flux at a depth of 8cm (G_{8cm}) was measured directly at each of the surface energy flux stations (Malek *et al.* 1997). G_{8cm} was converted to surface heat flux (G_{sfc}) (Hanks and Ashcroft 1980; Oke 1987; Malek 1994) using the following relationship (Malek *et al.* 1997) (n=518, r=0.96):

$$G_{sfc} = 1.615G_{8\,cm} \tag{4.11}$$

This is an obvious soil-specific realtionship and its validity here without modification is unknown. Because of the strong relationship between soil

heat flux, net radiation, and vegetation cover, G_{sfc} can be defined on the basis of reflectivity in the red and near-infrared wavebands via a regression relationship as

$$G_{sfc} = [0.774 - 0.324(\rho(\text{NIR})/\rho(\text{Red}))]R_n \tag{4.12}$$

where ρ(NIR) and ρ(Red) are the reflectances in TM bands 4 and 3, respectively, and R_n is remotely sensed.

Sensible and latent heat flux

The sensible heat flux can be defined using the surface-air temperature difference in a bulk resistance approach analogous to Ohm's Law (Monteith 1973)

$$H = \rho C_p(T_s - T_a)/r_{ah} \tag{4.13}$$

where ρ is the density of air at 1700 m, Cp the specific heat of air (ρCp is the volumetric heat capacity), T_s the remotely sensed surface temperature, and T_a is the air temperature at height z above the surface. The bulk resistance to heat transfer across a single surface-atmosphere layer, r_{ah}, or aerodynamic resistance, is determined by Monin-Obukhov surface layer similarity theory as

$$r_{ah} = \{\ln[(z-d_{0m})/z_{0m}] - \psi_m\}\{\ln[(z-d_{0h})/z_{0m}] + \ln(z_{0m}/z_{0h}) - \psi_h\}/k^2u \tag{4.14}$$

where d_{0m}, z_{0m}, d_{0h}, and z_{0h} are the zero-plane displacement heights and roughness lengths for momentum and heat, respectively, k is von Karman's constant (~0.4), and u is the wind speed measured at the reference height, z. ψ_m and ψ_h are the stability correction functions for wind and temperature, respectively.

There is little experimental evidence to suggest that d_{0m} and d_{0h} differ significantly (Kustas 1990) and were, therefore, treated with the same value defined hereafter as simply d_0. Reasonable estimates of z_{0m} and d_0 have been obtained for vegetation on flat uniformly covered surfaces with several empirical relationships based on vegetation height. After Monteith (1973), displacement was defined as d_0=*2/3h*. In contrast to displacement, theoretical and experimental evidence exists for significant differences in the values of scalar versus momentum roughnesses due mainly to differences in transfer processes near the soil and vegetation surfaces (see Thom 1972). Consequently, an added resistance to heat transfer results in $z_{0h}<z_{0m}$ and suggests that z0h can be taken as a fraction of z_{0m}. Studies reported in Kustas *et al.* (1989b) suggest z_{0h} is 1/10 to 1/5 of z_{0m}. Chamberlain (1968) expressed the relationship between z_{0h} and z_{0m} in the form

$$kB^{-1} = \ln(z_{0m}/z_{0h}) \tag{4.15}$$

Experimental data and physical models of vegetated surfaces suggest a constant $kB^{-1} \approx 2$ (see Garratt and Hicks 1973), but analytical data show it to be sensitive to plant structure (Massman 1987). Brutsaert (1982) suggests that kB^{-1} can vary from 2 to 10. Little data have been collected for sparsely vegetated surfaces as in Goshute Valley. With exception, Kustas *et al.* (1989a) suggests $kB^{-1} \approx 2$ or 3 is reasonable for sparse canopy in the arid southwest. In addition, Kustas *et al.* (1989a) found that analysis over sparse canopy cover required that kB^{-1} be a function of the thermometric surface temperature observed from a nadir-viewing, thermal infrared sensor to obtain satisfactory results, and gave the following relationship:

$$kB^{-1} = 0.17u(T_s - T_a) \tag{4.16}$$

Historically, the stability functions have been determined by the MoninObukhov similarity theory, which holds that the diffusion coefficients for momentum and heat are equivalent (e.g. Paulson 1970). The assumptions in the basic aerodynamic approach of neutral stability and similarity of all coefficients are restrictive. Its applicability, however, can be extended by incorporating adjustments that depend upon stability and that include empirical terms to account for non-similarity of the diffusion coefficients. The Richardson number is a convenient way of categorizing atmospheric stability in the surface layer (Panofsky and Dutton 1984; Oke 1987). The Richardson number, *Ri,* is given by

$$Ri = Ri_b/s^2 \tag{4.17}$$

where Ri_b is the bulk Richardson number is given by

$$Ri_b = [g\bar{z}^2/\overline{T}]\left[\frac{(\Delta T/\Delta z) + \gamma_d}{U^2}\right] \tag{4.18}$$

where g is the acceleration due to gravity, $\Delta T/\Delta z$ is the temperature gradient, γ_d is the dry adiabatic lapse rate, and U is the mean wind speed from the flux stations. The value of s in equation (4.17) is given by

$$s = \frac{\phi_m}{\ln[(\bar{z} - d_0)/z_{0m}] - \psi_m} \tag{419}$$

where ϕ_m and ψ_m are the shear and profile functions, respectively, and are given in terms of the dimensionless height, z/L. Högström (1988) suggests that the Businger–Dyer formulations (Businger *et al.* 1971; Dyer 1974) give satisfactory results for shear and profile functions for stable and unstable conditions.

Stable ($0 \leq Ri \leq 0.2$):

$$z/L = Ri/(1 - 5Ri) \tag{4.20}$$

$$\phi_m = \phi_h = 1 + 5(z/L) \tag{4.21}$$

$$\psi_m = \psi_h = -5(z/L) \tag{4.22}$$

Unstable ($Ri < 0$):

$$z/L = Ri \tag{4.23}$$

$$\phi_m = [1 - 16(z/L)]^{-1/4} \tag{4.24}$$

$$\psi_m = 2\ln\left[\tfrac{1}{2}(1+x)\right] + \ln\left[\tfrac{1}{2}(1+x^2)\right] - 2\tan^{-1}(x) + \pi/2 \tag{4.25}$$

$$\psi_h = 2\ln\left[\tfrac{1}{2}(1+x^2)\right] \tag{4.26}$$

where

$$x = [1 - 16(z/L)]^{1/4} \tag{4.27}$$

Spatially distributed values of H are estimated based on spatially distributed values of T_s and mean vegetation height based on a classification (discussed below). In this study, z_{0m} and z_{0h} could not be defined on the basis of the wind and temperature profiles because these parameters were only measured at two heights. Thus, a fundamental weakness of this method is the need to assume that aerodynamic resistance parameters were spatially invariant. With kB^{-1} and z_{0m} unknown, z_{0h} was determined for five of the flux station sites by an iterative method using equations (4.17)–(4.27), and observed wind and temperature until the calculated H was in agreement with the observed value. In this way, z_{0m} and z_{0h} were defined as 0.17 and 0.035, respectively. These values are comparable with values for other semi-arid regions reported by Stewart *et al.* (1994).

The calculation of spatially distributed H then proceeded with two nested iterations beginning with a neutral profile (ϕ=1, ψ=0), and initial estimates of H and LE from surface energy balance flux stations. Calculation of Ri was used to define whether to use the stable or the unstable case. The Richardson number was calculated using equations (4.17)–(4.19), then either (4.20)-(4.22) or (4.23)–(4.26), depending on Ri. The process was completed with a solution for equations (4.14) and (4.13) and a new value of LE was computed from a rewritten form of equation (4.1). Calculation of Ri was repeated with the new values of ϕ_m and ψ_m, and the rest of the process was repeated with new values of LE and H. Iteration on H continues until additional changes in Ri are negligible (<0.02).

4.3.5 Vegetation classification

The three major shrubs that dominate the valley landscape are big sagebrush *(Artemesia tridentata wyomingensis)*, black greasewood *(Sarcobatus vermiculatus)*, and shadscale *(Atriplex convertifolia)* (Boettenger, pers. comm.) (Figure 4.4). Other minor shrubs, forbs, and perennial and annual grasses, including Gardner's saltbush *(Atriplex gardneri)*, gray molly *(Kochia americana)*, winter fat *(Ceratoides lanata)*, halogeton *(Halogeton glomerata)*, squirrel tail *(Elymus elymoides)*, Indian rice grass *(Oryzopsis hymenoides)*, and cheat grass *(Bromus tectorum)*, occur in varying amounts depending on soil type and disturbance history (Boettenger, pers. comm.).

False color images of Goshute Valley were used in the field during the June 1994 observing period to identify the types of vegetation present in the valley and to locate classification training sites. Both supervised and unsupervised classifications utilizing different combinations of bands were attempted with mixed success. Difficulties arose because of the low contrast in reflectivity among the red and near-infrared bands and because of the presence of a microphytic crust on the ground surface of generally low reflectivity. These initial tests were used to define the number of classes that

Figure 4.4 Photograph of the shrub vegetation (greasewood, saltbush, shadscale) typical of the Goshute Valley lake plain. The cracking results from desiccation and the darkening of the soil surface is due to the presence of microphytic crust. Both phenomena decrease albedo.

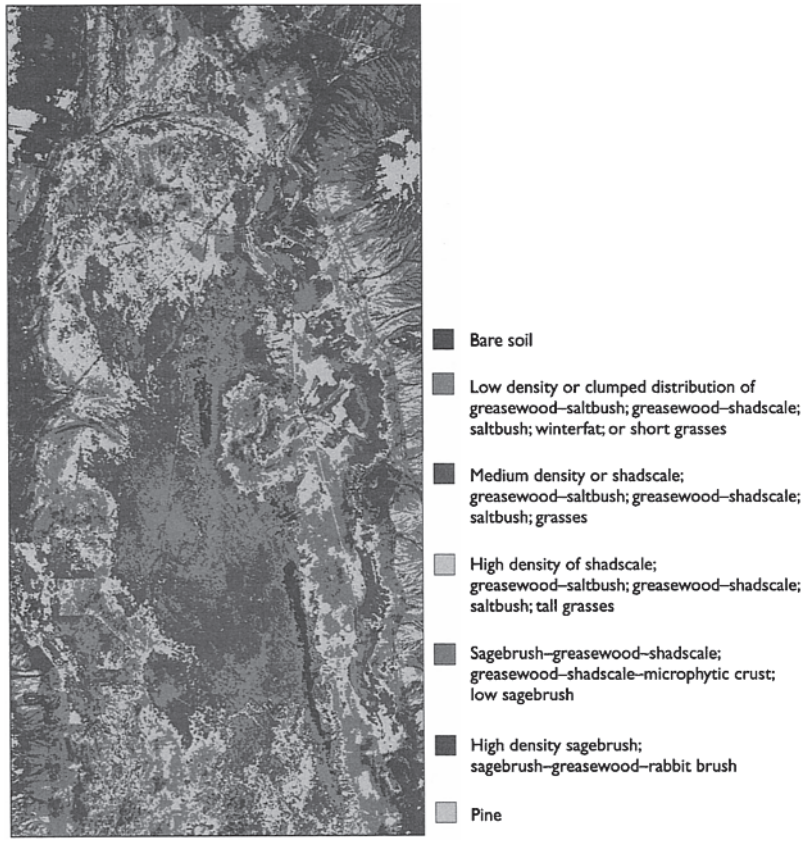


Figure 4.5 Classification of land cover types in Goshute valley. Many desert plant species coexist in assemblages. Most vegetation classes reflect changes in assemblage members or differences in plant density due to changes in soil salinity and moisture availability. Similarities in plant structure and large plant spacing relative to image resolution make classification of desert vegetation extremely difficult (see Colour Plate XXIII).

could be satisfactorily discriminated. A successful vegetation classification was produced using unsupervised, competitive training with bands 2, 3, 4, 5, and 7, yielding seven classes (Figure 4.5). These classes represent assemblages of the dominant vegetation types as well as differences in plant density.

4.4 Results and discussion

Broadband albedo derived from remote sensing using the six reflected TM bands is in good agreement with values measured at the flux stations with

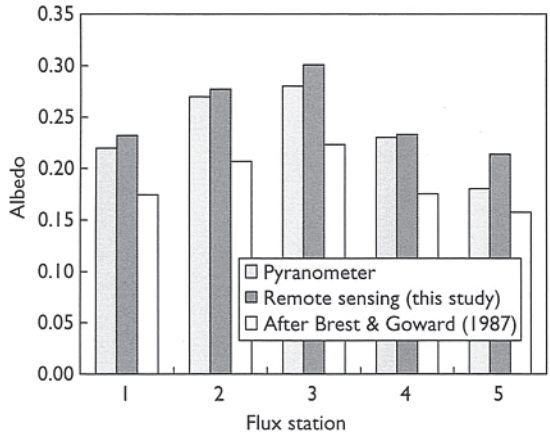


Figure 4.6 Comparison of broadband albedo derived from pyranometer measurements at the flux stations, a remote sensing method using all six reflected Landsat TM bands (this study), and the two-band method of Brest and Goward (1987).

hemispherical pyranometers (Malek *et al.* 1997) (Figure 4.6). The remotely sensed values of aBB are within 1–15% of the pyranometer measurements. This technique of using all six reflected TM bands gives much better agreement with surface measurements than the two-band method (cf. Brest and Goward 1987). The two-band method yielded results consistently about 25% lower than the six-band method used in this study. This difference is, in part, because of the low contrast in reflectivity between the red and near-infrared bands for the sparsely vegetated surface of this semi-arid region. Soils of arid regions generally have a very high albedo. In addition, the presence of microphytic crust, which is fairly extensive throughout the valley, has an effect on aBB independent of vegetation cover because of its dark color. Use of the additional bands incorporates more spectral information about the surface composition not available in the two-band method.

Measurements of surface temperature with Infrared Thermometers (IRT) were only available at four of the six flux stations. In each case, the remotely sensed T_s was 2-7°C lower than the observed values (Figure 4.7). These results contradict several other investigations in which Landsat 5-derived temperatures systematically overestimated coincident surface measurements or remotely sensed measurements from other satellite instruments by several degrees (Schott and Volchok 1985; Wukelic *et al.* 1989; Sugita and Brutsaert 1993; Goetz *et al.* 1995). Other investigations have noted that temperatures observed by IRTs are sensitive to the viewing angle and wind speed (Lhomme

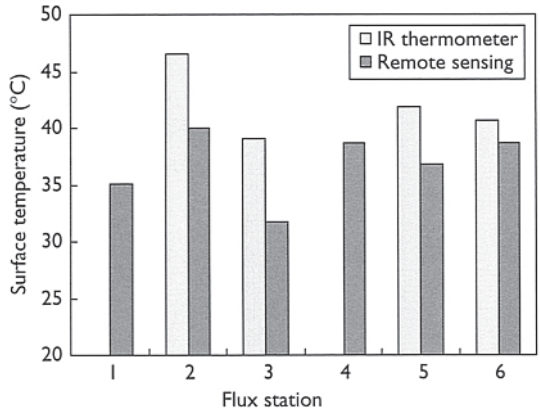


Figure 4.7 Comparison of surface temperatures measured at the flux stations using a tower-mounted infrared thermometer and those obtained with remote sensing.

et al. 1988; Vining and Blad 1992; Kohsiek *et al.* 1993; Stewart *et al.* 1994). Without further information, no correction was made for these effects in this study. Such difficulties associated with retrieval of surface fluxes have been discussed by Hall *et al.* (1992).

Figure 4.8 shows the spatial distribution of R_n, G_{sfc}, H, and LE based on the Landsat TM data. The relationship among geomorphology, soil, and vegetation is readily apparent on these maps and is summarized in Figure 4.9. Table 4.2 gives the mean flux values for each vegetation class. Remotely sensed estimates of net radiation involve using all seven Landsat TM bands through the derivation of estimates of a_{BB}, ε_s, and T_s. The values of $R_s\downarrow$ and $R_1\downarrow$ among the flux stations vary by no more than 5% because of their spatial homogeneity, therefore, it is reasonable to use a mean value for each. Although the correlation between observed and remotely sensed R_n is good, remotely sensed R_n is biased high by about 8% (Figure 4.10). This results largely because the remotely sensed estimates of T_s underestimate values of T_s measured at the flux stations. Although remotely sensed T_s may be slightly underestimated, the error is systematic and the apparent spatial structure of R_n is not affected. R_n is in the range of 350–520 Wm^{-2} where vegetation is least dense on the playa and lake plain, and where vegetation cover is continuous but short as in the chained pasture and extensive areas of saltbush and winterfat on the middle part of the fans (Figure 4.8(a)). R_n is greater than 500 Wm^{-2} in areas of dense greasewood on the lower part of the fans and in areas covered with sagebrush on the highest reaches of the alluvial fans.

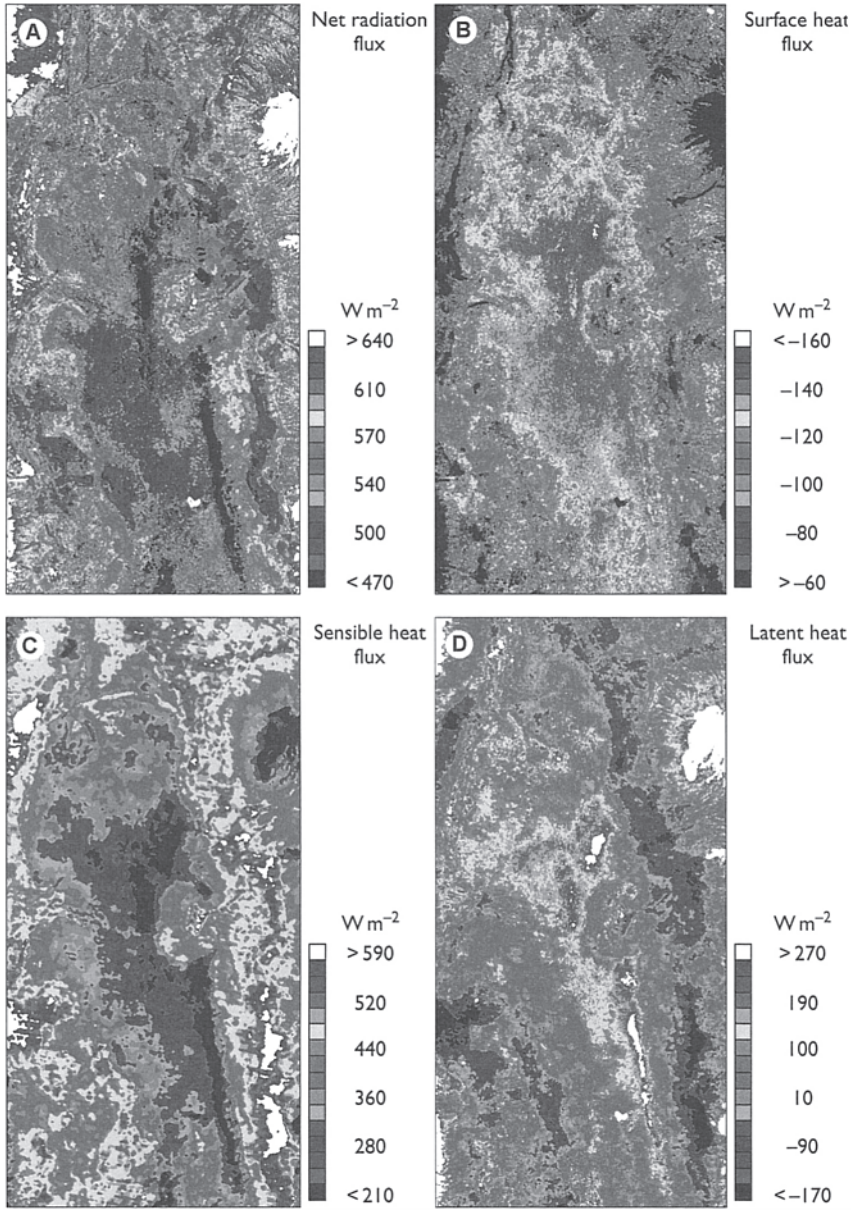


Figure 4.8 Maps showing the spatial distribution of instantaneous surface energy fluxes derived from assimilation of surface meteorological and remote sensing data. See text for discussion (see Colour Plate XXIV).

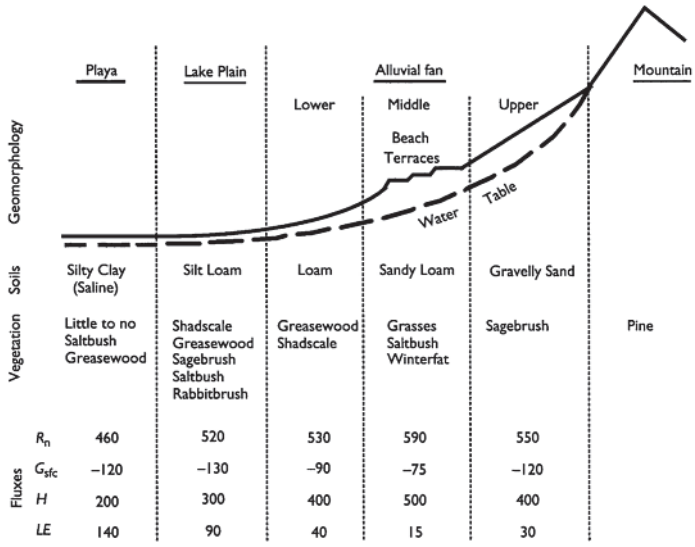


Figure 4.9 Generalized cross-sectional profile from the playa in the central part of the valley to the adjacent mountains, clarifying the terminology used in the text and the relationship between the geomorphology, soils, vegetation, and mean fluxes of Goshute Valley. Fluxes are balanced at the pixel scale and not necessarily for area means shown here.

Table 4.2 Mean flux values for vegetation classes comprised of single vegetation types and assemblages of plants

Class	*Description*	*% Area*	R_n	G_{sfc}	H	LE
1	Bare soil	2	494	−113	343	100
2	Low density or clumped distribution of greasewood–saltbush, greasewood–shadscale, saltbush, winterfat, short grasses	11	493	−122	299	74
3	Medium density of shadscale, greasewood–saltbush, greasewood–shadscale, saltbush, grasses	22	520	−117	373	30
4	High density of shadscale, greasewood–saltbush, greasewood–shadscale, saltbush, tall grasses	24	550	−109	422	20
5	Sagebrush–greasewood–shadscale, greasewood–shadscale–microphytic crust, low sagebrush	24	579	−96	449	34
6	High density sagebrush, sagebrush–greasewood–rabbit brush	16	615	−79	447	85
7	Pine	1	661	−52	332	164

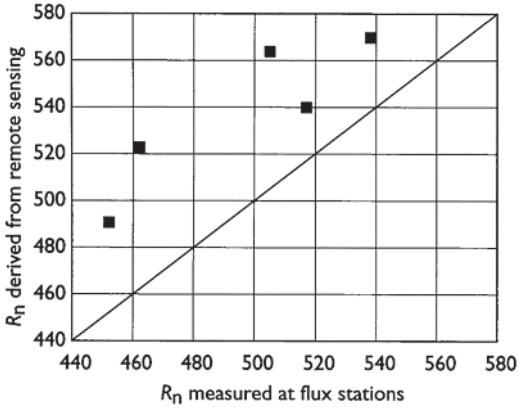


Figure 4.10 Comparison of net radiation measured directly at the flux stations and net radiation derived from remote sensing estimates of emissivity, surface temperature, and albedo. For visual reference, the line depicts the 1:1 relationship. See text for discussion.

The spatial distribution of surface heat flux, G_{sfc}, is shown in Figure 4.8(b) with its obvious relationship to the distribution of vegetation density. As with R_n, the remotely sensed estimates of G_{sfc} overestimate the values measured at the flux stations by about 8 %. This deviation may be largely inherited from the overestimate of R_n. On the lake plain, surface heat flux averages about –130 W m^{-2} whereas on the higher slopes of the fans, where sagebrush is dominant, values range from about –130 to –110Wm^{-2}. Surface heat fluxes are smallest, less than 100 Wm^{-2}, on the middle to lower reaches of the fans.

The sensible heat flux component of the surface energy balance, H, is the most difficult to ascertain using remote sensing data and is consequently the largest source of error in estimating ET. Errors in H are introduced in the estimation of T_s, aerodynamic resistance parameters, and through use of area-mean T_a because information is lacking about the spatial variability of T_a. Accurate calculation of energy and water fluxes requires that errors in estimate of surface temperature to be small compared to the differences in surface-air temperatures. The larger the difference in surface-air temperature, the more insensitive the method is to errors in remotely sensed estimates of T_s. Fortunately, in this study, the difference in the surfaceair temperature at the flux stations using a spatially distributed T_s and the mean T_a is large (mean difference=11°C). Typically, this difference is on the order of less than 2°C. Kustas *et al.* (1994a) described several methods of estimating roughness parameters, but these techniques require additional measurements of the wind profile and vertical velocity not available in this study. Alternatively, z_{0m} and z_{0h} were defined through an iterative process using mean values from

the flux stations and constraints imposed by the literature. Although many solutions for z_{0m} and z_{0h} exist at any one location, very few solutions exist that can universally satisfy a solution for H at all pixels within the area of interest. Once the roughness parameters were defined, the process was inverted to solve for H at each pixel using remote sensing estimates of a_{BB}, e_s, T_s, R_n, and G_{sfc} (Figure 4.8(c)). The lowest values of H, ranging from about 50 to 350 Wm^{-2} (mean = 245 Wm^{-2}) occur in the playa and lake plain. Higher values occur on the alluvial fans, ranging from about 350 to 600 Wm^{-2}, with the highest being in the middle reaches of the fans. Because of the difficulties noted above, these values may be in error by as much as±40%, but the spatial structure of the sensible heat flux field and its relationship to the geomorphology, soils, and vegetation of the valley is at least qualitatively correct.

The latent heat flux represents the energy that is exchanged between the land and the atmosphere in vapor form during the processes of evaporation and transpiration. *LE* is calculated as a residual of the energy balance and as such, errors in the estimation of R_n, G_{sfc}, and H are compounded in the estimation of LE. In this case, errors in *LE* can be on the order of one or even two orders of magnitude. For example, the mean *LE* measured at the flux stations is about 10 Wm^{-2}, whereas the mean remotely sensed estimate of *LE* is about 100 Wm^{-2}. In general, however, the sign of the flux vector remains true, and maps of *LE* are believed to represent the relative magnitude of spatial variations in *LE*. Although it is computationally possible to obtain negative values for *LE*, theoretically this is not reasonable for mid-morning in the desert; values less than zero should be treated as small fluxes.

In general, *LE* is low throughout the study area (Figure 4.8(d)). Even though Figure 4.8(d) only depicts instantaneous values of *LE*, the remotely sensed estimates of *LE* are consistent with measures of the annual cycle of actual ET derived from the flux stations, which indicate very low to negligible values of actual ET at the time of year corresponding to acquisition of the remote sensing data (Malek *et al.* 1997). *LE* is lowest (negligible) on the middle part of the fans. These areas are covered predominantly with small shrubs (shadscale), forbs (winterfat), and grasses. In this reach of the fan, the depth of the water table depth is greatest (Figure 4.9). This combination of small plants and deep water table results in low ET. In June 1994, when the TM image was acquired, the grasses were well into senescence having already browned-out. Intermediate values of *LE* occur on the upper and lower reaches of the fans and in the outer areas of the lake plain where the larger shrubs (greasewood and sagebrush) occur in higher density. Presumably, the water table is within reach of these deep-rooting plant species. *LE* is greatest in the central part of the lake plain, particularly in the playa. This *LE* is dominated by evaporation because these areas have the lowest vegetation density. Investigation of the soil in the playa revealed significant moisture just below the surface. The fine-grained soil is conducive to capillary rise and recharge

of water from below in response to water potential gradients imposed by surface evaporation. Rangeland scientists refer to this as the "inverse texture principle"; although fine-grained soils can hold more water, they also wick more water compared to sandier soils.

4.5 Summary and conclusions

Making ground-based measurements of hydrologic and micrometeorologic processes is nearly routine today. These measurements, however, provide little information about regional hydrology because the measurements are only representative of a very small area around the sensors. The variability of the land surface precludes using these measurements to make inferences about processes that occur over an area of a hectare, much less the size of an entire valley. Recent developments have demonstrated an increasing capability to estimate the spatial distribution of hydrologic surface fluxes for very large areas utilizing remote sensing data. The Goshute Valley research program provided the necessary instrumentation to further study the use of remote sensing for measuring surface properties and processes, and to relate these processes to the geomorphological setting of the Great Basin.

The distribution of R_n is consistent with the spatial distribution of vegetation type and density. Generally, R_n is lowest in the playa, the central part of the lake plain, and in the middle reaches of the alluvial fans where vegetation density is lowest. In contrast, R_n is highest on the lower and upper reaches of the fans where vegetation density is highest and dominated by shrubs. G_{sfc} is lowest on the alluvial fans and highest in the lake plain. Because the emissivity of bare soil is lower than that of vegetation, areas of lower vegetation density have a cooler surface temperature and, therefore, a lower H. Thus, H is lowest in the playa and increases gradually in the lake plain outward from the playa to the surrounding fans. Somewhat surprisingly, H is highest in the middle reaches of the fans. Conversely, LE is lowest in the middle reaches of the fans where plants are small presumably because the water table is deep, resulting in low evaporation and transpiration. Highest ET occurs in the center of the valley, particularly in the playa, where little to no vegetation occurs. We infer that in the playa evaporation is relatively high because of a shallow water table and the presence of silty clay soil capable of large capillary water movement. In contrast, intermediate values of LE associated with large shrubs are presumably dominated by transpiration.

This investigation was an experimental attempt to estimate instantaneous regional-scale ET using Landsat TM data. Well-developed pointbased models of surface energy and water balance fluxes were applied to individual pixels of the remotely sensed image. The method requires certain assumptions be made about the spatial distribution of several physical parameters. In some instances, remotely sensed proxies were used; in others, spatial averages were

assumed. For example, (a) we assumed that downwelling short- and longwave radiation and one or more aerodynamic resistance parameters are invariant over the study area, (b) that for nadirviewing instruments the land surface behaves as a Lambertian reflector, and (c) that the emissivity-NDVI relationship is indeed linear.

Although the current state of the technique is imprecise, the results herein and in other studies suggest that it is possible to utilize remote sensing to scale from point measurements of environmental state variables to regional estimates of energy exchange to obtain an understanding of the spatial relationship between these fluxes and landscape variables. The newest generation of thermal remote sensing instruments [i.e. Enhanced Thematic Mapper Plus (ETM+), Moderate-resolution Imaging Spectroradiometer (MODIS), Advanced Spaceborne Thermal Emission and Reflection (ASTER) Radiometer, Multispectral Thermal Imager (MTI)] offer much potential for improving this technique. Some of these instruments possess more than one thermal channel, which facilitiates using a split window technique for atmospheric correction. Shorter revisit time for surface energy studies is another advantage provided by some instruments. ETM+, ASTER, and MTI are all higher-resolution thermal sensors than TM band 6. In addition to improvements in sensor characteristics, new techniques are also being developed to address spatial heterogeneity of surface properties at the subpixel scale. The combination of these factors should lead to improved surface energy balance estimation.

4.6 Acknowledgments

This research was conducted under a subcontract to NASA Marshall Space Flight Center from grant NAS5–2043 from NASA's Earth Science Enterprise to Utah State University, G.Bingham, PI.We wish thank C.Justus of Computer Services Corporation for use of his SPECTRL radiative transfer model and for insightful discussions regarding its use and implementation in this research. We also thank E.Malek and G.McCurdy for flux station data, L.Hipps for the emissivity data, D.Groeneveld and C.Elvidge for vegetation density data, J.Boettenger for plant identification, B.Howell for assistance in image processing, W.Crosson for discussions regarding this research, and reviewer's comments that improved this work.

References

Brest, C.L. and S.N.Goward (1987) Deriving surface albedo measurements from narrow band satellite data. *Int. J. Remote Sens.* 8:351–67.

Brutsaert, W. (1975) On a derivable formula for long-wave radiation from clear skies. *Water Resour. Res.* 14:742–4.

Brutsaert, W.H. (1982) *Evaporation into the Atmosphere.* D Reidel, Norwell, MA, 199 pp.

Businger, J.A., J.C.Wyngaard, Y.Izumi, and E.F.Bradley E.F. (1971) Flux-profile relationships in the atmospheric surface layer. *J. Atmos. Sci.* 28:181–9.

Chamberlain, A.C. (1968) Transport of gases to and from surfaces with bluff and wave-like roughness elements. *Q. J. R. Meteorol. Soc.* 94:318–32.

Chevez, P.S. (1989) Radiometric calibration of Landsat Thematic Mapper multispectral images. *Photogramm. Eng. Remote Sens.* 55:1285–94.

Choudhury, B.J. (1991) Multispectral satellite data in the context of land surface heat balance. *Rev. Geophys.* 29:217–36.

Clothier, B.E., K.L.Clawson, P.J.Pinter, Jr., M.S.Moran, R.J.Reginato, and R.D.Jackson (1986) Estimation of soil heat flux from net radiation during the growth of alfalfa. *Agric. For. Meteorol.* 37:319–29.

Dyer, A.J. (1974) A review of the flux-profile relationships. *Bound–Layer Meteorol.* 7: 363–72.

Garratt, J.R. and B.B.Hicks (1973) Momentum, heat and water vapour transfer to and from natural and artificial surfaces. *Q. J. R. Meteorol. Soc.* 104:491–502.

Geotz, S.J., R.N.Halthorne, F.G.Hall, and B.L.Markham (1995) Surface temperature retrieval in a temperate grassland with multiresolution sensors. *J. Geophys. Res.* 100:25397–25410.

Gilabert, M.A., C.Conese, and F.Maselli (1994) An atmospheric correction method for the automatic retrieval of surface reflectances from TM images. *Int. J. Remote Sens.* 15:2065–86.

Goodin, D.G. (1995) Mapping the surface radiation budget and net radiation in a Sand Hills wetland using a combined modeling/remote sensing method and Landsat Thematic Mapper Imagery. *Geocarto Int.* 10:19–29.

Gordon, H.R. (1978) Removal of atmospheric effects from satellite imagery of the oceans. *Appl. Opt.* 17:1631–36.

Groeneveld, D.P. (1997) Vertical point quadrat sampling and an extinction factor to calculate leaf index. *J. Arid Environ.* 36:475–85.

Hall, F.G., K.F.Huemmrich, S.J.Goetz, P.J.Sellers, and J.E.Nickerson (1992) Satellite remote sensing of surface energy balance: success, failures, and unresolved issues in FIFE. *J. Geophys. Res.* 97:19061–89.

Hanks, R.J. and G.L.Ashcroft, G.L. (1980) *Applied Soil Physics*. Springer-Verlag, New York, 159 pp.

Hill, J. and B.Sturm (1991) Radiometric correction of multitemporal Thematic Mapper data for use in agricultural land-cover classification and vegetation monitoring. *Int. J. Remote Sens.* 12: 1471–91.

Hipps, L.E. (1989) The infrared emissivities of soil and *Artemisia tridentata* and subsequent temperature corrections in a shrub-steppe ecosystem. *Remote Sens. Environ.* 27: 337–42.

Högström, U. (1988) Non-dimensional wind and temperature profiles in the atmospheric surface layer: a re-evaluation. *Bound.–Layer Meteorol.* 42:55–78.

Humes, K.S., W.P.Kustas, and M.S.Moran (1994) Use of remote sensing and reference site measurements to estimate instantaneous surface energy balance components over a semiarid rangeland watershed. *Water Resour. Res.* 30: 1363–73.

Humes, K.S., W.P.Kustas, and T.J.Schmugge (1995) Effects of soil moisture and spatial resolution on the surface temperature/vegetation index relationships for a semiarid watershed. In *Proceedings, of the Conference on Hydrology,* Boston, MA, American Meteorology Society, pp. 147–51.

Jackson, R.D., P.J.Pinter, Jr, and R.J.Reginato (1985) Net radiation calculated from remote multispectral and ground station meteorological data. *Agric. For. Meteorol.* 35:153–64.

Jackson, R.D., M.S.Moran, P.N.Slater, and S.F.Biggar (1987) Field calibration of reference reflectance panel. *Remote Sens. Environ.* 25:145–58.

Justus, C.G. and M.V.Paris (1985) A model for solar spectral irradiance and radiance at the bottom and top of a cloudless atmosphere. *J. Clim. Appl. Meteorol.* 24: 193–205.

Justus, C.G. and M.V.Paris (1987) Modeling solar spectral irradiance and radiance at the bottom and top of a cloudless atmosphere. Unpublished Technical Report, Georgia Institute of Technology, Atlanta, GA, 99 pp.

Kohsiek, W., H.A.R.De Bruin, H. The, and B.van der Hurk (1993) Estimation of the sensible heat flux of a semi-arid area using surface radiative temperature measurements. *Bound.–Layer Meteorol.* 63:213–30.

Kustas, W.P. (1990) Estimates of evapotranspiration with a one- and two-layer model of heat transfer over partial canopy cover. *J. Appl. Meteorol.* 29:704–15.

Kustas, W.P., B.J.Choudhury, M.S.Moran, R.J.Reginato, R.D.Jackson, L.W.Gay, and H.L.Weaver (1989a) Determination of sensible heat flux over sparse canopy using thermal infrared data. *Agric. For. Meteorol.* 44:197–216.

Kustas, W.P., R.D.Jackson, and G.Asrar (1989b) Estimating surface energy-balance components from remotely sensed data. In G.Asrar (ed.) *Theory and Applications of Optical Remote Sensing.* John Wiley, New York, 743 pp.

Kustas, W.P. and C.S.T.Daughtry (1990) Estimation of the soil heat flux/net radiation ratio from spectral data. *Agric. For. Meteorol.* 49:205–23.

Kustas, W.P., J.H.Blanford, D.I.Stannard, C.S.T.Daughtry, W.D.Nichols, and M.A.Weltz (1994a) Local energy flux estimates for unstable conditions using variance data in semiarid rangelands. *Water Resour. Res.* 30:1351–61.

Kustas, W.P., M.S.Moran, K.S.Humes, D.I.Stannard, P.J.Pinter Jr, L.E.Hipps, E.Swiatek, and D.C.Goodrich (1994b) Surface energy balance estimates at local and regional scales using optical remote sensing from an aircraft platform and atmospheric data collected over semiarid rangelands. *Water Resour. Res.* 30: 1241–59.

Kustas, W.P., E.M.Perry, P.C.Doraiswamy, and M.S.Moran (1994c) Using satellite remote sensing to extrapolate evaportranspiration estimates in time and space over a semiarid rangeland basin. *Remote Sens. Environ.* 49:275–86.

Kustas, W.P., R.T.Pinker, T.J.Schmugge, and K.H.Humes (1994d) Daytime net radiation estimated for a semiarid rangeland basin from remotely sensed data. *Agric. For. Meteorol.* 71:337–57.

Kustas, W.P., M.S.Moran, R.D.Jackson, L.W.Gay, L.F.W.Duell, K.E.Kunkel, and A.D.Matthias (1990) Instantaneous and daily values of the surface energy balance over agricultural fields using remote sensing and a reference field in an arid environment. *Remote Sens. Environ.* 32: 125–41.

Kustas, W.P., L.E.Hipps, and K.S.Humes (1995) Calculation of basin-scale surface fluxes by combining remotely-sensed data and atmospheric properties in a semiarid landscape. *Bound.–Layer Meteorol.* 73: 105–24.

Labed, J. and M.P.Stoll (1991) Spatial variability of land surface emissivity in the thermal infrared band: spectral signature and effective surface temperature. *Remote Sens. Environ.* 38:1–17.

Lhomme, J.-P., N.Katerji, A.Perrier, and J.-M.Bertolini (1988) Radiative surface temperature and convective flux calculation over crop canopies. *Bound.–Layer Meteorol.* 43:383–92.

Malek, E. (1994) Calibration of the Penman wind function using the Bowen ratio energy balance method. *J. Hydrol.* 163:289–98.

Malek, E. and G.E.Bingham (1997) Partitioning of radiation and energy balance components in an inhomogeneous desert valley. *J. Arid Environ.* 37:193–207.

Malek, E., G.Bingham, D.Or, and G.McCurdy (1997) Annual mesoscale study of water balance in a Great Basin heterogeneous desert valley. *J. Hydrol.* 191: 223–44.
Markham, B.L. and J.L.Barker (1986) Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. *EOSAT Landsat Technical Notes,* Vol. 1, pp. 3–8.
Markham, B.L. and J.L.Barker (1987a) Thematic Mapper bandpass solar exoatmospheric irradiances. *Int.J. Remote Sens.* 8:517–23.
Markham, B.L. and J.L.Barker (1987b) Radiometric properties of U.S. processed Landsat MSS data. *Remote Sens. Environ.* 22:39–71.
Markham, B.L., R.N.Halthore, and S.J.Geotz (1992) Surface reflectance retrieval from satellite and aircraft sensors: Results of sensor and algorithm comparisons during FIFE. *J. Geophys. Res.* 97:785–95.
Massman, W.J. (1987) Heat transfer to and from vegetated surfaces: an analytical method for bulk exchange coefficients. *Bound.–Layer Meteorol.* 40:269–81.
Monteith, J.L. (1973) *Principles of Environmental Physics.* Elsevier, New York, 241 pp.
Moran, M.S., R.D.Jackson, L.H.Raymond, L.W.Gay, and P.N.Slater (1989) Mapping surface energy balance components by combining Landsat Thematic Mapper and ground-based meteorological data. *Remote Sens. Environ.* 30: 77–87.
Moran, M.S. and R.D.Jackson (1991) Assessing the spatial distribution of evaoptranspiration using remotely sensed inputs. *J. Environ. Qual.* 20:725–37.
Moran, M.S., T.R.Clarke, W.P.Kustas, M.Weltz, and S.A.Amer (1994) Evaluation of hydrologic parameters in a semiarid rangeland using remotely sensed spectral data. *Water Resour. Res.* 30:1287–97.
Nicodemus, F.E., J.C.Richmond, J.J.Hsia, I.W.Ginsberg, and T.Limperis (1977) Geometrical considerations and nomenclature for reflectance. *NBS Monograph 160,* National Institute of Standards and Technology, Gaithersburg, MD, 52pp.
Oke, T.R. (1987) *Boundary Layer Climates.* 2nd edn, Routledge, New York, 435 pp.
Ottlé, C. and D.Vidal-madjar (1994) Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-MOBILHY region. *J. Hydrol.* 158:241–64.
Panofsky, H.A. and J.A.Dutton (1984) *Atmospheric Turbulence.* John Wiley, New York, 397 pp.
Paulson, C.A. (1970) The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteorol.* 9: 857–61.
Reginato, R.J., R.D.Jackson, and P.J.Printer (1985) Evaporation calculated from remote multispectral and ground station meteorological data. *Remote Sens. Environ.* 19:75–89.
Saunders, R.W. (1990) The determination of broad band surface from AVHRR visible and near-infrared radiances. *Int. J. Remote Sens.* 11:49–67.
Schott, J.R. and W.J.Volchok (1985) Thematic Mapper thermal infrared calibration. *Photogramm. Eng. Remote Sens.* 51:1351–8.
Starks, P.J., J.M.Norman, B.L.Blad, E.A.Walter-Shea, and C.L.Walthall (1991) Estimation of shortwave hemispherical reflectance (albedo) from bidirectionally reflected radiance data. *Remote Sens. Environ.* 38:123–34.
Stewart, J.B., W.P.Kustas, K.S.Humes, W.D.Nichols, M.S.Moran, and H.A.R.De Bruin (1994) Sensible heat flux-radiometric surface temperature relationship for eight semiarid areas. *J. Appl. Meteorol.* 33:1110–17.
Sturm, B. (1981) The atmospheric correction of remotely sensed data and the quantitative determination of suspended matter in marine water surface layers, In A.P.Cracknell (ed.) *Remote Sensing in Meteorology, Oceanography and Hydrology.* John Wiley, New York, pp. 163–96.

Sugita, M. and W.Brutsaert (1993) Comparison of land surface temperatures derived from satellite observations with ground truth during FIFE. *Int. J. Remote Sens.* 14: 1659–76.

Thom, A.S. (1972) Momentum, mass and heat exchange of vegetation. *Q. J. R. Meteorol. Soc.* 98:124–34.

Tueller, P.T. (1994) Great Basin annual vegetation patterns assessed by remote sensing, In S.B.Monsen and S.G.Kitchen (eds) *Proceedings—Ecology and Management of Annual Rangelands*. USDA Forest Service, Intermountain Research Station, General Technical Report INT-GTR-313, USDA Forest Service, Intermountain Research Station, Ogden, UT, pp. 126–31.

Vidal, A., C.Devaux-Ros, and M.S.Moran (1994) Atmospheric correction of Landsat TM thermal band using surface energy balance. In *Proceedings of the ISPRS 6th International Symposium on Physical Measurements and Signatures in Remote Sensing,* Val d' Isere, France, 17–21 January (Avignon: INRA), pp. 159–66.

Vining, R.C. and B.L.Blad (1992) Estimation of sensible heat flux from remotely sensed canopy temperatures. *J. Geophys. Res.* 97: 18951–4.

Wukelic, G.E., D.E.Gibbons, L.M.Martucci, and H.P.Foote (1989) Radiometric calibration of Landsat Thematic Mapper thermal band. *Remote Sens. Environ.* 28: 339–47.

Chapter 5

Coupling thermal infrared and visible satellite measurements to infer biophysical variables at the land surface

Robert R. Gillies and Bekele Temesgen

5.1 Introduction

The measurements of surface soil moisture, net radiation, sensible heat flux, and latent heat flux have many applications in different disciplines such as hydrology, agriculture, meteorology, and forestry. These important biophysical variables are either measured or estimated using the different methods available. For example, surface fluxes can be measured using lysimeters, Bowen ratio and, eddy correlation systems, water balance methods, gas exchange chambers, and micrometeorological methods (Dugas *et al.* 1991). The soil moisture can be measured using gravimetric methods, neutron probes, tensiometers, and hygrometric methods (Schmugge *et al.* 1980). However, point measurements produce reliable results only for the site at which the measurements were taken and are not spatially representative at larger regional scales. Furthermore, the use of point measurements as a typical value for a larger area is likely to introduce significant errors because of the heterogeneity of most land surfaces. Striking differences in these variables occur over short distances even within the same field. For example, Davenport and Hudson (1967) measured evaporation rates from open-water, using evaporimeter dishes, and found a decrease of about 30% in latent heat flux with in a 60 m distance of the upwind edge of a cotton field. Other researchers (Burman *et al.* 1975; Holmes 1970) report similar results.

Remote sensing of certain surface parameters, such as surface radiant temperature and surface reflectance, has enabled researchers to obtain estimates of soil moisture, net radiation, sensible heat flux, and latent heat flux at larger spatial scales. Remote sensing methods produce spatial averages of surface parameters. Therefore, the error that is introduced as a result of this averaging depends on the spatial resolution of the sensors. However, even for the coarser resolutions, spatial averaging of parameters is less likely to result in significant errors compared to taking point measurements as a representative value.

Remote sensing methods also have limitations. Remotely sensed data, for example, are instantaneous and therefore do not account for temporal

variations. However, under stable atmospheric conditions, the instantaneous measurements can be extrapolated to daily values with a reasonable degree of confidence (Jackson *et al.* 1983; Kustas *et al.* 1989). The other obvious problem with remote sensing methods is the difficulty of deriving biophysical variables from the remotely sensed parameters. Although there are many reported methods for deriving biophysical variables from the remotely sensed data, none has yet been accepted as a standard.

This chapter focuses on the use of Soil-Vegetation-Atmosphere-Transfer (SVAT) models in deriving biophysical variables. SVAT models are usually coupled with remotely sensed data so that the biophysical variables are derived using what is referred to as an inverse technique. The inverse technique is a method in which a measured variable is fitted to a simulated (modeled) one such that their equivalence defines a solution (Gillies *et al.* 1997). There are many SVAT models that are used for these purposes. The emphasis in this chapter is on the methods and application of a SVAT model that was developed over many years, at The Pennsylvania State University (Carlson and Boland 1978; Taconet *et al.* 1986; Gillies 1994). From now onwards in this chapter, this model is referred to as the ABL-SVAT, or simply the SVAT model, ABL representing the fact that the SVAT included a dynamic mixing layer as part of the atmospheric boundary layer.

One of the more important applications of the ABL-SVAT model is the derivation of surface soil moisture at a regional scale. The surface soil moisture, in turn, has many uses and applications. In this chapter, we present the methods used to derive the surface soil moisture and the use of the derived surface soil moisture as an index to assess the prevalence of certain diseases, such as malaria and filariasis. The surface soil moisture was used as an index because the mosquito vectors that carry the parasite need standing water for breeding (Crombie *et al.* 1999). Standing water implies higher values of surface soil moisture. Details are presented using results of a recent study that was conducted along the Nile delta in Egypt.

5.2 Theory and background

5.2.1 Surface energy balance

In the determination of biophysical variables of land surfaces, boundary layer models have to be coupled with the remotely sensed measurements. In this process, the most critical step is the determination of the actual radiant energy available at the surface. This is calculated in the SVAT model by summing the various streams of radiation to and from the surface as represented in equation (5.1):

$$R_{\text{net}} = (1 - \alpha_s)R_{\text{solar}} + R_{\text{atmosphere}} - R_{\text{surface}} \tag{5.1}$$

where α_s is the reflectivity of the surface to solar radiation, called albedo, R_{solar} is the incident solar radiation (direct and diffuse) at the land surface,

$R_{atmosphere}$ is the infrared emission from the atmosphere toward the surface, and $R_{surface}$ is the infrared emission from the surface. The sum of these radiation streams represents the amount of energy available (R_{net}—the net radiation) to do work at the land surface. In other words, this energy will be used to power the various processes that drive the land surface processes that ultimately drive the climate system. The question arises, for what processes is this energy used? Ignoring smaller energy components such as friction, electrochemical interactions, and photosynthesis, the main processes are the evaporation of water (directly from the surface as well as from the interior of plants—a process termed transpiration—together these constitute the combined process of evapotranspiration), the flux of sensible heat between the land surface and the atmosphere, and the storage of heat in land and water masses. Therefore, the net radiation can be approximated using equation (5.2):

$$R_{net} = ET + H + G \tag{5.2}$$

where ET is the evapotranspiration from the surface, H the sensible heat, and G is the soil heat flux into the underlying surface (units are typically Wm^{-2}).

The energy flux components of equation (5.2) are related to the surface conditions; specifically, how wet or dry the surface is, to what extent a particular area is vegetated, as well as the properties of the soil. For most natural systems, ET is the largest component where the land surface is wet and/or covered with well-watered plants. The operative word here is wet, because it indicates that soil moisture is a control in the system. This soil moisture content directly affects ET through evaporation from the soil surface and indirectly through the plants that transpire. The ET over any vegetated surface will therefore be a function of (a) the amount of bare surface and its degree of wetness and (b) the extent to which any area is vegetated, as well as on vegetation type and characteristics. Such combined surface characteristics will govern into which processes the radiative energy will be transformed.

5.2.2 Link to remote sensing

Since both equations (5.1) and (5.2) are expressions for R_{net}, it is evident then that the association between the radiation streams, noted in equation (5.1), are closely linked to the energy flux terms, as given in equation (5.2), which in turn are linked to the surface characteristics of the land.

Our interest is in determining the surface energy fluxes (right-hand side of equation 5.2) at the regional scale. Passive remote sensing provides distributed values of most of the radiation streams given in equation (5.1). The question then is, how to utilize these values in order to obtain estimates of surface energy fluxes? This is achieved by coupling a ABL-SVAT model with the remotely sensed data. The coupling is necessary because the radiation

streams, the components of which can be remotely sensed, drive the surface energy fluxes.

In terms of the radiation streams given in equation (5.1), R_{solar} lies in the solar-reflective spectral range where the energy supplied at the surface by the sun exceeds that emitted by the earth. $R_{atmosphere}$ and $R_{surface}$ are emitted thermal radiation from the atmosphere and the earth, respectively. These radiation sources define surface properties of reflectance and temperature.

As stated in Section 5.2.1, to first approximation, the amount of vegetation cover and wetness of the surface are two of the most important factors in estimating the surface fluxes. The issue, now is how to use the reflected component of R_{solar} and the emitted component ($R_{surface}$) to determine the extent to which an area is vegetated and the degree of wetness of the soil. The reflective properties of a surface and the relative contribution of specific wavebands of radiation are used to detect the presence of vegetation. Such modulation ratios, generally referred to as vegetation indexes, are commonly used to exploit the spectral signatures of vegetation. The "Normalized Difference Vegetation Index" (NDVI) is just such an index:

$$\mathrm{NDVI} = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1} \tag{5.3}$$

where ρ_1 and ρ_2 are the measured surface reflectances in the visible red and near-infrared spectral regions, respectively. (Wavelength ranges are typically $\rho_1=\rho_{red}=\rho_{0.6–0.7\mu m}$, $\rho_2=\rho_{NIR}=\rho_{0.7\text{-}1.1\mu m}$.) Clearly, $-1 \leq \mathrm{NDVI} \leq 1$, with highly vegetated surfaces characterized by values of NDVI ~ 0.8; this is because, for healthy vegetation, ρ_{NIR} is large and ρ_{red} is small, and NDVI ~ 0.1 for bare soil surfaces. NDVI, though, is not an intrinsic physical quantity and there is often confusion when trying to relate NDVI to physical or biophysical variables, such as Leaf Area Index (LAI) or the fractional vegetation cover (*Fr*). Likewise, there is similar confusion as to what constitutes a fully vegetated surface. To clarify such definitions and relationships, the reader is directed to the discussion in Carlson and Ripley (1997). NDVI, as defined in equation (5.3) (as measured at the surface), is often referred to as the "true" NDVI as compared to what is actually measured from space, where differential attenuation between the red and near-infrared wavelengths occurs due to radiative interaction with atmospheric aerosols. While the definition of NDVI is undoubtedly important (Liu and Huete 1995; Leprieur *et al.* 1996), Carlson and Ripley (1997) show that the relationship between NDVI and *Fr* holds equally well whether NDVI is corrected for atmospheric attenuation or not. The practical significance of this will become apparent.

The emitted component of the earth ($R_{surface}$) defines the surface property of temperature, more correctly defined as the surface radiant temperature (T_o). Radiant temperature over a non-vegetated surface exhibits sensitivity with soil water content (Idso *et al.* 1975); this is due to the greater amount of energy required to change the water from a liquid to a vapor state. For any

increase in water content, a greater amount of evaporation will take place that will be manifest as a lower radiant temperature. However, what is meant by what has been loosely described as soil water content? Obviously, any definition of soil water content should be tied closely to evaporation. A familiar term that does this and is extensively used in the literature is the soil moisture availability (M_o). Treatment of this very important variable goes beyond the scope of this chapter but an outline is given in Appendix 5.A1. The soil moisture availability then links the evaporation at the surface to the surface radiant temperature and, through the radiation laws, to $R_{surface}$.

In summary, the argument posed is that certain biophysical variables cannot be directly remotely sensed and have to be interpreted through alternate means. The crux of the argument is simply relating the radiation streams to the fluxes with the land surface properties: starting from remotely sensed measurements of reflectivity and emission one can calculate NDVI and To, from which M_o and *Fr* are inferred through coupling the remotely sensed surface parameters to a ABL-SVAT model, while, at the same time, the scheme can also output *H* and ET. In the literature, R_{solar} and $R_{surface}$ are referred to as primary remotely sensed variables, whereas NDVI and T_o are considered secondary remotely sensed variables. The groupings (*Fr,* M_o, *H*, and ET) are customarily referred to as biophysical variables.

Finally, it is important to realize that an inherent complexity exists in using a radiant temperature (Norman and Becker 1995) over a heterogeneous surface comprising vegetation and bare soil. While the temperature of the bare soil will vary with the soil water content, the temperature of healthy, well-watered vegetation will not change by much. Consequently, as fractional vegetation cover increases in any spatial domain, the radiant temperature measurement from the vegetation increasingly modulates the overall radiant temperature, and thereby becomes rather insensitive to the actual soil surface moisture conditions. Therefore, in any scheme that uses radiant temperature as a prognostic variable, such circumstances will have to be considered.

5.3 Determination of biophysical variables using the triangle method

5.3.1 Satellite data depiction and theoretical interpretation

As is the case for the derivation of biophysical fields, an intermediate physical model (^{i}ABL-SVAT) is generally applied in the interpretation of the secondary remotely sensed fields. This has the potential to introduce uncertainty in the derived products: the triangle technique reduces such uncertainties as it identifies physical limits found in the secondary remotely sensed data that are subsequently used to constrain the model solution for obtaining the biophysical variables. The generated biophysical fields are surface moisture

availability (M_o), fractional vegetation cover (*Fr*), sensible heat flux (*H*), and evapotranspiration (ET).

The triangle method is by definition, a multispectral method that combines measurements of surface radiant temperature, T_o, and reflectance (red and near infrared). The reflectance measurements are used to calculate the NDVI. NDVI is subsequently plotted as a function of T_o and the observed association between the two variables is illustrated in Figure 5.1.

The physical limits that constrain the solution for the biophysical variables are defined from the scatterplot, shown at high resolution in Figure 5.1. It relies theoretically upon observing the complete spectrum of *Fr* and M_o at the ground surface. By definition then, the observations must be bounded by their physical limits and will manifest themselves as boundaries—visually characterized in the observations by "distinct" edges.

Such edges for *Fr* are straightforward, where a low NDVI extreme and a narrow vertex at the high NDVI extreme represent zero (bare soil) to 100% (complete canopy) vegetation cover, respectively. (The relationship between NDVI and *Fr* is non-linear and is described in a later section.) The edges associated with the limits of M_o are somewhat more difficult to visualize and

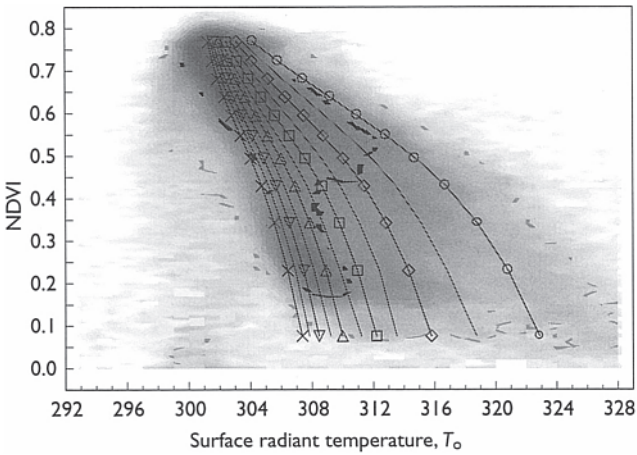


Figure 5.1 NS001-Thematic Mapper Simulator derived T_o-NDVI scatterplot (gray spectral scaling) at a 5-m spatial resolution for a 7×3 km² area of the Mahantango Watershed, Pennsylvania, July 18, 1990, 11:45 h LST. T_o is in Kelvin. Isopleths representing moisture availability index M_o are overlaid with the legend, o=0.0 ("warm" edge), ◊=0.2, ˇ=0.4, Δ=0.6, Δ0.8, and×=1.0 ("cold" edge). A relatively few number of observations at temperatures higher than the warm edge and lower than the cold edge exist due to terrain effects (slope and aspect) which are not considered in the analysis that follows. The consequence of terrain on the NDVI/T_o relationship is detailed in Temesgen (2001).

describe. This is due to (a) the nature of the spectral plot, as shown in Figure 5.1, which makes it difficult to isolate relatively few extraneous pixels that mask the edge, and (b) the physics specifying the temperature soil water content (M_o) relationship. Careful observation (with the aid of the model output of M_o as overlaid) helps delineate (a). For (b), the work of Idso *et al.* (1975), who describes the physics, is translated by Goward *et al.* (1985), Price (1990), and Carlson *et al.* (1994), who apply the theory to NDVI/To space. The relevant point that emerges from their interpretation of the physics is that the warmest pixels, despite variations in *Fr*, correspond to the driest surfaces and define an edge—the so-called "warm edge" that represents a minimum limit of M_o. At the other extreme, the opposite applies where a "cold edge" represents a maximum limit of M_o.

From theoretical considerations alone, within the domain of the scatterplot there must exist all possible permutations of *Fr* and M_o. The issue, then, is to map the domain appropriately. This is where the remote observations are coupled to a physically based algorithm (usually a model) and the conceptual interpretation (biophysical variables) the result. A review and categorization of various techniques (for determining evapotranspiration) is given by Kustas and Norman (1996).

5.3.2 Exploiting the NDWI/T_o relationship with the triangle method

A first step in the process is to establish the relationship between *Fr* and NDVI. First NDVI is scaled between its fully vegetated (maximum NDVI) and bare soil (minimum NDVI) values creating a re-scaled NDVI or N^* where

$$N^* = \frac{\mathrm{NDVI} - \mathrm{NDVI_o}}{\mathrm{NDVI_s} - \mathrm{NDVI_o}} \tag{5.4}$$

in which $\mathrm{NDVI_s}$ and $\mathrm{NDVI_o}$ are, respectively, the maximum and minimum values of NDVI. N^* varies from 0 to 1.

Fr is established from a transform between N^* and *Fr*. This relationship was obtained by fitting the ABL-SVAT model output to satellite data along the "warm edge" with M_o set to zero for two conditions: 100% vegetative cover with the maximum NDVI (known a priori) and bare soil conditions knowing the minimum NDVI (Gillies and Carlson 1995; Gillies *et al.* 1997). The procedure involved using ancillary data (including a morning sounding, vegetation and soil type information); whereupon, the model was tuned until the modeled and measured T_o closely matched for both cases. The model was subsequently run in increments of fractional vegetation cover so that $T_o = f(Fr)$. By representing the warm edge as a regression of N^* against T_o, $T_o = f(Fr)$, the two relationships were used to associate N^* with *Fr*. Applying

this method to satellite data from the Mahantango Creek watershed, PA, Konza Prairie, KS, and Newcastle-on-Tyne, United Kingdom, a simple relationship (Gillies *et al.* 1997) was derived, where

$$Fr = N^{*2} \tag{5.5}$$

This is similar to a relationship by Choudhury *et al.* (1994) where they related the LAI to NDVI. Other independent support (radiation transfer calculations) for this simple square root relationship is to be found in Carlson and Ripley (1997). Of further significance, their work suggests that this scaling eliminates errors in the reflectance data attributed to sensor calibration and atmospheric correction. With the Fr/N^* relationship, it is a simple matter to transform the remote observations from NDVI/T_o space to Fr/T_o space.

The second step is to derive M_o. The same iteration scheme as described earlier for M_o set to zero is adopted, with the exception that M_o is now also cycled over its theoretical range (i.e. 0–1). The output of this process is a matrix of data that assigns the modeled biophysical variable (M_o) to the remote measurements (T_o), but most importantly, the biophysical variable is constrained by the physical limits implied by the remote observations as evidenced by the M_o overlay in Figure 5.1. This matrix can be expanded to include model derived sensible heat (H) and ET, which are model output variables. The only step left in the process is to mathematically derive the coefficients necessary to fit the data, for example, ET=$f(Fr, M_o)$. The logistics for generating relationships for M_o, ET, and H are given in Appendix 5.A2.

An alternate modification to the triangle method negates the requirement of iterating over M_o as previously described. A temperature normalization procedure (Owen *et al.* 1998) uses a scaled temperature for both the model and remotely sensed values of T_o.

Similar to N^*, this scaled temperature, $\hat{T}$, is expressed as

$$\hat{T} = \frac{T_o - T_o(Fr = 1, M_o = 1)}{T_o(Fr = 0, M_o = 0) - T_o(Fr = 1, M_o = 1)} \tag{5.6}$$

so that the temperature, like NDVI, is scaled between its maximum and minimum values. This is done between the maximum value for bare soil and the minimum value for a wet vegetated surface. The reader will notice that knowledge of the vegetation fraction is crucial to the method. Unlike the case for the NDVI, this scaling reduces but does not eliminate calibration and correction issues. The advantage of normalization, for example, is that it permits the application of a so-called nomogram for a satellite scene for which the necessary data to initialize the SVAT model was not available—reference to this is made in Section 5.4.

One final point to be made in all of this, is that because of errors inherent in calculating T_o, and because the triangle narrows along the temperature axis, the usable area of the triangle domain is below 80% *Fr* (Gillies and Carlson 1995).

5.3.3 Further explanation and information about the biophysical variables

Fractional vegetation cover (Fr)

The derivation of *Fr* is described in the previous section. Though *Fr* is exclusively a function of NDVI, the extremes of NDVI used to calculate N^* require information taken from the scatterplot. In the event that a scatterplot is not complete, a value for $NDVI_o$ can be extracted from neighboring urban or bare soil pixels. In cases where the NDVI is not expected to represent 100% vegetation in a given scene, the value of $NDVI_s$ can be approximated by extrapolating a polynomial representation of the "warm edge" in T_o/NDVI space into a T_o commensurate with the air temperature. A working value of $NDVI_s$ can then be estimated.

Moisture availability index (M_o)

The surface moisture availability is also related to the surface soil water content. For bare soil, M_o is based on a ratio of the surface soil water content to its value at field capacity above which the soil water evaporates at the potential rate. This parameter is also referred to in the literature as *b*. For a vegetation canopy, M_o continues to refer to the soil below the canopy. Soil water through the vegetation itself is extracted from the lower of the ABL-SVAT's two soil layers. Validation of the triangle method as conducted by Gillies *et al*. (1997) found good agreement between the gravimetric measurements and NS001 derived estimates obtained from FIFE (First ISLSCP (International Satellite Land Surface Climatology Project) Field Experiment) and Walnut Gulch, AZ. However, linking tangible estimates of soil water content from satellite data has been shown to be problematic due to the high spatial variability of the surface soil moisture, especially under drying conditions (Capehart and Carlson 1997). At such coarse resolution as is indicative of NOAA AVHRR satellite data (1.1 km at Nadir), thermally derived M_o may have little physical meaning beyond that of an index.

Evapotranspiration (ET)

Unlike soil moisture availability, evapotranspiration is an output from the model as is T_o. Isopleths of ET are presented in Figure 5.2. Though one would at first anticipate these isopleths to be similar to M_o in their position

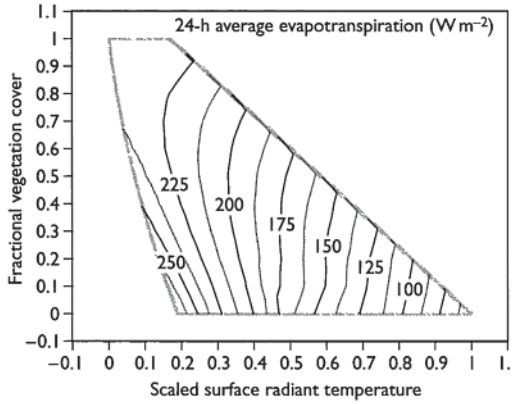


Figure 5.2 Isopleths of 24 h evapotranspiration (Wm-2) for the July 18, 1990, Mahantango NS001 image overlaid on the triangle domain in $\hat{T}/Fr$ space.

in the triangle, the contribution of the vegetation component of transpiration does not cause this to happen. As the deeper rooting zone below the soil is typically not subject to the high temporal and spatial variability in the soil water profile, the vegetation canopy has a moderating effect on ET and the triangle's structure. This is seen in the narrow vertex (high NDVI or *Fr* end) of the triangle representing a small range in surface radiant temperature whose minimum value is close to the air temperature. ET estimates, like the soil moisture, verified well against *in situ* observations at FIFE and Walnut Gulch.

Sensible heat flux (H)

A third output variable from the ABL-SVAT used in the remote sensing method is the sensible heat flux. It should be noted here that the T_o output from the model is not the skin temperature used in sensible heat flux estimate. Also, while the net radiation, ET, and ground fluxes are calculated explicitly within the model, the sensible heat flux is calculated both explicitly *and* residually to maintain the surface energy balance. Validations for *H* in the Gillies *et al.* (1997) study were satisfactory, but did not correlate as well with *in situ* observations as did the soil moisture and latent heat flux.

5.4 The triangle method—algorithmic development and application

5.4.1 Data preparation and algorithms

Given that many users purchase level 1 satellite data for cost reasons, they must first perform corrections and calibrations as needed for their application

Step no.	*Processing step*	*Description*
1	Calculate NDVI	Note that NDVI can be calculated using surface or apparent values of reflectance
2	Plot NDVI and T_o	Graphic visualization of the NDVI/T_o triangle
3	Calculate N^* and Fr	Scale NDVI between its bare soil and dense vegetated values for N^*. Square N^* for Fr
4	Calculate $\hat{T}$	Estimate dry, bare temperature by inspection of the NDVI/T_o triangle. The wet, vegetated surface can be approximated using the NDVI/T_o triangle or by using a large scale air temperature (Owen *et al.* 1998)
5	Set up the solution polynomial	Preliminary sensitivity tests suggest that the triangle is insensitive to initial conditions such that a single polynomial (equation 5.7, Table 5.2) can represent a wide range of surface conditions
6	Spatially map biophysical variable	Code Equation 5.6 (or variants) within appropriate software (e.g. image processing packages) to map M_o, ET, or H Fr can be mapped directly with the simple square root relationship

Table 5.1 Systematic processing steps for the triangle method for multispectral imagery

(e.g. Markam and Barker 1986) before performing reliable data analysis. Ancillary geometric and radiometric calibration data are used to map the data to a suitable projection as radiance. Atmospheric calibration techniques are then necessary to correct the imagery for atmospheric effects to obtain physical units of reflectance and surface radiant temperature. At this point the user may follow steps 1 through 3 and use the custom algorithm (Appendix 5.A2, model operation) that constitutes the triangle method. If the ancillary data needed to perform a SVAT run are not available, temperature normalization is performed as a means to generate values of M_o. The alternative sequence of steps (4 and 5) are laid out in Table 5.1.

5.4.2 Example application—filariasis in the Nile delta

Numerous studies have demonstrated the usefulness of remote sensing data in mapping environmental risk factors such as diurnal temperature variations and land use practices that control, in part, the distribution of tropical diseases, including schistosomiasis, trypanosomiasis, malaria, and filariasis (Rodgers and Randolph 1991; Beck *et al.* 1994; Malone *et al.* 1994). As an example of the application of remote sensing, a case study citing the prevalence of filariasis in the Nile delta of Egypt is highlighted (Crombie *et al.* 1999). Filariasis is a disfiguring and disabling disease caused by a parasite that is carried by mosquitoes. The full effect of the disease is realized when one

considers that over 120 million people worldwide are affected (Ottesen and Ramachandran 1995).

In a prior remote sensing study on filariasis, Thompson *et al.* (1996) used Advanced Very High Resolution Radiometer (AVHRR) satellite data to map the diurnal temperature fluctuations (dT) of the Nile delta as a proxy for soil water content. Statistical comparison of filariasis prevalence and diurnal temperature variation showed a good correlation between the median dT values of 10 km^2 areas and three filariasis prevalence level groupings (5, 15, and 25%). Their study, however, used bulk surface radiant temperature measurements that, by definition, can only proxy soil moisture content in a very crude way because such measurements do not implicitly account for the effect of vegetation. The triangle method, however, differentiates the vegetation radiant temperature from the soil radiant temperature and should in theory be able to resolve a more representative soil water content, and so, improve upon the singular approach of Thompson *et al.* (1996).

The example that follows illustrates how Landsat TM data (August 1990) were used to map the surface moisture availability (M_o) in the southern Nile delta of Egypt (Figure 5.3), a region where filariasis is focally endemic and re-emergent (Harb 1993). The data were subsequently related to filariasis

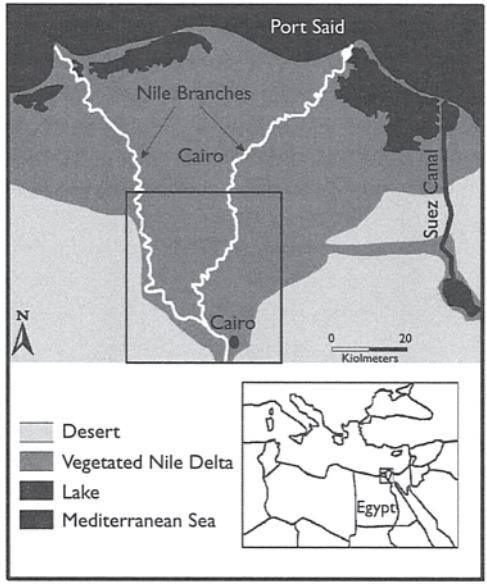


Figure 5.3 Sketch map of the Nile delta, with the southern Nile delta study area identified within the box. The lagoons at the north of the delta (i.e. Bahra el Manzals, Burullus, and Idku) are easily identifiable.

prevalence data collected by the Egyptian Ministry of Health over the period 1985–1991 for 173 villages located in the study area in the southern Nile delta (Crombie *et al.* 1999).

$$M_o = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij}\hat{T}^i Fr^j \tag{5.7}$$

Table 5.2 Polynomial coefficients for polynomial expression for M_o

a_{ij}	$j=0$	$j=1$	$j=2$	$j=3$
$i=0$	2.5661	−2.0602	0.7707	−0.3062
$i=1$	−7.7404	−0.3920	−2.8252	−0.2436
$i=2$	8.5222	7.5239	13.5438	9.2252
$i=3$	−3.3493	−6.1453	−8.1908	−30.0037

Source: Courtesy of Capehart (1996).

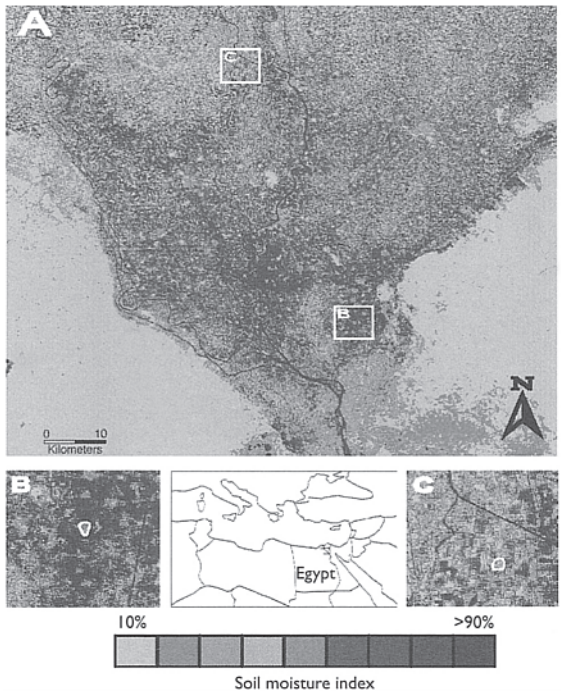


Figure 5.4 Map and subsets of regional surface soil moisture availability (M_o) for the Nile delta study area. Dark colors indicate high soil moisture and light colors indicate low surface soil moisture availability (see Colour Plate XXV).

The temperature normalization triangle method was applied by Crombie *et al.* (1999), for the southern Nile delta (as identified in Figure 5.3), to analyze the region in terms of M_o. The steps followed were those as outlined in Table 5.1. A polynomial (equation 5.7 and Table 5.2), typical of eastern US summertime conditions was used to map M_o for the region. The end result of the process is shown in Figure 5.4. The insets defined as B and C are, respectively, full resolution maps of surface soil moisture availability near the village El-Kashish, which had high soil moisture availability and high filariasis infection rate and, the village El-Aysha, which is surrounded by areas of low surface soil moisture availability and had low filariasis infection rate. The correlation of filariasis infection rate with M_o was not as clearly delineated. The linear correlation coefficient (R^2) between filariasis prevalence in the population and M_o was 0.37. A partial linear correlation with the effects of different villages removed was still about 0.37. The conclusion reached was that there was only a threshold effect in which M_o below 0.2 keeps infections way down.

5.5 Concluding remarks

This chapter presents an analysis method (the triangle method) that uses remote measurements in the infrared and the visible bands to estimate biophysical variables of evapotranspiration (ET), surface soil moisture (M_o), and sensible heat flux (H). The technique is one of information generation to produce spatially representative values of these land surface parameters. The satellite data, as plotted in the NDVI and T_o domain, are used to impose limits on a solution to the surface conditions as prescribed by an ABL-SVAT model. Two approaches to generating the biophysical variables are detailed. The regular approach is direct convolution with a SVAT model to generate the fields. The alternate (temperature scaling) is used when initial conditions are difficult to prescribe or obtain. In this context, it may be possible to use a single polynomial for M_o, ET, or H to represent a wide range of surface climate conditions that compose different land surface types.

The implications for this technique are wide and varied and have direct application in the fields of hydrology, agriculture, meteorology, and forestry. The application of remotely generated data to resolving health issues is, however, a very recent innovation. The triangle method permits reliable assessments of temporal variability in surface moisture as well as across very large areas. The epidemiological significance lies in exploiting such modeling of satellite data that have statistical correlates in the field (e.g. schistosomiasis, Malone *et al.* 1994). The further example cited here is where M_o (defined more rigorously and at higher spatial resolution) is used as a surrogate to identify high-risk situations for contracting the disease filariasis.

Appendix A

5.AI Concept of soil moisture availability

The soil moisture availability (M_s) is defined here as a fraction of "field capacity." It refers to a surface layer of some depth (subscript s) measured from the surface; the subscript o refers to a skin surface layer. In terms of remote sensing and its measurement using thermal infrared radiation, the relevant layer is a skin surface layer. Others may define moisture availability differently, for example, as the fraction of "extractable" soil water content, where some residual amount of soil water is considered to be un-extractable by plants.

In terms of how soil water content is used in the ABL-SVAT model, let us consider soil water content expressed as the volumetric water content (w), where this term represents the volume of soil water content per unit volume of soil. A related expression to w is the atmospheric moisture availability M_a. To explain what is meant by this expression consider equation (5.A1), which is simply a variation of the resistance law but for water vapor flux:

$$L_e E = \rho L_e \frac{\overline{\Delta q}}{R_v} \tag{5.A1}$$

where L_{eE} is the latent heat flux (W m-2), ρ is the density of water (kg m^{-3}), L_e the latent heat of vaporization (Jkg^{-1}), Δq is the vertical time averaged specific humidity difference between two layers, and R_v is the resistance (sm^{-1}) pertaining to the same vertical interval as Δq.

For a flux within the atmospheric surface layer—the layer of "constant" flux—we can write the specific humidity difference as that between the surface and the top of the surface layer (or any other pair of levels in that surface layer), so long as the resistance in the formula pertains to that same vertical interval. It is quite feasible to measure the specific humidity at any level above the surface. The problem, however, is to specify (let alone measure) the specific humidity at the bottom of an imaginary column, that is, at the ground surface, which is hypothetically in the air directly touching the surface. A compromise is to try to calculate the flux of water vapor by specifying specific humidity q at two ends of an elevated vertical column, that is, at two levels above the ground. There are computational problems in doing so. First, though, we want to tie the flux of moisture at the surface to the amount of substrate water and the specific humidity at the ground surface which is at the level in the atmosphere most closely tied to the soil water content. Second, were we to choose one level slightly above the surface and one at the surface, we would be faced with the problem that the bottom layer of the column is indeterminate and there is no equivalent surface specific humidity corresponding to the surface radiant temperature, which can be more easily

measured than specific humidity at the surface. If we did choose to measure the specific humidity at the bottom of the column at a level just above the surface, we would still need to relate the specific humidity at that level to that at the soil surface. Such a system would be difficult to close mathematically.

Having chosen a trial value of q_o (specific humidity at the earth's surface) for the bottom of the layer (equation 5.A2) and q_a, that at the top of the surface layer (subscript a) for the upper value, we are still faced with the problem of determining q_o. To do this, one can resort to an artifact, which is as follows: let us define a "saturation" specific humidity q_s at the surface radiant temperature (T_o). The fractional relative humidity at the surface h_o is thus approximately (noting that relative humidity is strictly defined in terms of vapor pressure) equal to $q_o/(q_s.\, T_o)$ which results in equation (5.A3). This type of formulation, although used by some modelers, still leaves a highly variable and difficult to prescribe unknown parameter, that of h_o:

$$L_e E = \rho L_e \left[\frac{q_o - q_a}{r_{oa}} \right] \tag{5.A2}$$

$$L_e E = \rho L_e \left[\frac{h_o \cdot q_s(T_o) - q_a}{r_{oa}} \right] \tag{5.A3}$$

The problem of linking q_o or h_o to the substrate water content has not yet been rigorously solved, although some expedients for circumventing the whole issue exist. For now, though, let us proceed by defining a "potential" evaporation $L_e E_{pot}$ There are various ways of defining this quantity. For the purposes here, it is simply that evaporation occurring over a saturated surface of radiant temperature T_o, using equation (5.A3), for 100% relative humidity at the surface $h_o = 1$. This is equation (5.A4). We can now define the quantity, M_a, the atmospheric moisture availability, which is simply the ratio of actual evaporation to potential evaporation. Thus, we can write equation (5.A1) as (5.A5). Intuitively, M_o should be related to soil water content, so, to proceed, we set M_a equal to M_o, where M_o is the soil water content as a fraction of field capacity in the soil surface layer. Because there is very little else one can do, this equivalence between M_a and M_o has been adopted by many modelers. There is some justification for such an assumption based on field measurements, but the exact form of the relationship between M_a and M_o undoubtedly varies with soil type, among other things. It is merely a first-order approximation, necessary to make the calculations feasible. Nevertheless, the evaporation will vary from zero for $M_a = 0$ (perfectly dry soil) to potential evaporation for perfectly wet soil. All that can be said for these two somewhat related variables is that both vary between 0 and 1.0, probably in some non-linear fashion with each other. When the soil is absolutely dry, both variables will have the value zero. When at field capacity,

both M_a and M_o will equal one. Simply equating one to the other may not be exactly correct, but the differences can never be great because both are constrained to vary in more or less the same way between the same limits.

$$L_e E_{pot} = \rho L_e \left[\frac{q_s(T_o) - q_a}{r_{oa}} \right] \tag{5.A4}$$

$$L_e E = \rho L_e \left[\frac{M_a(q_s(T_o) - q_a)}{r_{oa}} \right] \quad \text{where } M_a = \frac{L_e E}{L_e E_{pot}} \tag{5.A5}$$

Equation (5.A5) ties evaporation closely to the soil water content. When $M_a = 0$ the evaporation is totally suppressed. For M_o not equal to zero, however, a decrease in M_a, while certainly causing a decrease in evaporation, also corresponds to an increase in the sensible heat flux, and therefore, a rise in surface radiant temperature. However, a rise in surface radiant temperature corresponds to an increase in the saturation specific humidity at the surface $qs(T_o)$ and so, to an increase in the vertical gradient in q and therefore to potential evaporation. Thus, a compensation (negative feedback) exists that attempts to brake the decrease in evaporation caused by the decrease in the soil water content. The result is that a 10% decrease in M_a may correspond in some instances to far less than a 10% decrease in evaporation. Clearly, since a 100% change in M_a causes a 100% change in evaporation, there must be ranges of M_a in which evaporation changes less rapidly with changing M_a and ranges where evaporation changes more rapidly than at an equivalent change in M_a.

5.A2 Generic SVAT model inversion

Introduction

SVAT models vary depending on how they parameterize the soil, atmosphere, and vegetation components. For example, some SVAT models parameterize numerous soil layers while others specify fewer layers, that is, a thin surface layer and a deeper layer. Similarly, for some models, the atmospheric boundary conditions are prescribed, implying no dynamic interaction between the surface and mixing layers, whereas for others there is a dynamic interaction between the surface fluxes and the atmosphere (Gottschalck *et al.* 2001). The differences are even more pronounced when it comes to the parameterization of vegetation in that some use a simple Deardorff s model whereas others use the Ball-Berry method to estimate stomatal resistances. Despite these differences, the end result tends to be the same, that is, the simulation of surface energy fluxes, surface temperature, surface soil moisture, etc.

In the context of the inversion algorithms, we propose a one-dimensional ABL-SVAT model (i.e. one that includes mixing layer physics) to be the best application. Such models require an initial set of conditions that include information on time (year, month, day), location (latitude, longitude, time zone), soils, vegetation, and most importantly, a localized atmospheric sounding. These initial sets of conditions are subsequently used to calculate surface energy fluxes (net radiation, sensible heat, and latent heat) and surface temperature, which are in turn used to estimate the profiles of temperature (subsurface, surface, and air), humidity, and wind speed for the next time step. The process continues over 24 h, or until the time of the satellite overpass/aircraft over-flight, as needed. The outputs from the SVAT model are coupled with remotely sensed data and correlations developed between the different parameters of the model output. These correlations are then used to map such variables as surface temperature, surface soil moisture availability (M_o), and surface energy fluxes (net radiation, sensible heat, and latent heat fluxes) and, in some models, thermal inertia.

Model framework

During the daytime, generic ABL-SVAT models of the type proposed in this chapter have a total of four layers: a mixing layer (whose depth varies throughout the day), a surface layer (about 50m high from the surface), a thin transitional layer, and a soil layer (about 1.5 m deep). During the nighttime, the mixed layer collapses and becomes a residual layer, whereas the lower part of the mixed layer (usually about 200 m) will become a stable boundary layer. The surface layer is a region where turbulent fluxes are nearly constant and therefore similarity theory is applied. The transitional layer is the layer in which molecular diffusivity is significant. The soil layer has two moisture layers: a thin slab of surface layer (about 10 cm) and the bottom layer (root zone layer) and five temperature layers. The bottom layer supplies moisture to plants and also recharges the surface layer during evaporation loss.

Input data

As mentioned before, input parameters vary for most SVAT models. The SVAT models of the types proposed here generally require data on: time and location, soil and vegetation parameters, and atmospheric sounding. Although the type and accuracy of the input data used in the SVAT model can have a significant effect on the accuracy of the estimated parameters, we believe it is the methodology that is used to derive the parameters that is more important. In the following section, we demonstrate the specifics of one such method, known as the "triangle method," for mapping surface energy fluxes and surface soil moisture.

Model operation

Operating the SVAT model involves many steps. This section demonstrates those used by the ABL-SVAT model. However, it is stressed that any user with similar models and output parameters can adopt these steps. The steps are:

1 Obtain fractional vegetation cover (*Fr*) versus surface radiant temperature (T_o) plot from the remotely sensed data. If observations contain a complete spectrum of vegetation amount and surface soil water content, the distribution of the data points resembles a triangle, with its vertex at higher *Fr,* for biophysical reasons.
2 Couple the model with remotely sensed data. The first model run is made for dry, bare soil conditions. The resulting surface temperature is compared against the corresponding value from the remotely sensed data. If there are significant differences, some of the input parameters are refined until agreement is reached. This process is known as tuning the input data and is necessitated by the uncertainties in the true values of the input parameters. Once an agreement is reached, the model run is repeated for increments of *Fr* (until *Fr* = 1.0 is achieved) to define the warm edge. The same procedure is again repeated for wet surfaces (M_o = 1) to define the cold edge.
3 Convolute the model. Once the coupling process is complete, the model has to be convoluted over possible ranges in values of *Fr* (0–100%) and the soil moisture availability (0–1). The results (including surface soil moisture, fractional vegetation cover, surface radiant temperature, and surface energy fluxes) from each model run are written to an output file for the time of the remotely sensed data acquisition. Table 5.A1 demonstrates an example of such an output for the Upper Sheep Creek subbasin at the Reynolds Creek watershed in Southwestern Idaho. The airborne data for the study site was collected on August 18, 1993, at 11:04h MST.
4 Develop relationships between biophysical variables and surface parameters using data from the model output. One of the most important objectives of coupling the SVAT model with remotely sensed data is the development of relationships between the biophysical variables and surface parameters using the model output. These relationships are then used to derive surface energy fluxes and surface soil moisture from the remotely sensed *Fr* and T_o. The following example demonstrates how these relationships are developed from the model outputs of the type shown in Table 7.A1. This particular example is presented to show how M_o is derived from *Fr* and T_o. Similar methods are used to derive surface energy fluxes from M_o and *Fr.* Gillies and Carlson (1995) proposed the

Table 5.AI An example of an output file from the SVAT model for the Reynolds Creek watershed in Idaho. The output was made for a time that corresponds to the remotely sensed data acquisition, that is, 11:00 h MST

	A	B	C
1	Mo	FR	To(C)
2			
3	0.0	0.00	35.34
4	0.0	0.10	31.11
5	0.0	0.20	27.97
6	0.0	0.30	25.63
7	0.0	0.40	23.87
8	0.0	0.50	22.51
9	0.0	0.60	21.39
10	0.0	0.70	20.48
11	0.0	0.80	19.71
12	0.1	0.00	32.53
13	0.1	0.10	29.29
14	0.1	0.20	26.8
15	0.1	0.30	24.89
16	0.1	0.40	23.41
17	0.1	0.50	22.2
18	0.1	0.60	21.2
19	0.1	0.70	20.35
20	0.1	0.80	19.64
21	0.2	0.00	30.52
22	0.2	0.10	27.92
23	0.2	0.20	25.89
24	0.2	0.30	24.3
25	0.2	0.40	23.03
26	0.2	0.50	21.94
27	0.2	0.60	21.03
28	0.2	0.70	20.23
29	0.2	0.80	19.57
	.	.	.
	.	.	.
	.	.	.
84	0.9	0.00	24.66
85	0.9	0.10	23.71
86	0.9	0.20	22.9
87	0.9	0.30	22.13
88	0.9	0.40	21.39
89	0.9	0.50	20.73
90	0.9	0.60	20.16
91	0.9	0.70	19.66
92	0.9	0.80	19.23
93	1.0	0.00	24.34
94	1.0	0.10	23.47
95	1.0	0.20	22.69
96	1.0	0.30	21.94
97	1.0	0.40	21.26
98	1.0	0.50	20.63
99	1.0	0.60	20.08
100	1.0	0.70	19.59
101	1.0	0.80	19.18

following regression equation for M_o:

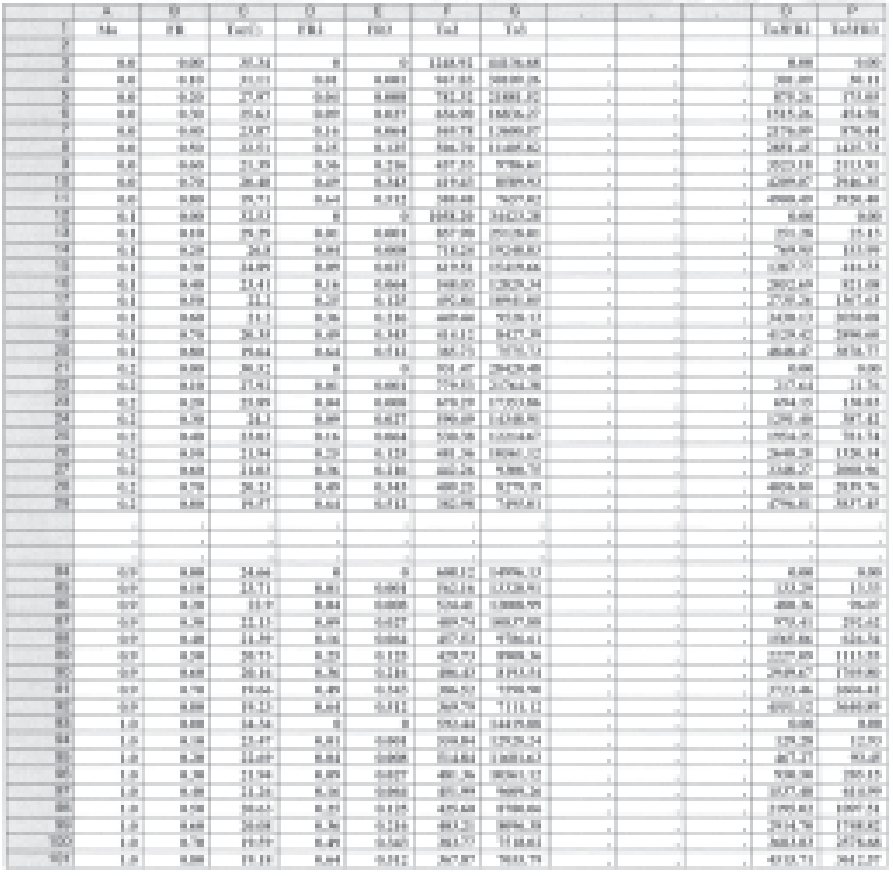
$$M_o = \sum_{p=0}^{3}\sum_{q=0}^{3} a_{pq} T_o^p FR^q \tag{5.A6}$$

where a_{pq} is the matrix of coefficients of the polynomial regression. Equation (5.A6) is expanded as:

$$M_o = a_{00} + a_{01}Fr + a_{02}Fr^2 + a_{03}Fr^3 + a_{10}T_o + \cdots + a_{33}T_o^3 T_o^3 \tag{5.A7}$$

Equation (5.A7) is used to derive the coefficients a00–a33 using the convoluted Fr and M_o values and the calculated T_o values from Table

Table 5.A2 Data points in Table 5.A1 expanded to obtain each of the terms in equation (5.A2). The numbers next to the variables indicate powers



5.A1. Table 5.A2 shows a partial view of the expanded form of Table 5.A1.

Multiple linear regression is then used to derive the polynomial coefficients by entering column A (M_o) as a dependent variable and columns B through P as independent variables. The regression output is shown at the bottom of Table 5.A3 for the example used. The use of these correlation coefficients results in the following equation with some terms omitted.

$$M_o = 38.13 + 185.87Fr + 435.41Fr^2 + 1418.96Fr^3 - 3.46T_o + \cdots - 0.18T_o^3T_o^3 \quad (5.A8)$$

Table 5.A3 Regression coefficients derived from Table 5.A2 using multiple linear regression

	A	B	C	D	E	F	G				O	P
1	Mo	FR	To(C)	FR2	FR3	To2	To3				To3FR2	To3FR3
2												
3	0.0	0.00	35.34	0	0	1248.92	44136.68				0.00	0.00
4	0.0	0.10	31.11	0.01	0.001	967.83	30109.26				301.09	30.11
5	0.0	0.20	27.97	0.04	0.008	782.32	21881.52				875.26	175.05
6	0.0	0.30	25.63	0.09	0.027	656.90	16836.27				1515.26	454.58
7	0.0	0.40	23.87	0.16	0.064	569.78	13600.57				2176.09	870.44
8	0.0	0.50	22.51	0.25	0.125	506.70	11405.82				2851.45	1425.73
9	0.0	0.60	21.39	0.36	0.216	457.53	9786.61				3523.18	2113.91
10	0.0	0.70	20.48	0.49	0.343	419.43	8589.93				4209.07	2946.35
11	0.0	0.80	19.71	0.64	0.512	388.48	7657.02				4900.49	3920.40
12	0.1	0.00	32.53	0	0	1058.20	34423.28				0.00	0.00
13	0.1	0.10	29.29	0.01	0.001	857.90	25128.01				251.28	25.13
14	0.1	0.20	26.8	0.04	0.008	718.24	19248.83				769.95	153.99
15	0.1	0.30	24.89	0.09	0.027	619.51	15419.66				1387.77	416.33
16	0.1	0.40	23.41	0.16	0.064	548.03	12829.34				2052.69	821.08
17	0.1	0.50	22.2	0.25	0.125	492.84	10941.05				2735.26	1367.63
18	0.1	0.60	21.2	0.36	0.216	449.44	9528.13				3430.13	2058.08
19	0.1	0.70	20.35	0.49	0.343	414.12	8427.39				4129.42	2890.60
20	0.1	0.80	19.64	0.64	0.512	385.73	7575.73				4848.47	3878.77
21	0.2	0.00	30.52	0	0	931.47	28428.48				0.00	0.00
22	0.2	0.10	27.92	0.01	0.001	779.53	21764.38				217.64	21.76
23	0.2	0.20	25.89	0.04	0.008	670.29	17353.86				694.15	138.83
24	0.2	0.30	24.3	0.09	0.027	590.49	14348.91				1291.40	387.42
25	0.2	0.40	23.03	0.16	0.064	530.38	12214.67				1954.35	781.74
26	0.2	0.50	21.94	0.25	0.125	481.36	10561.12				2640.28	1320.14
27	0.2	0.60	21.03	0.36	0.216	442.26	9300.75				3348.27	2008.96
28	0.2	0.70	20.23	0.49	0.343	409.25	8279.19				4056.80	2839.76
29	0.2	0.80	19.57	0.64	0.512	382.98	7495.01				4796.81	3837.45
	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.
84	0.9	0.00	24.66	0	0	608.12	14996.13				0.00	0.00
85	0.9	0.10	23.71	0.01	0.001	562.16	13328.91				133.29	13.33
86	0.9	0.20	22.9	0.04	0.008	524.41	12008.99				480.36	96.07
87	0.9	0.30	22.13	0.09	0.027	489.74	10837.88				975.41	292.62
88	0.9	0.40	21.39	0.16	0.064	457.53	9786.61				1565.86	626.34
89	0.9	0.50	20.73	0.25	0.125	429.73	8908.36				2227.09	1113.55
90	0.9	0.60	20.16	0.36	0.216	406.43	8193.54				2949.67	1769.80
91	0.9	0.70	19.66	0.49	0.343	386.52	7598.90				3723.46	2606.42
92	0.9	0.80	19.23	0.64	0.512	369.79	7111.12				4551.12	3640.89
93	1.0	0.00	24.34	0	0	592.44	14419.88				0.00	0.00
94	1.0	0.10	23.47	0.01	0.001	550.84	12928.24				129.28	12.93
95	1.0	0.20	22.69	0.04	0.008	514.84	11681.63				467.27	93.45
96	1.0	0.30	21.94	0.09	0.027	481.36	10561.12				950.50	285.15
97	1.0	0.40	21.26	0.16	0.064	451.99	9609.26				1537.48	614.99
98	1.0	0.50	20.63	0.25	0.125	425.60	8780.06				2195.02	1097.51
99	1.0	0.60	20.08	0.36	0.216	403.21	8096.38				2914.70	1748.82
100	1.0	0.70	19.59	0.49	0.343	383.77	7518.02				3683.83	2578.68
101	1.0	0.80	19.18	0.64	0.512	367.87	7055.79				4515.71	3612.57
102		Regression Output:										
103	Constant			38.13389								
104	Std Err of Y Est			0.013274								
105	R Squared			0.998465								
106	No. of Observations			99								
107	Degrees of Freedom			83								
108		F	T(C)	F2	F3	T2	T3	TF	TF2	TF3	T2F	T2F2	T2F3	T3F	T3F2	T3F3
109	X Coefficient(s)	185.8736	-3.45765	435.405	1418.964	0.105804	-0.00109	-20.1963	-49.8361	-218.775	0.725236	1.706739	11.09337	-0.00875	-0.01627	-0.18484
110	Std Err of Coef.	28.06724	0.317367	90.61053	124.0308	0.010789	0.000121	3.120603	12.05624	18.80215	0.116059	0.527508	0.959954	0.001443	0.007611	0.01642

Note

i ABL-SVAT:Atmospheric Boundary Layer-Soil Vegetation AtmosphereTransfer Model. Carlson and Boland (1978), Taconet *et al.* (1986) and Lynn and Carlson (1990).

5 Use the relationship to map surface energy fluxes and soil moisture. Equation (5.A8) is used to map M_o using such tools as the ERDAS Imagine® modeler. Similar equations are developed for net radiation, sensible heat, and latent heat fluxes. For the fluxes, one only needs to replace the T_o column by M_o and use the same methods to correlate surface energy fluxes to *Fr* and M_o.

Acknowledgments

The authors are indebted to the reviewers for their time and for their in-depth suggestions on improving the preliminary manuscript. The authors would also like to thank Drs Larry Hipps, Dani Or, and Mike White at Utah State University for their interaction in the preparation of the manuscript.

References

Beck, L.R., M.H.Rodriguez, S.W.Dister, A.D.Rodriguez, E.Rejmankova, A. Ulloa, R.A.Meza, D.R.Roberts, J.K.Paris, M.A.Spanner, R.K.Washino, C. Hacker, and L.J.Legters (1994) Remote sensing as a landscape epidemiological tool to identify villages at high risk for malaria transmission. *Am. J. Trop. Med. Hyg.,* 51 (3): 271–80.

Burman, R.D., J.L.Wright, and M.E.Jensen (1975) Changes in climate and estimated evaporation across a large irrigated area in Idaho. *Trans. ASCAE* 18 (6): 1089–93.

Capehart, W.J. (1996) Issues regarding the remote sensing and modeling of soil moisture for meteorological applications. PhD Thesis, Department of Meteorology, The Pennsylvania State University.

Capehart, W.J. and T.N.Carlson (1997) Decoupling of surface and near-surface soil water content: a remote sensing perspective. *Water Resour. Res.* 33 (6): 1383–95.

Carlson, T.N. and Boland (1978) Analysis of urban-rural canopy using a surface heat flux/temperature model. *J. Appl Meteorol.* 17:998–1013.

Carlson, T.N. and D.A.J.Ripley (1997) On the relationship between NDVI, fractional vegetation cover and leaf area index. *Remote Sens. Environ.* 62:241–52.

Carlson, T.N., R.R.Gillies, and E.M.Perry (1994) A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. *Remote Sens. Rev.* 9:161–73.

Choudhury, B.J., N.U.Ahmed, S.B.Idso, R.J.Reginato, and C.S.T.Daughtry (1994) Relationships between evaporation coefficients and vegetation indices studied by model simulations. *Remote Sens. Environ.* 50:1–17.

Crombie, M.K., R.R.Gillies, R.E.Arvidson, P.Brookmeyer, G.J.Weil, M.Sultan, and M.Harb (1999) An application of remotely derived climatological fields for risk assessment of vector-borne diseases—A spatial study of filariasis prevalence in the Nile delta, Egypt. *Photogramm. Eng. Remote Sens.* 65 (12): 1401–9.

Davenport, D.C. and J.P.Hudson (1967) Changes in evaporation rates along a 17-km transect in the Sudan Gezira. *Agric. Meteorol.* 4:339–52.

Dugas, W.A., L.J.Fritschen, L.W.Gay, A.A.Held, A.D.Matthias, D.C.Reicosky, P.Steduto, and J.L.Steiner (1991) Bowen ratio, eddy correlation, and portable

chamber measurements of sensible and latent heat flux over irrigated spring wheat. *Agric. For. Meteorol.* 56:1–20.

Gillies, R.R. (1994) A physically based land-use classification scheme using remote solar and thermal infrared measurements suitable for describing urbanization. PhD Thesis, University of Newcastle.

Gillies, R.R. and T.N.Carlson (1995) Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. *J. Appl. Meteor.* 34:745–56.

Gillies, R.R., J.Cui, T.N.Carlson, W.P.Kustas, and K.S.Humes (1997) Verification of a method for obtaining surface energy fluxes of surface soil moisture availability from remote measurements of *NDVI* and surface radiometric temperature. *Int. J. Remote Sens.* 18 (15): 3145–66.

Gottschalck, J.C., R.R.Gillies, and T.N.Carlson (2001) The simulation of canopy transpiration under doubled CO2: the evidence and impact of feedbacks on transpiration in two 1-D soil-vegetation-atmosphere-transfer models. *Agric. For. Meteorol.* 106:1–21.

Goward, S., G.D.Cruickshanks, and A.Hope (1985) Observed relation between thermal emission and reflected spectral radiance of a complex vegetated landscape. *Remote Sens. Environ.* 18:137–46.

Harb, M., R.Faris, A.M.Gad, O.N.Hafez, R.Ramzy, and A.A.Buck (1993) The resurgence of lymphatic filariasis in the Nile delta. *Bull. WHO* 71 (1): 49–54.

Homes, R.H. (1970) Mesoscale effects of agriculture and a large prairie lake on the atmospheric boundary layer. *Agron. J.* 63:546–9.

Idso, S.B., T.J.Schmugge, R.D.Jackson, and R.J.Reginato (1975) The utility of surface temperature measurements for the remote sensing of surface soil water status.*J. Geophys. Res.* 80 (21): 3044–9.

Jackson, R.D., J.L.Hatefield, R.J.Reginato, S.B.Idso, and P.J.Pinter (1983) Estimation of daily evapotranspiration from one time-of-day measurements. *Agric. Water Manage.* 7:351–62.

Kustas, W.P., R.D.Jackson, and G.Asrar (1989) Estimating surface energybalance components from remotely sensed data. In: G. Asrar (ed.), *Theory and Applications of Optical Remote Sensing.* John Wiley, New York, pp. 604–27.

Kustas, W.P. and J.M.Norman (1996) Use of remote sensing for evapotranspiration monitoring over land surfaces. *Hydrol. Sci.* 41 (4): 495–516.

Leprieur, C., Y.H.Kerr, and J.M.Pichon (1996) Critical assessment of vegetation indicies from AVHRR in a semi-arid environment. *Int. J.Remote Sens.* 17:2594–63.

Lui, Q. and A.Huete (1995) A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* 33: 457–65.

Lynn, B. and T.N.Carlson (1990) A stomatal resistance model illustrating plant versus external control of transpiration. *Agric. For. Meteorol.* 52:1483–93.

Malone, J.B., O.K.Huh, D.P.Fehler, P.A.Wilson, D.E.Wilensky, R.A.Holmes, and A.El Magdoub (1994) Temperature data from satellite imagery and the distribution of shistosomiasis in Egypt. *Am. J. Trop. Med. Hyg.* 50 (6): 714–22.

Markam, B.L. and J.L.Barker (1986) Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. *EOSAT Landsat Tech. Note* 1:3–8.

Norman, J.M. and F.Becker (1995) Terminology in thermal infrared remote sensing of natural surfaces. *Remote Sens. Rev.* 12:159–73.

Owen, T.W., T.N.Carlson, and R.R.Gillies (1998) An assessment of satellite remotely sensed land cover parameters in quantitatively describing the climate effect of urbanization. *Int. J. Remote Sens.* 19 (9): 1663–81.

Ottensen, E.A. and C.P.Ramachandran (1995) Lymphatic filariasis infection and disease. Control strategies. *Parasitol. Today* 11:129–31.

Price, J.C. (1990) Using spatial context in satellite data to infer regional scale evapotranspiration. *IEEE Trans. Geosci. Remote Sens.* GE-28:940–8.

Rodgers, D.J. and S.E.Randolph (1991) Mortality rates and population density of tsetse flies correlated with satellite imagery. *Nature* 351:739–41.

Schmugge, T.J., T.J.Jackson, and H.L.McKim (1980) Survey of methods for soil moisture determination. *Water Resour. Res.* 16 (6): 961–79.

Taconet, O., T.Carlson, R.Bernard, and D. Vidal-Majar (1986) Evaluation of a surface/vegetation parameterization using satellite measurements of surface temperature.*J. Clim. Appl Meteorol.* 25:1752–67.

Temesgen, B. (2001) Terrain and biome effects in geophysical variables derived from boundary layer models coupled with airborne data. PhD Thesis, Utah State University, Logan, UT, 137 pp.

Thompson, D.F., J.B.Malone, A.A.Buck, and B.L.Cline (1996) Bancroftian filariasis distribution in the southern Nile delta: correlation with diurnal temperature differences from satellite imagery. *Emerging Infect. Dis.* 2:234–5.

Chapter 6

Rapid soil drying and its implications for remote sensing of soil moisture and the surface energy fluxes

Toby N. Carlson, David A. J. Ripley and Thomas J. Schmugge

6.1 The problem

Soil drying under the influence of sunlight is often detectable by an increase in surface radiant temperature. While this is true in the general sense, all other factors being equal, a problem arises in trying to determine a correct value of soil water content for a given application, such as for atmospheric prediction, hydrology, or agriculture.

Affixing a correct level or depth for a soil moisture estimate is necessary, not only for practical applications, but also for making comparisons with and for assessing the value of soil water content derived by differing techniques, such as from *in situ* or microwave measurements. Uncertainty arises from lack of agreement between measurements made by differing techniques and from the abstract notion of soil moisture as used in land surface models. This uncertainty has led to an unfortunate disparagement of the surface radiant temperature as a means for deriving either the surface turbulent energy fluxes or the soil water content and it has tended to obscure serious investigations relating surface energy fluxes and substrate hydrology. A question that is seldom asked, however, is: *which* soil water content does one wish to obtain and for what purpose? Indeed, one can speak of surface soil water content and root zone soil water content without being very specific as to the fact that evaporation and transpiration draw water from different layers in the soil in a way that is uniquely related to the soil type, vegetation type and amount, rooting depth, and the current vertical profile of soil water content.

Simply stated, the problem as posed above does not resolve itself by determining which method yields the most accurate results but of knowing what each measurement means and how it can be used. An indirect soil water estimate, consisting of an entire vertical profile or vertically integrated soil water content, cannot be obtained with any known remote sensing technique, as each method has its limitations and each pertains to a different facet of the soil water profile. Indeed, a point to be made in this chapter is

that differing indirect techniques may reveal only parts of the whole, and, therefore, a particular estimate of soil water content, however, accurate within its own context, may be inappropriate for some applications and useful for others. We will illustrate the problem with some measurements of soil water content and soil temperatures, including the surface radiant temperatures.

6.2 Measurements of soil water content and surface radiant temperature

6.2.1 Evidence of rapid surface drying

At the heart of the problem lies the fact that temperature and soil water content vary somewhat independently with depth. The problem is most pronounced in space (horizontal and vertical) and time variability at the soil-air interface. Even with *in situ* methods, the matter of determining the soil water content profile accurately within the top several centimeters of the surface is difficult, as most soil moisture sensors are incapable of resolving soil water content in layers less than 1 or 2 cm in depth. With care, gravimetric methods can be used to achieve such resolution, although such measurements in the top 0.5 or 1 cm are fairly rare.

Jackson (1973) provides some detailed and highly resolved vertical measurements of soil water content near the surface of a common agricultural soil (Adelanto loam). Using gravimetric sampling, he showed the time variation of soil water content in the top 0.5 cm layer and at 1-cm intervals below that level to 5 cm, and thereafter at 2-cm intervals down to 9 cm. He also showed the profile of vertical water flux and the surface evaporation. What Jackson found was that the vertical gradient of soil water content was largest just below the surface and that the soil water content in the top 0.5 cm decreased very rapidly with time to values less than 0.05 by volume within a few days following irrigation.

Importantly, the largest vertical gradients in soil water content occurred not when the soil was initially very wet (about 0.35 by volume) or later when it had dried to the extent that the soil water content at 9 cm had decreased to 0.15 by volume, but during an intermediate period when the values between 5 and 9 cm were between 0.20 and 0.25 by volume. Jackson (1973) identified these three stages of drying, pointing out that the soil water content during the middle phase of drying, in which the soil was neither very dry nor very wet, depends on the soil's ability to conduct water to the surface and not on atmospheric conditions. Jackson also showed that the vertical fluxes of soil water were also much smaller below 5 cm than in the top 2 cm, which is an indication that evaporation removes a proportionately larger amount of water from the top 2 cm than from deeper layers. Similarly, the top 2 cm dries out the most rapidly because the water from below is unable

to resupply the surface at a fast enough rate. Similar results were obtained by Ek and Cuenca (1994).

Equally evident is the fact that the implied water flux divergence from the surface layer cannot continue indefinitely. Ultimately, the surface layer desiccates, leaving a surface crust that may cap an underlying wet layer. Because hydraulic conductivity is so sensitively dependent upon soil water content, a decrease in the latter from 0.35 to 0.05 by volume causes a decrease of the hydraulic conductivity by orders of magnitude (Capehart and Carlson 1997). Consequently, rapid drying impedes the re-supply of liquid water from deeper layers, so that the evaporation flux decreases rapidly with time until the surface layer is almost completely desiccated. Jackson's measurements showed that, despite surface desiccation, the soil water content at 9 cm remained above the wilting point even after 34 days following irrigation!

Capehart and Carlson (1997), using a surface hydrology model, illustrated differential drying between the surface and substrate, as shown in Figure 6.1. They showed that the drying rates at 5–10 cm below the surface were almost identical and slowly decreasing under strong sunlight, but that the drying rate at 0.5 cm rapidly increased after the first three days and then decreased to zero as the soil entered the dry phase as referred to by Jackson (1973). They called this phenomenon "decoupling" because the soil water content near the surface was no longer a predictor of the soil water content at 5-cm depth and below. Their simulations showed that soil water content at 5- and 10-cm depths remained almost constant at about 30% of saturation (about 0.14 by volume) during the decoupling and desiccating stages in the surface layers. A purpose of this chapter is to illustrate decoupling and how it affects the interpretation of remote measurements.

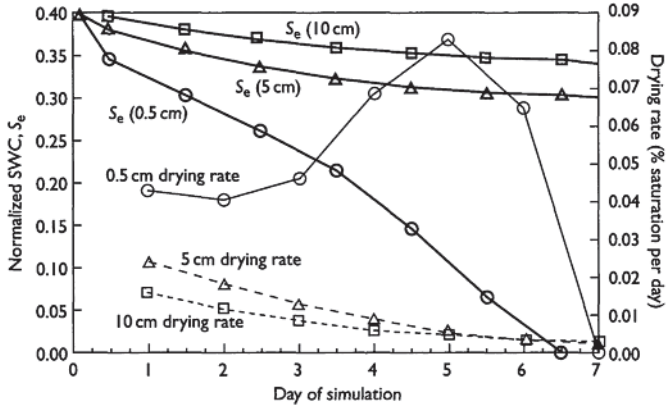


Figure 6.1 Normalized soil water content expressed as a fraction of saturation (S_e) and soil drying rate; the right-hand axis is expressed as a % change per day of the fraction of saturation and the lower axis is time. Graphs pertain to 0.5, 1.0, 5.0, and 10.0 cm depths. The drying rate is omitted for the 1.0-cm level. These simulations were made with a hydrological model for idealized sunlit conditions (from Capehart and Carlson 1997).

Studies by Ek and Cuenca (1994) and Capehart and Carlson (1997) showed that soil water content estimates determined from surface radiant temperatures can be poorly correlated with those measured over deeper layers, which tend to possess larger values. Perry and Carlson (1988) and Carlson *et al.* (1995) present examples showing a large scatter of points plotted on graphs of soil water content derived from microwave measurements, which sample a depth typically about 3–5 cm (Schmugge and Jackson 1994; Jackson *et al.* 1997), and from thermal-infrared measurements. Essentially no significant correlation was found between the two types of measurements in these studies, except by Perry and Carlson when the thermal data were heavily smoothed. This lack of correlation between thermal and microwave estimates of soil water content is not only due to the large vertical gradients, but also due to the enormous spatial variability of *surface* temperature and surface soil water content, which depend much more on the microscale variability of the soil type, texture, exposure, and surface debris content than does the deeper layer soil water content. Large variability in surface soil water content, however, is captured by the surface temperatures, which nevertheless can be unrepresentative of the deeper layer soil water content while relating very closely to the surface fluxes of heat and moisture.

Gillies *et al.* (1997) note that high-resolution imagery from aircraft (5m resolution) consistently show a full range of surface radiant temperatures over drying soil and, therefore, a full range of soil water content from dry to moist. Such local variations in soil water content is an indication that the heterogeneity of natural soils—and especially of the hydraulic conductivity in the surfaces layers—is as large within a particular (classical) soil type as that between differing soil types.

6.2.2 Radiometry at infrared and microwave frequencies

Measurement of the thermally emitted radiation from the earth's surface at various wavelengths can yield useful information about parameters, such as surface temperature and surface soil water content. To estimate surface temperatures, radiation at wavelengths around 10μm is used because the peak intensity of thermally emitted radiation, as described by the Planck equation, occurs at these wavelengths for terrestrial temperatures (≈ 300 K) and the atmosphere is relatively transparent. Therefore, variations in the observed intensity of infrared radiation are mainly related to surface temperature variations. Nevertheless, it is not possible to obtain accuracy much better than about plus or minus 1–1.5°C in surface temperature when the information is derived from the thermal channels of satellites.

In contrast to microwave measurements, emitted thermal radiation from the soil originates within the top few tenths of centimeters of soil. Moreover, over vegetation, thermal radiances emitted are more apt to contain a blend of energy originating over vegetation and bare soil than microwave radiances.

Over dense vegetation, infrared surface temperatures tend to be very close to that of the leaves, although shadowing may result in a temperature somewhat below that of a given sunlit leaf. In general, the radiometric temperature of a dense vegetation canopy is typically only one or two degrees higher than that of the air just above the canopy.

At microwave frequencies, the most striking feature of the emission from the earth's surface is the large contrast between water and mineral material. This emissivity contrast is due to the large difference between the dielectric constant of water (≈80) and that of dry soils (≈5). Thus, a mixture of water and dry soil had a dielectric constant between these two extremes, affording a mechanism for the remote sensing of soil moisture at microwave frequencies. This variation in the soil's dielectric constant produces a range of emissivities from 0.95 for dry soils to less than 0.6 for wet soils, which is easily observable with a microwave radiometer

6.2.3 Vegetation and surface energy fluxes

Vegetation constitutes an additional source of uncertainty in using surface radiant temperatures to determine soil water content. Until the mid-1980s, remote methods for determining soil water content using surface radiant temperatures (as measured by satellite) made no distinction between soil surface and vegetation surface radiant temperatures. It became possible to distinguish one type of surface from the other with an increased knowledge of vegetation, particularly the vegetation amount, which can be inferred from indices based on multi-spectral measurements in the visible and nearinfrared.

In order to determine unique temperatures for both the vegetated and bare soil fractions of a pixel (Figure 6.2), we make a series of simple but reasonable assumptions (Gillies and Carlson 1995): (a) the radiant temperature pertains to a surface consisting of sunlit leaves and sunlit bare soil; (b) the normalized difference vegetation index (NDVI) is closely related to fractional vegetation cover, such that the surface is 100% covered by vegetation where NDVI is large (e.g. ˜0.6) and bare where NDVI is small (e.g. zero); (c) the temperature of the vegetation is a constant over an image or field of view. The latter assumption is based on extensive observations with satellite imagery, which show little spatial variability in the surface radiant temperature over dense vegetation, at least for images with pixels sizes of several meters or more.

While individual sunlit leaf temperatures may increase well above air temperature, we also find from inspection of many thermal images that vegetation canopies, which consist of a large ensemble of leaves, exhibit little elevation in temperatures above the ambient air temperature. Simulations that we have made of crop canopy temperature support this observation, showing a very slow increase in surface radiant temperature with decreasing soil water content until the latter reaches values approaching

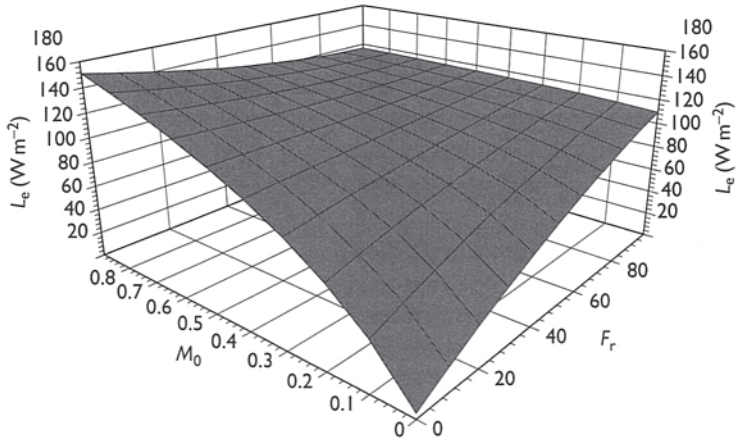


Figure 6.2 Sensitivity of latent heat flux (vertical axis labeled L_e; W m^{-2}) to soil moisture availability (M_0) and fractional vegetation cover (F_r; %), as simulated with a soil/vegetation/atmosphere transfer model.

the so-called "wilting point" in the root zone. Indeed, even when the soil water content is reduced to values below the wilting point, plant canopies react to water stress, not so much by increasing the ensemble leaf temperature, but by undergoing a change in leaf orientation and shape, such that more solar radiation reaches the soil and less solar radiation is intercepted by the leaves. In the extreme case, the leaves may simply drop off. Transient elevations in leaf temperature due to water stress occur for a short time during the middle part of the day when the plant is subject to a period of water depletion (Lynn and Carlson 1990; Olioso *et al.* 1996). For the most part, however, increased radiometric temperatures of vegetation canopies during dry conditions depend on an increased fraction of bare soil visible to the radiometer, rather than a substantial rise in leaf temperature.

We wish, therefore, to emphasize that variations in surface soil temperature and the fraction of surface covered by vegetation, and not the leaf temperature, produce most of the spatial variability in surface radiant temperature during periods of soil drying. Partial plant canopies modify the temperature of a sunlit surface and impose patterns of surface radiant temperature that depend partly upon vegetation cover as well as upon soil surface wetness. Of course, the vegetation behaves differently from bare soil. Vegetation extracts soil water from deep in the root zone, so that soil drying in the presence of vegetation may produce a greater decrease in root zone water content than in the absence of plants and possibly more than at the surface, as the latter would remain shaded by the leaves. Unlike rapid surface drying, water loss at deeper layers would be undetectable in the surface radiant temperatures.

In contrast with soil water content, the surface turbulent energy fluxes are rather sensitively dependent on surface radiant temperature (see Chapter 7). Indeed, fluxes can be determined without any explicit knowledge of the soil water content, given the surface radiant temperature and some supporting data, such as air temperature (Gillies and Carlson 1995). Because the surface turbulent energy fluxes depend directly on the surface radiant temperature, they can be determined with less uncertainty than the soil water content. It is fair to say that current methods for estimating these fluxes from surface radiant temperatures can achieve an accuracy of ± 20–40 W m^{-2} for both types of fluxes and ± 10–30% of their maximum values, with latent heat flux errors corresponding more to the lower part of these ranges and sensible heat fluxes more to the higher end.

The two most important factors governing the partition of net radiation into sensible and latent energy are found to be the fractional vegetation cover and the soil surface wetness (moisture availability—defined here as the ratio of soil water content to that at field capacity in a surface layer). Figure 6.2 is based on simulations with a land surface model (Gillies and Carlson 1995), which uses a "force-restore" method similar to that of Deardorff (1978) to calculate the vertical transfer of water in the soil. The three soil layers consist of a surface layer (set at 10cm), a transition layer, and a root zone layer (set at 50cm). Transpired water is drawn from the root zone and surface evaporation originates in the surface layer. Water can move from one layer to another depending on the vertical gradient of water content, but the hydraulic conductivity does not vary with soil water content.

As shown in Figure 6.2, sensitivity of evapotranspiration to these parameters is not uniformly distributed over the range of moisture availability and fractional vegetation cover. Rather, significant sensitivity of the fluxes to surface moisture availability and vegetation cover occur only when these two factors are both less than 0.5 (expressed as 50% in Figure 6.2), and they become quite large when the surface moisture availability is less than 0.1. It is worth repeating that the root zone soil water content, which was held constant in the simulations used to produce Figure 6.2, is not a major factor in the surface flux balance for bare soil, except insofar as it is able to slowly re-supply the surface layer with water. Figure 6.2 remains unaffected in these simulations when the root zone soil water content is varied over a wide range of values. The importance of the surface becomes increasingly obvious as the surface layer in the model is reduced in thickness.

6.2.4 A soil experiment

In order to study the drying process in relation to the surface radiant temperature, we conducted a simple field experiment. Each of four wooden boxes, approximately 55 cm deep by 60 × 60 cm^2 of top surface area, was filled with locally obtained soils. The boxes were situated on the roof of

Walker Building at Penn State University, about 20 m above ground level, and were exposed to normal insolation and wetting by precipitation. Holes were drilled in the bottom of the boxes to allow infiltrating rainwater to seep downward and exit. The soil surface was made flush with the top of the boxes so as to eliminate shadowing by the raised sides of the boxes. Each box of soil was divided into two sections of approximately 25 × 50 cm^2 by a wooden partition. The soils used are called Murrill Channery silt loam (box 1), Hagerstown silt loam, A horizon (box 2), Hagerstown silt loam, B horizon (box 3) and Buchanan Channery loam, B horizon (box 4). Their arrangement is shown in Figure 6.3.

Two types of soil water probes along with copper-constantan thermocouples were implanted at soil depths of 1, 2, 5, 10, and 20 cm on both sides of the partition of each box. One soil probe was a commercial product, gypsum blocks made by Delmhorst™; the other was a grid mesh construction of our own design. This latter instrument closely resembles the one described by Amer *et al.* (1994). Delmhorst blocks are wine cork-sized plugs made of gypsum enclosed around a wire mesh through which an induced current is passed from a proprietary meter made by Delmhorst with which electrical conductance of the soil block is measured. The meshes consist of thin perforated wafers of non-conducting ceramic ("perfboard") material about 2 × 2cm^2 on a side and about 1.5-mm thick, to which stainless steel wire meshes are attached by nylon strands on either side of the wafer. Soil

Figure 6.3 Photograph of the four soil boxes on the roof of Walker Building, Penn State University.

surrounds and fills the holes, allowing an electrical current to pass across the mesh. All probes were installed a week or two prior to making the outdoor measurements.

Before beginning the outdoor measurement program, calibrations were performed indoors for both the gypsum blocks and the grid meshes. Both meshes and gypsum blocks were calibrated in separate soil pots (approximately 15cm in diameter and 15cm deep). Each pot was filled with soils identical to those in the boxes and implanted with similar probes. Soil pots were wetted to field capacity (drainage ceases) and allowed to dry naturally or in a drying oven in stages. Pot and soil were periodically weighed with an electrical balance and the temperature of the soil measured. Electrical resistance measurements were made for both probes at each stage of drying. The Delmhorst meter was used to calibrate the gypsum blocks for all soil types. Meshes were also calibrated with the Delmhorst meter as if they were gypsum blocks.

Measurements were carried out during three summers, approximately June through August of 1995, 1996, and 1997. Except for the gypsum blocks, which were not placed at the 1-cm level because of their size, measurements were taken for all five levels for each type of probe on each side of the four partitioned boxes every day near noon, with the exception being weekends and during rainy periods. Meter readings were converted into soil water content via a set of polynomials that were developed to fit the calibration data. Soil temperatures were calculated directly from measured current using a standard ammeter; surface radiant temperatures were measured with an Everest™(Model 100) portable radiometer. Air temperature was also measured with the radiometer by sighting a shaded surface near the boxes. Precipitation was measured routinely by Penn State Weather Station personnel in the Walker building. Weather and the visual appearance of the soil surfaces were noted at the experiment site.

Calibration curves obtained for the sensors are similar to those published by Amer *et al.* (1994) (their Figure 3a), except that a temperature correction was made to both block and mesh data, as it was found that soil resistance varied significantly with both soil water content and temperature. The sensitivity of the mesh data to soil water content was highly non-linear and apparently not very stable. Amer *et al.* (1994) showed the largest variation in resistance as a function of soil water content occurred over a narrow range of soil water content (0.1–0.2 by volume) with very small variations in resistance for large changes in soil water content outside this range. This response of the soil water content made accurate calibration of the meshes very difficult, ultimately requiring us to change calibration strategies for the grid meshes.

Because the first year of operations was extremely dry and the second year unusually wet, only data from 1997 are presented. During this third summer, soil water content fluctuated between moderately dry and wet values.

It was found that the soil water content values for the meshes appeared unrealistically low. Yet it was clear that the meshes were able to capture, at least in a relative sense, the large variations in soil water content that occur in the top 2 cm. Initial calibrations for the meshes were, therefore, discarded in favor of a method that tied the soil water content values to those obtained with the aid of the gypsum blocks. In order to assign reasonable soil water content to the mesh data, we scaled the raw meter readings by setting the highest values equal to the soil water content measured by the gypsum blocks in the deeper layers of the soil and during the wettest periods and we set the lowest meter readings equal to zero. This was done individually for boxes 1, 3, and 4. Box 2 appeared to need no such adjustment and no scaling was made for that soil. Our impression is that this scaling produced similar mesh and block values, except in the mid-range where the former tended to exceed the latter.

6.3 Results of the soil experiments

We now present significant results from the field measurements. The purpose here is to illustrate that decoupling does occur under conditions of rapid drying and strong sunlight. Figure 6.4 consists of an eight-panel series of soil water content and temperature profiles in box 1 during 1997. Each data point corresponds to an average of two measurements, one on each side of the partition. The horizontal scale represents both volumetric soil water content (%) and temperature (°C). Solid curves with shaded circles pertain to the grid meshes (W_g), the dashed curves with small triangles to the gypsum blocks (W_b), and the heavy solid curve with blackened squares to the temperature of the soil or soil surface. Arrows at the top denote the air temperature at the time of measurement. Except for July 21, all measurements shown in Figure 6.4 were made under strong, direct sunshine.

Drying and warming in the top 5-cm layer is clearly evident after June 19, the day after a rain event, which deposited more than 2.5 cm of precipitation. Except for some very light rain showers during the next three weeks, no significant precipitation occurred again until July 9. During this drying period, the soil temperatures increased with time, so that by June 28 a shallow desiccated surface layer is evident in the top 5 cm. After two more light precipitation events during the next two weeks about 4.0 cm of rain fell during several days just after July 21, so that overcast and wet conditions are again in evidence on July 25.

A comparison between measurements made on different days and in different boxes is shown in Figures 6.4 and 6.5. Soil water content profiles in boxes 2–4 exhibit no remarkable differences from box 1 and henceforth will not be shown in detail except for June 28 and July 21 (Figure 6.5(a) and (b)). Regardless of whether differences between boxes shown in these

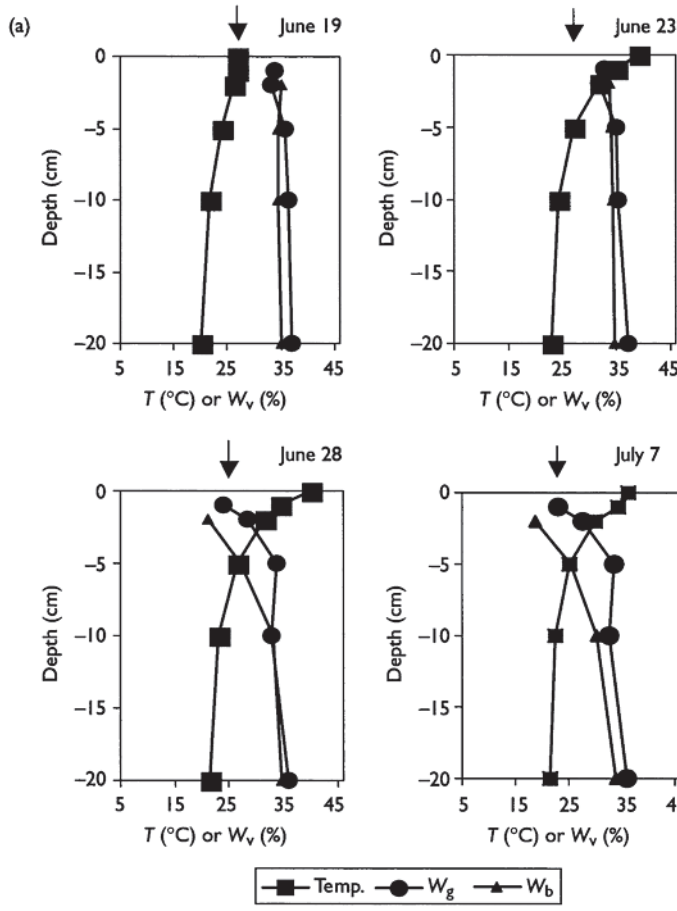


Figure 6.4(a) Vertical profiles of volumetric soil water content for the gypsum blocks (W_b; triangles) and grid meshes (W_g; circles) and temperature (°C; squares) as a function of depth (cm) for 8 days in 1997. The heavy solid arrow on top denotes the air temperature.

figures are intrinsic to the soil or are simply due to random spatial heterogeneity in the soil and to measurement inaccuracy, the emergence of a shallow surface drying layer (decoupling) is evident for all soils, which desiccate noticeably above 5 cm but show less change in soil water content below 5 cm.

The effect of decoupling on temperature is illustrated in Figure 6.6. Temperature profiles show maximum vertical gradients near the surface on warm, dry days and small vertical gradients on the wetter day (June 19); larger differences between soils occur on the dryer day (July 21).

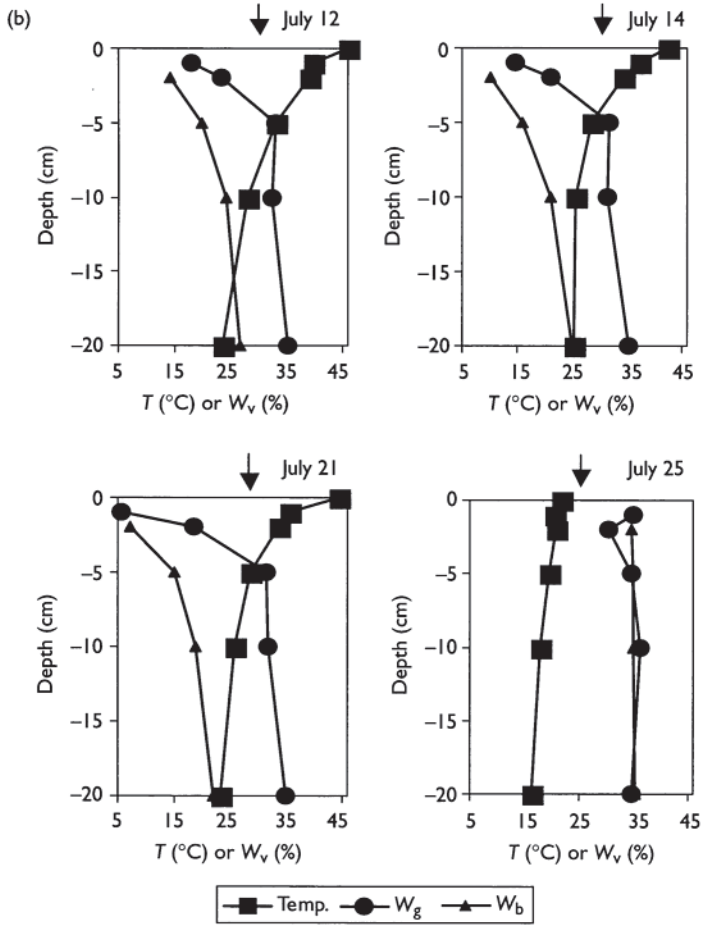


Figure 6.4(a) (Continued)

It is possible to estimate from these data the effect of decoupling on the surface sensible heat flux. A useful indicator of sensible heat flux is the difference between the surface radiant temperature and the surface-air temperature, a large difference implying a large surface sensible heat flux and small evaporative flux. The proportionality factor between the surface minus air temperature differences and the surface sensible heat flux is not unique, of course, as the relationship depends on environmental factors such as wind speed. Nevertheless, one might look for some relationship between the surface minus air temperature difference and the surface sensible heat flux and an inverse relationship with soil water content. However, Figure

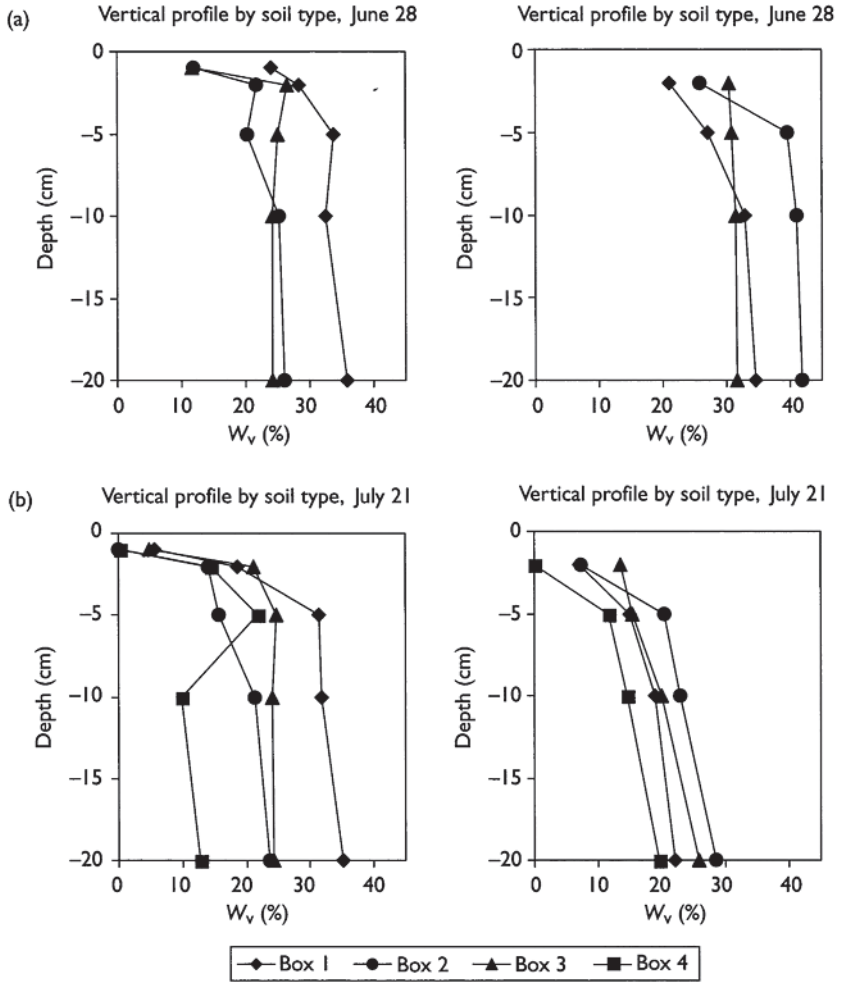


Figure 6.5 Vertical profiles of volumetric soil water content (W_v) as a function of depth (cm) on June 28 (a) and July 21, 1997 (b) for different soil boxes. Soil water content values derived from grid mesh are on the left and those derived from the gypsum blocks are on the right.

6.7(a) and (b) shows that the sensitivity of the surface-air temperature difference increases with decreasing soil water content only when the former exceeds some threshold. Moreover, the relationship between surface minus air temperature and soil water content is weaker at 5 cm than at 2 cm and

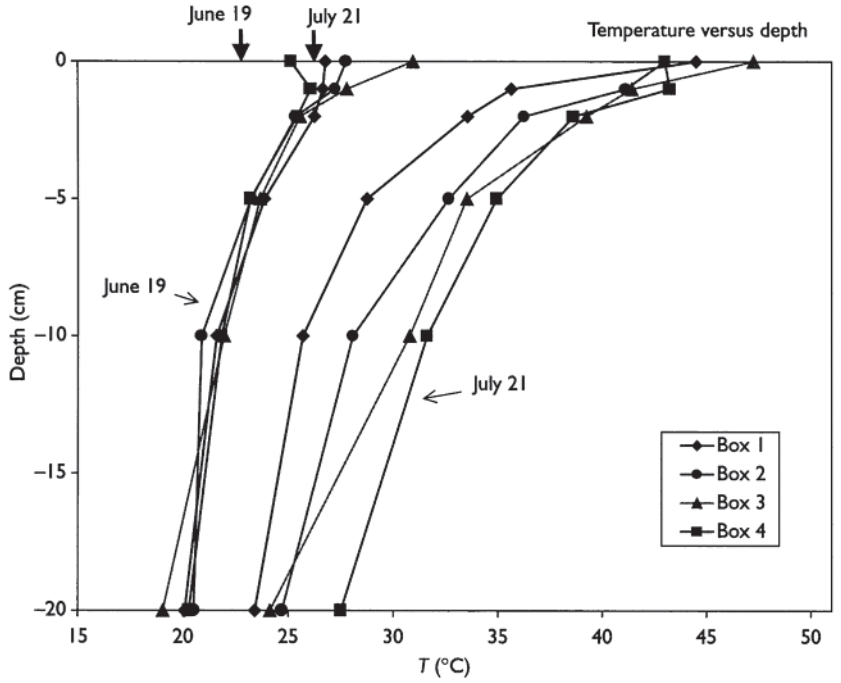


Figure 6.6 Soil temperature profiles (°C) as a function of depth (cm) for different soil boxes on June 19 and July 21, 1997. Arrows at the top denote air temperatures. The measurement at zero depth is the surface radiant temperature; all others were obtained from thermocouple readings at 1, 2, 5, 10, and 20 cm depths.

the surface minus air temperature threshold is higher at 5 cm (Figure 6.7b) than at 2 cm, respectively, about 6 and 12°C.

The virtual absence of sensitivity of surface-air temperature difference to soil water content below some threshold in surface minus air temperature difference suggests that surface heat flux may not always be closely related to soil water content except at the very soil surface. This is due to the differences in the propagation rates for the thermal and drying fronts in the soil. Clearly, the drying front had not yet penetrated to 2 cm in Figure 6.7(a) or to 5 cm in Figure 6.7(b) when the surface minus air temperature differences were below threshold. Once the drying front had reached these two levels, the temperature differences between surface and air increased with decreasing soil water content.

Figure 6.8 shows that the profile of temperature change within the ground is especially sensitive to soil water content near the surface. The various line profiles refer to simulations made with the soil/vegetation/atmosphere transfer model, referred to with regard to Figure 6.2; the other symbols refer to

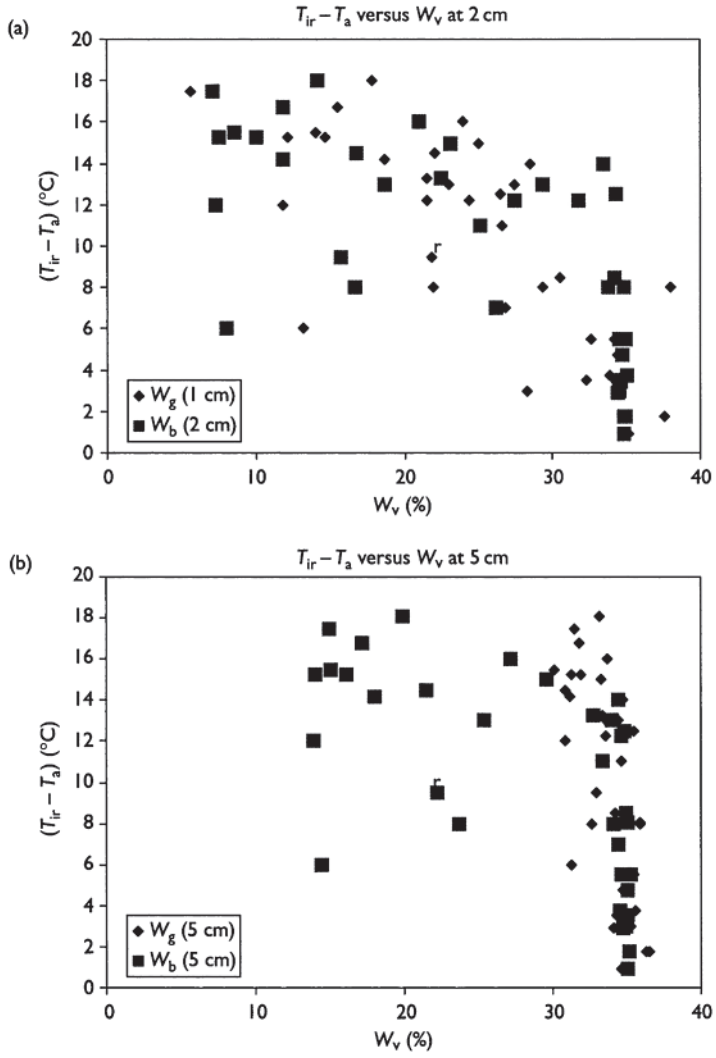


Figure 6.7 Surface radiant temperature minus air temperature (T_{ir}–T_a; ºC) differences versus volumetric soil water content (W_v; %) measured by the grid meshes and gypsum blocks at (a) 1 or 2 cm depth and (b) 5 cm depth for all measurements made in 1997.

measurements made in this soil experiment on a day in July 1995. Several days of soil drying had taken place prior to these measurements so that the soil surface was visually quite dry, in conformity with the soil water content measurements (not shown). Temperatures close to the surface vary over a

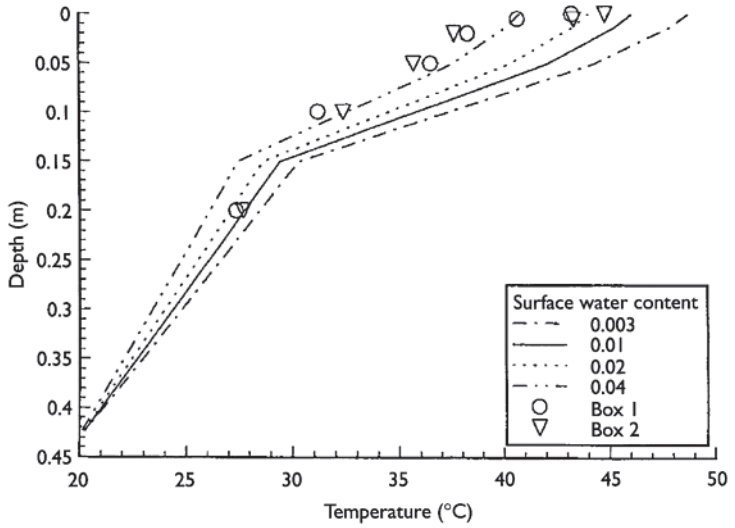


Figure 6.8 Measured soil and surface radiometric temperatures (circles and triangles for boxes 1 and 2) and simulated (continuous curves) soil and surface radiometric temperatures for different soil water content at approximately mid-day for a case with strong sun and dry soil in mid-July 1995.

range of nearly 10°C for a surface soil water content variation from 0.003 to 0.04 by volume. A close fit between simulated and measured temperatures in the soil is thus achieved with only a slight adjustment of the surface soil water content. What the figure shows is that small differences in the soil water content in the surface desiccation layer greatly affect the surface radiant temperature, even though such changes in soil water content are not measurable by current techniques.

More to the point, while a wide range of elevated surface radiant temperatures correspond to near desiccation in the surface layer, the water content of that drying layer however accurately measured, is of very little use as a predictor of soil water content at any depth below the top few centimeters. As we have indicated, however, the implied fluxes are highly sensitive to the temperature and water content of this shallow surface layer leading to the paradoxical conclusion that the surface radiometric temperatures may be useful for estimating the surface energy balance but not the total soil water content.

6.4 Interpretation of thermal and microwave measurements

Let us now consider some ambiguities implied in the estimation of soil water content using surface radiant temperature. The schematic in Figure 6.9 shows

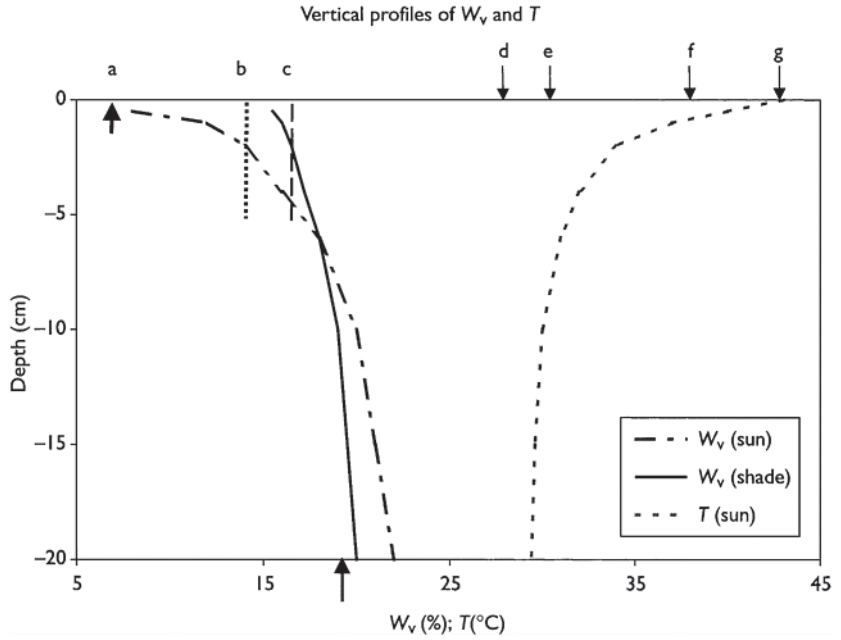


Figure 6.9 Schematic illustration showing volumetric soil water content (W_v; dot dash and solid curves) and temperature (ºC; dashed curve) as a function of depth (cm) for a case of strong sunlight and drying soil. The dot-dashed curve pertains to sunlit bare soil, the solid soil water profile to bare soil shaded by sunlit vegetation and the dotted curve to the temperature of sunlit bare soil. Letters (a-c) on top denote soil water content values obtained indirectly from the surface radiant temperature, respectively: (a) estimated from the surface radiant temperature over sunlit bare soil (marked g); (b) estimated from a microwave radiometer capable of sampling the top 5 cm in the sunlit bare soil areas (dotted segment); (c) same as (b) but in an area of vegetation with an underlying shaded surface (dashed segment). Letters d, e, f, and g, respectively, denote: (d) the air temperature; (e) the radiometric temperature of sunlit vegetation; (f) the radiometric temperature of a mix of sunlit vegetation and bare soil, and (g) the radiometric temperature of the sunlit bare soil. The arrow along the bottom axis denotes the column average soil water content between the surface and 20 cm.

typical profiles of soil water content (left-hand curves) and temperature (right-hand curve) as a function of depth. Let us imagine that the schematic situation depicted here pertains to strong sunlight with partial vegetation cover and that soil drying has occurred for at least a couple of days. Soil surfaces below the vegetation are shaded; the remaining bare soil patches are sunlit. Surface radiometric temperatures are obtained corresponding to three temperatures marked by the letters (g)—sunlit bare soil between vegetation patches, (f)—an ensemble of bare soil and vegetation, and (e)—the temperature of leaves

in the areas of dense, sunlit vegetation. Air temperature is denoted by the letter (d).

Now, consider the soil water content associated with these temperatures. The value of soil moisture corresponding to the temperature of the sunlit bare soil (g) is marked by the letter "a". This temperature applies to the surface of a shallow desiccation layer, whose soil water content is about 0.06 by volume (the short bold-faced arrow below the point labeled a). Suppose we were to measure a surface radiant temperature "f" over a partial vegetation canopy in which the range of surface radiant temperature might vary from "e" to "g". We might infer from this large variation in radiant surface temperature that the soil water content is varying greatly in space. Yet, the soil water content may not be changing and the surface radiant temperature is simply a function of vegetation amount. Clearly the measured temperature "f" is not an appropriate temperature for determining the correct profile of soil water content, although it might actually yield a reasonable estimate of the surface turbulent energy fluxes.

Vegetation itself introduces a spurious component in the derived soil water content using surface radiant temperatures. It is evident from Figure 6.9 that the average 5-cm values correlate poorly with the column average soil water content (the vertical arrow at the bottom of the graph) and with the soil water content closer to the surface. Differing vertical profiles of soil water content between sunlit bare patches (dot-dashed curve labeled b) and vegetation clumps (solid curve labeled c), shown in Figure 6.9, have differing average values over the top 5 cm, as shown, respectively, by the dotted and dashed segments suspended from the top axis. Shaded by plants, the profile exhibits no distinct surface drying layer, while the root zone, which supplies most of the transpiration, dries faster than comparable levels in the sunlit areas.

Interpretation of the radiometric signals for thermal and microwave radiometers differs over partial vegetation cover, as the microwave signal can penetrate through moderately dense vegetation (Wang *et al.* 1989). With its characteristically larger footprint, microwave measurements yield a spatial average of the two surface regimes, vegetated and bare. Typically, a microwave radiometer is capable of sensing the average soil water content over the top few centimeters with a horizontal resolution of tens of meters. If the sampling depth is, say, 5 cm, the vertical average of soil water content for such a signal would fall between points b and c, or between 0.14 and 0.17 depending on the amount of vegetation cover. By comparison, the actual soil water content averaged over the column from 0 to 20 cm is 0.19 by volume, as indicated by the short arrow at the bottom of Figure 6.9. The soil water content at the surface is, as previously stated, 0.06 by volume.

Finally, one can imagine that the surface desiccation layer would persist for a time after an onset of cloudy weather, although the soil temperature and especially the surface radiant temperature would fail to show the

characteristic surface minus air temperature differences of Figure 6.7. One would expect that such conditions would foster the diffusion of water to the surface at a rate which would be faster than the surface evaporation, leading to a gradual removal of the desiccation layer. On the other hand, a persistent cloudy period, more precisely sky conditions that would favor reduced surface evaporations, would not give rise at all to a surface desiccation layer even as the whole soil column dries out.

6.5 Conclusions

Rapid soil drying in the presence of strong sunlight produces a shallow drying layer whose radiometric temperature can become quite elevated despite moderately wet soil conditions at 5cm and below. This behavior, called decoupling by Capehart and Carlson (1997), can render soil water content values obtained with the aid of the surface radiant temperature inappropriate as estimates of the soil water content in a deep column, such as over the top 5cm, yet useful within its own context and more appropriate than a column average for determining the surface turbulent energy fluxes.

In the absence of a complete soil water content profile, estimates of surface energy fluxes are of limited use by themselves, as they represent only one measurement of a rapidly changing quantity. The same might be said for surface soil water content. A more useful parameter would be one that is slowly varying in time and intrinsic to the surface. Such a parameter would be more appropriate for inclusion as a land surface parameter in an atmospheric model and would provide a more representative measure of the surface turbulent energy fluxes and the net daily water loss. The Bowen Ratio might constitute one such parameter in that it avoids specifying the soil water content or vegetation cover, yet it is appropriate for calculating useful flux estimates. This ratio would stand on its own as a valid representation of a heterogeneous landscape, though requiring frequent updates due to varying rainfall and vegetation amount. Ideally, however, a combination of multiwavelength microwave measurements and surface radiant temperature might yield a more complete soil water content profile and serve a greater range of purposes than any single sensor.

Acknowledgements

We would like to thank the USDA under cooperative agreement 58-1270–3-030 for their support of this research and Douglas Miller for his help in setting up the roof experiment and for his editorial comments on our manuscript. We would also like to thank our field assistants, Jan Lukens, Brian Cosgrove, and Jeanne Jagodzinski, for their diligent help with the

measurements, often under a broiling sun, respectively, during the summers of 1995, 1996, and 1997. Finally, we are also indebted to Jim Breon and Richard Thompson for their patience in helping us with the electronics.

References

Amer, S.A., T.O.Keefer, M.A.Weltz, D.C.Goodrich, and L.B.Bach (1994) Soil moisture sensors for continuous monitoring. *Water Resour. Bull.* 30:69–83.

Capehart, W.J. and T.N.Carlson (1997) Decoupling of surface and near-surface soil water content: A remote sensing perspective. *Water Resour. Res.* 33:1383–95.

Carlson, T.N., R.R.Gillies, and T.J.Schmugge (1995) An interpretation of methodologies for indirect measurement of soil water content. *Agric. For. Meteorol.* 77: 191–205.

Deardorff, J.W. (1978) Efficient prediction of ground surface temperature and moisture with inclusion of a layer of vegetation. *J. Geophys. Res.* 83:1889–904.

Ek, M. and R.H.Cuenca (1994) Variation in soil parameters: implications for modeling surface fluxes and atmospheric boundary layer development. *Bound–Layer Meteorol.* 70:369–83.

Gillies, R.R., T.N.Carlson, J.Cui, W.P.Kustas, and K.S.Humes (1996) A verification of the "triangle" method for obtaining surface soil water content and energy fluxes from remote measurements of Normalized Difference Vegetation Index (NDVI) and surface radiant temperature. *Int. J. Remote Sens.* 18: 3145–66.

Jackson, R.D. (1973) Diurnal changes in soil water content during drying. In: R.R.Bruce, K.Flach, and H.M.Taylor (eds), *Field Soil Water Regime,* Chapter 3. Soil Science Society of America, Madison, WI, pp. 37–55.

Jackson, T.J., P.E.O'Neill, and C.T.Swift (1997) Passive microwave observation of diurnal surface soil moisture. *IEEE Trans Geosci. Remote Sens.* 35:1210–22.

Lynn, B.H. and T.N.Carlson (1990) A stomatal resistance model illustrating plant vs. external control of transpiration. *Agric. For. Meteorol.* 52:5–43.

Olioso, A., T.N.Carlson, and N.Brisson (1996) Simulation of diurnal transpiration and photosynthesis of a water stressed soybean crop. *Agric. For. Meteorol.* 81: 41–59.

Perry, E.M and T.N.Carlson (1988) Comparison of active microwave soil water content with infrared surface temperatures and surface moisture availability. *Water Resour. Res.* 24:1818–24.

Schmugge, T.J. and T.J.Jackson (1994) Mapping surface soil moisture with microwave radiometers. *Meteorol. Atmos. Phys.* 54:213–23.

Wang, J.R., J.C.Shiue, T.J.Schmugge, and E.T.Engman (1989) Mapping surface soil moisture with L-band radiometric measurements. *Remote Sens. Environ.* 27: 906–14.

Mapping surface energy fluxes with radiometric temperature

William P.Kustas, John M.Norman, Thomas J.Schmugge and Martha C.Anderson

7.1 Introduction

Directional radiometric surface temperature, $T_R(\phi)$, from a zenith view angle ϕ has been used to estimate surface sensible heat flux with varying degrees of success (Kustas and Norman 1996). The use of $T_R(\phi)$ frequently involves the controversial assumption that it is equivalent to the so-called "aerodynamic temperature," T_0, of the surface. T_0 is the temperature that satisfies the bulk transport expression having the form

$$H = \rho C_p \frac{(T_0 - T_A)}{R_A + R_{EX}} = \rho C_p \frac{(T_0 - T_A)}{R_{AH}} \tag{7.1}$$

where H is the sensible heat flux (Wm^{-2}), ${}_\rho C_p$ is the volumetric heat capacity of air ($J\ m^{-3}\ K^{-1}$), T_A is the air temperature at some reference height above the surface (K), R_{EX} is an excess resistance associated with heat transport, and R_A is the aerodynamic resistance ($s\ m^{-1}$), which has the following form in the surface layer (Brutsaert 1982):

$$R_A = \frac{[\ln((z_U - d_O)/z_{OM}) - \Psi_M][\ln((z_T - d_O)/z_{OM}) - \Psi_H]}{k^2 u} \tag{7.2}$$

In this equation, d_O is the displacement height, u is the wind speed measured at height z_U, k is von Karman's constant (»0.4), z_T is the height of the T_A measurement, ψ_M and ψ_H are the Monin-Obukhov stability functions for momentum and heat, respectively, z_{OM} is the roughness length for momentum transport. The excess resistance is often related to a roughness length for heat so that $R_{EX} = [\ln(z_{OM}/z_{OH})]/[ku_*]$, where z_{OH} is the roughness length for heat transport and u_* is the friction velocity; $u_* = u\ k/[\ln(z_U - d_O)/z_{OM} - \psi_M]$. T_0 cannot be measured, so it is often replaced with an observation of $T_R(\phi)$ in equation (7.1). However, for sparse canopies differences between T_0 and $T_R(\phi)$ can be > 10°. This has forced many users of this bulk transport or single-source approach to adjust z_{OH} or the ratio $\mathrm{In}(z_{OM}/z_{OH}) = kB^{-1} = ku_*R_{EX}$ (Garratt and Hicks 1973) to obtain good agreement with measured H. Most approaches have been empirical (e.g. Kustas *et al.* 1989; Stewart *et al.* 1994;

Kubota and Sugita 1994) and therefore difficult to apply a priori to different surface types. Indeed, the testing of various formulations for z_{OH} or the kB^{-1} parameter in single-source models with experimental data indicates that this is not a viable approach for partial canopy covered surfaces (Sun and Mahrt 1995; Kustas *et al.* 1996; Verhoef *et al.* 1997; Troufleau *et al* 1997). Blyth and Dolman (1995), using a two-source modeling approach, show the dependence of z_{OH} on surface conditions, including fractional vegetation cover and soil and vegetation resistances, as well as the available energy or net radiation less soil heat flux (i.e. $R_N - G$), and humidity deficit. A similar result was obtained by Lhomme *et al.* (1997) using the two-source model originally developed by Shuttleworth and Wallace (1985). For this reason, others have tried to account for the difference between T0 and TR using twosource models to account for the effect of soil and vegetation temperatures and resistances on both T_0 and T_R (e.g. Lhomme *et al.* 1994; Chebhouni *et al.* 1996).

Vining and Blad (1992) showed that the viewing angle of the sensor, ϕ, can significantly affect the computation of H when $T_R(\phi)$ replaces T_0 in equation (7.1). Other theoretical and observational studies suggest that $T_R(\phi)$ observations at multiple viewing angles may have the potential to account for the variability of z_{OH} (Brutsaert and Sugita 1995; Sugita and Brutsaert 1996). Using a detailed multilayer model, Matsuhima and Kondo (1997) find that optimum viewing angle for single-source approaches is between 50 and 70° from nadir.

A recent review of two-source models by Zhan *et al.* (1996) suggests that the Simplified Two-Source (STS) model proposed by Norman *et al.* (1995) can yield satisfactory estimates of sensible and latent heat flux, LE, over different surfaces and is relatively insensitive to the expected errors associated with estimating many of its input parameters and variables, except for $T_R(\phi)$ and T_A. Because the STS model was designed to use input data primarily from satellite observations, several simplifying assumptions about energy partitioning between the soil and vegetation reduce both computational time and input data required to characterize surface properties. While the model has been shown to satisfactorily predict surface fluxes when compared to field observations, it is not known how well the model realistically simulates the separate contributions from the soil-surface and vegetation.

This can be evaluated reliably using a comprehensive Plant-Environment (PE) model such as Cupid (Norman and Campbell 1983; Norman and Arkebauer 1991), which simulates radiation exchange, turbulent fluxes, and $T_R(N)$ for plant canopies. Cupid accommodates all the generalities inherent in a comprehensive PE model by using parameterizations of important processes at the leaf level (cm) and integrating mechanistic equations to the canopy level (10–100m). Cupid is applied to field data collected from a semiarid rangeland containing partial vegetation cover randomly distributed

over the landscape. The simulated $T_R(\phi)$ values computed from Cupid are used as input to the STS model for computing the energy balance of the soil and vegetation. These flux estimates are compared to Cupid output.

The simplified parameterizations of energy partitioning between the soil and vegetation with the STS model are evaluated and implications of their utility for application to different surfaces is discussed. Issues of how to estimate model parameters and key input variables related to vegetation properties on a regional basis are also discussed. An example of running the STS model for computing large-scale spatially distributed fluxes with remotely sensed surface temperature images of the semiarid rangeland landscape is presented.

For regional scale applications using satellite data, the STS model may be operational because its input requirements can be obtained primarily from the satellite data; information for all input parameters required by detailed PE models such as Cupid would not be available. This means many of the parameters in PE models would need to be specified from educated guesses, and if the parameter specification is unreliable, the overall model performance of the PE model deteriorates. As stated by Giorgi and Avissar (1997) discussing soil-vegetation-atmosphere transfer schemes (SVATS)

> …increased physical complexity and realism of SVATS may actually result in poorer model performance. Availability of observed data may in fact provide useful insights concerning the optimal level of complexity in SVATS in terms of the comprehensiveness of biophysical and hydrological representation on the one hand and model performance and verificability on the other.

Another issue in the application of satellite data for large area mapping of fluxes is the effect of heterogeneity of surface conditions at the subpixel scale and its impact on the fluxes. Methods for dealing with heterogeneity effects are being addressed in the hydrologic and atmospheric modeling communities. Giorgi and Avissar (1997) provide a detailed review of methodologies for dealing with subgrid scale heterogeneity. Interestingly, observational work on the effects of surface heterogeneity on surface flux aggregation using remote sensing with SVATS suggest that using simple averaging rules to define surface parameters for length scales on the order of 1–10 km causes relatively small errors for land surfaces where heterogeneity exists at length scales <10km. The simulations from Cupid under the various surface conditions will be used for testing the effect of heterogeneity in surface wetness, vegetation stress, and roughness. These preliminary results will consider more extreme cases of landscape variability and thus provide an upper bound to potential errors caused by subpixel heterogeneity.

7.2 Cupid model description

Cupid is a detailed PE model that simulates a wide variety of physiological and environmental processes simultaneously. The vegetation is a central emphasis in Cupid so that above-ground processes are formulated around plant-atmosphere interactions and below-ground processes are described by plant-soil interactions. Thus, the central emphasis of Cupid is the transport of energy, mass, and momentum between plants and their environment. For above-ground processes, the transfers between individual leaves and their local environment are described (Norman 1979); then the collective effect of all the leaves is integrated to obtain the response of the entire vegetative canopy. The canopy is divided into horizontal layers and leaves in each layer are arranged with appropriate position and orientation distributions. Transfer of energy, mass, and momentum is assumed to occur only in the vertical dimension, and this transport is described by turbulent diffusion with leaves in each layer acting as sources or sinks of various quantities (Norman and Campbell 1983). The below-ground transport of heat and mass provides a description of the soil environment that surrounds the roots and incorporates the exchanges between these roots and the soil system.

The interface between the above- and below-ground regions, namely the soil-surface, represents one of the most difficult parts of the system to simulate. Many processes occur at the soil/canopy interface; for example, absorption of radiation and momentum by the soil-surface, convective transport of heat and water to the atmosphere, conduction of heat, water, and CO_2 from lower in the soil to the surface, uptake of water by roots near the soil-surface, and infiltration of rainfall, irrigation water, or water that drips from the canopy as a result of interception or dew. All these processes are simulated in Cupid.

Characterization of the dependence of leaf physiological properties (photosynthetic rate, respiration rate, and stomatal conductance) on environmental factors (light, temperature, humidity, and soil water status) is essential to simulating leaf energy and mass exchanges. The leaf model combines the response of photosynthetic rate and stomatal conductance (Collatz *et al.* 1991,1992) to solve the leaf energy budgets and is described in Norman and Polley (1989). Canopy exchange rates are estimated by combining equations that describe leaf exchange rates with a characterization of canopy architecture, with boundary measurements of ambient environment above the canopy and below the root zone, and with equations that describe convective, conductive, and radiative exchange processes throughout the soil-plantatmosphere system. A description of canopy architecture includes the vertical distribution of stem and leaf areas, leaf angle distribution, canopy height, and some information about the horizontal distribution of leaf area such as random or clumped. Ambient atmospheric conditions may be obtained at every time step in the model from measurements of air temperature,

humidity, wind speed, solar radiation, and precipitation some meters above the canopy. Ambient soil boundary conditions consist of temperature and moisture content near the bottom of the root zone (0.5–2 m depth).

The influence of vertical gradients throughout the soil-plant-atmosphere system is included by using an iterative-solution technique that simultaneously solves the leaf energy budget for all leaves and the vertical flux-gradient equations. Such a solution requires conductances throughout the soil and atmospheric system; including aerodynamic conductances above and within the canopy (Goudriaan 1977), convective transfer coefficients at the soil surface (Sauer *et al.* 1995), leaf boundary-layer conductances, and soil thermal and hydraulic conductances (Campbell 1985).

The Cupid model has been used for numerous applications: (a) predicting canopy photosynthesis and light-use-efficiency from leaf characteristics in corn (Norman and Arkebauer 1991); (b) simulating evapotranspiration and CO_2 flux from cranberry (Bland *et al.* 1996) and a native prairie (Norman and Polley 1989; Norman *et al.* 1992); (c) predicting the evapotranspiration, drainage, and soil moisture changes of chisel-plow corn, no-till corn, and a replanted prairie (Brye *et al.* 2000); (d) estimating bidirectional reflectance factors for plant canopies (Norman *et al.* 1985); (e) characterizing the water budget of irrigated crops (Norman and Campbell 1983; Thompson *et al.* 1993); (f) quantifying the pest-microenvironment interaction for spider mites on corn (Toole *et al.* 1984); (g) characterizing light penetration in corn (Norman 1980, 1988), predicting leaf wetness duration from dew fall, and distillation in snap beans (Weiss *et al.* 1989); and (h) evaluating various definitions for "surface" temperature (Norman *et al.* 1990; Norman and Becker 1995).

Cupid provides a useful platform for studying the relationship between aerodynamic temperature, which is related to the sensible heat flux from a canopy (cf. equation 7.1) and cannot be measured directly, and the radiometric temperature, which can be measured with thermal radiometers or infrared thermometers. The aerodynamic temperature of a surface is that temperature, which when combined with the air temperature and a resistance calculated from the log-profile theory, provides an estimate of the surface sensible heat flux (Norman and Becker 1995). The radiometric temperature is based on the infrared radiance emanating from a canopy. The directional radiometric temperature is calculated from the radiance measured by a narrow-fieldof-view infrared radiometer, and is actually referred to as the "ensemble directional radiometric surface temperature" (Norman and Becker 1995). The equations used in Cupid are outlined in Appendix A along with a comparison of model versus measured brightness temperatures supporting Cupid algorithms (Figure 7.A1). Numerous surface temperatures can be defined (Norman and Becker 1995). Converting the raw, calibrated infrared thermometer measurement of brightness temperature to a directional radiometric temperature requires a directional emissivity. Unfortunately, two

directional emissivities can be defined: a directional r-emissivity and a directional e-emissivity (Norman and Becker 1995). The directional r-emissivity is one minus the hemispherical-directional reflectance, which can be computed by various reflectance models (Verhoef 1984; Norman *et al.* 1985). This directional r-emissivity is based on the assumption that the canopy/soil system is isothermal; a condition that frequently does not occur, especially in sparse canopies such as those described in this chapter. The directional e-emissivity is the ratio of the spectral radiance of a particular canopy to the spectral radiance of the same canopy with the same temperature distribution but with each element being a black body. Both the directional r-emissivity and e-emissivity can be computed with the Cupid model.

A quantitative description of the relationship between convective and radiative fluxes can begin with energy budgets of all the individual vegetative and soil elements of the plant/soil system. The dominant vegetative component is usually the leaf, so the leaf energy budget must be evaluated for all layers and leaf angle classes (Norman 1979; Campbell and Norman 1997); including radiation and wind penetration into the vegetation (Goudriaan 1977), and physiological controls over stomatal conductance (Collatz *et al.* 1991, 1992). The dependence of leaf-boundary-layer conductance on leaf size, shape, and local wind speed must be known and is the source of some uncertainty (Grace 1981). The emissivity of individual leaves must also be known and a value of 0.97 appears suitable for most leaves.

The partitioning of the radiation absorbed at the soil-surface between conduction into the soil and convection into the canopy space is critical to the relation between aerodynamic and radiative temperatures; especially in sparse canopies. This occurs because a hot soil surface tends to contribute more to a radiometric temperature than an aerodynamic temperature. Although conduction of heat and water in the soil can be simulated reasonably using variations on the approach of Campbell (1985), convective exchange at the soil-surface beneath a canopy has proven troublesome. Recently, based on the work of Sauer *et al.* (1995), Kustas and Norman (1999a,b) suggested the following relation for the boundary layer conductance of the soil-surface beneath a canopy (cf. equation 7.B19):

$$g_S = 2.5\left(\sqrt[3]{(T_S - T_{AC})}\right) + 12(u_S) \tag{7.3}$$

where g_S is in mm s^{-1}, T_S is the soil surface temperature (°C), T_{AC} is the mean air temperature (°C) in the canopy space (often approximated by the mean canopy temperature), and u_S (m s^{-1}) is the wind speed above the soil at a height where the drag from the soil roughness is negligible (typically a few centimeters to a few tens of centimeters). Although g_S is expected to depend on surface roughness (Sauer *et al.* 1995), the above equation works well because beneath most canopies soil-surfaces are relatively smooth and wind speeds are relatively low. Soilsurface emissivities are more variable than leaf

emissivities (Salisbury and D'Aria 1992). Although some ground-based brightness temperature measurements are made with infrared thermometers sensitive to the 8–12 μm wavelength band, most aircraft and satellite brightness temperature measurements are made in the 10–12μm band where a soil emissivity of 0.96 is reasonable.

In Cupid, aerodynamic temperature is computed by several methods, but the most widely accepted method is described by equations (24) and (26) in Norman and Becker (1995), which uses an excess resistance for heat that is added to the aerodynamic resistance for momentum (cf. equation 7.1). The calculation of sensible heat flux in Cupid, which is necessary to calculate aerodynamic surface temperature, is described by Norman and Campbell (1983).

7.3 Cupid model validation

Predictions of various quantities with the Cupid model can be compared with measurements from the Lucky Hills site (Site 1) of the Monsoon 90 experiment (Kustas and Goodrich 1994); in particular, the energy balance components, the component temperatures of the vegetation and soil, the canopy/soil emissivity, and the soil-surface evaporation. Soil, canopy, and weather inputs for the Cupid model were obtained from published measurements for the Monsoon 90 experiment, and Table 7.1 contains a list of parameter values. One modification was made in the Ball *et al.* (1987) equation for stomatal conductance that is used in Cupid; namely the index given by A^*h_S/C_s was replaced by $A^*\phi(h_S)/C_s$ where

$$f(h_S) = \frac{\sqrt{h_S^2 + h_{S,MIN}^2}}{\sqrt{1 + h_{S,MIN}^2}} \tag{7.4}$$

and A^* is the leaf assimilation rate (μmol m^{-2} s^{-1}), h_S is the relative humidity at the leaf surface, and C_s is the CO_2 concentration at the leaf surface. The influence of leaf-surface relative humidity on stomatal conductance becomes negligible at $h_{S,MIN}$. This generalization of the Ball *et al.* (1987) approach provides for the possibility that leaf-surface humidity may be non-linearly related to stomatal conductance, and alleviates the well-known failure of the model at very low surface humidity; humidity that is likely in the Monsoon 90 experiment. By setting $h_{S,MIN} = 0$, the modified form of the Ball *et al.* (1987) index becomes identical to the original.

7.3.1 Energy balance components

The primary energy balance flux components are net radiation, soil heat conduction, latent heat, and sensible heat. Figure 7.1(a)-(d) contains comparisons of the flux components and the results indicate that model and

Table 7.1 Parameter values used in the Cupid model for simulations with Lucky Hills observations

Parameter	*Source*	*Value*
Maximum velocity of carboxylation	Gutschick (1996)	81 μmol m^{-2} s^{-1}
Slope of Ball *et al.* (1987) stomatal conductance curve	Gutschick (1996)	11
Intercept of stomatal conductance curve	Ball *et al.* (1987)	0.04
Minimum humidity for stomatal conductance effect ($h_{S,MIN}$)		0.5
Saturated hydraulic conductivity	Flerchinger (pers. comm.)	38 mm h^{-1}
Slope for soil moisture release curve	Flerchinger (pers. comm.)	4.35
Air entry potential	Flerchinger (pers. comm.)	-1.1 J kg^{-1}
Texture 70% sand, 20% silt, 10% clay	Flerchinger (pers. comm.)	
Bulk density	Flerchinger (pers. comm.)	1.35 Mg m^{-3}
Leaf area index	Daughtry *et al.* (1991)	0.5
Fraction of vegetation that is green		0.8
Clumping factor		0.7
Height of vegetation		0.5 m
Spherical leaf angle distribution		
Displacement height	Raupach (1994)	0.22 m
Roughness length	Raupach (1994)	0.08 m
Soil emissivity	Humes *et al.* (1994)	0.96
Leaf emissivity		0.97
Leaf size		0.01 m
Leaf absorptivity in PAR		0.85
Leaf absorptivity in near-infrared		0.15
Soil reflectivity in PAR		0.15
Soil reflectivity in near-infrared		0.25

micro-meteorological measurements described by Stannard *et al.* (1994) are in reasonable agreement. Root mean square difference (RMSD) values (Willmott 1982) are 20, 25, 30, and 40 W m^{-2} for R_N, G, H, and LE, respectively. The largest difference occurs with the latent heat when the soil-surface is wet and the Cupid model tends to predict greater evaporation fluxes from the surface than the eddy covariance measurements indicate. Using the original form of the equation relating stomatal conductance to other factors (Ball *et al.* 1987) results in predictions of transpiration being about 20% less than the results shown in Figure 7.1(c).

7.3.2 Component temperatures of vegetation and soil

The individual temperatures of the vegetated canopy and soil-surface were measured for several time periods in the Monsoon 90 experiment using infrared thermometers directed toward the appropriate surfaces (Norman *et*

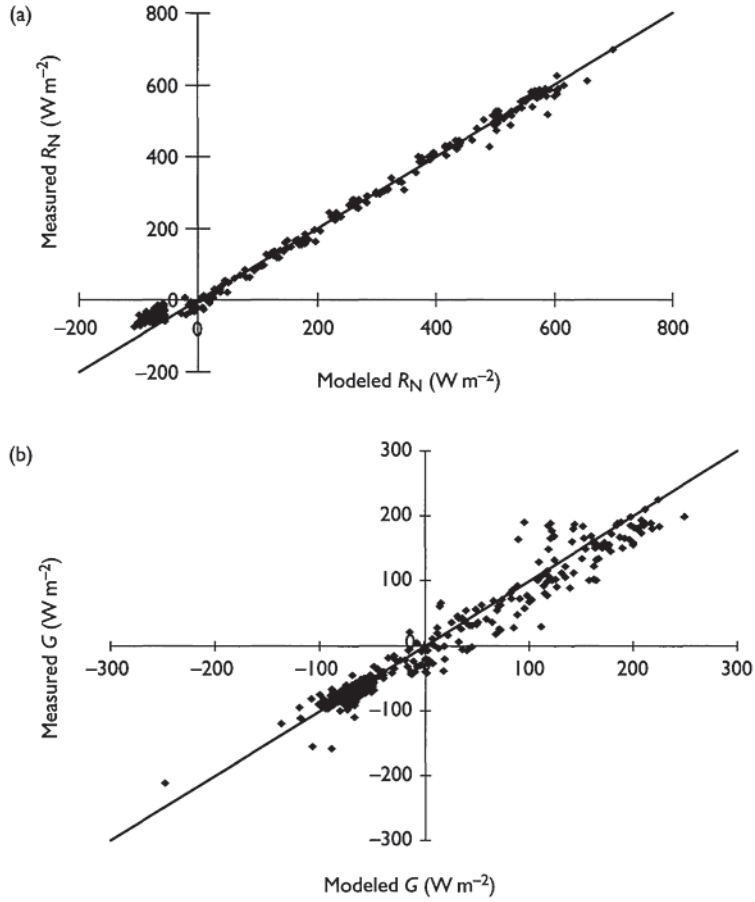


Figure 7.1 Comparison of (a) net radiation and (b) soil heat flux measurements with predictions from the Cupid model for the Lucky Hills site. Comparison of eddy covariance measurements of (c) latent heat and (d) sensible heat fluxes with predictions from the Cupid model for the Lucky Hills site.

al. 1995). Figure 7.2 contains a comparison of predicted component temperatures from Cupid with field measurements. Although some scatter is apparent, the agreement appears to be reasonable. The large temperature differences between the vegetation and soil (> 20°C) surface are typical of sparse vegetation with dry soil surfaces.

7.3.3 Canopy/soil emissivity

The measured soil emissivity of 0.96 (Humes *et al.* 1994) was used as an input in the Cupid model. Assuming the leaf emissivity to be 0.97, an estimate

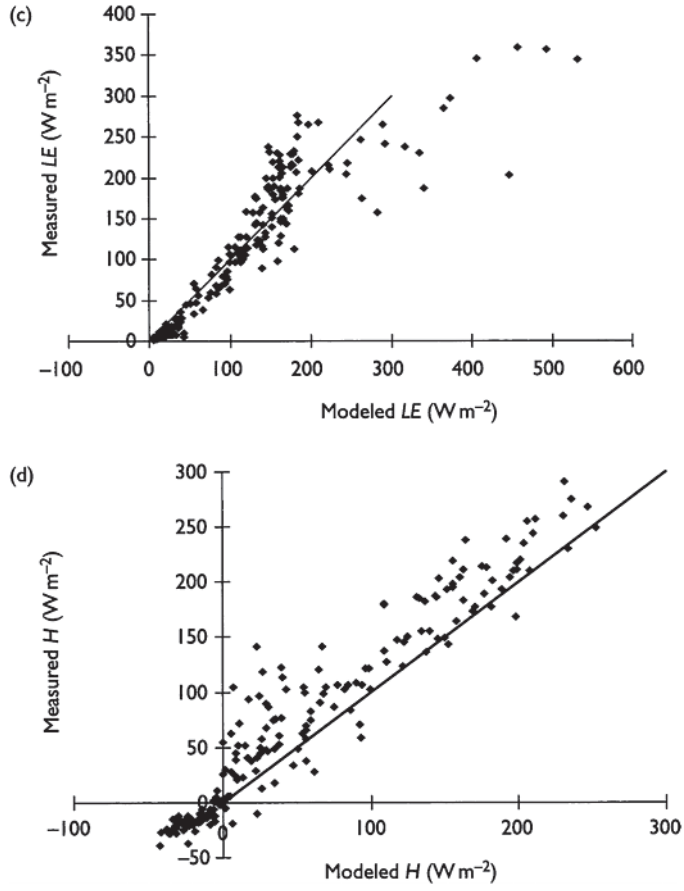


Figure 7.1 (Continued).

of the emissivity of the vegetation/soil system from Cupid can be compared with measurements from Humes *et al.* (1994). The emissivity estimate from Cupid is 0.97. Humes *et al.* (1994) estimated composite emissivity values by two methods and got 0.97 and 0.98. This agreement within 0.01 is probably within the accuracy of the measurement method.

7.3.4 Soil and canopy evaporative fluxes

During Monsoon 90, chamber measurements of soil and vegetation evaporative fluxes were made using a device and procedure described by Stannard (1988). By combining these chamber measurements with the eddy

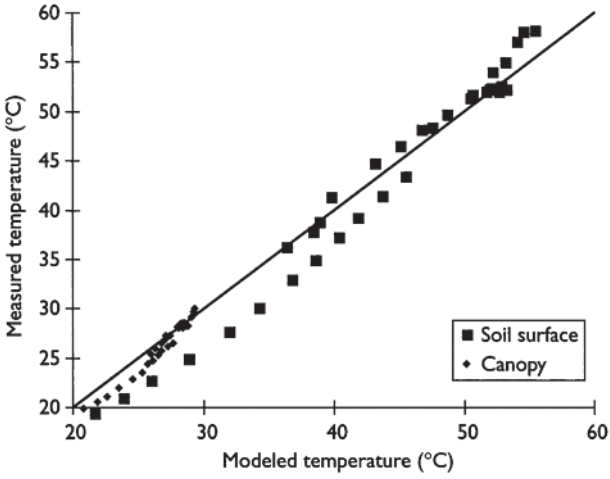


Figure 7.2 Comparison of predicted soil-surface and vegetation canopy temperatures from Cupid with measurements for the Lucky Hills site.

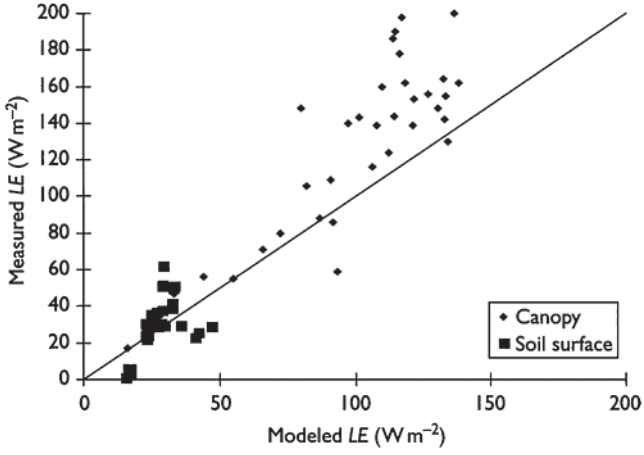


Figure 7.3 Comparison of soil-surface evaporation and canopy transpiration from Cupid with measurements from the Lucky Hills site.

covariance latent heat flux measurements above the canopy, Stannard (1999) was able to separate the soil-surface evaporation from the vegetative-canopy transpiration. Figure 7.3 shows a preliminary comparison and indicates reasonable agreement between the Cupid model and measurements, although there is a tendency for the model to predict lower transpiration than the

measurements. This discrepancy between modeled and measured transpiration is under further investigation.

7.4 Evaluation of radiometric versus aerodynamic temperature using cupidΩ

The brightness temperature (related to the radiometric temperature) can be measured directly with infrared radiometers; relating this measurement to vegetation/soil sensible heat flux is exceedingly valuable because such infrared measurements are routinely made on the ground, from aircraft and from satellites. Relating radiometric and aerodynamic surface temperatures to each other offers the possibility of translating maps of surface brightness temperature into maps of surface fluxes.

Using the validated Cupid model (with the same parameterizations as used in the validation), brightness, radiometric, and aerodynamic temperatures were simulated for 36 cases: two radiometer view angles (ϕ = 0° or nadir and 55°), two wind speeds (1 and 5 m s^{-1}), LAI = 0.5 (a clumping factor, $\Omega = 0.7$, a green fraction, $f_G = 0.8$, $z_{OM} = 0.08$ m), LAI = 1.5 ($\Omega = 0.7$, $f_G = 0.8$, $z_{OM} = 0.05$ m), LAI = 3.0($\Omega = 1$, $f_G = 1$, $z_{OM} = 0.05$ m), unstressed vegetation with a dry soil surface, unstressed vegetation with a moist soil surface, and stressed vegetation with a dry soil surface. In addition, there was an unstressed vegetation case with LAI = 3.0, $h_C = 5$ m, $\Omega = 1$, $f_G = 1$, z_{OM} = 0.5 m and leaf width = 0.05 m, which are typical values for the riparian vegetation. The weather data at Site 1 from Day Of Year (DOY) 209 was used (except for the wind speed); the solar radiation, S, varied from 120 to 990 Wm^{-2}, the vapor pressure varied from 0.85 to 1.25 kPa, and the air temperature, T_A, varied from approximately 28.4 to 31.5°C. Values of the aerodynamic temperature, T_0, were estimated using equations (7.1) and (7.2) with $z_{OM} = z_{OH}$ because of the numerous uncertainties associated with parameterization of z_{OH} (Verhoef *et al.* 1997).

In Figure 7.4(a), the relation between $T_R(\phi) - T_0$ and $T_S - T_C$ for the 18 cases with a nadir viewing angle are illustrated, and in Figure 7.4(b) the results for the 18 cases with = 55° off nadir are shown. The difference between radiometric and aerodynamic temperatures is highly variable and clearly indicates the challenge associated with any simple scheme to relate remotely sensed temperature to surface heat fluxes. Also note the negative values of $T_R(\phi) - T_0$, which are predominately from wet soil conditions; this has been observed in experimental data (Sun and Mahrt 1995). For many of the partial canopy cover cases with LAI < 3 there is a strong linear relationship between $T_R(\phi) - T_0$ and $T_S - T_C$; however, the slope varies with stress and moisture condition supporting the simulations of Blyth and Dolman (1995). Moreover, comparing Figure 7.4(a) and (b), the slopes change with radiometer viewing angle. For the higher vegetation cover, LAI = 3, there appears to be little

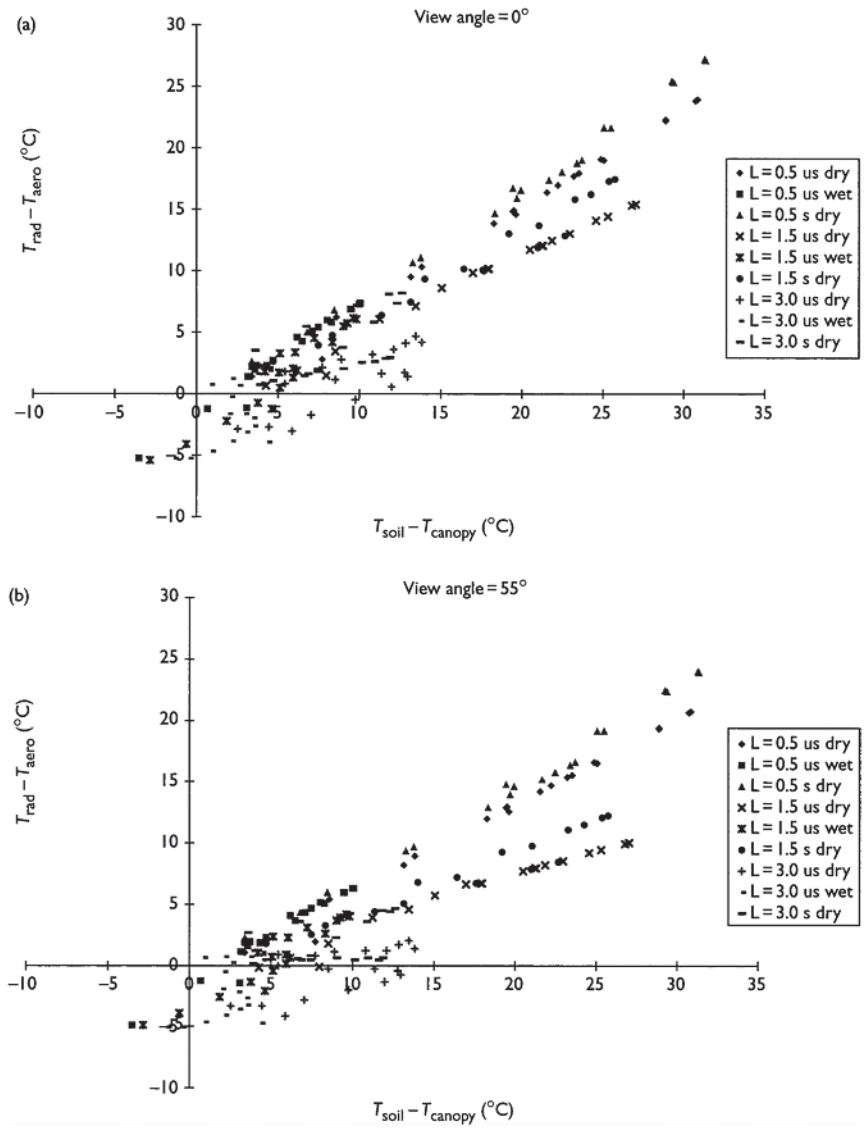


Figure 7.4 Radiometric-aerodynamic temperature differences ($T_R(\varphi) - T_0$) plotted against soil surface-canopy temperature differences ($T_S - T_C$) for various Cupid model runs (L = LAI, us = unstressed, s = stressed, dry = dry surface moisture, wet = wet surface moisture). Each symbol type contains a range of solar radiation values and two wind speeds (u = 1 and 5 m s^{-1}). The aerodynamic temperature calculation is based on $z_{OH} = z_{OM}$ (cf.equation 7.2). The view angle for radiometric temperature is (a) nadir ($\varphi = 0°$) and (b) 55º from nadir ($\varphi = 55º$).

dependency of $T_R(\phi) - T_0$ on $T_S - T_C$, with $T_R(\phi) - T_0$ values generally within ± 2 K. Small differences between $T_R(\phi)$ and T_0 under these conditions indicate that they may be highly correlated, which has been observed experimentally for dense grassland cover (Sun 1999; Kustas *et al.* 2001).

7.5 Two-source models accommodating differences between T_0 and $T_R(\phi)$

In regional applications with remote sensing data, a model such as Cupid has input parameters that are not routinely available. This has lead to the development of the STS model requiring a minimum number of parameters that can be obtained from remote sensing. This model, however, accommodates differences between $T_R(\phi)$ and T_0 and, therefore, can consider factors affecting their relationship illustrated in Figure 7.4.

Several versions/forms of the STS model have been developed, which take advantage of the type of $T_R(\phi)$ observations available. These include having $T_R(\phi)$ data at several viewing angles and multiple $T_R(\phi)$ observations over the day. Although the different versions of the STS model have been evaluated with flux observations, the Cupid simulations used in generating Figure 7.4 gives us an opportunity to evaluate the limitations of the various forms of the model under more extreme environmental conditions.

7.5.1 Simplified two-source model

A detail description of the original STS model using a single $T_R(\phi)$ observation (1ANGLE_PT) can be found in Norman *et al.* (1995)—hereafter referred to as N95. A brief description of the model formulations is given in Appendix B, and includes modifications to several of the original N95 formulations to account for temporal variations in net radiation divergence through the canopy layer and in the soil heat flux-soil net radiation ratio (Kustas *et al.* 1998). ln addition, experience has revealed that the exponential extinction of net radiation (i.e. equations (7.B8) and (7.B9)) is only appropriate for canopies of nearly full cover and contains significant systematic errors for sparse canopies with relatively hot soil surfaces. These errors occur because the contribution of soil thermal radiation to net radiation depends on soil surface temperature, which can be more than 20° C above vegetation or air temperature; hence, exponential equations such as equation (7.B8) or (7.B9) do not account for such surface conditions. For a sparse canopy having a leaf area index (LAI) ~0.5, with differences in soil and vegetation temperatures on the order of 20° C, net radiation absorbed by the soil surface and canopy calculated from equations (7.B8) and (7.B9) can be in error by over 50 W m^{-2} resulting in relative errors of ~15 and ~40% for the soil and canopy, respectively.

In the present analysis, a more physically based algorithm for estimating the divergence of R_N was constructed requiring incident solar radiation observations and formulations for the transmission of direct and diffuse shortwave radiation and for the transmission of long-wave radiation through the canopy (Campbell and Norman 1997). Since the reflection and absorption of radiation in the visible and near-infrared wavelengths is markedly different for vegetation and soils, the visible and near-infrared reflectances of the soil and vegetation were evaluated separately before combining to give an overall shortwave albedo. The equations for estimating the transmission and reflection of direct and diffuse shortwave radiation are described in Chapter 15 of Campbell and Norman (1997); hence, the net shortwave radiation balance for the soil ($S_{N,S}$) and canopy ($S_{N,C}$) are computed separately from the net long-wave radiation balance for the soil ($L_{N,S}$) and canopy ($L_{N,C}$). The long-wave balance for the soil-vegetation-atmosphere system is derived by calculating diffuse radiation transmission through the canopy (Ross 1975). A simpler formulation of the net long-wave radiation balance than that described in Ross (1975) was derived where a single exponential equation is used for estimating the transmission for both the soil and canopy,

$$L_{N,S} = \exp(-\kappa_L \mathrm{LAI}) L_{sky} + [1 - \exp(-\kappa_L \mathrm{LAI})] L_C - L_S \tag{7.5a}$$

$$L_{N,S} = [1 - \exp(-\kappa_L \mathrm{LAI})][L_{sky} + L_S - 2L_C] \tag{7.5b}$$

where the extinction coefficient for diffuse radiation depends on LAI, and if LAI = 0.5, $\kappa_L = 0.95$ (Campbell and Norman 1997); L_C, L_S and L_{sky} are the long-wave emissions from the canopy, soil, and sky, respectively. L_C, L_S, and L_{sky} are computed from the Stefan-Boltzmann equation using canopy temperature, soil temperature, and shelter level air temperature and vapor pressure (Brutsaert 1982). Thus, equations (7.B8) and (7.B9) are replaced by visible and near-infrared radiation penetration equations from Chapter 15 of Campbell and Norman (1997) combined with equations (7.5a) and (7.5b) (i.e. $R_{N,S} = S_{N,S} + L_{N,S}$ and $R_{N,C} = S_{N,C} + L_{N,C}$).

The radiative exchange algorithms used in the model apply to vegetative canopies with leaves randomly distributed over the surface. When the leaves are not randomly distributed over the surface but clumped as in the case of row crops, they may only intercept 70–80% of the radiation in comparison to the same crop randomly distributed over the surface (Campbell and Norman 1997). Models to estimate radiation extinction for clumped vegetation have been developed (e.g. Gijzen and Goudriaan 1989 for a row crop), but are rather complex and require additional information about the surface that will not be available operationally. An alternative is to use the same formulations described above, but with LAI multiplied by a clumping factor Ω, namely ΩLAI, (Chen and Cihlar 1995). Campbell and Norman

(1997) suggest for strongly clumped canopies, Ω is a function of the solar zenith angle, θ_S, and can be estimated by the following expression:

$$\Omega(\theta_S) = \frac{\Omega(0)}{\Omega(0) + [1 - \Omega(0)]\exp[-2.2(\theta_S)^p]}, \quad p = 3.8 - 0.46D \quad (7.6)$$

where Ω(0) is the clumping factor when the canopy is viewed at nadir and D is the ratio of vegetation height versus width. The value of Ω(0) can be estimated with general knowledge of LAI and the fractional cover of the canopy. For the shrub site used in this study, LAI = 0.5 and fractional cover, $f_C \approx 0.28$. If the vegetation were randomly distributed and the leaf angle distribution approximated a spherical distribution, the canopy gap fraction from the zenith would be exp(–0.5 LAI) ≈ 0.78. In actuality, the vegetation is clumped so the field-scale LAI of 0.5 corresponds to a local LAI (LAI_L) within the vegetated region of $LAI_L \approx LAI/f_C \approx 1.80$. If all the leaves contained within the vegetated region are randomly distributed, then the transmission of this vegetated region is $f_C \exp(-0.5\ LAI_L + (1 - f_C) \approx 0.83$; therefore, exp(–0.5ΩLAI) = 0.83 so that Ω = 0.7. For canopies with low LAI, Ω is almost independent of angle until very large zenith angles are reached. A more general approach to obtaining clumping factors from canopy architecture is given by Kucharik *et al.* (1999).

7.5.2 *Model formulations using dual-angle radiometric observations*

The model accounts in a simplified way for directional or view angle effects on the radiometric temperature (cf. equations 7.B1 and 7.B2), so that one can use $T_R(\phi)$ observations at multiple viewing angles and avoid the need for the Priestley-Taylor parameterization used for estimating canopy transpiration in the original model; otherwise by using the Priestley-Taylor parameterization the revised model does not require a measurement of T_A (Kustas and Norman 1997).

Use of the Priestley-Taylor assumption and the need for an estimate of f_G, which is difficult to estimate from remote sensing, can be avoided by having $T_R(\phi)$ estimated from two view angles because T_C and T_S can be obtained from the simultaneous solution of two equations and two unknowns. For example, with two view angles $\phi_1 = 0°$ and $\phi_2 = 55°$, T_S can be solved via equation (7.B1) as follows:

$$T_S = \left[\frac{f(\phi_2)T_R^4(\phi_1) - f(\phi_1)T_R^4(\phi_2)}{f(\phi_2)[1 - f(\phi_1) - f(\phi_1)(1 - f(\phi_2))]}\right]^{1/4} \quad (7.7)$$

which permits T_C to be computed by the expression

$$T_C = \left[\frac{T_R^4(\phi_1) - (1 - f(\phi_1))T_S^4}{f(\phi_1)}\right]^{1/4} \quad (7.8)$$

This approach is similar to the method described by Kimes (1983) to separate canopy and soil-surface temperatures. Thus, H_C and H_S can be computed directly from equations (7.B12)–(7.B13) or (7.B14)– (7.B15). With an estimate of H_S, LE_S can be computed as the residual in equation (7.B20) and similarly LE_C can be computed as a residual, namely $LE_C = R_{N,C} - H_C$.

The Along-Track Scanning Radiometer (ATSR) is capable of making two nearly simultaneous measurements of brightness temperature from two different view angles, at nadir (0°) and 55° along the satellite path, so that this version of the model, 2ANGLE, may have practical applications for estimating surface energy fluxes. A land surface temperature algorithm developed by Prata (1993) and tested with *in situ* field data indicates RMSD values within ±1.5K can be achieved (Prata 1994). The ±1.5K uncertainty using ATSR data was verified independently by Ottle and Francois (1994). With the appropriate ground measurements of the meteorological variables u and T_A, basic information concerning vegetation type and cover, and brightness temperature from two view angles, H can be estimated without requiring empirically determined "adjustment" factors for R_{AH} or the assumptions used in estimating LE_C via equation (7.B21).

The 2ANGLE model accounts for the difference between T_0 and $T_R(N)$, avoids using the Priestley-Taylor assumption for the vegetation, and obviates the need for estimating f_G. A major difficulty with the STS model, and many other models driven by the $T_R(\phi) - T_A$ difference, is the requirement for a T_A observation. T_A is not measured in many regions and where it is measured, it only represents local conditions near the site of the measurement and not at each satellite pixel. With most current satellite observations of $T_R(\phi)$ at the 1 km pixel scale, significant variations in near-surface meteorological conditions may exist depending on surface conditions. Methods using satellite data indicate at least a ≈ 3 K uncertainty in the estimate of T_A when compared to standard weather station observations (Goward *et al.* 1994; Prince *et al.* 1998). Zhan *et al* (1996) and Anderson *et al.* (1997) showed that two-source models are generally more sensitive to errors in $T_R(\phi) - T_A$ than to most other model parameters, thus it is a major advantage for a model not to require a measurement of T_A. This is one of the main advantages of the Two-Source-Time-Integrated (TSTI) model described below (Anderson *et al.* 1997).

If the Priestley-Taylor approach for partitioning $R_{N,C}$ between H_C and LE_C via equations (7.B7) and (7.B21) and two view angles of $T_R(\phi)$ are used, then a measurement of T_A is not required. With the "parallel" resistance network, this is derived by substituting the expression for H_C from equation (7.B13) and LEc from equation (7.B21) into the energy budget equation for the vegetation, equation (7.B7), and by using T_C computed via equation (7.8); thus equation (7.B13) can be rewritten to yield an estimate of T_A, $T_{A,E}$, namely,

$$T_{A,E} = T_C - \frac{R_{AH} R_{N,C}[1 - \alpha_{PT} f_G(\Delta/(\Delta + \gamma))]}{\rho C_p} \tag{7.9}$$

This estimate of the air temperature can be used to calculate H_S in equation (7.B12), which then yields an estimate of the total sensible heat flux, H. Finally, LE_S can be calculated using equation (7.B20) while LE_C is computed directly from equation (7.B21). This version of the model, 2ANGLE_PT, has potential for operational applications because air temperature is not required. The model requires only a nominal estimate of wind speed. With the series resistance network, the temperature in the canopy air space, T_{AC}, is estimated instead of T_A (Kustas and Norman 1997).

The results of the comparison between modeled and observed heat fluxes and the sensitivity analysis to model input variables and parameters indicates that by using the semi-empirical Priestley-Taylor parameterization and computing T_A the 2ANGLE_PT model satisfactorily predicted the heat fluxes. Furthermore, predictions are essentially unaffected by the 1–2 K error in estimating $T_R(N)$ from satellites and errors in extrapolating T_A from a sparse network of meteorological observations to each satellite pixel, a very unreliable approach (Goward *et al.* 1994).

The semi-empirical Priestley-Taylor parameterization is based on results of the model of Ball *et al.* (1987), which relates assimilation and stomatal conductance. It supports the hypothesis embodied in this approach; however, the Priestley-Taylor constant α_{PT} of ~1.3 may vary with vegetation type and environmental condition (Kustas and Norman 1999a,b). To better understand the variability and limitations in the Priestley-Taylor parameterization, the output of Cupid described above is used with data from a semiarid shrub dominated semiarid rangeland site during Monsoon 90. These data have already been used to evaluate the STS model (Norman *et al.* 1995). The Cupid-simulated heat fluxes from the canopy and soil as well as $T_R(\phi)$ values are used to evaluate the effect on STS model flux predictions. The results provide a guide to assess the validity of parameterizations in the STS model, with special attention to the Priestley-Taylor approach and the potential variability of the Priestley-Taylor constant α_{PT}.

7.5.3 Two-source-time-integrated model formulation

The TSTI model of Anderson *et al.* (1997) (presently called AtmosphereLand-EXchange-Inverse, ALEXI, Mecikalski *et al.* 1999) provides a practical algorithm for using a combination of satellite data, synoptic weather data, and ancillary information to map surface energy flux components on a continental scale (Mecikalski *et al.* 1999). The ALEXI approach builds on the earlier work with the STS model (Norman *et al.* 1995; Kustas and Norman 1996) by using remote brightness temperature observations at two times in

the morning hours and considering planetary boundary layer processes. The methodology removes the need for a measurement of near-surface air temperature and is relatively insensitive to uncertainties in surface thermal emissivity and atmospheric corrections on the GOES brightness temperature measurements. Anderson *et al.* (1997) and Mecikalski *et al.* (1999) have shown that surface fluxes retrieved from the ALEXI approach compare well with measurements. The ALEXI approach is a practical means to operational estimates of surface fluxes over continental scales with 5–10 km pixel resolution.

7.6 Comparison of three versions of the STS model (I ANGLE_PT, 2ANGLE, and 2ANGLE_PT) flux predictions versus cupid

In the comparisons that follow between STS and Cupid model predictions of the fluxes, the output for the 1030 local time will be used. This is the approximate time of coverage of the Landsat-7 and the EOS-Terra satellite supporting the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) instrument, which will be used for surface flux monitoring (Schmugge *et al.* 1998; French *et al.* 2002). The comparisons are for all the various conditions outlined in Section 7.4. This yields 22 values of the energy balance components for a given time. The ALEXI model could not be evaluated using the Cupid simulations since atmospheric boundary-layer data are not available.

Based on preliminary results with the Cupid simulations and results from Kustas and Norman (1999a,b) in the application of the STS model with radiometric temperature observations over a sparse cover of irrigated cotton crop, there were some additional modifications made to several of the model algorithms. The first has to do with the fact that the simple formulation for soil heat flux (equation 7.B10) had to be modified in order to account for very low wind speeds near the soil surface. Under these conditions, Cupid predicts a value of $G/R_{N,S} > 0.5$ (see Figure 7.5). Therefore, an additional modification to the value of c_G was made by adjusting the wind speed near the soil surface u_S when it was less than 0.25 m s^{-1} with the empirical curve fit to the data in Figure 7.5.

The second modification was to create a new algorithm for adjusting the a_{PT} parameter. This is based in part on the results obtained by Kustas and Norman (1999a,b) for an irrigated cotton canopy with an LAI ~ 0.4. With T_C and T_S observations over the cotton field, Kustas (1990) applied a two-source modeling scheme to derive canopy and soil heat fluxes. The model indicated that heat was being advected towards the canopy from the surrounding hot bare soil surface yielding $H_C \sim -100$W m^{-2}. To obtain this

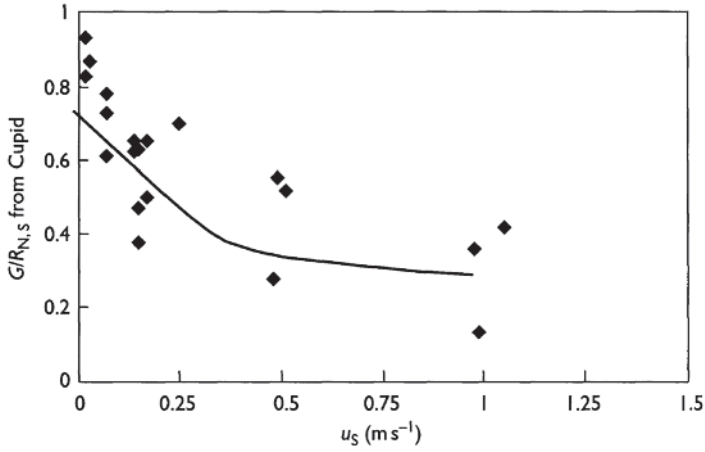


Figure 7.5 Comparison of $G/R_{N,S}$ from Cupid versus u_S from the STS model. The curve is an empirical fit to the data.

result for the canopy and physically realistic values of the other components, particularly LE_S, a value of $\alpha_{PT} \sim 2$ had to be adopted. For natural vegetated surfaces the effect of advection from the surrounding bare soil surface is not likely to be as extreme, and therefore $\alpha_{PT} < 2$. Therefore, an algorithm was incorporated so that the model would initially assume $\alpha_{PT} = 2$, but in cases where $T_R(\phi)$ was high due to dry soil and elevated canopy temperatures (e.g. when $T_C > T_A$), the value of α_{PT} would be allowed to decrease by 0.1 increments until a reasonable solution is obtained. From experience we have found that when $T_R(\phi) - T_A$ is significant (i.e. nominally $\gtrsim$15 K with LAI 0.5) a high value for $\gtrsim T_S$ 60°C is estimated such that $R_{NS} - H_S = LE_S < 0$; this is not a physically realistic solution under daytime conditions. This new procedure would allow for a variable α_{PT} value, which would implicitly consider variations in canopy resistance of vegetation types (McNaughton and Jarvis 1991).

7.6.1 Results using 1ANGLE_PT

The STS model output of the component fluxes for R_N, H, and LE using the 1ANGLE_PT version compared to Cupid is illustrated in Figure 7.6. In general, closer agreement between Cupid and the STS flux predictions was obtained using the "series" versus "parallel" resistance network described in Appendix B (see also Norman *et al.* 1995). Therefore, only the results using the series approach are described. The simulated fluxes for stressed and unstressed vegetation conditions are denoted with different symbols

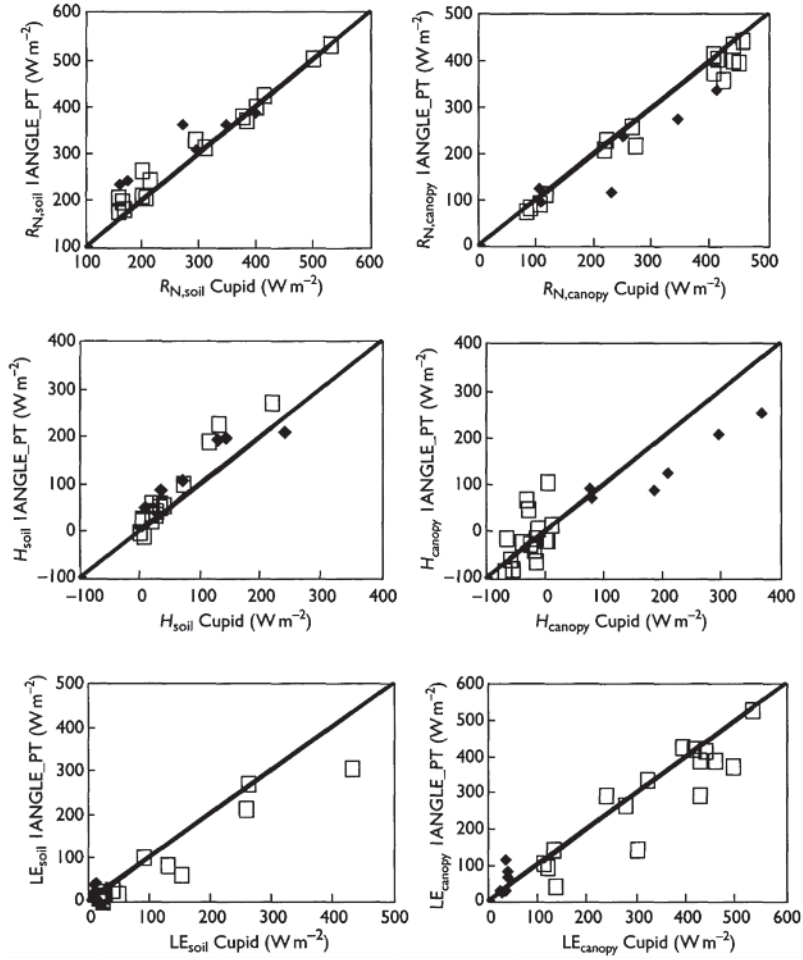


Figure 7.6 Comparison of component (canopy and soil) net radiation, sensible, and latent heat fluxes from Cupid versus values from the 1ANGLE_PT version of the STS model. The open square symbol represents unstressed vegetation and the solid diamond symbol represents stressed vegetation conditions.

in order to show that other factors (wind speed and amount of cover or LAI) contribute to the wide range in simulated heat fluxes from Cupid. In general, the component fluxes predicted by STS is in fair agreement with the Cupid simulated values. The RMSD values for the component fluxes are listed in Table 7.2 and the indicated discrepancies are ~50W m^{-2}. The combined canopy and soil net radiation, soil heat flux, and sensible and latent heat fluxes along with the RMSD values are illustrated in Figure 7.7. The RMSD values

Table 7.2 Statistical results comparing Cupid flux output of the component heat fluxes with the STS model using the 1ANGLE_PT version under the various vegetation, surface of soil moisture, and wind speed condition described in Section 7.4[a]

FLUX ($W\ m^{-2}$)	*Root mean square difference, RMSD,* ($W\ m^{-2}$)
$R_{N,S}$	38
$R_{N,C}$	40
G	49
H_S	43
H_C	38
LE_S	58
LE_C	62

Note

a The following constant meteorological conditions existed for the comparisons with Cupid: solar radiation (S)= 882 W m^{-2}, *TA* = 28.4ºC, and relative humidity (RH)= 33%.

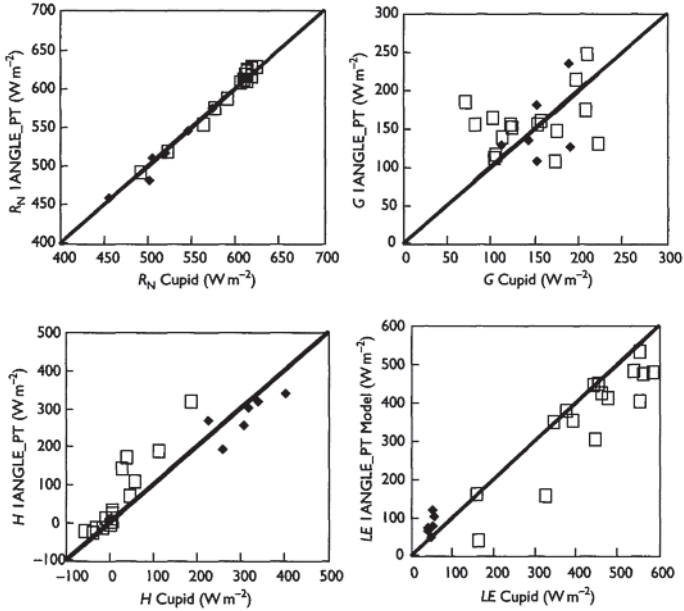


Figure 7.7 Comparison of net radiation, soil, sensible, and latent heat fluxes from Cupid versus values from the 1ANGLE_PT version of the STS model. The open square symbol represents unstressed vegetation and the solid diamond symbol represents stressed vegetation conditions.

for H and LE are higher at 60 and 70 W m^{-2} than what is generally obtained when comparing to observations (i.e. ~50 Wm^{-2}). However, the r^2 values for H and LE are 0.85 and 0.90, indicating that the STS model is accounting for a significant amount of the variation in the heat fluxes.

7.6.2 Results using 2ANGLE_PT & 2ANGLE model versions

In contrast to the results using the 1ANGLE_PT version, generally closer agreement between Cupid and the STS flux predictions was obtained using the "parallel" versus "series" resistance approach with the 2ANGLE and 2ANGLE_PT versions. This was also the result obtained by Kustas and Norman (1997) using $T_R(\phi)$ observations and comparing with flux measurements over a heterogeneous grassland.

As one might expect, given the assumptions that go into the 2ANGLE_PT model version, it cannot compute reliable heat fluxes under stressed conditions (Figure 7.8). However, under non-stressed conditions the values of $T_R(\phi)$

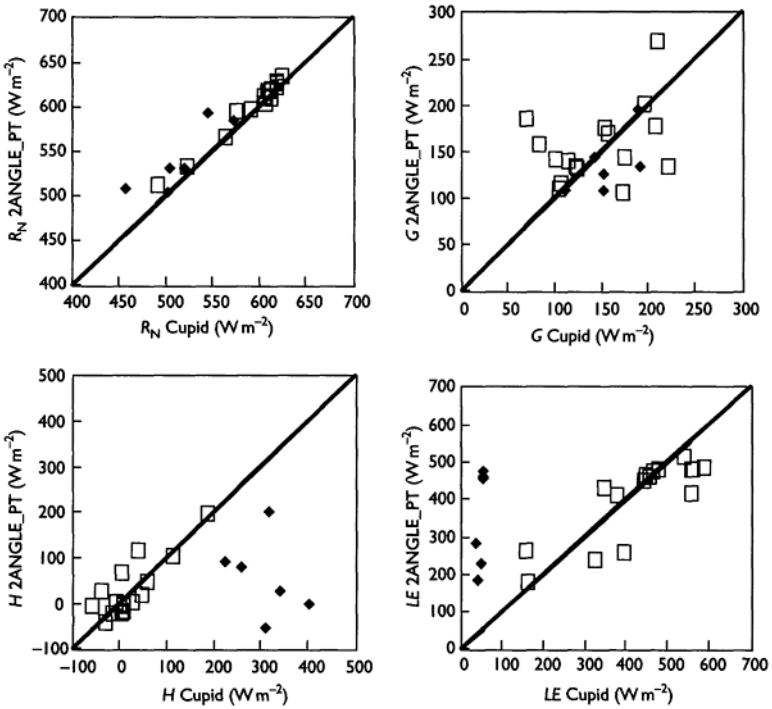


Figure 7.8 Comparison of net radiation, soil, sensible, and latent heat fluxes from Cupid versus values from the 2ANGLE_PT version of the STS model. The open sqaure symbol represents unstressed vegetation and the solid diamond symbol represents stressed vegetation conditions.

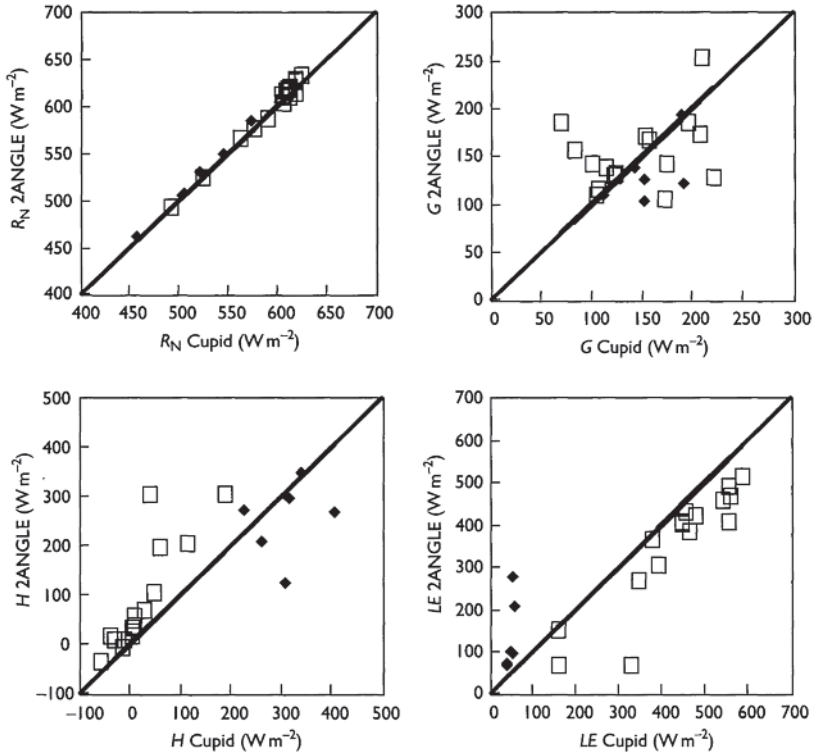


Figure 7.9 Comparison of net radiation, soil, sensible, and latent heat fluxes from Cupid versus values from the 2ANGLE version of the STS model. The open square symbol represents unstressed vegetation and the solid diamond symbol represents stressed vegetation conditions.

from Cupid for ϕ= 0 and 55° indicated reasonable results can be obtained even when the cover is high (i.e. LAI = 3). RMSD values for *H* and LE for the unstressed cases are 40 and 75 Wm^{-2}, respectively. This is comparable to what was obtained with the 1ANGLE_PT version.

With the 2ANGLE version of the STS model, the results are generally better than 2ANGLE_PT, and the 2ANGLE can predict more realistic heat fluxes under the stressed vegetation case (Figure 7.9). However, the results are not very encouraging since the scatter between Cupid and the 2ANGLE estimates of the heat fluxes is significantly larger than the 1ANGLE_PT version with RMSD ~ 100 W m^{-2} for *H* and LE.

These results are similar to what was obtained using $T_R(\phi)$ observations from prairie grassland sites containing a range of cover, stress, and wind conditions (Kustas and Norman 1997). In that study they found

discrepancies between predicted and observed H and LE were on the order of 70–80 W m^{-2}.

7.6.3 Summary of Cupid versus STS flux predictions

The results above suggest that the original version of the STS model, namely the 1ANGLE_PT version, yields the most reliable predictions of the heat fluxes over a wide range of conditions. It appears that using two sensor view angles for $T_R(\phi)$ can give reliable results under unstressed conditions using the 2ANGLE_PT approach; however, this version is not suitable for stressed conditions. Although the 2ANGLE approach does not require the Priestley-Taylor assumption, it will be difficult to extract reliable canopy and soil temperatures in order for this approach to yield reliable values of H and LE; the uncertainty of this approach is on the order of 100 W m^{-2}, which is obviously unacceptable.

7.7 Evaluating the effects of subpixel heterogeneity on pixel-derived fluxes

The pixel resolution of present and future satellite radiometers providing $T_R(\phi)$ images ranges from ~100m (e.g. Landsat TM) to ~5km (GOES); therefore, pixels will contain a mixture of vegetation and soil. In many cases riparian areas are only tens to hundreds of meters wide so that satellite $T_R(\phi)$ observations will commonly contain a mixture of riparian vegetation and the surrounding area, which for semiarid areas is typically sparsely vegetated. Kustas and Norman (2000) ran Cupid under the present set of environmental conditions and found that one of the greatest contrasts in energy flux partitioning and in the magnitude of $T_R(\phi)$ comes from the simulations using stressed shrub vegetation under low cover and dry soil versus unstressed riparian vegetation with wet surface soil moisture conditions. The output from 1ANGLE_PT for these two surfaces is listed in Table 7.3 for the solar radiation, air temperature, and relative humidity existing at 1030 Mountain Standard Time (MST) on DOY 209 for both the 1 and 5ms^{-1} wind speed conditions. These flux predictions from 1ANGLE_PT are similar to the Cupid-simulated output (Kustas and Norman 2000).

Subpixel heterogeneity from mixtures of riparian trees and stressed shrubs will cause an error in the calculated flux because the relationship between $T_R(\phi)$ and the heat fluxes is non-linear. However, energy balance models using remote sensing observations over heterogeneous surfaces have found the effects of subgrid or subpixel variability to be relatively minor (Sellers *et al.* 1995; Kustas and Humes 1996; Friedl 1997). These results may have been obtained under conditions where the variability of surface characteristics was not great enough to cause significant errors. The set of conditions

Table 7.3 1ANGLE_PT output of the surface fluxes for stressed shrub (SS) and riparian vegetation (RIP) under meteorological conditions at 1030 MST on DOY 209 (S = 882 W m^{-2}, T_A = 28.4ºC, and RH = 33%) with u = 1 and 5 m s^{-1}

Surface condition	R_N (W m^{-2})	G (W m^{-2})	H (W m^{-2})	LE (W m^{-2})	$T_R(0°)$ (°C)
Stressed Shrub (SS), dry soil; LAI = 0.5, h_C = 0.5 m, u = 1 m s^{-1}	460	126	310	24	55.4
Riparian vegetation (RIP), wet soil; LAI = 3, h_C = 5 m, u = 1 m s^{-1}	626	153	14	459	29.9
Stressed shrub (SS), dry soil; LAI = 0.5, h_C = 0.5 m, u = 5 m s^{-1}	504	147	344	13	48.6
Riparian vegetation (RIP), wet soil; LAI = 3, h_C = 5 m, u = 5 m s^{-1}	629	157	−11	483	28.4

simulated by Cupid provide an extreme case that might exist naturally in a semiarid environment, or be imposed via agricultural practices (e.g. an irrigated pecan orchard). With the data from Lucky Hills used for validating Cupid, the shrub site was defined as having a fractional vegetation cover, f_C ~ 0.3 with an LAI ~ 0.5 and a canopy height, h_C ~ 0.5 m (see Table 7.1). The riparian area was assumed to have f_C ~ 0.8 with LAI ~ 3 and h_C ~ 5 m (Kustas and Norman 2000).

One case considered 50% of a pixel comprising of the stressed shrub while the other half had unstressed riparian vegetation. The composite value of $T_R(\phi)$ T_{COM}, for the mixed pixel was estimated as

$$T_{COM} \approx \left[f_{RIP} T_{RIP}^4 + f_{SS} T_{SS}^4\right]^{1/4} \tag{7.10}$$

where T_{RIP} is the radiometric temperature of the riparian surface with corresponding fractional area of the pixel f_{RJP} and T_{SS} is the radiometric temperature of the stressed shrub vegetation with corresponding fractional area of the pixel f_{SS} such that $f_{SS} + f_{RIP} = 1$. The LAI for the composite scene, LAI_{COM}, was calculated by first computing the average fractional cover, which is simply $f_{C,COM} = f_{RIP} * f_{C,RIP} + f_{SS} * f_{C,SS}$. With $f_{C,COM} \approx 0.55$, LAI_{COM} was estimated from equation (7.17) using β = 0.5; this yields $LAI_{COM} \approx 1.6$. The value of h_C for the composite or mixed pixel, $h_{C,COM}$, was computed the same way as $f_{C,COM}$, by simply weighting h_C of the shrub and riparian vegetation, which yielded $h_{C,COM} \approx 2.8$ m. With $LAI_{COM} \approx 1.6$ and $h_{C,COM} \approx 2.8$ m the equations of Raupach (1994) yielded $z_{OM} \approx 0.3$ and $d_O \approx 1.9$.

The 1ANGLE_PT version was run using the meteorological data listed in Table 7.3 and the value of T_{COM} computed from equation (7.10) with u = 1

Table 7.4 1ANGLE_PT output of the surface fluxes for a mixed pixel of 50% stressed shrub (SS) and 50% riparian vegetation (RIP) under meteorological conditions at 1030 MST on DOY 209 (S = 882 W m^{-2}, T_A = 28.4ºC, and RH = 33%) with u = 1 and 5 m s^{-1}

Aggregation procedure	R_N ($W\,m^{-2}$)	G ($W\,m^{-2}$)	H ($W\,m^{-2}$)	LE ($W\,m^{-2}$)	T_{COM} (°C)
$FLUX_{AVG}$ from equation (7.11), $u = 1\,m\,s^{-1}$	543	140	162	242	—
$FLUX_{COM}$ using 1ANGLE_PT, $u = 1\,m\,s^{-1}$	540	225	245	70	43.4
$FLUX_{AVG}$ from equation (7.11), $u = 5\,m\,s^{-1}$	566	152	166	248	—
$FLUX_{COM}$ using 1ANGLE_PT, $u = 5\,m\,s^{-1}$	587	99	154	334	38.95

and 5 m s^{-1}. The output of the composited fluxes predicted by the STS model is compared to the "true" flux average, which is computed simply as

$$FLUX_{AVG} = f_{RIP}FLUX_{RIP} + f_{SS}FLUX_{SS} \tag{7.11}$$

where FLUX represents the energy balance components R_N, G, H, and LE. The results are listed in Table 7.4. For the light wind case where u = 1 m s^{-1}, there is significant differences between $FLUX_{AVG}$ and $FLUX_{COM}$, especially between the turbulent fluxes H and LE where differences >100W m^{-2}. However, for the high wind case where u = 5 m s^{-1}, the differences are significantly reduced, especially in H. This suggests that although subpixel variability can cause a significant departure of the composite fluxes estimated with T_{COM} compared to the "true" average values, the discrepancies between $FLUX_{AVG}$ and $FLUX_{COM}$ can be significantly reduced by higher wind speeds (i.e. $u \gtrsim 5$ m s^{-1}). Wind speeds significantly greater than 1 m s^{-1} were commonly observed during Monsoon 90 where hourly averaged winds around 1030 MST were between 3 and 5 m s^{-1} suggesting that subpixel variability for this region may not cause large discrepancies between $FLUX_{AVG}$ and $FLUX_{COM}$.

A more likely scenario using larger pixel data such as from GOES or NOAA-AVHRR is for only ~10% of the pixel being comprised of riparian vegetation. With f_{RIP} = 0.1 and f_{SS} = 0.9, this yields as input an $LAI_{COM} \approx 0.8$, a $h_{C,COM} \approx$ 1m and a T_{COM} derived from equation (7.10) listed in Table 7.5. Comparison of $FLUX_{AVG}$ and $FLUX_{COM}$ values (see Table 7.5) indicates that differences of less than 50 W m^{-2} are observed for both wind speed conditions. For H and LE less than a 50 Wm^{-2} discrepancy is within the typical uncertainty in micro-meteorological measurements and model predictions. For this case, wind speed did not alter the results.

Table 7.5 1ANGLE_PT output of the surface fluxes for a mixed pixel of 90% stressed shrub (SS) and 10% riparian vegetation (RIP) under meteorological conditions at 1030 MST on DOY 209 (S = 882 W m^{-2}, T_A = 28.4ºC and RH = 33%) with u = 1 and 45 m s^{-1}

Aggregation procedure	R_N ($W\,m^{-2}$)	G ($W\,m^{-2}$)	H ($W\,m^{-2}$)	LE ($W\,m^{-2}$)	T_{COM} (°C)
$FLUX_{AVG}$ from equation (7.11), $u = 1\,m\,s^{-1}$	477	129	280	67	—
$FLUX_{COM}$ using 1ANGLE_PT, $u = 1\,m\,s^{-1}$	484	116	294	74	53.1
$FLUX_{AVG}$ from equation (7.11), $u = 5\,m\,s^{-1}$	516	148	309	60	—
$FLUX_{COM}$ using 1ANGLE_PT, $u = 5\,m\,s^{-1}$	524	133	343	48	46.7

This preliminary analysis of the error introduced by subpixel variability on composite flux estimates suggests that under a relatively extreme case with f_{SS} = 0.5 (50% of the surface containing stressed shrub vegetation with low cover and canopy height) and f_{RIP} = 0.5 (50% having unstressed riparian vegetation with high cover and canopy height) the difference between $FLUX_{AVG}$ and $FLUX_{COM}$ can be significant. Yet, even under this extreme case, a higher wind speed condition can reduce the differences between $FLUX_{AVG}$ and $FLUX_{COM}$, although the discrepancies may still be relatively large. The same effect of higher wind speeds was not evident for the case where f_{SS} = 0.9 and f_{RIP} = 0.1. However, in this case the differences between $FLUX_{AVG}$ and $FLUX_{COM}$ were relatively small. A broader range of mixed-pixel conditions analyzed by Kustas and Norman (2000) suggests that, in general, errors on the order of 50 W m^{-2} or smaller are expected in mixed-pixel cases having less than 20% of an extreme surface condition, such as a riparian wetland. However, it will be difficult to define conditions a priori where the subpixel heterogeneity may cause a significant error in flux calculations using T_{COM}.

7.8 Evaluating the effect of surface heterogeneity on STS model surface flux predictions for an actual $T_R(\phi)$ image over a semiarid rangeland watershed

7.8.1 Experimental data

The data set for running the STS model over a region was collected during the Monsoon 90 field experiment conducted in the Walnut Gulch Experimental Watershed (31.5°N 110°W) maintained by the Southwest

Watershed Research Center in Tucson, AZ. Details of the experiment, conducted during a 2-week period in the summer rainy season, are given in Kustas and Goodrich (1994). This semiarid rangeland environment supports desert steppe and grassland communities, both of which are contained in the watershed. The vegetation cover is highly variable ranging from 25 to 60%. The elevation in the watershed ranges from about 1300m at the west end and outlet to 1800 m at the northeast corner. The terrain is mildly hilly with ridge to valley heights on the order of 10m for the west and central portion of the watershed, and reaching 15–20m for the eastern half. Typical distances between ridges is on the order of 500 m.

The aircraft data used in this analysis were acquired with the NS001 sensor mounted in a NASA C-130 aircraft and are described by Humes *et al.* (1997). Of the three days, DOY 213, 216, and 221 when images were processed, DOY 221 contained the most significant variability in surface moisture conditions as evidence from the $T_R(\phi)$ observations and surface fluxes (Kustas and Humes 1996; Schmugge *et al.* 1998). The surface soil moisture (0–5 cm) was uniformly dry for DOY 213. By DOY 216, several recent rainfall events resulted in relatively wet conditions in the study area; but by DOY 221, some drying of the surface had occurred resulting in some significant variability in moisture conditions.

Humes *et al.* (1997) evaluated the overall quality of radiometric temperatures derived from the NS001 multispectral scanner data by correcting for atmospheric effects using LOWTRAN-7 and comparing the resulting temperature with simultaneously acquired, ground-based remote sensing measurements over two large target areas. The results illustrated from Humes *et al.* (1997) suggest that the aircraft-based radiometric temperatures generally agree with the ground-based temperature measurements to within approximately 1–2°C. The original 6m pixel resolution of the NS001 image was degraded to 90m by Schmugge *et al.* (1998), commensurate with the resolution of the future ASTER instrument on the EOS Terra satellite.

Meteorological and energy flux (METFLUX) data were collected at eight sites within the watershed. Comparisons between predicted and observed heat fluxes from the METFLUX sites using the STS model with the aircraftderived $T_R()$ data indicate the model provides satisfactory estimates (Kustas and Humes 1996; Schmugge *et al.* 1998). The METFLUX sites covered the main plant communities in the region (Weltz *et al.* 1994). A Landsat-5 TM scene from September 1990 was used together with the vegetation data (Weltz *et al.* 1994) to develop a land use map for the region defining the five major land types (see Table 7.6).

An atmospheric boundary layer sounding was available near the time of the NS001 overpass; this was used to define mixed-layer atmospheric variables to be used with the STS model in a mosaic-type approach for estimating the fluxes (Kustas and Humes 1996). This avoids the need for screen-level

Table 7.6 Surface parameters used in the 1ANGLE_PT version of the STS model for the five land cover classifications with estimated fractional area f_i of the scene determined from Landsat TM data

Land cover type	f_i	h_C (m)	z_{OM} (m)	d_O (m)	l_W (m)
Grass	0.41	0.2	0.01	0.1	0.005
Shrub	0.52	0.5	0.05	0.25	0.01
Riparian	0.03	4.0	0.5	2	0.05
Very sparse	0.04	0.1	0.008	0.05	0.01
Bare	0.001	—	0.005	0.05	—

meteorological data, which can vary significantly over a region whereas mixed-layer wind speed and air temperature are much more uniform.

7.8.2 Parameter estimation

The Landsat-5 TM radiance data were used to compute the normalized difference vegetation index (NDVI) as follows:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{7.12}$$

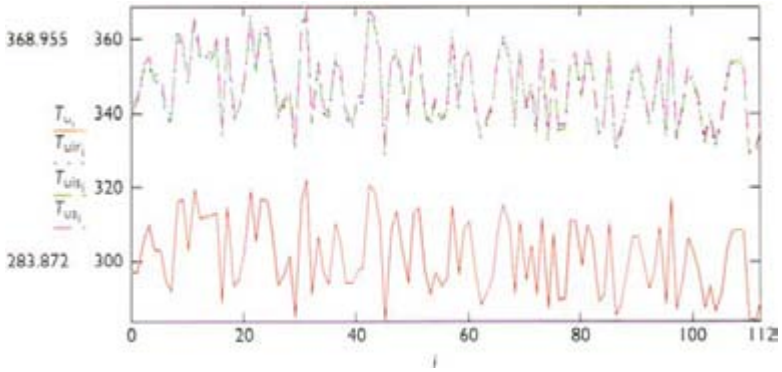
where RED is the reflectance in the red wavelength and NIR is the reflectance in the near-infrared. The fractional cover was estimated using a normalized NDVI quantity, N^*, as described by Gillies and Carlson (1995). This normalized parameter N^* is scaled between values of NDVI for the two limits of bare soil, $f_c = 0$, and full vegetation cover, $f_c = 1$:

$$N^* = \frac{\mathrm{NDVI} - \mathrm{NDVI}_0}{\mathrm{NDVI}_\mathrm{M} - \mathrm{NDVI}_0} \tag{7.13}$$

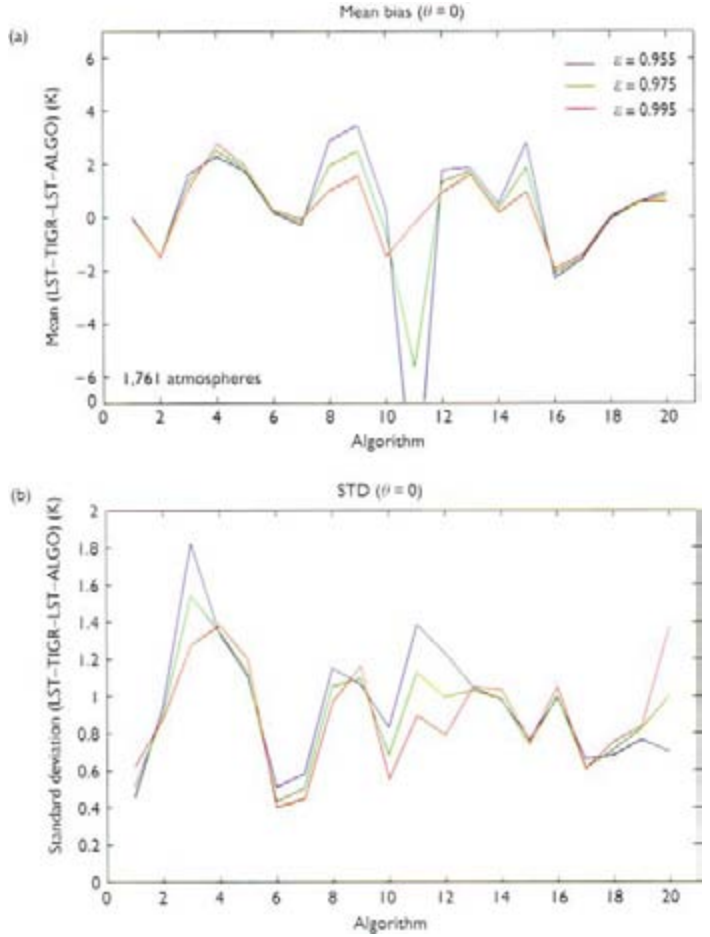
where $\mathrm{NDVI}_0 \sim 0.2$ is typical for bare soil surfaces and $\mathrm{NDVI}_m \sim 0.7$ is typical for the full canopy cover (Carlson and Ripley 1997). However, both upper and lower values are vegetation and soil moisture dependent. Carlson and Ripley (1997) show that there is a simple relationship between N^* and f_c,

$$f_C \approx N^{*2} \tag{7.14}$$

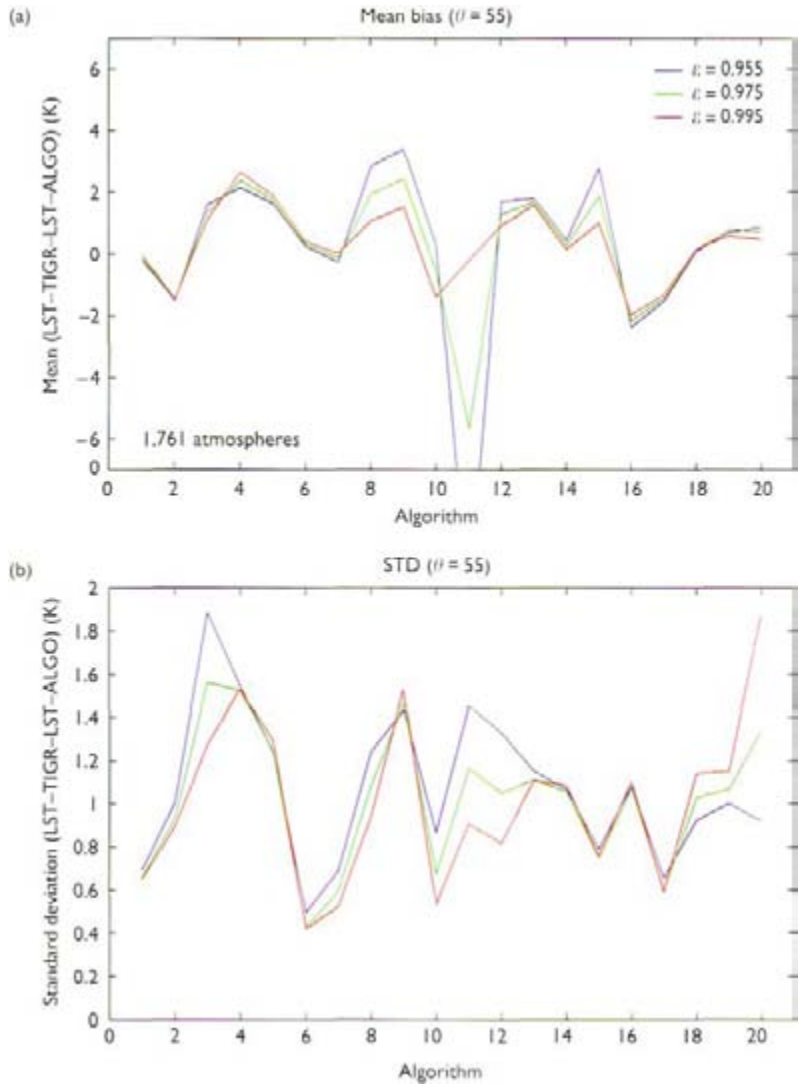
A similar result was obtained by Choudhury *et al.* (1994) except their equation has a more linear form, which is supported by a review of studies (Gutman and Ignatov 1998)



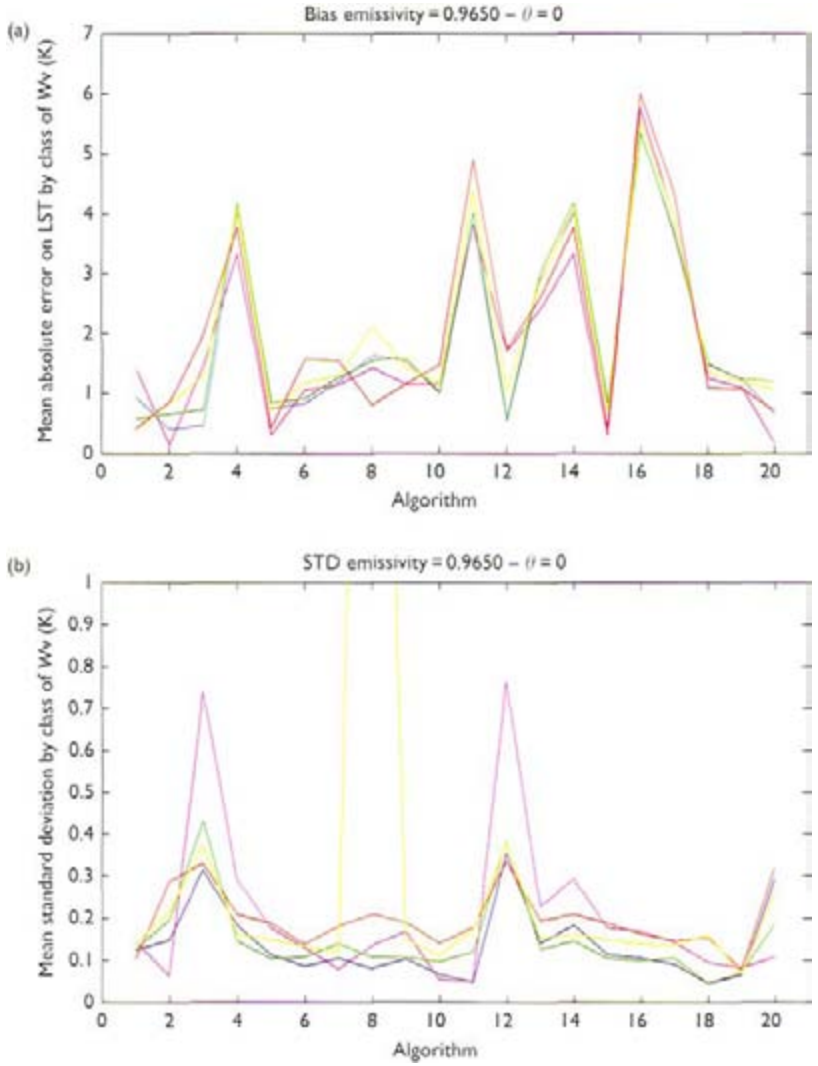
Colour Plate I Different results obtained with emissivities of bare soil (red), or emissivities given in Table 2.1 (dotted lines). The *x*-axis represents the ground data number (arbitrary), the *y*-axis is the temperature in K (see Figure 2.2).



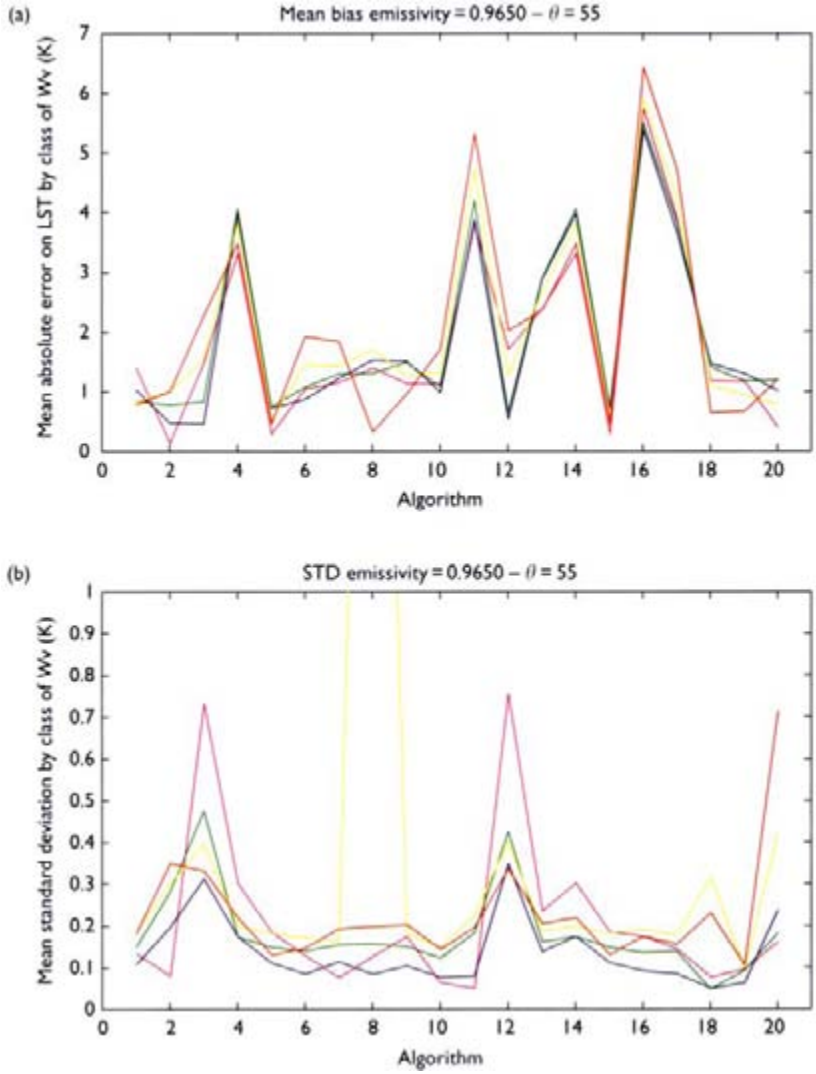
Colour Plate II LST retrievals comparison with TIGR global statistics (see Figure 2.3).



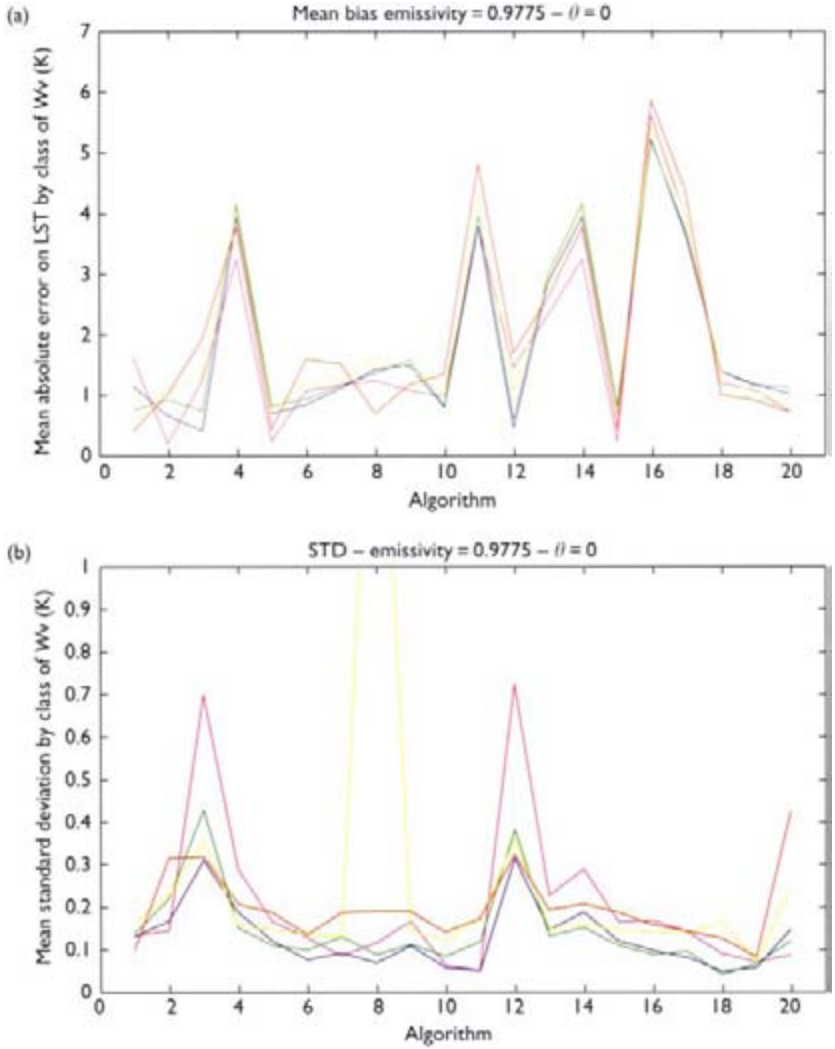
Colour Plate III LST retrievals comparison with TIGR global statistics (see Figure 2.4).



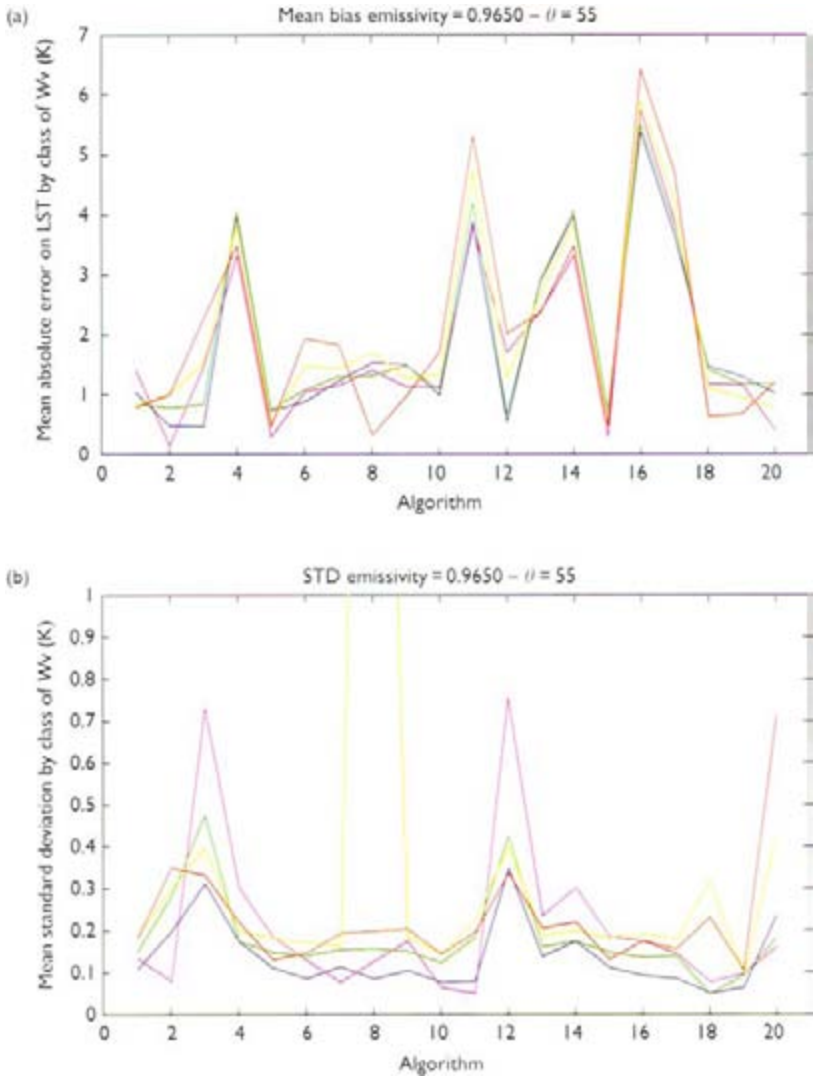
Colour Plate IV For the 20 algorithms, for $\theta = 0$ and an emissivity of 0.9650 absolute errors on LST retrievals (mean values per water vapor category) resulting from an error on water vapor and emissivity (see text) (see Figure 2.6).



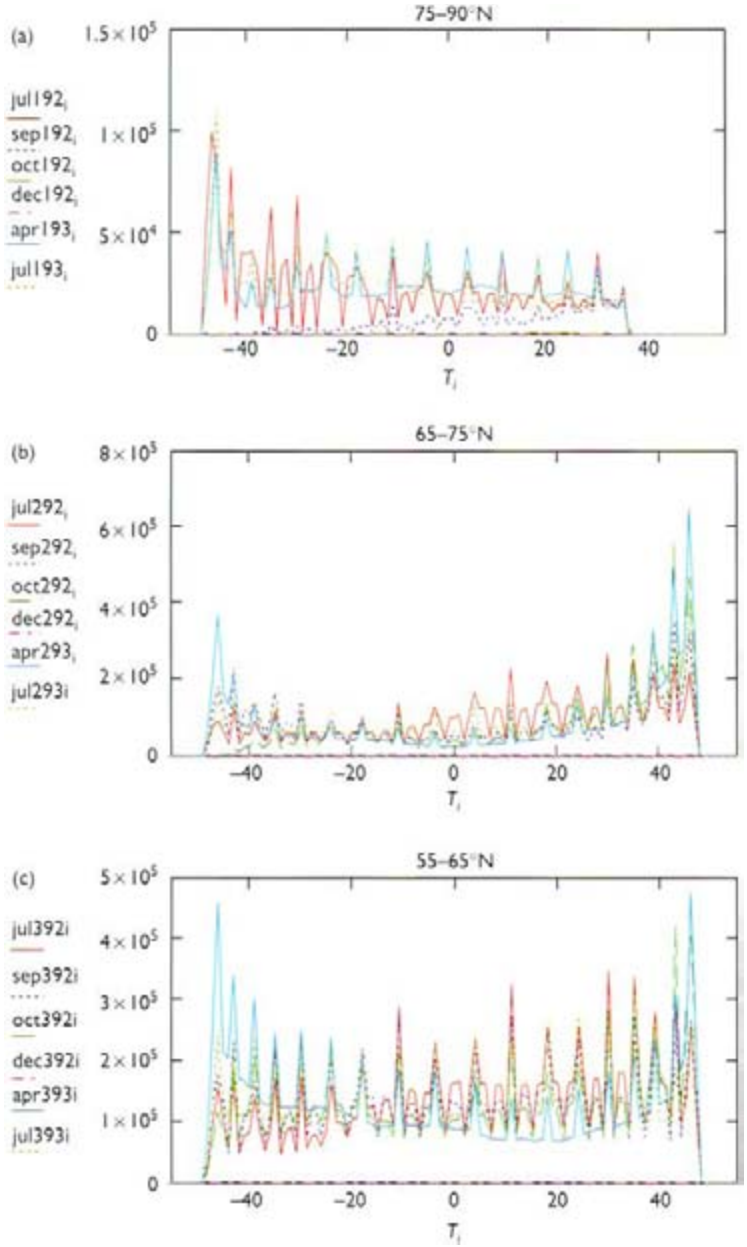
Colour Plate V For the 20 algorithms, for $\theta = 55$ and an emissivity of 0.9650 absolute errors on LST retrievals (mean values per water vapor category) resulting from an error on water vapor and emissivity (see text) (see Figure 2.7).



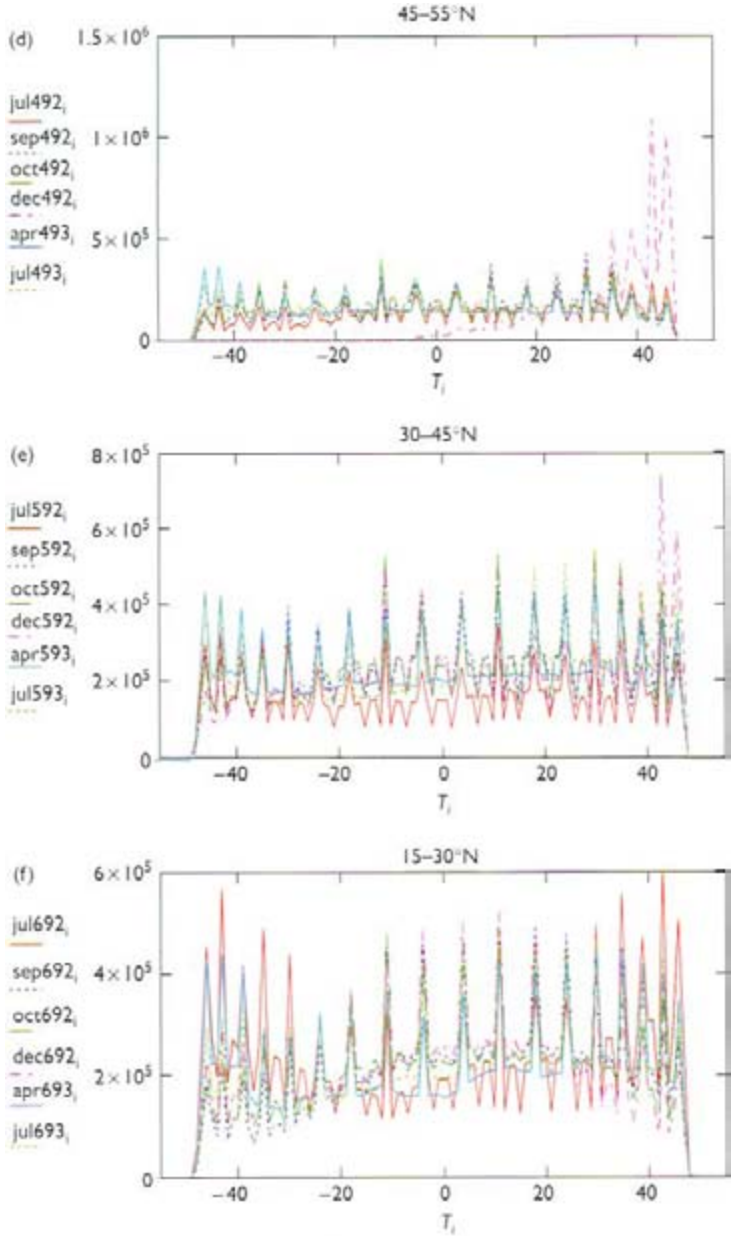
Colour Plate VI For the 20 algorithms, for $\theta = 0$ and an emissivity of 0.9775 absolute errors on LST retrievals (mean values per water vapor category) resulting from an error on water vapor and emissivity (see text) (see Figure 2.8).



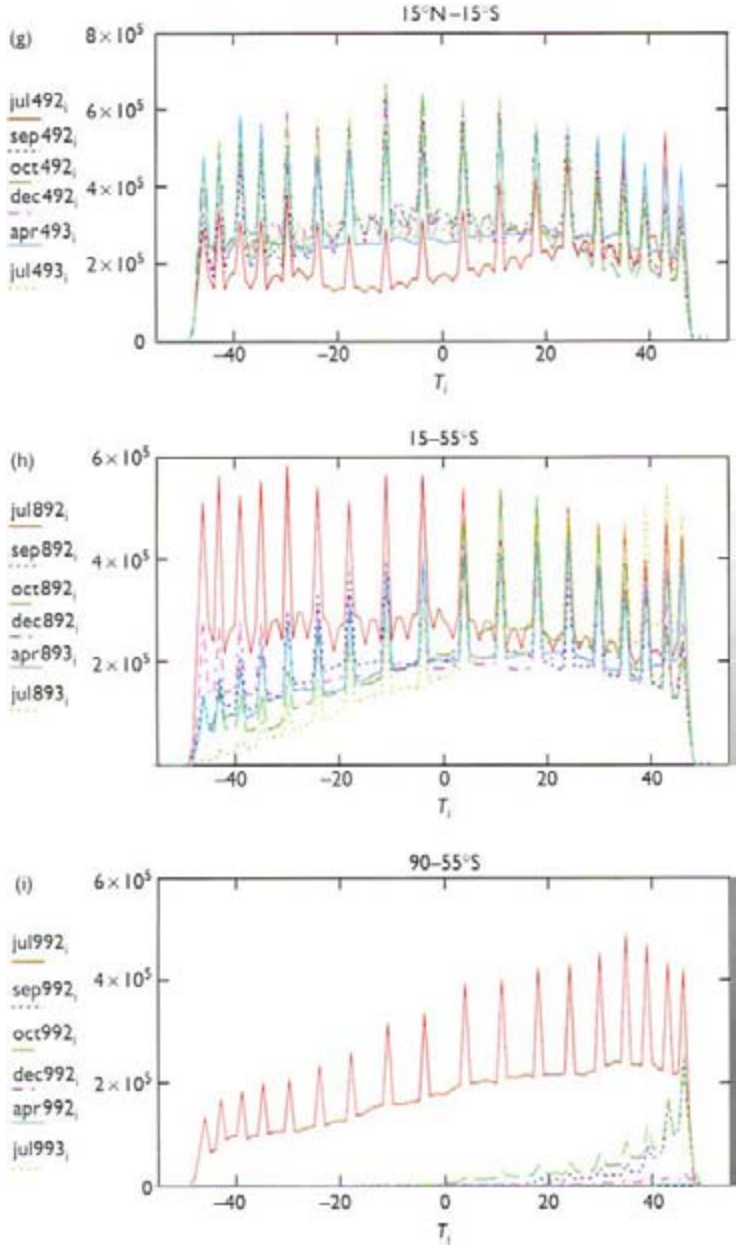
Colour Plate VII For the 20 algorithms, for $\theta = 55$ and an emissivity of 0.9775 absolute errors on LST retrievals (mean values per water vapor category) resulting from an error on water vapor and emissivity (see text) (see Figure 2.9).



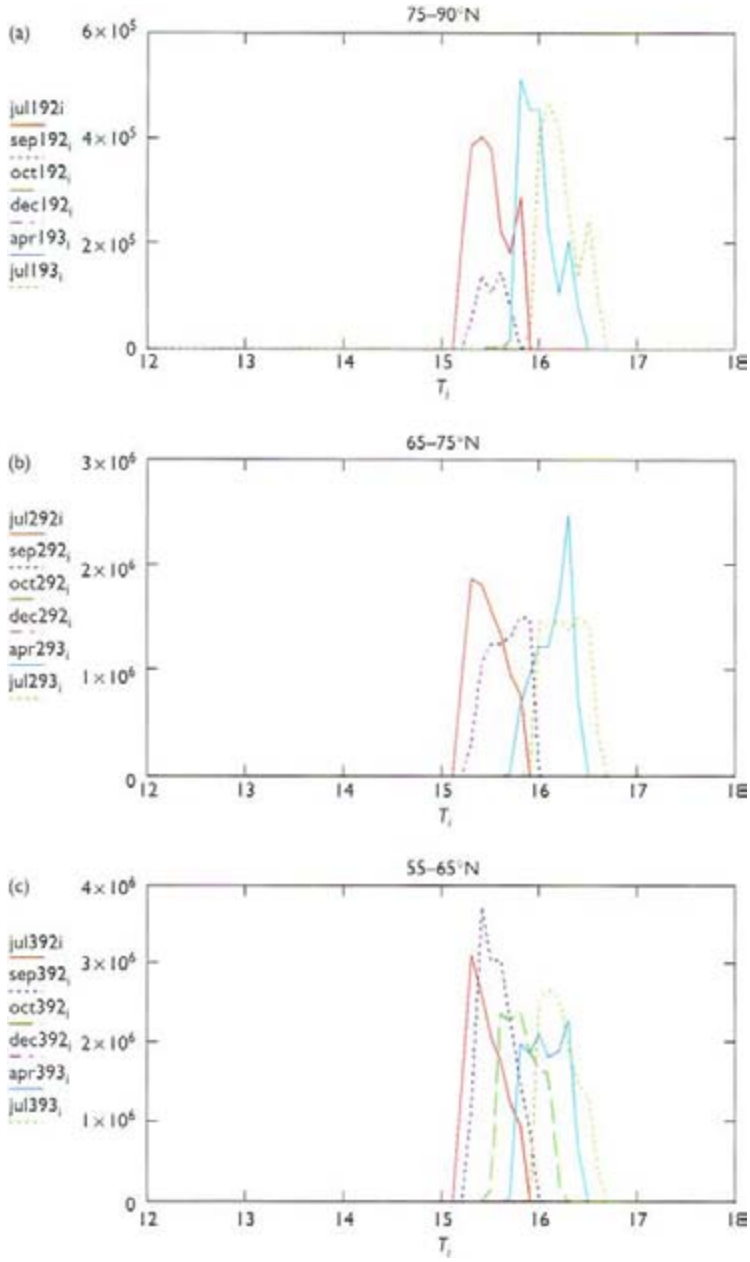
Colour Plate VIII Histograms (10-day composites) of view angles for different decades: (a) latitudinal band 75–90° N; (b) 65–75° N; (c) 55–65° N (see Figure 2.10a-c).



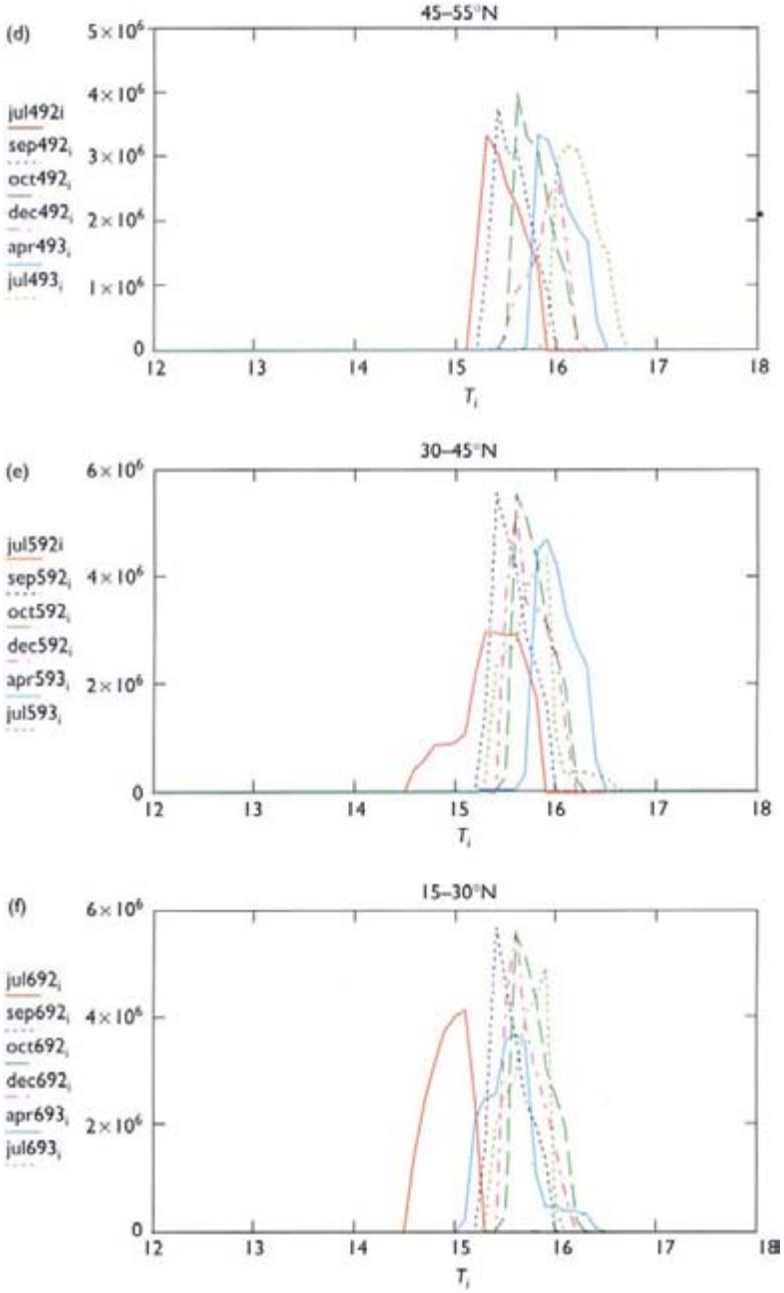
Colour Plate VIII (d) Latitudinal band 45—55° N; (e) 3—45° N; (f) 15—30° N (see Figure 2.10d-f).



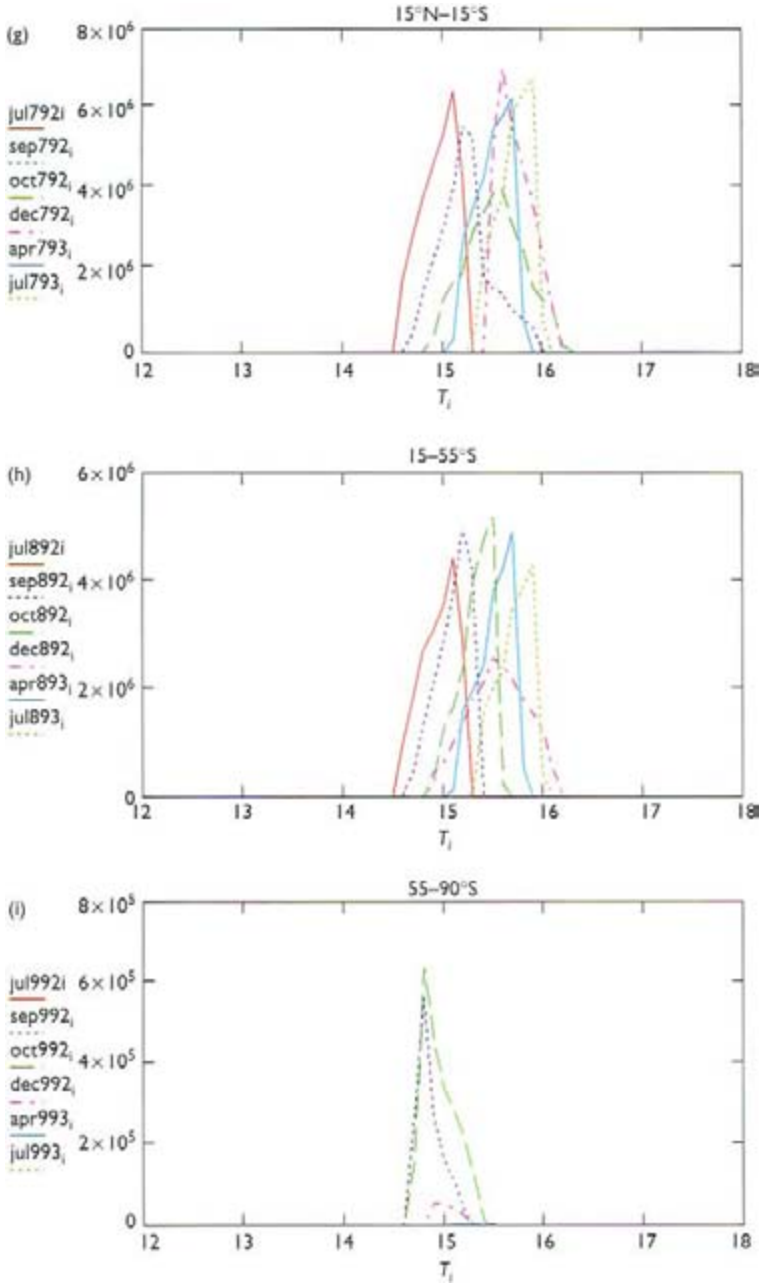
Colour Plate VIII (g) Latitudinal band 15°S–15°N; (h) 55-15°S; (i) 90–55°S (see Figure 2.10g-i).



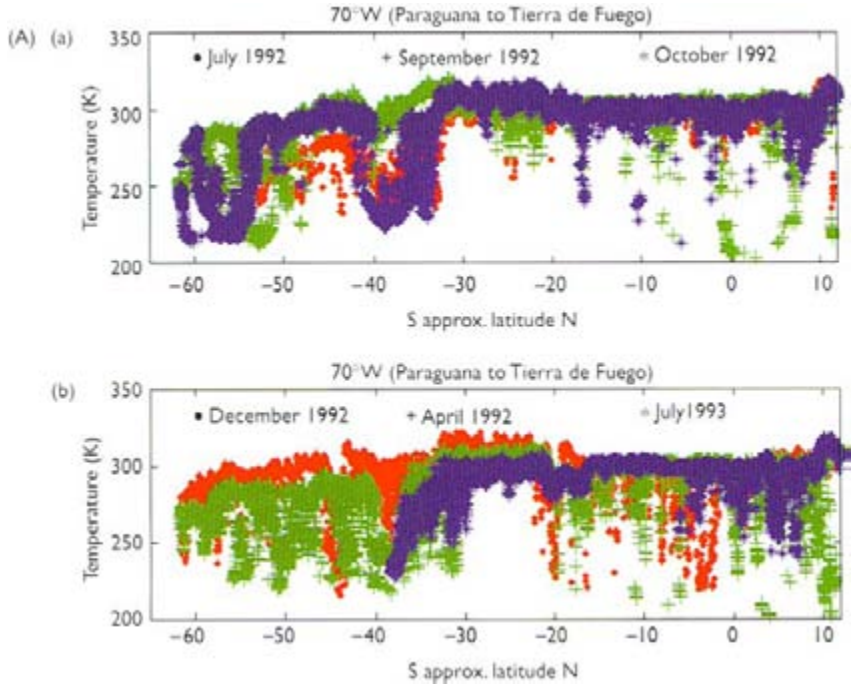
Colour Plate IX Histograms (10-day composites) of the local solar time for different decades. *x*-Axis is Local Solar Time: (a) latitudinal band 75–90° N; (b) 65–75° N;(c) 55–65° N (see Figure 2.11a-c).



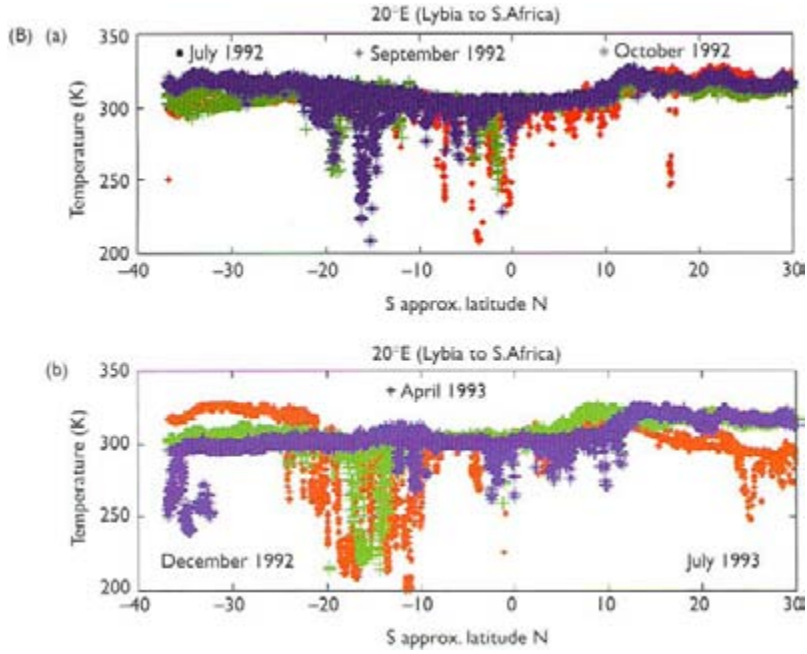
Colour Plate IX (d) Latitudinal band 45–55° N; (e) 3–45° N; (f) 15–30° N (see Figure 2.11d-f).



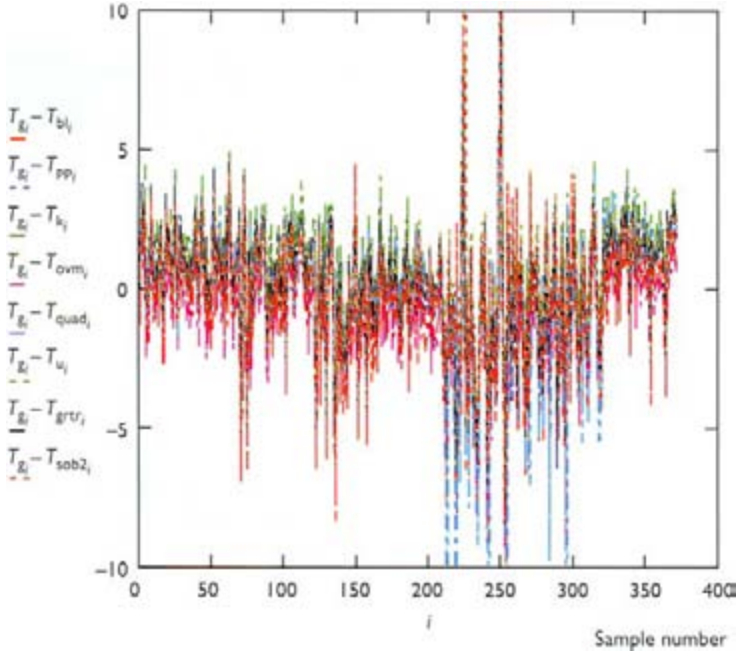
Colour Plate IX (g) Latitudinal band 15° S-15° N; (h) 55—15° S; (i) 90—55° S (see Figure 2.11g-i).



Colour Plate X (A) Transects at longitude 70° W for different decades: (a) July 1992 (red), September 1992 (green), October 1992 (blue); (b) December 1992 (red), April 1993 (green) July 1993 (blue) (see Figure 2.12A).

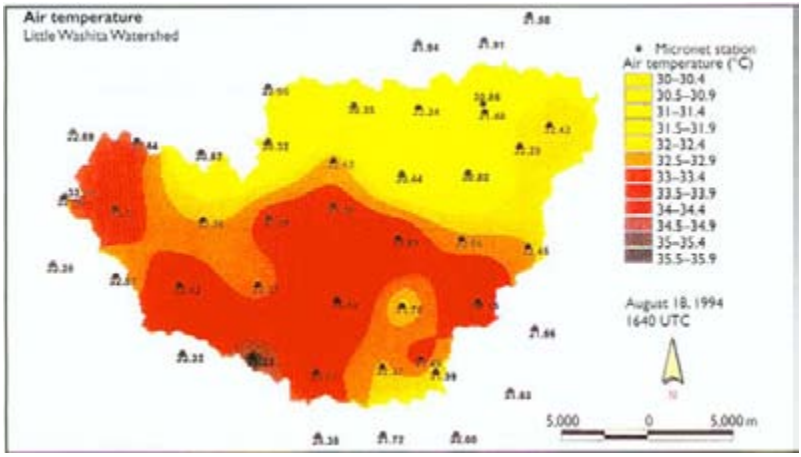


Colour Plate X (B) Transects at longitude 20° E for different decades: (a) July 1992 (red), September 1992 (green), October 1992 (blue); (b) December 1992 (red), April 1993 (green) July 1993 (blue) (see Figure 2.12B).



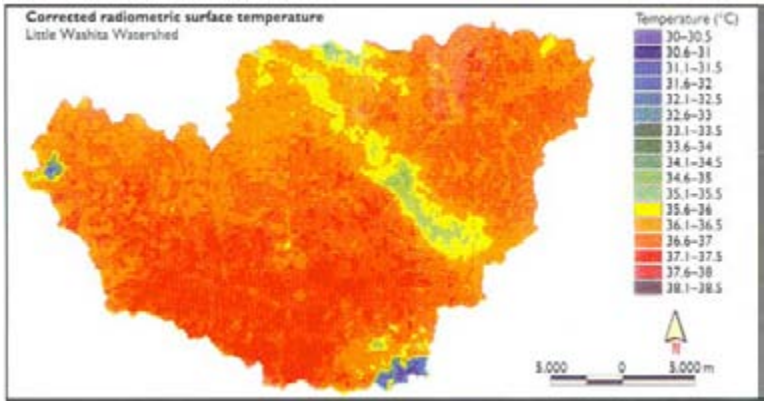
Colour Plate XI Difference between algorithm output and ground measurements for the complete data set used (see Figure 2.13).

Colour Plate XII False color composite image from the Landsat TM sensor for the LWW from August 18, 1994 (see Figure 3.1).

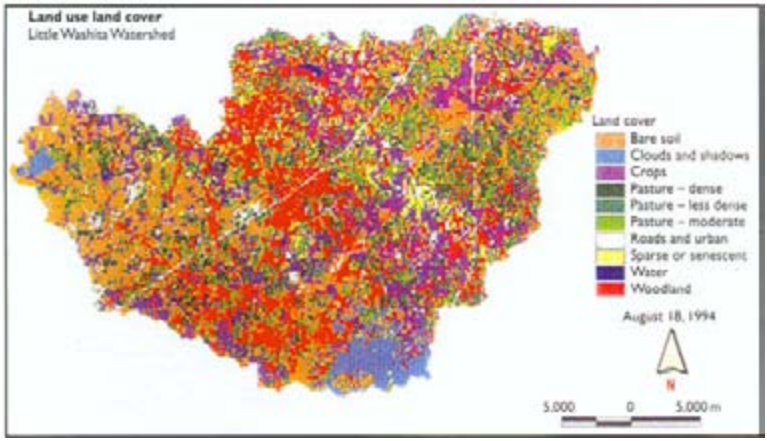


Colour Plate XIII Gridded field of air temperature 2 m above the surface interpolated from measurements at Micronet stations (see Figure 3.3).

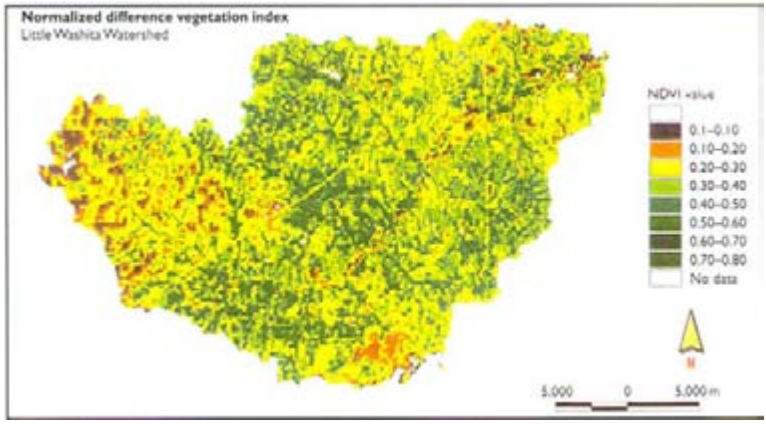
Colour Plate XIV Gridded field of incoming solar radiation measurements interpolated from measurements at Micronet stations (see Figure 3.4).



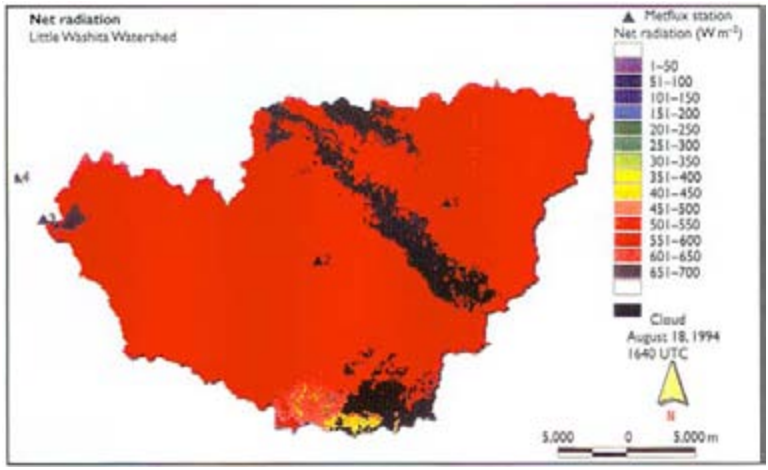
Colour Plate XV Atmospherically corrected radiometric surface temperature derived from a Landsat 5 TM scene acquired over the Little Washita Watershed on August 18, 1994. The dark areas in the east-central portion of the image corresponds to contamination by cirrus clouds, and the dark spots in the far southern and western edges of the watershed correspond to contamination by cumulus clouds (see Figure 3.5).



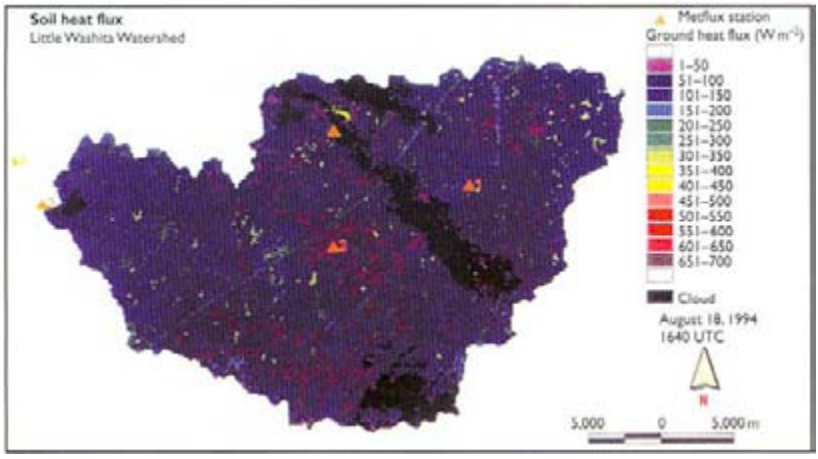
Colour Plate XVI Land cover map derived from the unsupervised classification of data from six of the Landsat TM bands from the August 18, 1994, image (see Figure 3.6).



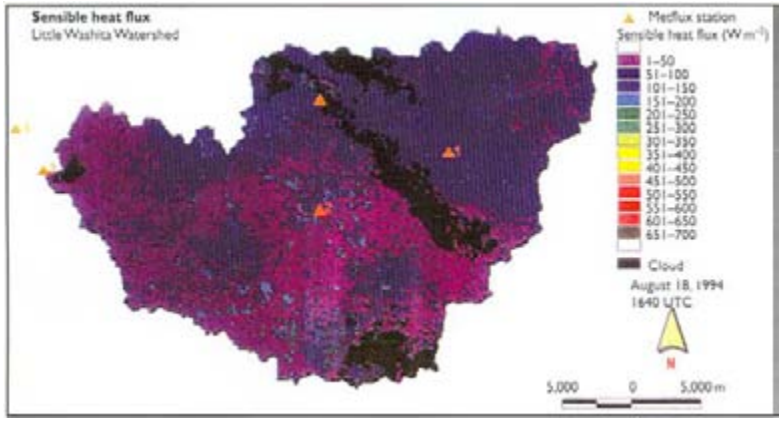
Colour Plate XVII NDVI derived from the August 18, 1994, TM image (see Figure 3.7).



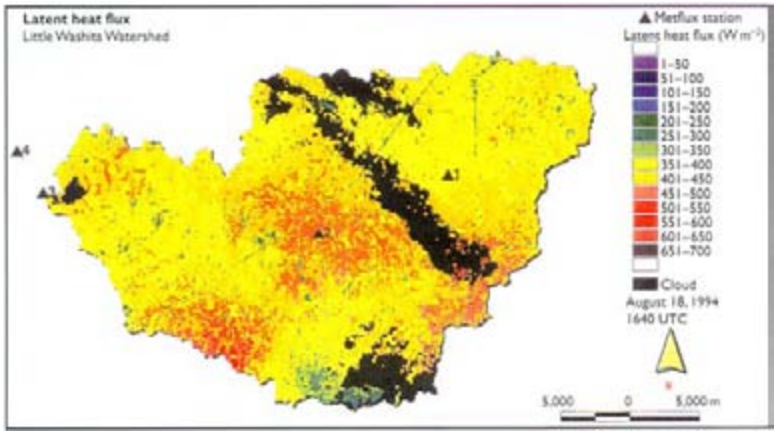
Colour Plate XVIII Map of net radiation (R_{net}) over the watershed computed with the Norman *et al.* (1995) model and Landsat TM data (see Figure 3.8).



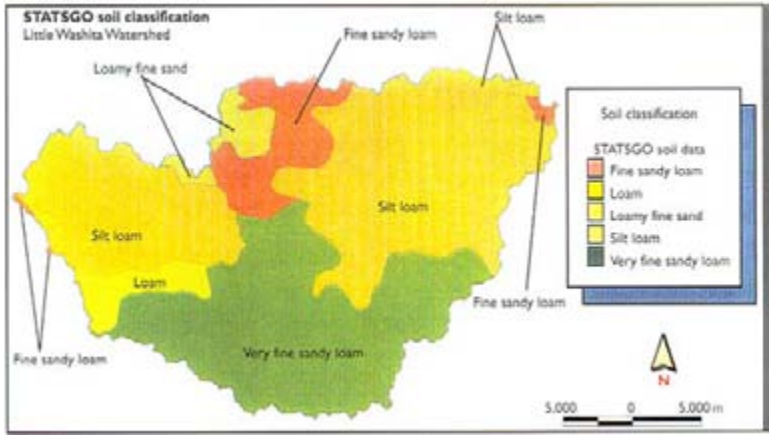
Colour Plate XIX Map of soil heat flux (*G*) over the watershed computed with the Norman *et al.* (1995) model and Landsat TM data (see Figure 3.9).



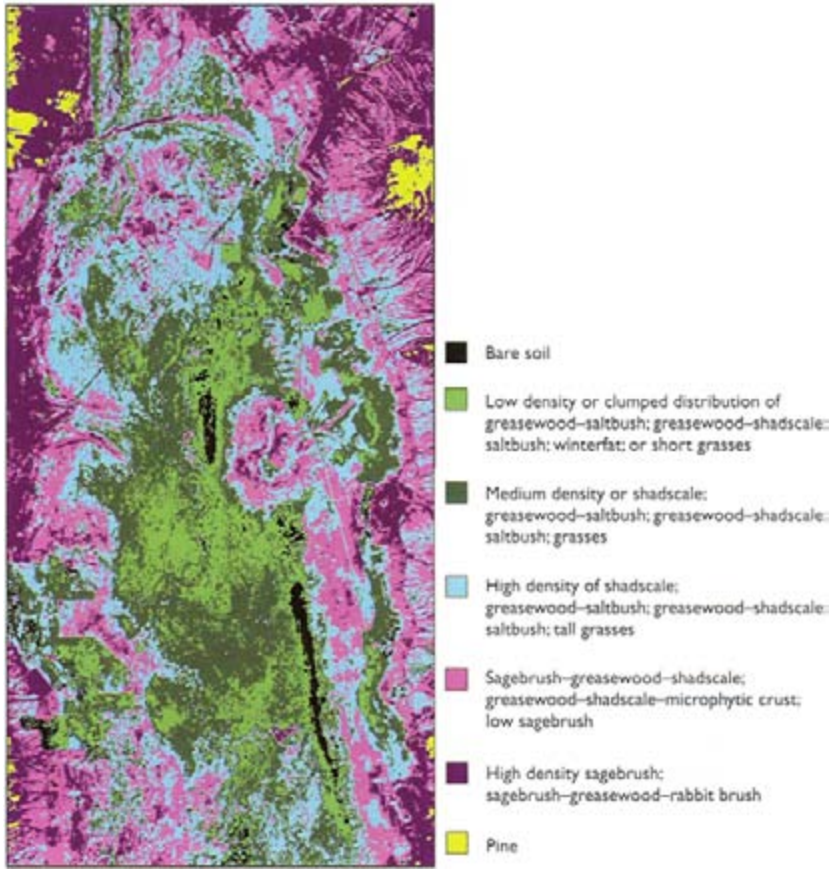
Colour Plate XX Map of sensible heat flux (*H*) over the watershed computed with the Norman *et al.* (1995) model and Landsat TM data (see Figure 3.10).



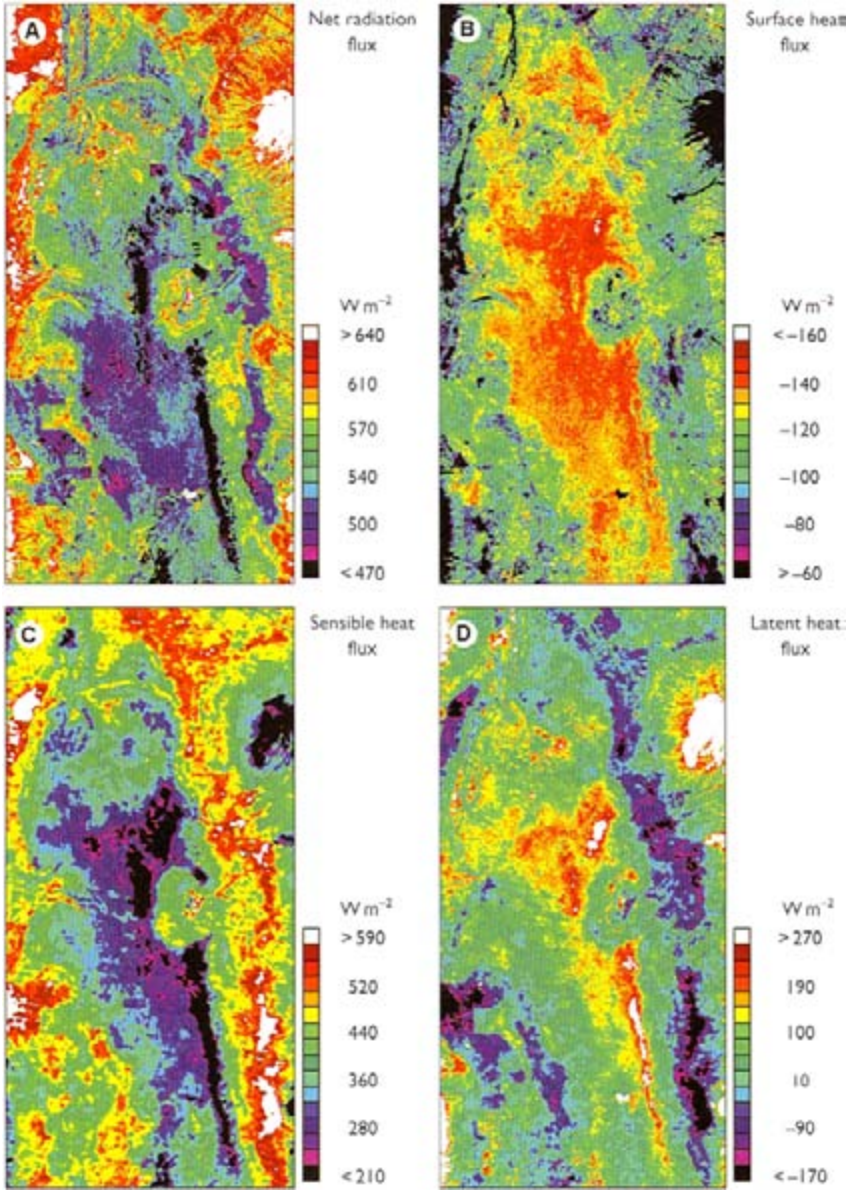
Colour Plate XXI Map of latent heat flux (*LE*) over the watershed computed with the Norman *et al.* (1995) model and Landsat TM data (see Figure 3.11).



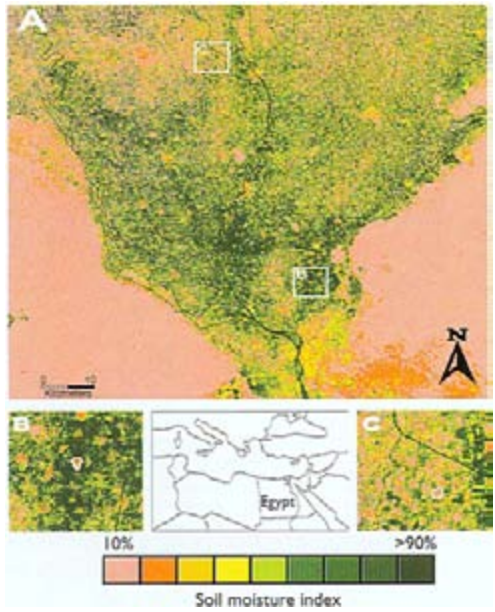
Colour Plate XXII Map of surface soil texture from STATSGO database (see Figure 3.13).



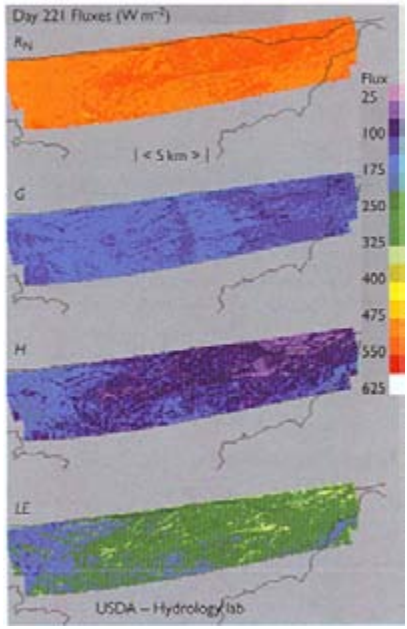
Colour Plate XXIII Classification of land cover types in Goshute valley. Many desert plant species coexist in assemblages. Most vegetation classes reflect changes in assemblage members or differences in plant density due to changes in soil salinity and moisture availability. Similarities in plant structure and large plant spacing relative to image resolution make classification of desert vegetation extremely difficult (see Figure 4.5).



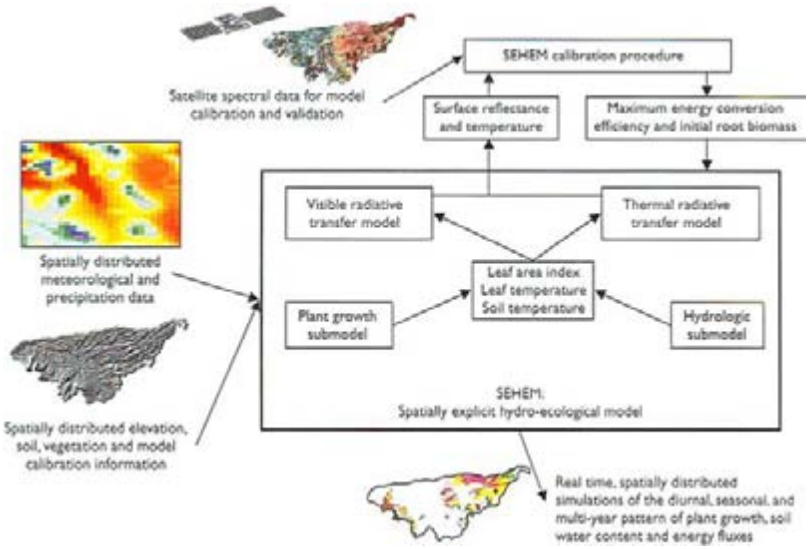
Colour Plate XXIV Maps showing the spatial distribution of instantaneous surface energy fluxes derived from assimilation of surface meteorological and remote sensing data. See text for discussion (see Figure 4.8).



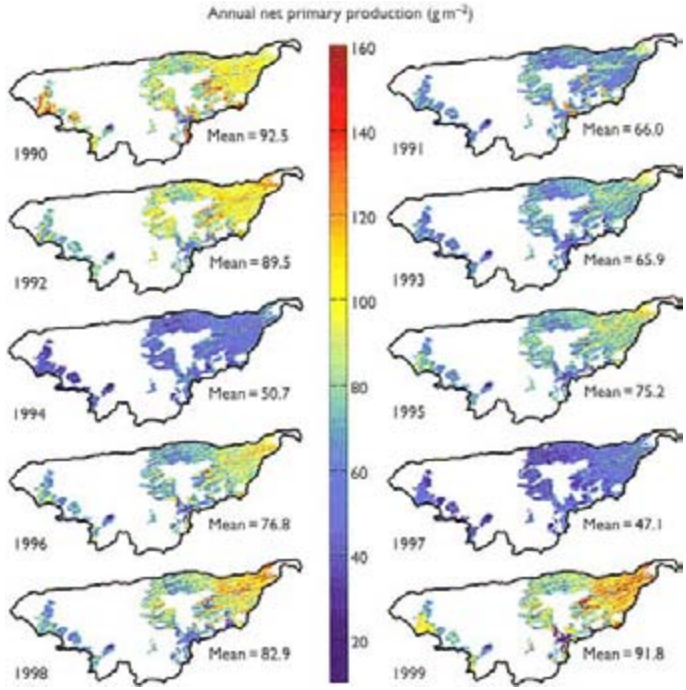
Colour Plate XXV Map and subsets of regional surface soil moisture availability *(M_o)* for the Nile delta study area. Dark colors indicate high soil moisture and light colors indicate low surface soil moisture availability (see Figure 5.4).



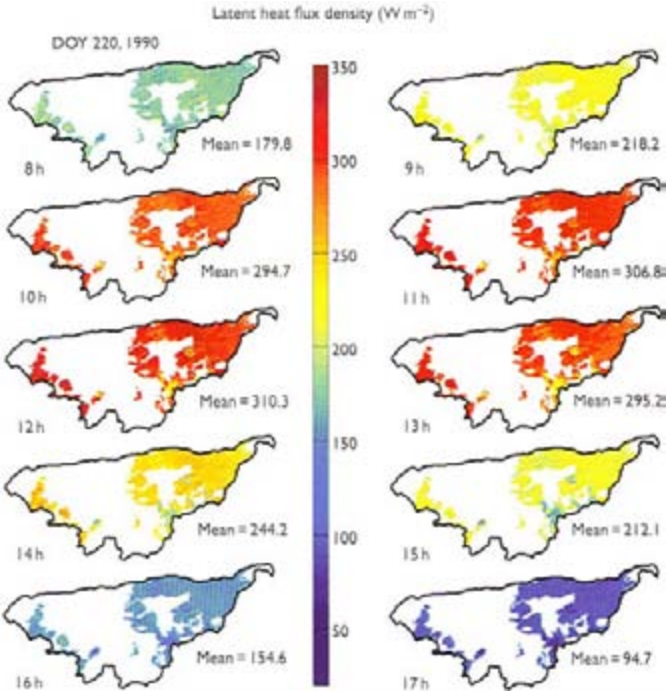
Colour Plate XXVI Output of surface fluxes using the 1ANGLE_PT version of the STS model with radiometric surface temperature imagery at 90 m resolution from Monsoon 90. Boundary of the Walnut Gulch Experimental Watershed is delineated in black (see Figure 7.10).



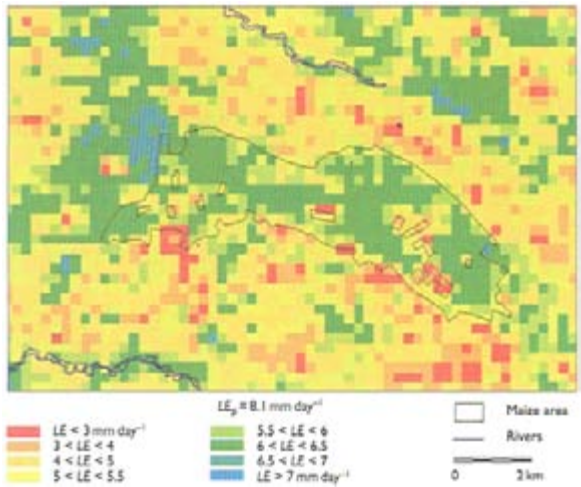
Colour Plate XXVII Diagrammatic representation of the SEHEM showing the inputs of surface and meteorological information, the SEHEM calibration procedure based on satellite spectral data, and the potential outputs of plant growth, soil water content, and energy fluxes. The images represent the USDA ARS Walnut Gulch Experimental Watershed (WGEW) in Southeast Arizona (see Figure 8.5).



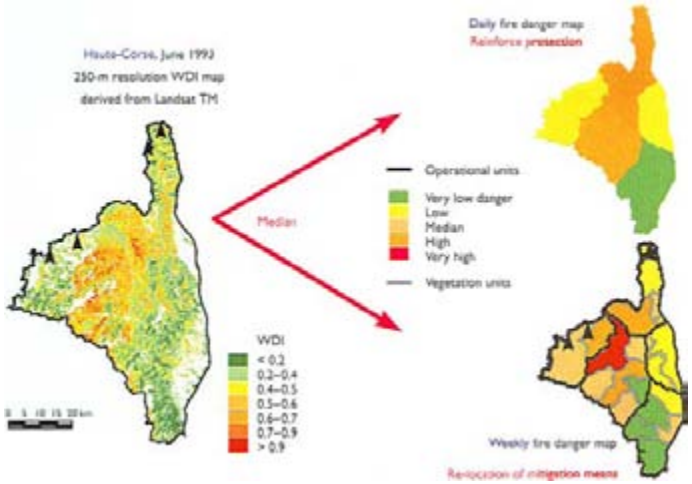
Colour Plate XXVIII SEHEM simulations of annual net primary production over a 10-year period for grassland regions with the Walnut Gulch Experimental Watershed. Output validation was presented by Nouvellon *et al.* (2000a) (see Figure 8.6).



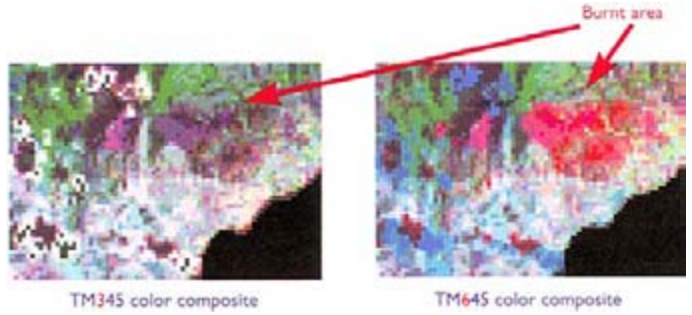
Colour Plate XXIX SEHEM simulations of hourly latent heat flux density for grassland regions within the Walnut Gulch Experimental Watershed. Output validation was presented by Nouvellon *et al.* (2000a) (see Figure 8.7).



Colour Plate XXX Daily evapotranspiration map obtained from MUST-simulated thermal data using equation (1 1.4) (Orthez–France) (see Figure 11.2).



Colour Plate XXXI Daily and weekly fire risk index on the right part are the results of subsampling a full scale risk index obtained from MUST-simulated thermal data (on the left), useful for the establishment of 1/50,000 long-term risk maps (see Figure 11.4).



Colour Plate XXXII Classiflcation of fire damaged areas using different bands of a Landsat TM image. Respectively, red, NIR, SWIR (on the left), and thermal infrared, NIR and SWIR (on the right). The latter provides a higher accuracy (see Figure 11.5).

$$f_C \approx 1 - N_C^{*p} \tag{7.15}$$

with

$$N_C^* = \frac{NDVI_m - NDVI}{NDVI_m - NDVI_0} \tag{7.16}$$

and p ranges from 0.5 to 0.7 for dark and bright soils, respectively. Using a simple radiative transfer model Carlson and Ripley (1997) also show that these relations are applicable for an NDVI image uncorrected for atmospheric attenuation. With the NDVI values computed from the Landsat-5 TM data and a summary of the ground surveys of vegetation cover of METFLUX sites in the watershed (Daughtry *et al* 1991; Weltz *et al.* 1994), $NDVI_o$ = 0.15 and $NDVI_m$ = 0.6 were used in equations (7.13) and (7.16). For relatively homogeneous canopies there is an exponential relationship between f_C and LAI (Choudhury 1987):

$$f_C = 1 - \exp(-\beta LAI) \tag{7.17}$$

where β is a function of the leaf angle distribution (e.g. β = 0.5 for randomly distributed leaves). Choudhury *et al.* (1994) estimated a mean β = 0.67 from β values for 18 crops presented by Ross (1975).

The land cover map from the Landsat-5 TM image together with analyses of Kustas *et al.* (1994) and Menenti and Ritchie (1994) help to define momentum roughness parameters. The vegetation characterization by Weltz *et al.* (1994) provided the vegetation parameters, including leaf width, l_w, and mean vegetation height, h_c, for the STS model (see Table 7.6). The original 30-m Landsat-5 TM data were degraded to the 90-m resolution of the $T_R(\phi)$ data.

The flux output of the model using 90-m pixel resolution image is shown in Figure 7.10. The areas with lower sensible heat fluxes and higher latent heat fluxes are in the regions, which received more recent precipitation. The sensible heat fluxes varied from ~ – 175 to 255 W m^{-2} and the latent heat fluxes ranged from ~85 to 670 W m^{-2}. The area-average and standard deviation of the four flux component values over the whole image, namely $FLUX_{AVG}$ and $FLUX_{STD}$, are listed in Table 7.7. The remotely sensed NDVI and $T_R(\phi)$ were also aggregated up to the whole image producing a single value of NDVI and $T_R()$ for the image (i.e. $NDVI_{COM}$ and T_{COM}). These values were used in the STS model for computing area-average $FLUX_{COM}$ values. The model parameters for the whole image were specified by simply weighting by the fractional area, f_i, comprised the five land cover types. The composite roughness parameters ($d_{O,COM}$ and $z_{OM,COM}$) for the whole image were derived by using the following weighting scheme (Shuttleworth *et al.* 1997):

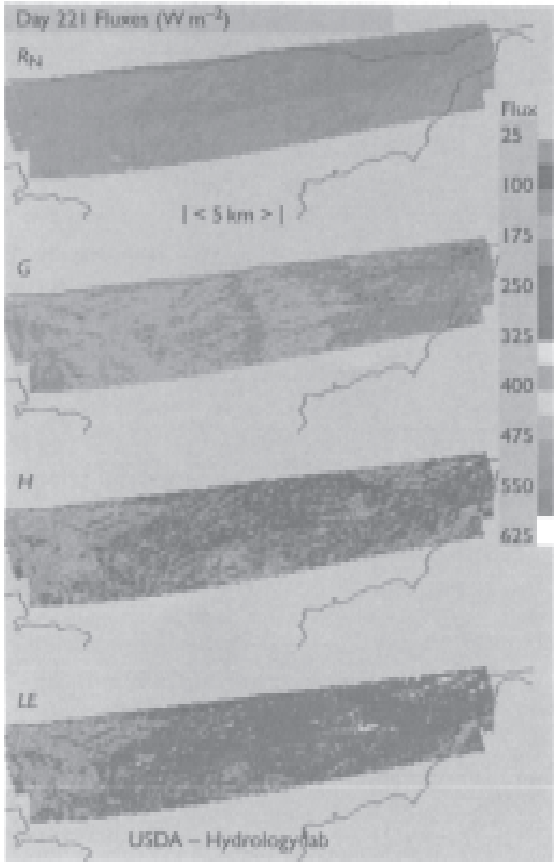


Figure 7.10 Output of surface fluxes using the 1ANGLE_PT version of the STS model with radiometric surface temperature imagery at 90 m resolution from Monsoon 90. Boundary of the Walnut Gulch Experimental Watershed is delineated in black (see Colour Plate XXVI).

$$d_{\mathrm{O,COM}} = \sum_i f_i d_{\mathrm{O},i}$$

$$\ln^{-2}\left(\frac{z_{\mathrm{B}} - d_{\mathrm{O,COM}}}{z_{\mathrm{OM,COM}}}\right) = \sum_i f_i \ln^{-2}\left(\frac{z_{\mathrm{B}} - d_{\mathrm{O}}, i}{z_{\mathrm{OM},i}}\right) \tag{7.18}$$

where $d_{\mathrm{O,i}}$ and $z_{\mathrm{OM,i}}$ are for the different land cover types in Table 7.6, and z_{B} is the "blending height" (Wieringa 1986) where meteorological variables

Table 7.7 $FLUX_{AVG}$ values and the corresponding standard deviations ($FLUX_{STD}$) derived from $T_R(\phi)$ image at 90m resolution from DOY 221 (see Figure 7.10). In addition, the values of $FLUX_{COM}$ computed with T_{COM} averaged for the whole image (see text)

Aggregation procedure	R_N ($W\,m^{-2}$)	G ($W\,m^{-2}$)	H ($W\,m^{-2}$)	LE ($W\,m^{-2}$)
$FLUX_{AVG}$	525	151	116	257
$FLUX_{STD}$	±12	±12	±38	±54
$FLUX_{COM}$	526	142	125	259

can be treated as uniform over the grid or pixel. Sounding data suggested that $z_B \sim 20$ m but the model is relatively insensitive to a range of z_B values from 20 to 150 m (Kustas and Humes 1996). The resulting model parameters defined for the whole image were $h_C = 0.4$ m, LAI = 0.65, $d_{O,COM} = 0.2$ m, and $z_{OM,COM} = 0.03$ m.

The comparison of $FLUX_{AVG}$ and $FLUX_{COM}$ values in Table 7.7 suggest that the use of very coarse resolution data for this site had minimal effect on the model-derived area-averaged fluxes. This result may have been due, in part, to the fact that there was a relatively small spatial variation of $T_R(\phi)$, probably due, in part, to very few pixels classified as riparian (Table 7.6). Although there were a few pixels with $T_R(\phi) \approx 26$°C and others reaching ≈ 40°C, the average $T_R(\phi)$ was ≈ 33°C with a standard deviation of only ± 2°C.

7.9 Summary and conclusions

A detailed PE model (Cupid) was first validated with experimental data from a semiarid rangeland environment and then used to investigate the aerodynamic-radiometric temperature relationship for a wide range of environmental conditions and fractional canopy covers. The results indicate that the aerodynamic-radiometric relationship cannot be described adequately using single-source approaches that rely on using a roughness length for heat, because the relationship depends strongly on the soil-canopy temperature differences. A simple STS model was evaluated using output from the Cupid model. Performance of the STS model, which used a single radiometric temperature (1ANGLE_PT version), in predicting energy flux partitioning between soil and vegetation under a wide range of conditions was satisfactory; although under certain conditions differences with Cupid were significant. Discrepancies in the flux predictions were significantly larger when the STS model was modified to make use of two radiometric temperatures at nadir and 55° from nadir (2ANGLE_PT and 2ANGLE versions) for deriving canopy and soil temperatures. The differences were in H and LE were on the order of 100 W m^{-2}, which is clearly unsatisfactory.

The 1ANGLE_PT version of the STS model was used to evaluate errors caused by subpixel variability in surface properties. An extreme case was evaluated by assuming that 50% of the area was stressed shrub vegetation under low cover and 50% unstressed riparian vegetation. Differences in the aggregated heat fluxes from these two surfaces versus the composite fluxes computed using an effective surface temperature were significant (i.e. >100 W m^{-2}) at the low wind speed case of 1 m s^{-1}. However, the results were improved when the model was applied under the higher wind speed case of 5 m s^{-1}, where differences were on the order of 50 W m^{-2}. A less extreme case was also tested where the riparian zone comprised only 10% of the pixel. In this case the discrepancies were significantly smaller, namely less than 35 W m^{-2} for both wind speed conditions. These preliminary findings are supported by a more thorough analysis of errors caused by mixed-pixel variations where errors on the order of 50 W m^{-2} or smaller were obtained for mixed-pixel cases having less than 20% of an extreme surface condition (Kustas and Norman 2000).

An example of energy flux maps generated by the STS model with $T_R()$ images over a heterogeneous semiarid landscape was presented. In this particular case the resolution of the remotely sensed data was degraded from 90 m to ~10 km or the whole image. Although the area covered by the image had a wide range of computed heat fluxes, the difference in average fluxes for the entire image was minor, indicating that degree of variability in surface temperature and surface conditions was not significant enough to cause problems in using very coarse resolution data.

A review of the literature indicates that the application of radiometric temperature for surface flux estimation has yielded inconsistent results (Kustas and Norman 1996). Hall *et al.* (1992) concluded that schemes employing radiometric temperature will yield unreliable heat flux predictions due to significant discrepancies between aerodynamic and radiometric temperatures. Results from this chapter indicate that approaches such as the one used in Hall *et al.* (1992), which do not consider factors causing differences between $T_R(\phi)$ and $T_{0,}$ support their result. However, it is shown here that the STS model adequately addresses most of these factors and provides a practical solution that can be used operationally with satellite data.

With satellite observations, atmospheric effects and emissivity variations are likely to cause a 1–2K error in $T_R(\phi)$. Compounding this error is the spatial variability of atmospheric forcing variables, particularly air temperature, T_A. This uncertainty in screen level air temperature increases the errors in $T_R(\phi)$ - T_A over a satellite image. Using a ±3K error in $T_R(\phi)$ - T_A, Kustas and Norman (1997) find that this can lead to errors on the order of 50% in H and LE using the STS model. The TSTI/ALEXI scheme is essentially insensitive to errors in $T_R(\phi)$ since it uses a timedifferencing in $T_R(\phi)$ coupled to atmospheric boundary-layer dynamics. Thus, T_A is actually simulated by the model so observations are not required. The 2ANGLE_PT version of the STS scheme is also essentially insensitive to errors in $T_R(\phi)$ - T_A; however, it

can only be applied under non-stressed vegetation conditions. A relatively simple time differencing scheme using regional weather station data (referred to as the Dual-Temperature-Difference, DTD, approach) and employing the STS modeling framework was recently developed and tested by Norman *et al.* (2000). It is shown to be considerably less sensitive to errors in $T_R(\phi)$ and T_A, and can compute regional fluxes using satellite data that are comparable to ALEXI.

Although the GOES is capable of providing multiple $T_R(\phi)$ observations used by ALEXI and the DTD approach, the pixels are relatively course (~5 × 5 km). At this scale, results from this chapter indicate subpixel variability under certain landscape conditions may cause significant error in heat flux predictions using pixel-average input data. Higher-resolution $T_R(\phi)$ data from the EOS Terra platform (i.e. ASTER) could be used for investigating the effects of subpixel variability on heat flux predictions at the coarser resolutions with aggregation techniques described in this chapter.

In summary, it appears that reliable heat flux predictions at regional scales with satellite $T_R(\phi)$ observations are possible with models that (a) consider the main factors affecting the non-unique relationship between $T_R(\phi)$ and T_0, and (b) are relatively insensitive to errors in $T_R(\phi)$ - T_A. Methods for addressing the impact of subpixel variability on pixel-average heat flux predictions should be an area of active research.

Appendix A: radiometric temperature calculation in Cupid

The calculation of brightness temperature and ensemble directional radiometric temperature requires first calculating thermal radiance. Technically, thermal radiance should be based on integrations of the radiometer response over the blackbody spectrum between the wavelengths of sensitivity. Fortunately, using absolute temperature to a power n (typically $n = 4$ is adequate) usually provides a close approximation to the blackbody integral (Becker and Li 1990) and is simpler to do. If a power other than four is used, then s must be replaced with a constant that is appropriately adjusted. The equation that describes the thermal radiance of a vegetative canopy is analogous to equation (7.15) in Norman *et al.* (1985), which describes the shortwave radiance for a canopy. The following equations use the same symbolism and definitions as Norman *et al.* (1985), so the reader is referred to that paper for more details. The radiance $[R_c(\theta_v\phi_{,v})]$ from a canopy in a particular viewing direction $(\theta_v\phi_{,v})$ can be approximated as follows if each layer in the canopy is of equal incremental leaf area index:

$$R_c(\theta_v,\phi_v) = \sum_{j=1}^{N} \int_{\alpha}\int_{\beta} \{C_j\varepsilon\sigma T_{j,\mathrm{sun}}^{n}(\alpha,\beta) + (1-C_j)\varepsilon\sigma T_{j,\mathrm{shade}}^{n} + E_j(\alpha)\}\,|f_v|\,W_j\,\mathrm{d}\beta\,\mathrm{d}\alpha \tag{7.A1}$$

where

$g(\alpha, ß)$	fraction of leaf area inclined at angle a and oriented at azimuth angle $ß$
C_j	fraction of leaf area in layer j that is sunlit
ε	leaf emissivity
σ	Stefan-Boltzman constant - 5.67×10^{-8} W m^{-2} K^{-4}
$T_{j,\text{sun}}$ $(\alpha, ß)$	thermodynamic temperature of a sunlit leaf at inclination angle α and azimuth angle $ß$
$T_{j,\text{shade}}$	thermodynamic temperature of shade leaf
$E_j(\alpha)$	reflected diffuse irradiance from leaf inclined at angle a and in layer j
$\|f_v\|$	fraction of leaf area projected into the direction of view, which depends on the inclination (α) and azimuth $(ß)$ angles of the leaf and view zenith (θ_v) and azimuth (ϕ_v) angles—see equation (9) in Norman *et al.* (1985)
W_j	factor that weights the contribution of various layers to the view from a particular direction—see Norman *et al* (1985)

The contribution of the soil layer is given by the following:

$$R_g(\theta_v, \phi_v) = (\varepsilon_{s\sigma} T_s^4 + D_1 \rho_s) W_0 \tag{7.A2}$$

where ε_s is the soil thermal emissivity in the appropriate wavelength band, ρ_s is the soil thermal reflectivity ($\varepsilon_s = 1 - \rho_s$), D_1 is the downward thermal flux density just above the soil, W_0 is the fraction of the radiometer view occupied by the soil, and

$$W_0 = 1 - \sum_{j=1}^{N} W_j \tag{7.A3}$$

Profiles of thermal radiant flux density and leaf temperature distributions are obtained from the simultaneous solution of radiative, convective, and conductive equations in Cupid (Norman and Campbell 1983).

If ε_s and ε are set to unity to represent a blackbody ($R_B(\theta_v, \phi_v)$) and downward thermal diffuse flux density from the sky is zero ($D_N = 0$) for the calculation of $E_j(\alpha)$, then

$$\varepsilon(\theta_v, \phi_v) = \frac{R_c(\theta_v, \phi_v)}{R_B(\theta_v, \phi_v)} \tag{7.A4}$$

The radiometric temperature of the soil/canopy system is given by

$$T_R(\theta_v, \phi_v) = \left[\frac{R_B(\theta_v, \phi_v)}{\sigma}\right]^{1/n} \tag{7.A5}$$

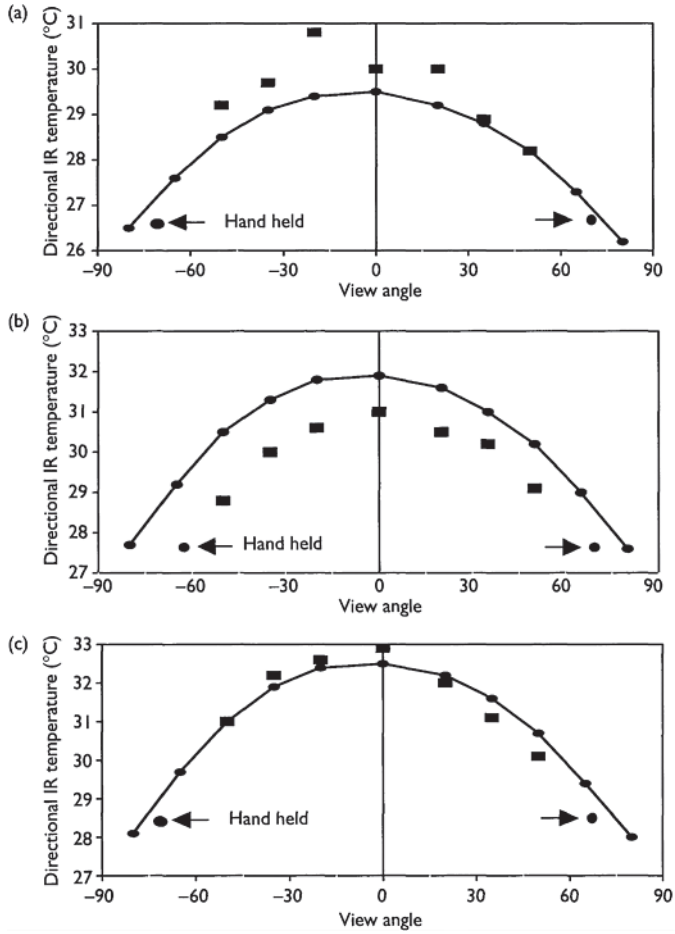


Figure 7.A1 Comparison of Cupid brightness temperature estimates (line with dots) with measurements (squares) from FIFE by B.L Blad (University of Nebraska, Lincoln, NE) using the Barnes Modular Multiband Radiometer at Site 11 on DOY 156 in 1987 at several times: (a) 1045, (b) 1145, (c) 1215 CST. Hand-held infrared thermometer measurements were also made at the same time and are shown in the figures. The LAI was approximately 1.5.

and the brightness temperature, or apparent infrared temperature senseddirectly by an accurate infrared thermometer, is given by

$$T_{\mathrm{brt}}(\theta_{\mathrm{v}}, \phi_{\mathrm{v}}) = \left[\frac{R_{\mathrm{c}}(\theta_{\mathrm{v}}, \phi_{\mathrm{v}})}{\sigma}\right]^{1/n} \tag{7.A6}$$

Although the directional emissivity depends on view angle, the dependence is slight for canopies of even moderate cover (LAI greater than 0.5) so that the directional and hemispherical emissivities are not significantly different. The ensemble directional radiometric temperature can depend strongly on view angle because of temperature differences between the soil and vegetation. Comparison of Cupid brightness temperature estimates with measurements from FIFE (Figure 7. Al) indicate that the formulations used in Cupid provide reliable directional temperature values.

Appendix B: overview of the N95 model

With the use of a single emissivity to represent the combined soil and vegetation, the ensemble directional radiometric temperature, $T_R(\phi)$, is related to the fraction of the radiometer view occupied by soil versus vegetation expressed as

$$T_R(\phi) \approx [f(\phi)T_C^n + (1 - f(\phi))T_S^n]^{1/n} \tag{7.B1}$$

where T_C and T_S are the thermodynamic temperatures of the vegetation canopy and soil surface, respectively, and are assumed to represent spatially weighted averages of the sunlit and shaded portions of the canopy and soil, respectively, and $n \sim 4$ (Becker and Li 1990). The fraction of the field of view of the infrared radiometer occupied by canopy, $f(\phi)$, depends upon the view zenith angle, ϕ, canopy type, and fraction of vegetative cover, f_C. For many vegetated surfaces, assuming a random canopy with a spherical leaf angle distribution is reasonable so that,

$$f(\phi) = 1 - \exp\left(\frac{-0.5\text{LAI}}{\cos\phi}\right) \tag{7.B2}$$

The use of $T_R(\phi)$ in a convective heat flux equation frequently involves the controversial assumption that $T_R(\phi)$ is equivalent to the so-called "aerodynamic temperature," T_0, of the surface. T_0 is the temperature that satisfies the bulk transport expression having the form

$$H = \rho C_p \frac{(T_0 - T_A)}{R_{AH}} \tag{7.B3}$$

where H is the sensible heat flux (W m^{-2}), ρC_p is the volumetric heat capacity of air (J m^{-3} K^{-1}), TA is the air temperature at some reference height above the surface (K), and RAH is the aerodynamic resistance to heat transport (s m^{-1}), which has the following form in the surface layer (Brutsaert 1982):

$$R_{AH} = \frac{[\ln((z_U - d_O)/z_{OM}) - \Psi_M][\ln((z_T - d_O)/z_{OH}) - \Psi_H]}{k^2 u} \quad (7.B4)$$

In this equation, d_O is the displacement height (m), u is the wind speed (m s^{-1}) measured at height z_U (m), k is von Karman's constant (≈ 0.4), z_T is the height (m) of the T_A measurement, ψ_M and ψ_H are the Monin-Obukhov stability functions for momentum and heat, respectively, and are functions of $(z - d_O)/L$ (see Brutsaert 1982) where $L = -/[k(g/T_A)(H_v/\rho C_p)]$ is the Monin-Obukhov length (m), u_* is the friction velocity (m s^{-1}), g is the acceleration of gravity (m s^{-2}), $H_v = (H + 0.61 T_A C_p E)$ is the virtual sensible heat flux (W m^{-2}), and E is the rate of surface evaporation (kg m^{-2} s^{-1}). The roughness parameter z_{OM} is the local roughness length (m) for momentum transport and z_{OH} is the local roughness length (m) for heat transport. T_0 cannot be measured, so it is often replaced with an observation of $T_R(\phi)$ in equation (7.B3).

The net energy balance of the soil-canopy system is given by (neglecting photosynthesis)

$$R_N = H + LE + G \quad (7.B5)$$

The system of equations for computing fluxes from the soil and canopy components, denoted by subscripts S and C, respectively, are listed below and will be used by all versions of the model. The energy budgets for the soil and vegetation are given by

$$R_{N,S} = H_S + LE_S + G \quad (7.B6)$$

$$R_N = H + LE + G \quad (7.B7)$$

with $R_N = R_{N,S} + R_{N,C}$. Similar to equation (7.B1) for estimating the contribution of soil and canopy temperatures to the observed radiometric temperature, equations (7.B8) and (7.B9) are used for partitioning net radiation, R_N, between the soil and vegetation in order to properly weight the contributions of sensible, H, and latent heat flux, LE, from the soil and vegetation, and estimate the soil heat flux, G:

$$R_{N,S} = R_N \exp(-\kappa \mathrm{LAI}/\sqrt{2\cos(\theta_S)}) \quad (7.B8)$$

$$R_{N,C} = R_N[1 - \exp(-\kappa \mathrm{LAI}/\sqrt{2\cos(\theta_S)})] \quad (7.B9)$$

Equations (7.B8) and (7.B9) are modifications from the original N95 formulations (where $\sqrt{2\cos(\theta_S)} = 1$ and $k = 0.45$) proposed by Anderson *et al.* (1997) based on simulations with a detailed soil-plant-atmosphere model, Cupid (Norman and Campbell 1983), where the net radiation divergence is found to be a function of the solar zenith angle θ_S. The value of $K = 0.6$ is used so that at low solar zenith angles (i.e. $\theta_S < 30°$) the quantity $K/$

$\sqrt{2\cos(\theta_S)}$ will have a value of 0.45 which is midway between its likely limits of 0.3 to 0.6 (Ross 1981).

Equations (7.B8) and (7.B9) are known to have large biases for low vegetation cover when soil surfaces are much hotter than vegetation. Thus, the STS model now uses equations (7.5a) and (7.5b) for attenuation of thermal radiation and the model of Goudriaan (1977) as described in Campbell and Norman (1997) for shortwave fluxes.

For computing G, the original formulation from N95 was simply

$$G = c_G R_{N,S} \tag{7.B10}$$

where the value of $c_G = 0.35$ (Choudhury *et al.* 1987). However, c_G is constant only for several hours around solar noon (Kustas and Daughtry 1990). Friedl (1996) included the effects of a temporally varying c_G by multiplying equation (7.B10) by $\cos(\theta_S)$. Another approach developed by Kustas *et al.* (1998) is based on time differences with the local solar noon quantified by the following non-dimensional time parameter, t_N:

$$t_N = \frac{|t_i - t_{SN}|}{t_{SN}} \tag{7.B11}$$

where t_i is the time nominally ≈5 h of the local time of solar noon, t_{SN}, and the | | represents the absolute value of the difference. Using experimental data to compute $G/R_{N,S}$ or c_G as a function of time t_i, an empirical function was fit to the relationship between $G/R_{N,S}$ and t_N. The results indicated that a constant $G/R_{N,S}$ could be used for $t_N < 0.3$ (i.e. several hours around solar noon) and linear least-squares regression equation between $G/R_{N,S}$ and t_N was needed for $t_N > 0.3$. Neither equation (7.B11) nor the approach suggested by Friedl (1996), however, considers the fact that G and R_N are not in phase, and hence the temporal change in the ratio $G/R_{N,S}$ will not be the same between the morning and afternoon. In addition, these approaches also do not account for possible variations in c_G due to soil moisture conditions (e.g. Friedl 1996) or the effect of low wind speeds near the soil surface, u_S (see equation 7.B17). From the Cupid simulations, $G/R_{N,S} > 0.5$ when $u_S < 0.25$; hence, an empirical equation was used to account for this effect (see text in main body of the chapter for details).

With $H = H_S + H_C$ and with the soil and vegetation taken in "parallel" (i.e. the resistance network provides for no interaction between the soil and vegetation), the heat fluxes from the soil and vegetation are computed by

$$H_S = \rho C_p \frac{T_S - T_A}{R_{AH} + R_S} \tag{7.B12}$$

$$H_C = \rho C_p \frac{T_C - T_A}{R_{AH}} \tag{7.B13}$$

With H_C and H_S taken in "series" (i.e. the resistance network allows for interaction between the soil and vegetation) yields:

$$H_S = \rho C_p \frac{T_S - T_{AC}}{R_S} \tag{7.B14}$$

$$H_C = \rho C_p \frac{T_C - T_{AC}}{R_X} \tag{7.B15}$$

where T_{AC} is related to T_0 in equation (7.1), namely,

$$H = \rho C_p \frac{T_{AC} - T_A}{R_A} \tag{7.B16}$$

See Figures 1 and 11 in N95 illustrating the "parallel" and "series" resistance networks.

R_S is the resistance to heat flow in the boundary layer immediately above the soil surface and is estimated from an empirical expression developed by Sauer *et al.* (1995) from extensive studies of this soil-surface resistance in a wind tunnel and beneath a corn canopy. R_X is the total boundarylayer resistance of the complete canopy of leaves (see Appendix A in N95) estimated with the wind speed in the canopy air space computed from the equations of Goudriaan (1977). R_{AH} is estimated via equation (7.B4); d_O and z_{OM} have been estimated simply as a fraction of canopy height, h_C (i.e. $d_O \approx 0.65h_C$; $z_{OM} \approx 0.13h_C$; see Brutsaert 1982). Raupach's (1994) generalized formulations as a function of LAI has recently been adopted since these may provide more accurate estimates over a wider range of cover conditions. For most vegetation, z_{OH} can be estimated as a fraction of z_{OM} as postulated by Garratt and Hicks (1973), namely $z_{OH} \approx z_{OM}/7$ or $kB^{-1} \approx 2$. R_A is computed from equation (7.B4) with $z_{OH} = z_{OM}$ (cf. equation 7.2). T_{AC} is the momentum aerodynamic temperature and only approximates the temperature in the canopy air space (see Appendix A in N95).

Although soil surface resistances depend on many factors, a reasonable, simplified equation has been developed where

$$R_S = \frac{1}{a + bu_S} \tag{7.B17}$$

In equation (7.B17), a 0.004 m $\approx^{-1}$ is the free convective velocity "constant", $b \approx$ 0.012, and u_S is the wind speed at a height above the soil surface where the effect of the soil surface roughness is minimal; typically 0.05–0.2 m. The coefficients in equation (7.B17) depend on turbulent length scale in the canopy, soil surface roughness and turbulence intensity in the canopy and are discussed by Sauer *et al.* (1995). The numerical value for the coefficient a was taken from data presented in Sauer (1993) as the mean intercept of plots of soil

surface transfer coefficients versus wind speed in the canopy. For the smooth aluminum plates used by Sauer (1993) the value of the parameter b was measured to be 0.007. The value of 0.012 for b used in equation (7.B17) was estimated from a combination of wind tunnel data for surfaces of various roughnesses and the field data on smooth plates to represent the more typical roughness that soil surfaces have.

From numerous experimental studies under free convective conditions, the free convective velocity U_{cv} is related to ΔT (Kondo and Ishida 1997),

$$U_{CV} = c(T_S - T_A)^{1/3} \tag{7.B18}$$

where Kondo and Ishida's (1997) experimental results indicate that c (m s^{-1} K$^{-1/3}$) ranges from 0.0011 for an aerodynamically smooth surface to 0.0038 for an aerodynamically rough surface. The laboratory and field data from Sauer (1993) indicate that U_{CV} or a (cf. equation (7.B17)), 0.004, is applicable to soil surfaces where $\Delta T \approx 3$–5 K. Substituting ΔT = 4 K and U_{CV} = 0.004 into equation (7.B17) yields $c \approx 0.0025$. This value of c may be more appropriate for soil surfaces associated with cultivated crops. Since the model derives both soil and canopy temperatures, as a first approximation, T_C was substituted for T_A in equation (7.B18). Therefore, equation (7.B17) was modified to the following form (cf. equation 7.3):

$$R_S = \frac{1}{c(T_S - T_C)^{1/3} + bu_S} \tag{7.B19}$$

with c = 0.0025 and b = 0.012. The effect on model output using c = 0.0038 in equation (7.5) will be discussed. Sauer and Norman (1995) indicate that for rough soil surfaces or higher turbulent intensities in the canopy air space, the magnitude of the coefficient b in equation (7.B17) for predicting R_S would be larger reaching an upper limit for $b \approx 0.024$. There is no direct way to estimate the change in b from routine observations. However, under the extreme temperature differences between soil and canopy that typically exist in open canopies under dry soil conditions where $T_S - T_C \approx 20$°C (see Figure 7.4), it is the free convective velocity U_{CV} that generally dominates the heat transfer from the soil surface.

Finally, for $LE = LE_S + LE_C$ the fluxes are estimated by the following expressions:

$$LE_S = R_{N,S} - G - H_S \tag{7.B20}$$

$$LE_C = \alpha_{PT} f_G \frac{\Delta}{\Delta + \gamma} R_{N,C} \tag{7.B21}$$

The Priestley-Taylor parameter, α_{PT}, is set equal to 1.26 (Priestley and Taylor 1972) for the green part of the canopy, Δ is the slope of the saturation vapor pressure-temperature curve at T_C (PaK^{-1}), and γ is the psychrometric constant

($\approx$66Pa K^{-1}). The fraction of LAI that is "green" or actively transpiring, f_G, may be obtained from knowledge of the phenology of the vegetation. If no information is available for estimating f_G, then it is assumed to equal unity.

Equation (7.B21) only provides an initial calculation of LE_C, and it can be overridden if the temperature difference between the soil-canopy system and the atmosphere is large causing erroneous flux estimates, such as negative LE_S or condensation during the daytime period. If the estimated radiometric temperature from equation (7.B1) is less than the measured $T_R()$, then the Priestley-Taylor approximation in equation (7.B21) will tend to overestimate the canopy transpiration rate because the water supply in the root zone is inadequate. Therefore, in N95 an iteration procedure was devised such that LE_C values were computed below estimates given by equation (7.B21) until values of T_C and T_S used in equation (7.B1) agree with the measured $T_R(N)$.

Recent experience with the STS model has revealed that a more general procedure for evaluating α_{PT} begins by assuming $\alpha_{PT} = 2$ instead of the conventional value of 1.26. This method is described in more detail in the main body of this chapter.

References

Anderson, M.C., J.M.Norman, G.R.Diak, W.P.Kustas, and J.R.Mecikalski (1997) A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote Sens. Environ.* 60:195–216.

Ball, J.T., I.E.Woodrow, and J.A.Berry (1987) A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. In J.Biggins (ed.) *Progress in Photosynthesis Research,* Vol. 4. Nijhoff, Dordrecht, The Netherlands, pp. 221–4.

Becker, F. and Z.-L.Li (1990) Temperature-independent spectral indices in the thermal infrared bands. *Remote Sens. Environ.* 32:161–173.

Bland, W.L., J.T.Loew, and J.M.Norman (1996) Evaporation from cranberry. *Agric. For. Meteorol.* 81:1–12.

Blyth, E.M. and A.J.Dolman (1995) The roughness length for heat of sparse vegetation. *J. Appl. Meteorol.* 34:583–5.

Brutsaert, W. (1982) *Evaporation into the Atmosphere.* D. Reidel, Dordrecht, Holland, 299 pp.

Brutsaert, W. and M.Sugita (1995) Sensible heat transfer parameterization for surfaces with anisothermal dense vegetation.*J. Atmos. Sci.* 53:209–16.

Brye, K.R., J.M.Norman, L.G.Bundy, and S.T.Gower (2000) Water budget evaluation of natural and managed ecosystems. *Soil Sci Soc. Am.J.* 64:715–24.

Campbell, G.S. (1985) *Soil Physics with Basic.* Elsevier, New York.

Campbell, G.S. and J.M.Norman (1997) *An Introduction to Environmental Biophysics.* Springer-Verlag, New York, 286 pp.

Carlson, T.N. and D.A.Ripley (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* 62:241–52.

Chehbouni, A., D.Lo Seen, E.G.Njoku, and B.M.Monteney (1996) Examination of

the difference between radiative and aerodynamic surface temperatures over sparsely vegetated surfaces. *Remote Sens. Environ.* 58:177–86.

Chen, J.M. and J.Cihlar (1995) Quantifying the effect of canopy architecture on optical measurements of leaf area index using two gap size methods. *IEEE Trans. Geosci. Remote Sens.* 33:777–87.

Choudhury, B.J. (1987) Relationships between vegetation indices, radiation absorption, and net photosynthesis evaluated by a sensitivity analysis. *Remote Sens. Environ.* 22:209–33.

Choudhury, B.J., N.U.Ahmed, S.B.Idso, R.J.Reginato, and C.S.T.Daughtry (1994) Relations between evaporation coefficients and vegetation indices studied by model simulations. *Remote Sens. Environ.* 50:1–17.

Choudhury, B.J., S.B.Idso, and R.J.Reginato (1987) Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evaporation by an infrared-temperature based energy balance equation. *Agric. For. Meteorol.* 39: 283–97.

Collatz, G.J., J.T.Ball, C.Grivet, and J.A.Berry (1991) Physiological and environmental regulation of stomatal conductance, photosynthesis, and transpiration: a model that includes a laminar boundary layer. *Agric. For. Meteorol.* 54:107–36.

Collatz, G.J., M.Ribas-Carbo, and J.A.Berry (1992) Coupled photosynthesisstomatal conductance model for leaves of C4 plants. *Aust. J. Plant Physiol.* 19: 519–38.

Daughtry, C.S.T., M.A.Weltz, E.M.Perry, and W.P.Dulaney (1991) Direct and indirect estimates of leaf area index. *AMS Preprints of the 10th Conference on Biometeorology and Aerobiology and Special Session on Hydrometeorology,* September 10–13, Salt Lake City, Utah, pp. 230–3.

French, A.N., T.J.Schmugge, and W.P.Kustas (2002) Estimating evapotranspiration over El Reno, Oklahoma with imagery. *Agronomie* 22:105–6.

Friedl, M.A. (1996) Relationships among remotely sensed data, surface energy balance, and area-averaged fluxes over partially vegetated land surfaces. *J. Appl. Meteorol.* 35:2091–103.

Friedl, M.A. (1997) Examining the effects of sensor resolution and sub-pixel heterogeneity on spectral vegetation indices: Implications for biophysical modeling. In D.A. Quattrochi and M.F. Goodchild (eds) *Scale in Remote Sensing and GIS.* CRC Lewis, New York, pp. 113–39.

Garratt, J.R. and B.B.Hicks (1973) Momentum, heat and water vapor transfer to and from natural and artificial surfaces. *Q.J. R. Meteorol. Soc.* 99: 680–7.

Gijzen, H. and J.Goudriaan (1989) A flexible and explanatory model of light distribution and photosynthesis in row crops. *Agric. For. Meteorol.* 48:1–20.

Gillies, R.R. and T.N.Carlson (1995) Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. *J. Appl Meteorol.* 34: 745–56.

Giorgi, F. and R.Avissar (1997) Representation of heterogeneity effects in earth system modeling: experience from land surface modeling. *Rev. Geophys.* 35: 413–37.

Goudriaan, J. (1977) *Crop Micrometeorology: A Simulation Study.* Center for Agric. Publ. and Doc., PUDOC, Wageningen, The Netherlands.

Goward, S.N., R.H.Waring, D.G. Dye, and J. Yang (1994) Ecological remote sensing at OTTER: satellite macroscale observations. *Ecol. Appl.* 4:322–43.

Grace, J. (1981) Some effects of wind on plants. InJ. Grace, E.D. Ford, and P.G.

Jarvis (eds) *Plants and their Atmospheric Environment*. Blackwell, Boston, MA, pp. 31–56.

Gutman, G. and A.Ignatov (1998) The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, *Int. J. Remote Sens.* 19:1533–43.

Gutschick, V.P. (1996) Physiological control of evapotranspiration by shrubs: scaling measurements from leaf to stand with the aid of comprehensive models. In J.R.Barrow, E.D.McArthur, R.E.Sosebee, and R.J.Tausch (eds) *Proceedings of Shrubland Ecosystem Dynamics in a Changing Environment*. Gen. Tech. Rep. INT-GTR-338, Ogden, UT; USDA Forest Ser., Intermountain Res. Station, pp. 214–19.

Hall, F.G., K.F.Huemmrich, S.J.Goetz, P.J.Sellers, and J.E.Nickeson (1992) Satellite remote sensing of surface energy balance: success, failures, and unresolved issues in FIFE.J. *Geophys. Res.* 97 (D17): 19061–89.

Humes, K.S., W.P.Kustas, M.S.Moran, W.D.Nichols, and M.A.Welz (1994) Variability of emissivity and surface temperature over a sparsely vegetated surface. *Water Resour. Res.* 30:1299–310.

Humes, K.S., W.P.Kustas, and D.C.Goodrich (1997) Spatially distributed sensible heat flux over a semiarid watershed. Part I: use of radiometric surface temperatures and spatially uniform resistance. *J. Appl. Meteorol.* 36: 281–92.

Kimes, D.S. (1983) Remote sensing of row crop structure and component temperatures using directional radiometric temperature and inversion techniques. *Remote Sens. Environ.* 13:33–55.

Kondo, J. and S.Ishida (1997) Sensible heat flux from the earth's surface under natural convective conditions. *J. Atmos. Sci.* 54: 498–509.

Kubota, A. and M.Sugita (1994) Radiometrically determined skin temperature and scalar roughness to estimate surface heat flux. Part I: parameterization of radiometric scalar roughness. *Boundary–Layer Meteor.* 69:397–416.

Kucharik, C.J., J.M.Norman, and S.T.Gower (1999) Characterizing the radiation regime of nonrandom forest canopies: Theory, measurements, modeling and a simplified approach. *Tree Physiol.* 19: 695–706.

Kustas, W.P. (1990) Estimates of evapotranspiration with a one- and two-layer model of heat transfer over partial canopy cover. *J. Appl. Meteorol.* 29: 704–15.

Kustas, W.P. and C.S.T.Daughtry (1990) Estimation of the soil heat flux/net radiation ratio from spectral data. *Agric. For. Meteorol.* 49:205–23.

Kustas, W.P. and D.C.Goodrich (1994) Preface, MONSOON'90 Multidisciplinary Experiment. *Water Resour. Res.* 30 (5): 1211–25.

Kustas, W.P. and K.S.Humes (1996) Variations in the surface energy balance for a semi-arid rangeland using remotely sensed data at different spatial resolutions. In J.B.Stewart, E.T.R.Engman, A.Feddes and Y.Kerr (eds) *Scaling up in Hydrology using Remote Sensing*. John Wiley, London, pp. 127–45.

Kustas, W.P. and J.M.Norman (1996) Use of remote sensing for evapotrans-piration monitoring over land surfaces. *Hydrol. Sci.* 41:495–516.

Kustas, W.P. and J.M.Norman (1997) A two-source approach for estimating turbulent fluxes using multiple angle thermal infrared observations. *Water Resour. Res.* 33:1495–508.

Kustas, W.P. and J.M.Norman (1999a) Evaluation of soil and vegetation heat flux

predictions using a simple two-source model with radiometric temperatures for a partial canopy cover. *Agric. For. Meteorol.* 94:13–29.

Kustas, W.P. and J.M.Norman (1999b) Reply to comments about the basic equations of dual-source vegetation-atmosphere transfer models. *Agric. For. Meteorol.* 94: 275–8.

Kustas, W.P. and J.M.Norman (2000) Evaluating the effects of subpixel heterogeneity on pixel average fluxes. *Remote Sens. Environ.* 74:327–42.

Kustas, W.P., B.J.Choudhury, M.S.Moran, R.D.Reginato, R.D.Jackson, L.W.Gay, and H.L.Weaver (1989) Determination of sensible heat flux over sparse canopy using thermal infrared data. *Agric. For. Meteorol.* 44:197–216.

Kustas, W.P., M.S.Moran, K.S.Humes, D.I.Stannard, P.J.Pinter Jr., L.E.Hipps, E.Swiatek, and D.C.Goodrich (1994) Surface energy balance estimates at local and regional scales using optical remote sensing from an aircraft platform and atmospheric data collected over semiarid rangelands. *Water Resour. Res.* 30 (5): 1241–59.

Kustas, W.P., K.S.Humes, J.M.Norman, and M.S.Moran (1996) Single- and dualsource modeling of surface energy fluxes with radiometric surface temperature. *J. Appl. Meteorol.* 35:110–21.

Kustas, W.P., X.Zhan, and T.J.Schmugge (1998) Combining optical and microwave remote sensing for mapping energy fluxes in a semiarid watershed. *Remote Sens. Environ.* 64:116–31.

Kustas, W.P., J.D.Albertson, T.M.Scanlon, and A.T.Cahill (2001) Issues in monitoring evapotranspiration with radiometric temperature observations. In M.Owe, K.Brubaker, J.Ritchie, and A.Rango (eds) *Remote Sensing and Hydrology 2000.* IAHS Publ. No. 267 (ISBN 1–901502–46–5), pp. 239–45.

Lhomme, J.-P., B.Monteny, and M.Amadou (1994) Estimating sensible heat flux from radiometric temperature over sparse millet. *Agric. For. Meteorol.* 68:77–91.

Lhomme, J.-P., D.Troufleau, B.Monteny, A.Chehbouni, and S.Bauduin (1997) Sensible heat flux and radiometric surface temperature over sparse Sahelian vegetation II. A model for the kB^{-1} parameter. *J. Hydrol.* 188–189:839–54.

Matsushima, D. and J.Kondo (1997) A proper method for estimating sensible heat flux above a horizontal-homogeneous vegetation canopy using radiometric surface observations.*J. Appl Meteorol.* 36: 1696–711.

McNaughton, K.G. and P.G.Jarvis (1991) Effects of spatial scale on stomatal control of transpiration. *Agric. For. Meteorol.* 54:279–301.

Mecikalski, J.R., G.R.Diak, M.C.Anderson, and J.M.Norman (1999) Estimating fluxes on continental scales using remotely-sensed data in an atmospheric-land exchange model.*J. Appl Meteorol.* 38:1352–69.

Menenti, M. and J.C.Ritchie (1994) Estimation of effective aerodynamic roughness of Walnut Gulch watershed with laser altimeter measurements. *Water Resour. Res.* 30 (5): 1329–37.

Norman, J.M. (1979) Modeling the complete crop canopy. In B.J.Barfield and J.F.Gerber (eds) *Modification of the Aerial Environment of Plants.* ASAE Monogr. Am. Soc. Agric. Engr., St Joseph, MI, pp. 249–77.

Norman, J.M. (1980) Interfacing leaf and canopy light interception models. In J.D.Hesketh and J.W.Jones (eds) *Predicting Photosynthate Production and use for Ecosystem Models,* Vol. II. CRC Press, West Palm Beach, FL, pp. 49–67.

Norman, J.M (1988) Synthesis of canopy processes. In G.Russell *et al.* (eds) *Plant*

Canopies: Their Growth, Form and Function. Soc. Exp. Biol., Seminar Series 31, Cambridge University Press, New York, pp. 161–75.

Norman, J.M. and G.S.Campbell (1983) Application of a plant- environment model to Problems in irrigation. In D.I.Hillel (ed.) *Advances in Irrigation,* Vol. II. Academic Press, New York, pp. 155–88.

Norman, J.M. and W.Polley (1989) Canopy photosynthesis. In W.R. Briggs (ed.) *Photosynthesis*. Alan R.Liss, New York, pp. 227–41.

Norman, J.M., and T.J.Arkebauer (1991) Predicting canopy light-use efficiency from leaf characteristics. In J.T.Ritchie and R.J.Hanks (eds) *Modeling Plant and Soil Systems*. ASA, Madison, WI, pp. 125–43.

Norman, J.M. and F.Becker (1995) Terminology in thermal infrared remote sensing of natural surfaces. *Remote Sens. Rev.* 12:159–73.

Norman, J.M., J.M.Welles, and E.A.Walter (1985) Contrasts among bidirectional reflectance of leaves, canopies and soils. *IEEE Trans. Geosci. Remote Sens.* GE-23: 659–77.

Norman, J.M., J.Chen, and N.S.Goel (1990) Thermal emissivity and infrared temperature dependence on plant canopy architecture and view angle. IGARSS Abstracts, IEEE Cat. no. 90CH2825–8, 3:1747–1750.

Norman, J.M., R.Garcia, and S.B.Verma (1992) Soil surface CO_2 fluxes and the carbon budget of a grassland. *J. Geophys. Res.* 97: 18845–53.

Norman, J.M., W.P.Kustas, and K.S.Humes (1995) A two-source approach for estimating soil and vegetation energy fluxes from observations of directional radiometric surface temperature. *Agric. For. Meteorol.* 77:263–93.

Norman, J.M., W.P.Kustas, J.H.Prueger, and G.R.Diak (2000) Surface flux estimation using radiometric temperature: a dual-temperature-difference method to minimize measurement errors. *Water Resour. Res.* 36:2263–74.

Ottle, C. and C.Francois (1994) Atmospheric corrections to ATSR-IR data. In *Workshop on Thermal Remote Sensing of the Energy and Water Balance over Vegetation in Conjunction with Other Sensors*. La Londe Les Maures, France, Cemegref Editions, ISBN 2–85362–371–8, pp. 79–82.

Priestley, C.H.B. and R.J.Taylor (1972) On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Rev.* 100:81–92.

Prata, A.J. (1993) Land surface temperatures derived from the AVHRR and ATSR I: theory.*J. Geophys. Res.* 89 (D9): 16689–702.

Prata, A.J. (1994) Angular effects on the determination of land surface temperatures. In *Workshop on Thermal Remote Sensing of the Energy and Water Balance over Vegetation in Conjunction with Other Sensors*. La Londe Les Maures, France, Cemegref Editions, ISBN 2–85362–371–8, pp. 113–22.

Prince, S.D., S.J.Goetz, R.O.Dubayah, K.P.Czajkowski, and M.Thawley (1998) Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very high-Resolution Radiometer satellite observations: comparison with field observations. *J. Hydrol.* 212–213: 230–49.

Raupach, M.R. (1994) Simplified expressions for vegetation roughness length and zero-plane displacement as functions of canopy height and area index, *BoundaryLayer Meteorol.* 71:211–16 (corrigenda 76:303–4).

Ross, J. (1975) Radiative transfer in plant communities. In J.L.Monteith (ed.) *Vegetation and the Atmosphere, Vol. 1,* Academic Press, New York, pp. 13–55.

Ross, J. (1981) The radiation regime and architecture of plants. In H.Lieth (ed.) *Tasks for Vegetation Sciences 3*. Dr. W.Junk, The Hague, Netherlands, 391 pp.

Salisbury, J.W. and D.M.D'Aria (1992) Emissivity of terrestrial materials in the 8–14 mm atmospheric window. *Remote Sens. Environ.* 42:83–106.

Sauer, T.J. (1993) Sensible and latent heat exchange at the soil surface beneath a maize canopy. PhD thesis, University of Wisconsin, Madison, WI, 292 pp.

Sauer, T.J. and J.M.Norman (1995) Simulated canopy microclimate using estimated below-canopy soil surface transfer coefficients. *Agric. For. Meteorol.* 75: 135–60.

Sauer, T.J., J.M.Norman, C.B.Tanner, and T.B.Wilson (1995) Measurement of heat and vapor transfer coefficients at the soil surface beneath a maize canopy using source plates. *Agric. For. Meteorol.* 75:161–89.

Schmugge, T.J., W.P.Kustas, and K.S.Humes (1998) Monitoring land surface fluxes using ASTER observations. *IEEE Trans. Geosci. Remote Sens.* 36:1421–30.

Sellers, P.J., M.D.Heiser, F.G.Hall, S.J.Goetz, D.E.Strebel, S.B.Verma, R.L.Desjardins, P.M.Schuepp, and J.I.MacPherson (1995) Effects of spatial variability in topography, vegetation cover and soil moisture on area-averaged surface fluxes: a case study using FIFE 1989 data. *J. Geophys. Res.* 100 (D12): 25607–29.

Shuttleworth, W.J. and J.S.Wallace (1985) Evaporation from sparse canopies—an energy combination theory. *Q.J. R. Meteorol. Soc.* 111:839–55.

Shuttleworth, W.J., Z.-L.Yang, and M.A.Arain (1997) Aggregation rules for surface parameters in global models. *Hydrol. Earth Sys. Sci.* 2:217–26.

Stannard, D.I. (1988) Use of a hemispherical chamber for measurement of evapotranspiration. US Geological survey Open File Report 88–452, 18 pp.

Stannard, D.I. (1999) Personnel communication.

Stannard, D.I., J.H.Blanford, W.P.Kustas, W.D.Nichols, S.A.Amer, T.J.Schmugge, and M.A.Weltz (1994) Interpretation of surface-flux measurements in heterogeneous terrain during the MONSOON 90 experiment. *Water Resour. Res.* 30 (5): 1227–39.

Stewart, J.B., W.P.Kustas, K.S.Humes, W.D.Nichols, M.S.Moran, and A.A.R.De Bruin (1994) Sensible heat flux-radiometric surface temperature relationship for eight semiarid areas. *J. Appl. Meteorol.* 33:1110–17.

Sugita, M. and W.Brutsaert (1996) Optimal measurement strategy for surface temperature to determine sensible heat flux from anisothermal vegetation. *Water Resour. Res.* 32:2129–34.

Sun, J. (1999) Diurnal variations of thermal roughness height over a grassland. *Boundary Layer Meteorol.* 92:407–427.

Sun, J. and L.Mahrt (1995) Determination of surface fluxes from the surface radiative temperature. *J. Atmos. Sci.* 52:1096–106.

Thompson, A.L., J.R.Gilley, and J.M.Norman (1993) A sprinkler water droplet evaporation and plant canopy model: II. Model application. *Trans. ASAE* 36: 743–50.

Toole, J.L., J.M.Norman, T.Holtzer, and T.Perring (1984) Simulating Banks grass mite population dynamics as a subsystem of a crop canopy-microenvironmental model. *Environ. Entomol.* 13:329–37.

Troufleau, D., J.P.Lhomme, B.Monteney, and A.Vidal (1997) Sensible heat flux and radiometric surface temperature over sparse Sahelian vegetation. I. An experimental analysis of the kB^{-1} parameter. *J. Hydrol.* 188–189:815–38.

Verhoef, W. (1984) Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sens. Environ.* 16:125–41.

Verhoef, A., H.A.R.De Bruin, and B.J.J.M.Van den Hurk (1997) Some practical notes on the parameter kB- for sparse vegetation. *J. Appl. Meteorol.* 36: 560–72.

Vining, R.C. and B.L.Blad (1992) Estimation of sensible heat flux from remotely sensed canopy temperatures.*J. Geophys. Res.* 97 (D17): 18951–4.

Weiss, A.D., L.Lukens, J.M.Norman, and J.R.Steadman (1989) Leaf wetness in dry beans under semi-arid conditions. *Agric. For. Meteorol.* 48:149–62.

Weltz, M.A., J.C.Ritchie, and H.D.Fox (1994) Comparison of laser and field measurements of vegetation height and canopy cover. *Water Resour. Res.* 30: 1311–20.

Wieringa, J. (1986) Roughness dependent geographical interpolation of surface wind speed averages. *Q.J. R. Meteorol. Soc.* 112:867–89.

Willmott, C.J. (1982) Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc.* 11:1309–13.

Zhan, X., W.P.Kustas, and K.S.Humes (1996) An intercomparison study on models of sensible heat flux over partial canopy surfaces with remotely sensed surface temperature. *Remote Sens. Environ.* 58:242–56.

Part II

Thermal infrared data for assessment of ecosystem health

Chapter 8

Thermal infrared measurement as an indicator of plant ecosystem health

M.Susan Moran

8.1 Introduction

The term ecosystem health has generally been used to indicate the proper functioning of a complex ecosystem. A healthy system would be one in which the biophysical processes are operating adequately to maintain the structure, organization, and optimal activity of the system over time. As a result, the healthy ecosystem would be capable of producing quality commodities and satisfying the values desired by society. For example, a healthy rangeland supplies such important economic goods as forage for livestock, wildlife habitat, energy, recreational opportunities, and such intangible products as natural beauty. A healthy cropland would allow us to grow copious, high-quality food and fiber using less water and pesticides. Although most would agree that healthy croplands and rangelands are desirable, there is less consensus on how to evaluate and monitor plant ecosystem health.

For example, rangeland health has been defined as *the degree to which the integrity of the soil and the ecological processes of rangeland ecosystems are sustained* (National Research Council 1994). The National Research Council suggested that determination of whether a rangeland is healthy, at risk, or unhealthy should be based on the evaluation of three criteria: (a) degree of soil stability and watershed function; (b) integrity of nutrient cycles and energy flows; and (c) presence of functioning recovery mechanisms. They proposed that these criteria could be measured by such indicators as soil surface characteristics (erosion and infiltration), transpiration rate, or photosynthetic period, and such plant community attributes as plant distribution, vegetation cover, plant vigor, and plant age class distribution.

On the other hand, cropland health is commonly evaluated as a *lack of stress,* where stress is defined as the downward deviation from the plant potential photosynthetic and transpiration rates, and ultimately a decrease in crop yield. For example, an indicator of cropland health, termed the Stress Degree Index (SDI), was proposed by Hiler and Clark (1971). The SDI is based on a measure of the degree and duration of plant water—and therefore photosynthetic—stress, and the plant's susceptibility to a given water stress,

$$SDI = \sum_{i=1}^{n} (SD_i \cdot CS_i) \tag{8.1}$$

where SD is the stress day factor, CS is the crop susceptibility factor, and n represents the number of growth periods (days) considered. SD is defined by the plant transpiration rate (Γ) and the potential transpiration rate (Γ_p), where

$$SD = 1 - (\Gamma / \Gamma_p) \tag{8.2}$$

CS represents the susceptibility of a crop to a given magnitude of SD and is a function of the species and stage of growth of the crop (Hiler and Clark 1971).

These disparate definitions of cropland and rangeland health reflect the differences in resource management. That is, in intensively managed croplands, crop health can be augmented by increasing water or fertilizer applications, modifying tillage or terracing practices, selecting hybrids, managing plant population, and applying agricultural chemicals. Rangelands, for the most part, do not receive such inputs and must depend on the availability of nutrients, the dynamics of the plant community, the quality of the soil profile, and management of the grazing pressure.

However, the definitions of cropland and rangeland health have two indicators in common—plant transpiration rate and plant photosynthetic rate. It is not coincidental that transpiration rate is also a primary indicator of forest health and forest fire risk (Vidal and Devaux-Ros 1995). Transpiration has such an important role in plant health that even a slight reduction in plant water content can impact both growth and such physiological functions as photosynthesis and respiration (Hatfield 1997). It is this importance that has led to extensive efforts to use thermal infrared (TIR) measurements to evaluate the spatial and temporal distribution of plant transpiration. The direct link between the process of transpiration and the plant thermal response (i.e. transpired water evaporates and cools the leaves) offers potential for the use of TIR measurements for monitoring and managing plant ecosystem health.

This chapter will include a history of developments in physics and technology that have led to development of TIR spectral indices of plant ecosystem health. Descriptions of several TIR indices will be provided with sufficient detail that readers can implement the algorithm with minimal reference to other publications. The chapter will conclude with a short discussion of the pitfalls associated with TIR measurements and algorithm implementation.

8.2 History

Parallel lines of research in physics and technology converged to define the current state of TIR measurement of thermal responses as an indicator of plant ecosystem health. One line of research was focused on the basic physics of evaporation and energy balance; another was focused on the measurements of leaf and canopy temperatures. These lines crossed with the work of Monteith and Szeicz (1962) in which canopy temperature was presented as a function of energy balance components. The historical developments leading to this pivotal work is the topic of this section.

Before continuing, a clarification of terms is in order. The term *evapotranspiration* (E_Γ) is used here to include the evaporation (E) of water from soil and leaf surfaces and the transpiration (Γ) of water through leaf stomata. Though in both cases water is "evaporating," and though many prefer to incorporate both evaporation and transpiration into the single term *evaporation,* the distinction is made here when necessary for discussion and presentation of theory. Furthermore, the distinction between the terms λE_Γ and E_Γ must be clarified; λE_Γ is evapotranspiration rate in energy flux density units (W m^{-2}) and EG is the same quantity converted to units of rate (mm h^{-1} or mm day^{-1}). In this discussion, λE_Γ and E_Γ are often used interchangeably.

It is also important to emphasize the differences between T_c, T_o, and T_s. T_c is the canopy temperature, defined by Norman *et al.* (1995) as the TIR temperature in which the "vegetation dominates the [measurement] field of view minimizing the effect of soil." T_o is the temperature of the soil surface. T_s is the surface composite temperature, defined by Norman *et al.* (1995) as the "aggregate temperature of all objects comprising the surface," which was shown by Kustas *et al.* (1990) to be a function of T_c and T_o, where

$$T_s^4 = f_c T_c^4 + (1 - f_c) T_o^4 \tag{8.3}$$

and f_c represents the fractional cover of the vegetation. When the surface is completely covered by vegetation, then $T_s = T_c$, and when the surface is bare soil, then $T_s = T_o$. Throughout this discussion, all temperatures are assumed to be kinetic values, that is, all radiometric temperature measurements have been corrected for surface emissivity. The scientific and technical notation, and selected acronyms included in this chapter are summarized in Table 8.1.

8.2.1 The physics of evaporation

Progress in understanding the physics of evaporation was slow until the twentieth century when Bowen (1926) showed how the partitioning of available energy between the fluxes of sensible and latent heat could be determined from gradients of temperature and humidity,

$$\lambda E_\Gamma = (R_n - G)/(1 + \beta) \tag{8.4}$$

Table 8.1 Summary of scientific and technical notation and selected acronyms

Energy balance	
R_n	net radiant heat flux density ($W\,m^{-2}$)
G	soil heat flux density ($W\,m^{-2}$)
H	sensible heat flux density ($W\,m^{-2}$)
λE_Γ	latent heat flux density ($W\,m^{-2}$)
λ	heat of vaporization ($J\,kg^{-1}$)
E_Γ	evapotranspiration rate ($mm\,h^{-1}$)
$E_{\Gamma p}$	potential evapotranspiration rate ($mm\,h^{-1}$)
E	evaporation rate ($mm\,h^{-1}$)
Γ	transpiration rate ($mm\,h^{-1}$)
Γ_p	potential transpiration rate ($mm\,h^{-1}$)
β	Bowen ratio (unitless)
K_h, K_v	eddy transfer coefficients for sensible and latent heat, respectively
W_f	wind function ($m\,s^{-1}$)
Temperature	
T_a	air temperature (°C)
T_c	canopy temperature (°C)
T_{cm}	minimum canopy temperature (°C)
T_{cx}	maximum canopy temperature (°C)
$T_{c'}$	optimum canopy temperature, defined by enzyme function (°C)
T_o	soil temperature (°C)
T_s	surface composite temperature (a weighted average of T_o and T_c)(°C)
T_q	temperature of an imitation paper leaf (°C)
Humidity	
VPD	$(e^* - e_a)$, vapor pressure deficit of the air at $(T_c - T_a)/2$ (kPa)
VPD_x	$(e_b^* - e_a)$, vapor pressure deficit of the air at $(T_a + f_0)$ (kPa)
e_a	vapor pressure of the air (kPa)
e^*	saturated vapor pressure at $(T_c + T_a)/2$ (kPa)
e_b^*	saturated vapor pressure at $(T_a + f_0)$ (kPa)
f_1, f_0	slope and offset of the relation between $(T_c - T_a)$ and VPD for a well-watered crop
Resistance	
r_a	aerodynamic resistance ($s\,m^{-1}$)
r_c	canopy resistance to vapor transport ($s\,m^{-1}$)
r_{cx}	maximum canopy resistance, associated with nearly complete stomatal closure ($s\,m^{-1}$)
r_{cp}	canopy resistance at potential evapotranspiration ($s\,m^{-1}$)
r_{cm}	minimum canopy resistance, where $r_{cm} = r_{cp}$ ($s\,m^{-1}$)
r_s	stomatal resistance ($s\,m^{-1}$)
r_{sx}	maximum stomatal resistance ($s\,m^{-1}$)
r_{sp}	stomatal resistance at potential evapotranspiration ($s\,m^{-1}$)
r_{sm}	minimum stomatal resistance, where $r_{sm} = r_{sp}$ ($s\,m^{-1}$)

(*continued*)

Table 8.1 (Continued)

Miscellaneous	
ρC_p	volumetric heat capacity of air ($J°C^{-1}\,m^{-3}$)
d_0	displacement height (m)
Δ	slope of the saturated vapor pressure–temperature relation ($kPa\,°C^{-1}$)
f_c	vegetation cover (ranging from 0 = no vegetation to 1 = full vegetation cover)
γ	psychrometric constant ($kPa°C^{-1}$)
γ^*	$\gamma(1 + r_{cp}/r_a)$ ($kPa°C^{-1}$)
k	von Karman's constant ($\approx$0.4)
K_m	minimum apparent Michaelis constant
LAI	leaf area index ($m^2\,m^{-2}$)
Φ_h	stability correction for heat (unitless)
Φ_m	stability correction for momentum (unitless)
ρ_{NIR}, ρ_{red}	near-infrared and red reflectances, respectively
u	wind speed ($m\,s^{-1}$)
z_{0m}	roughness length for momentum (m)
z_{0h}	scalar roughness for heat (m)
z	height above the surface at which u and T_a are measured (m)
Acronyms	
CTV	canopy temperature variability (Clawson and Blad 1982)
$CWSI_b$	baseline crop water stress index (Idso *et al.* 1981)
$CWSI_q$	three-temperature crop water stress index (Qiu *et al.* 1996a)
$CWSI_t$	theoretical crop water stress index (Jackson *et al.* 1981)
SAVI	soil-adjusted vegetation index (Huete 1988)
SDD	stress degree day (Idso *et al.*1977; Jackson *et al.* 1977)
SDI	stress degree index (Hiler and Clark 1971)
SEHEM	spatially explicit hydro-ecological model (Nouvellon *et al.* 2000)
TKW	thermal kinetic window (Burke *et al.* 1988)
TSD	temperature stress day (Gardner *et al.* 1981a)
WDI	water deficit index (Moran *et al.* 1994)

where λE_T is the latent heat flux ($W\,m^{-2}$), R_n is the net radiation flux at the surface ($W\,m^{-2}$), G is the sensible heat flux conducted to the soil ($W\,m^{-2}$), and ß is the Bowen ratio, the ratio of sensible heat flux density (H) to latent heat flux density. In equation (8.4), fluxes away from the surface are positive and those towards the surface are negative. The Bowen ratio can be derived from temperature and humidity measurements,

$$\beta = \gamma(K_h/K_v)(\Delta T/\Delta e) \tag{8.5}$$

where γ is the psychrometric constant (2.453 $MJkg^{-1}$ at 20°C), K_h and K_v are the eddy transfer coefficients for sensible and latent heat, respectively, and ΔT and Δe are the differences in temperature in °C and vapor pressure in kPa over the same vertical separation (Δz) of the sensors.

Following the work of Bowen (1926), Penman (1948) combined the thermal energy balance with certain aerodynamic aspects of evaporation and developed an equation for estimating evaporation that was soon adopted by hydrologists and irrigation specialists. The general form of the Penman combination equation is

$$\lambda E_\Gamma = [(\Delta/(\Delta+\gamma))(R_n - G) + (\gamma/(\Delta+\gamma))6.43\, W_f\, \mathrm{VPD}] \tag{8.6}$$

where Δ is the slope of the saturation vapor pressure-temperature curve (kPa $^oC^{-1}$), W_f is a wind function (generally, $a + b(u)$, where u is the wind speed in m s^{-1}), and VPD is Vapor Pressure Deficit (kPa).

The Penman formula was recast in terms of an aerodynamic resistance and a surface resistance for application to single leaves (Penman 1953) and vegetation canopies (Monteith 1965; Rijtema 1965). This result, now referred to as the Penman-Monteith equation, is probably the most universally used equation for calculating potential evaporation (Allen 1986),

$$\lambda E_\Gamma = [\Delta(R_n - G) + \rho C_p(\mathrm{VPD})/r_a]/[\Delta + \gamma^*] \tag{8.7}$$

where ρ is air density (kg m^{-3}), C_p is specific heat at constant pressure (kJ kg^{-1} $^oC^{-1}$), and the aerodynamic resistance, r_a(s m^{-1}) is

$$r_a = \{[\ln((z-d_0)/z_{0m}) + \ln(z_{0m}/z_{0h}) - \Phi_h][\ln((z-d_0)/z_{0m}) - \Phi_m]\}/k^2u \tag{8.8}$$

and z is the height above the surface at which u is measured (m), d_o is the displacement height (m), z_{0m} and z_{0h} are the roughness lengths for momentum and heat (m), respectively, Φ_h and Φ_m are the stability corrections for heat and momentum, respectively, and k is von Karman's constant (≈0.4). The integral stability functions were summarized by Beljaars and Holtslag (1991) for stable and unstable conditions. The value of γ^* (kPa $^oC^{-1}$) in equation (8.7) is a function of r_a and the canopy resistance to vapor transport (r_c, s m^{-1}), where

$$\gamma^* = \gamma(1 + r_{cp}/r_a) \tag{8.9}$$

Priestley and Taylor (1972) proposed a simplified version of the Penman combination equation for computation of potential evaporation heat flux density (λE_{Gp}), for a surface which has minimal resistance to evaporation. Under these conditions, the aerodynamic component was ignored and the energy component was multiplied by a coefficient,

$$\lambda E_{\Gamma p} = \alpha\ (\Delta/(\Delta+\gamma))(R_n - G) \tag{8.10}$$

where α = 1.26 for regions with no or low advective conditions. For an extensive history of evaporation research see the treatise by Monteith (1981) and the book by Brutsaert (1982).

During the four decades since Penman's pioneering work, theoretical, modeling, and experimental efforts have greatly expanded our knowledge of evaporation. Theoretical and modeling developments are dependent on experimental data for verification. Experimentally, the measurement of evaporation can be made accurately with equipment that evaluates Bowen's ratio (Spittlehouse and Black 1980; Gay and Greenberg 1985), with eddy correlation techniques (Kanemasu *et al.* 1979), with weighing lysimeters (van Bavel and Myers 1962), and with portable assimilation chambers (Reicosky 1981). A limitation of these techniques is that they yield essentially point values of evaporation, and therefore are applicable only to a homogeneous area surrounding the equipment that is exposed to the same environmental factors. This limitation has been the impetus for studies on the use of remotely sensed data, obtained by either aircraft- or spacecraft-based sensors, to obtain an estimate of the spatial distribution of evaporation.

8.2.2 The measurement of leaf and canopy temperatures

It has long been recognized that a direct measurement of some plant parameter would be superior to measurement of water status of the soil for monitoring the plant's response to its atmospheric and edaphic environment. Jackson (1982) reviewed the history of attempts to measure leaf and canopy temperatures and to relate them to plant stress conditions. That history is summarized here to give a perspective from which to understand the spectral plant stress indices presented in the next section.

One of the first measurements of leaf temperature of crop plants under natural field conditions was made by Miller and Saunders (1923) using a clamping device which, when closed, would make contact between the leaf and a thermocouple. This device required about 15s to make a single leaf measurement, and was criticized because the clamp shaded the leaf (Curtis 1936). Nonetheless, Miller and Saunders were perhaps the first scientists to show that leaf temperatures (in this case, alfalfa leaves) were cooler than air temperature under field conditions. These results were corroborated by Eaton and Belden (1929) who found that temperatures of cotton leaves in Arizona were 2-4° C below that of the air. Wallace and Clum (1938) found similar results for garden plants using thermocouples to record leaf and air temperatures. They reported leaf temperatures as much as 7°C below air temperature. Again, Curtis (1938) rebutted these findings with a theoretical proof that it would require impossibly high transpiration rates to lower the temperature of a leaf below that of air when in direct sunlight.

However, empirical evidence continued to mount that leaf temperatures could be cooler than air temperatures on clear, warm days (Waggoner and Shaw 1952; Asari and Loomis 1959). A theoretical proof based on energy balance was published by Gates (1964) stating that transpiration was indeed

important in the energy budget of plants, and his calculations indicated that shaded leaves could be cooler than air.

In the 1960s, empiricists began using crude infrared thermometers (IRT) to remotely sense leaf temperatures (for discussion of basic IR radiation thermometry, see Fuchs and Tanner 1966). Wiegand and Namken (1966) thus observed that cotton-leaf temperatures increased linearly with increasing insolation, and decreased linearly with increasing turgidity of the leaves. Based on these observations, they suggested a measurement protocol that is still in use today; that is, leaf temperature interpretation requires simultaneously measured radiation data, and the early afternoon is a good time of day for making leaf temperature measurements.

With the development of the IRT, research moved from studies of individual leaves to investigations of entire crop canopies. The conclusion that emerged from the early theoretical studies of leaf versus canopy temperatures was that reradiation was the major mode of heat transfer for individual leaves, but for canopies as a whole, transpiration was the dominant mechanism (Idso and Baker 1967). Thus, for a canopy, transpirational cooling played a major role in determining plant temperature. This conclusion supported the concept that TIR measurements would be useful for such canopy-scale management activities as irrigation scheduling and monitoring general plant ecosystem health.

The link between studies of the physics of evaporation (previous subsection) and the measurements of leaf temperatures (this subsection) was made by Monteith and Szeicz (1962) through their use of IRTs to measure plant canopy temperatures. Based on energy balance considerations, they derived an expression relating the canopy-air temperature difference to net radiation, wind speed, vapor pressure gradient, and the aerodynamic and canopy resistances, where

$$(T_c - T_a) = [r_a(R_n - G)/\rho C_p][\gamma(1 + r_c/r_a)/\{\Delta + \gamma(1 + r_c/r_a)\}] - [\mathrm{VPD}/\{\Delta + \gamma(1 + r_c/r_a)\}] \quad (8.11)$$

Equation (8.11), given in a slightly different form by Monteith and Szeicz (1962), was the basis for the theoretical development of a Crop Water Stress Index (CWSI), which will be discussed in the next section. In the decade following this landmark work, IR technology advanced rapidly (see other chapters in this book), leading to the handheld, airborne and satellite-based sensors, and scanners that are available today.

8.3 TIR indices, algorithms, and models

There has been a great deal of work directed at refining and validating methods for using TIR measurements for estimation of crop transpiration and soil evaporation (see review by Kustas and Norman 1996). A distinction

is made here between estimates of *absolute* plant transpiration rate, and stress indices, algorithms and models, which provide a *relative* measure of plant ecosystem health by normalizing biophysical activity (e.g. transpiration rate or enzyme function) to a known range of values, to a healthy system, or to another standard. The goal of this section is to review several of the latter approaches that have obtained scientific, and in some cases, commercial acceptance, and provide the reader with instructions and references for such algorithm implementation.

8.3.1 Stress degree day

The stress degree day (SDD) was defined by Idso *et al.* (1977) and Jackson *et al.* (1977) as

$$\mathrm{SDD} = \sum_{i=1}^{n} (T_c - T_a)_i \tag{8.12}$$

where T_c is plant canopy temperature measured 1–1.5 h after solar noon, T_a is the air temperature 1.5 m above the soil, and (T_c - T_a) is summed over n days beginning at day i. Idso *et al.* (1977) showed that in the arid climate of Phoenix, Arizona, wheat yields were reduced if the SDD became positive. Jackson *et al.* (1977) postulated that the SDD was directly related to water use and could be used to schedule crop irrigations and monitor periods of grassland stress to provide an estimate of yield. In a later work, Jackson (1982) acknowledged that the arbitrary division at SDD = 0 was not appropriate for all environmental conditions, particularly regions with high atmospheric humidity.

8.3.2 Canopy temperature variability

The canopy temperature variability (CTV) was defined by Clawson and Blad (1982) as the range (maximum minus minimum) of all IRT sensed canopy temperatures within a plot during a particular measurement period. They found that as the temperature difference between a water-stressed plot and a well-watered plot increased, so did CTV values until a critical stress level was reached beyond which CTV values stabilized. They postulated that the onset of plant stress could be signaled when CTV values exceeded 0.7°C. Like the SDD, the CTV did not account for differences in VPD; on the other hand, the CTV did not require an estimate of T_a and could possibly be implemented without atmospheric correction of T_c values acquired with an airborne thermal sensor. With our increased knowledge of plant thermal response and advances in TIR technology, this work deserves renewed attention.

8.3.3 Temperature stress day

Gardner *et al.* (1981a) attempted to account for the theoretical weaknesses in the SDD and CTV by taking the difference in temperatures of a stressed plot and well-watered plot, termed the temperature stress day (TSD). They postulated that the use of a well-watered or "healthy" standard would compensate for such environmental effects as T_a and VPD. As with the CTV and SDD, a threshold could be determined to signal the need for irrigation or the point of irreversible grassland degradation. In addition, Gardner *et al.* (1981b) demonstrated that TSD could also be used to predict the phenologic development stage of corn. A drawback to operational implementation of TSD for monitoring system health is that meteorological conditions necessitate that the TSD standard plot be in close proximity to the field or site in question.

In a re-examination of TSD, Clawson *et al.* (1989) reported that the TSD was not an environmentally independent index of crop water stress, but was limited by the same environmental dependencies as the SDD. They suggested a refinement of TSD to account for environmental factors while maintaining the measurement simplicity of the TSD. Their approach required estimates of R_n, T_a, and r_a in addition to the temperature measurements of the stressed and well-watered plots, and was similar to the CWSI described in the next subsection. This approach was recently validated in a study of the effects of stress on yield in irrigated cotton, corn, and soybeans (Wanjura and Upchurch 1998).

8.3.4 Crop water stress index

An important contribution to the use of plant temperature to assess plant ecosystem health was the creation of the Idso-Jackson CWSI (Idso *et al.* 1981; Jackson *et al.* 1981). Jackson *et al* (1981) derived the theoretical CWSI ($CWSI_t$) based largely on equation (8.7). Taking the ratio of actual (λE_Γ for any r_c) to potential ($\lambda E_{\Gamma p}$ for $r_c = r_{cp}$) latent heat flux density gives

$$\lambda E_\Gamma / \lambda E_{\Gamma p} = [\Delta + \gamma^*]/[\Delta + \gamma(1 + r_c/r_a)] \tag{8.13}$$

where r_{cp} is the canopy resistance at potential evapotranspiration. Jackson *et al.* (1981) defined the $CWSI_t$, ranging from 0 (ample water) to 1 (maximum stress), as

$$CWSI_t = 1 - \lambda E_\Gamma / \lambda E_{\Gamma p} = [\gamma(1 + r_c/r_a) - \gamma^*]/[\Delta + \gamma(1 + r_c/r_a)] \tag{8.14}$$

To solve equation (8.14), a value of r_c/r_a is obtained by rearranging equation (8.11) and assuming G is negligible for a full-cover canopy, where

$$\begin{aligned} r_c/r_a = {} & \{[\gamma r_a R_n/\rho C_p] - [(T_c - T_a)(\Delta + \gamma)] - \mathrm{VPD}\}/ \\ & \{\gamma[(T_c - T_a) - r_a R_n/\rho C_p]\} \end{aligned} \tag{8.15}$$

and r_c/r_a is substituted into equation (8.14) to obtain the $CWSI_t$.

Another equivalent approach for solution of equation (8.14) is to compute the theoretical upper and lower limits of T_c - T_a using equation (8.11) and combine these with the measured T_c - T_a value to compute $CWSI_t$ as

$$\begin{aligned} CWSI_t &= 1 - \lambda E_\Gamma / \lambda E_{\Gamma p} \\ &= [(T_c - T_a)_m - (T_c - T_a)_r]/[(T_c - T_a)_m - (T_c - T_a)_x] \end{aligned} \tag{8.16}$$

where the subscripts m, x, and r refer to the minimum, maximum and measured values, respectively. For full-cover, well-watered vegetation,

$$\begin{aligned} (T_c - T_a)_m = [r_a R_n/\rho C_p][\gamma(1 + r_{cm}/r_a)/\{\Delta + \gamma(1 + r_{cm}/r_a)\}] \\ - [VPD/\{\Delta + \gamma(1 + r_{cm}/r_a)\}] \end{aligned} \tag{8.17}$$

where $r_{cm} = r_{cp}$. For full-cover vegetation with no available water,

$$\begin{aligned} (T_c - T_a)_x = [r_a R_n/\rho C_p][\gamma(1 + r_{cx}/r_a)/\{\Delta + \gamma(1 + r_{cx}/r_a)\}] \\ - [VPD/\{\Delta + \gamma(1 + r_{cx}/r_a)\}] \end{aligned} \tag{8.18}$$

where r_{cx} is the canopy resistance associated with nearly complete stomatal closure ($r_{cx} \rightarrow \infty$). Monteith (1973) suggested the values of r_{cm} and r_{cx} could be obtained from measurements of stomatal resistance (r_s) and leaf area index (LAI), where

$$r_{cm} = r_{sm}/LAI \quad \text{and} \quad r_{cx} = r_{sx}/LAI \tag{8.19}$$

where LAI > 0. Values of minimum and maximum stomatal resistance (r_{sm} and r_{sx}, respectively) are published for many agricultural crops under a variety of atmospheric conditions. If values are not available, reasonable values of r_{sm} = 25-100 s m^{-1} and r_{sx} = 1,000–1,500 s m^{-1} will not result in appreciable error in equations (8.17) and (8.18). That is, when r_c is very large or very small (relative to r_a), its influence on the magnitude of (T_c - T_a) in equations (8.17) and (8.18) is small.

Though Jackson *et al.* (1981) provided a thorough theoretical approach for computation of CWSI, the concept is more universally applied using a semi-empirical variation proposed by Idso *et al.* (1981) based on the "nonwater-stressed baseline." This baseline is defined by the relation between (T_c - T_a) and VPD under non-limiting soil moisture conditions, that is, when the plants in question are transpiring at the potential rate (Figure 8.1). Such non-water-stressed baselines have been determined for many different crops, including aquatic crops and grain crops for both pre-heading and post-heading growth rates (Idso 1982).

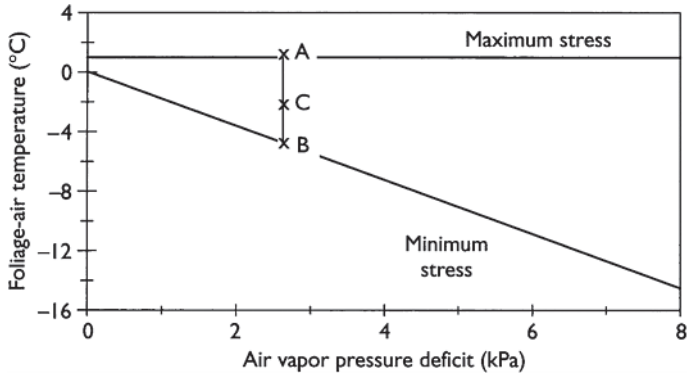


Figure 8.1 Canopy minus air temperature ($T_c - T_a$) versus VPD for well-watered and maximally stressed alfalfa based on measurements at various sites across the United States. The CWSIb is computed as the ratio of the distances CB and AB. Data from Idso *et al.* (1981).

Using the baseline slope [f_1 (°CkPa^{-1})] and intercept [f_0 (°C)] for a specific crop, equation (8.16) can be written as

$$\mathrm{CWSI_b} = \{(f_0 + f_1\mathrm{VPD}) - (T_c - T_a)_r\}/\{(f_0 - f_1\mathrm{VPD}) - (f_0 + f_1\mathrm{VPD_x})\} \tag{8.20}$$

where $(T_c - T_a)_r$ refers to on-site measurements of canopy and air temperature, $\mathrm{VPD_x} = e^*_b - e_a$, e^*_b is the saturated vapor pressure at $(T_a + f_0)$, and e_a is the vapor pressure of the air. VPD is defined by $\mathrm{VPD} = e^* - e_a$ and e^* is the saturated vapor pressure at $(T_c + T_a)/2$.

8.3.5 Water deficit index

Application of CWSI with satellite- or aircraft-based measurements of surface temperature is restricted to full-canopy conditions so that the surface temperature sensed is equal to canopy temperature. This limited the usefulness of CWSI for partial crop conditions when irrigation management decisions can be crucial. To deal with partial plant cover conditions, Moran *et al.* (1994) developed a Water Deficit Index (WDI), which combined measurements of reflectance (as expressed in equation (8.22)) with surface temperature measurements (a composite of both the soil and plant

temperatures), as expressed by

$$\mathrm{WDI} = 1-\lambda E_{\Gamma}/\lambda E_{\Gamma \mathrm{p}} = [(T_s-T_a)_m-(T_s-T_a)_r]/[(T_s-T_a)_m-(T_s-T_a)_x] \tag{8.21}$$

The WDI is operationally equivalent to the CWSI for full-cover canopies, where measurement of $T_s = T_c$. Graphically, WDI is equal to the ratio of distances AC/AB in the trapezoidal shape presented in Figure 8.2, where WDI = 0 for well-watered conditions and WDI = 1 for maximum stress conditions. That is, the left edge of the Vegetation Index/Temperature (VIT) trapezoid corresponds to $T_s - T_a$ values for surfaces evaporating at the potential rate; the right edge corresponds to $T_s - T_a$ values for surfaces in which no evaporation is occurring.

The Soil-Adjusted Vegetation Index (SAVI) is defined as

$$\mathrm{SAVI} = (\rho_{\mathrm{NIR}} - \rho_{\mathrm{red}})/(\rho_{\mathrm{NIR}} + \rho_{\mathrm{red}} + L)(1 + L) \tag{8.22}$$

where ρ_{NIR} and ρ_{red} are the near-infrared (NIR) and red reflectances, respectively, and L is a unitless constant assumed to be 0.5 for a wide variety of LAI values (Huete 1988). SAVI has been linearly related to f_c for a variety of crops; for example, for cotton $f_c = 0$ when SAVI ~ 0.1 and $f_c = 1$ when SAVI ~ 0.8.

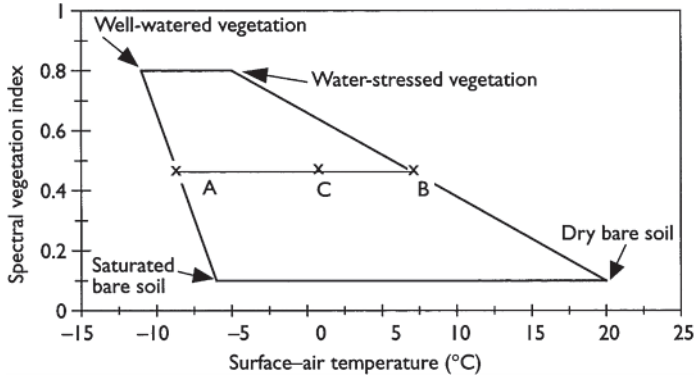


Figure 8.2 The trapezoidal shape that would result from the relation between surface minus air temperature ($T_s - T_a$) and the SAVI. With a measurement of $T_s - T_a$ at point C, the ratio of actual to potential evapotranspiration is equal to a ratio of the distances CB and AB. Data from Moran *et al.* (1994).

In practice, WDI utilizes the Penman-Monteith energy balance equation to define the four vertices of the VIT trapezoid that encompasses all possible combinations of SAVI and $T_s - T_a$ for one crop type on one day (Figure 8.2). That is, for full-cover, well-watered vegetation,

$$(T_s - T_a)_1 = [r_a(R_n - G)/\rho C_p][\gamma(1 + r_{cp}/r_a)/\{\Delta + \gamma(1 + r_{cp}/r_a)\}] - [\mathrm{VPD}/\{\Delta + \gamma(1 + r_{cp}/r_a)\}] \quad (8.23)$$

where r_{cp} is the canopy resistance at potential evapotranspiration and the subscript n of $(T_s - T_a)_n$ refers to the number of the vertex in Figure 8.2. For full-cover vegetation with no available water, represented by vertex 2 in Figure 8.2,

$$(T_s - T_a)_2 = [r_a(R_n - G)/\rho C_p][\gamma(1 + r_{cx}/r_a)/\{\Delta + \gamma(1 + r_{cx}/r_a)\}] - [\mathrm{VPD}/\{\Delta + \gamma(1 + r_{cx}/r_a)\}] \quad (8.24)$$

where r_{cx} is the canopy resistance associated with nearly complete stomatal closure. For saturated bare soil, where $r_c = 0$ (the case of a free water surface),

$$(T_s - T_a)_3 = [r_a(R_n - G)/\rho C_p][\gamma/(\Delta + \gamma)] - [\mathrm{VPD}/(\Delta + \gamma)] \quad (8.25)$$

and for dry bare soil, where $r_c = 8$ (analogous to complete stomatal closure),

$$(T_s - T_a)_4 = [r_a(R_n - G)/\rho C_p] \quad (8.26)$$

For any value of SAVI, it is possible to compute the maximum and minimum $(T_s - T_a)$, by

$$(T_s - T_a)_x = c_0 + c_1(\mathrm{SAVI}) \quad (8.27)$$

and

$$(T_s - T_a)_m = d_0 + d_1(\mathrm{SAVI}) \quad (8.28)$$

where c_0 and c_1 are the offset (°C) and slope (°C) of the line connecting points 2 and 4 in Figure 8.2; and d_0 and d_1 are the offset (°C) and slope (°C) of the line connecting points 1 and 3 in Figure 8.2. Evaluation of equations (8.27) and (8.28) using the VIT trapezoid provides an operational method for computation of WDI equation (8.21) for fields ranging from bare soil to fully vegetated.

The on-site measurements necessary to solve equations (8.23)-(8.26) and compute WDI are R_n, VPD, T_a, and u. A value of G can be estimated as a function of R_n and percent crop cover (or SAVI) (Clothier *et al.* 1986). It is also necessary to know the crop type and such characteristics as maximumpossible crop height, maximum-possible LAI, and maximum- and minimumpossible stomatal resistances (r_{sx} and r_{sp}). In many cases, these inputs are known or can be reasonably estimated. These inputs and the assumptions

associated with them are discussed in detail by Moran *et al.* (1994) and will not be addressed here.

8.3.6 Three-temperature method

Jackson's theoretical $CWSI_t$ (equation (8.14)) and the WDI are difficult to apply because they require a site-specific measure of aerodynamic resistance (r_a) which is challenging to derive for small fields of heterogeneous plant cover. Idso's baseline $CWSI_b$, (equation (8.20)) has been more successful in operational application because it does not require r_a; instead, it requires an estimate of the temperature of non-transpiring and non-water-stressed crops, which changes not only with the plant type but also with season. In a recent study, Qiu *et al.* (1996a) proposed a theoretically sound and practically simple approach that would circumvent the need for both r_a and the non-water-stressed baseline.

Qiu *et al.* (1996a) proposed a three-temperature method for detecting CWSI. Measurement of aerodynamic resistance was avoided by introducing the temperature of an imitation leaf, where the imitation leaf was made from green paper, cut in the shape of a plant leaf, and installed in the upper part of the plant canopy to avoid shading by the canopy. According to Qiu *et al.* (1996a), the temperature of the imitation leaf T_q was equivalent to the upper limiting canopy temperature, T_{cx}. Therefore, the three-temperature CWSI (termed here $CWSI_q$) was expressed as

$$\mathrm{CWSI_q} = [(T_c - T_a)_m - (T_c - T_a)_r]/[(T_c - T_a)_m - (T_q - T_a)] \quad (8.29)$$

They developed a method to estimate r_a by T_q as

$$r_a = \rho C_p (T_q - T_a)/R_{nq} \quad (8.30)$$

where R_{nq} is the net radiation of the imitation leaf. Substituting equation (8.30) into equation (8.11), assuming $G = 0$, and setting $r_c = 0$ (plants acting as a free water surface), the minimum canopy temperature (T_{cm}) was computed as

$$(T_c - T_a)_m = [R_n(T_q - T_a)/R_{nq}][\gamma/(\Delta + \gamma)] - [\mathrm{VPD}/(\Delta + \gamma)] \quad (8.31)$$

In natural conditions, where $r_{cm} \sim r_{cp}$, they suggested that

$$(T_c - T_a)_m = [R_n(T_q - T_a)/R_{nq}][\gamma'/(\Delta + \gamma')] - [\mathrm{VPD}/(\Delta + \gamma')] \quad (8.32)$$

where

$$\gamma' = \gamma\{1 + [r_{cp}R_{nq}]/[\rho C_p(T_q - T_a)]\} \quad (8.33)$$

In equations (8.31)-(8.33), values of R_n and R_{nq} were estimated by using measurements of temperature and solar radiation. The value of Δ was computed based on an approximation proposed by Jackson (1987), where

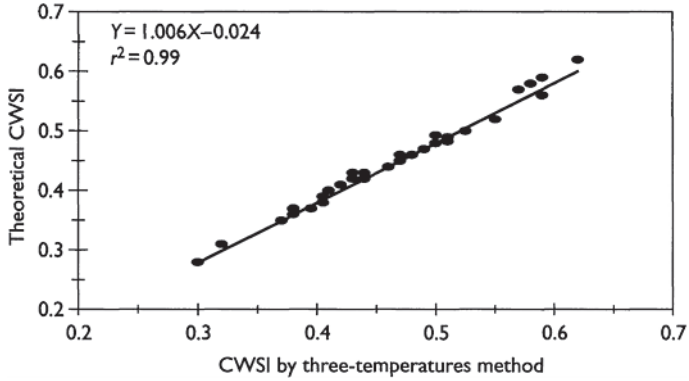


Figure 8.3 Comparison of CWSI estimated by the three-temperatures method ($CWSI_q$) and $CWSI_t$ estimated by Jackson's (Jackson 1982) theoretical method for sorghum. Data from Qiu *et al.* (1996b).

$$\Delta = \left(45.03 + 3.014\,T_v + 0.05345\,T_v^2 + 0.00224\,T_v^3\right) 10^{-2} \qquad (8.34)$$

and $T_v = (T_c + T_a)/2(^oC)$.

Thus, $CWSI_q$ was determined from three temperatures (T_c, T_q, and T_a), solar radiation, and VPD; the r_a value was eliminated and the complexity of application was reduced. Qiu *et al.* (1996b) demonstrated this approach in studies of grain sorghum in which T_c, T_q, and T_a were measured with thermocouples. They reported that the $CWSI_q$ agreed well with Jackson's $CWSI_t$ and both were correlated with the soil water status (Figure 8.3).

8.3.7 Thermal kinetic window

Quite distinct from the CWSI, Burke *et al.* (1988) developed a concept of thermal stress in plants which linked the biochemical characteristics of a plant with its optimal leaf temperature range. The thermal kinetic window (TKW) is the range of temperatures within which the plant maintains optimal metabolism. The TKW is the temperature range encompassing 200% of the observed minimum apparent Michaelis constant (K_m), and thus the optimal enzyme function, for a plant species. Plant temperatures either above or below the TKW result in stress that limits growth and yield. For example, the TKW for cotton growth is 23.5–32°C, with an optimum temperature of 28°C, and biomass production is directly related to the amount of time that canopy temperatures are within the TKW (Figure 8.4), provided insolation, soil moisture, and nutrients are non-limiting. Maintenance of plant temperatures within the TKW is controlled primarily by transpirational cooling during

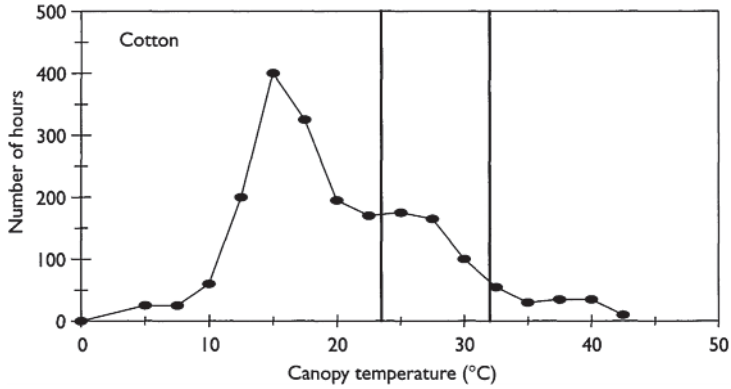


Figure 8.4 Seasonal canopy temperatures of cotton, where the vertical lines represent the temperature range that comprises the species-specific TKW as determined from the changes in the apparent K_m with temperature. Data from Burke et al. (1988).

Table 8.2 Summary of thermal kinetic windows and optimum canopy temperatures for a variety of crops and garden plants

Crop	*T_c range* (°C)	*T_c optimum* (°C)	*Reference*
Cotton	23.5–32.0	27.5	Burke *et al.* (1988); Burke and Upchurch (1995)
Wheat	17.5–23.0	20.0	Hatfield *et al.* (1987); Burke *et al.* (1988)
Cucumber	23.5–39.0	35.0	Burke (1990); Hatfield and Burke (1991)
Bell pepper	23.0–41.0	32.0	Hatfield and Burke (1991)
Tomato	20.5–24.5	22.5	Burke (1990)
Petunia	18.0–28.0	25.0	Burke (1990)
Potatoes	15.0–25.0	20.0	Ferguson and Burke (1991)
Soybean	15.0–30.0	25.0	Ferguson and Burke (1991)

the day, which is directly dependent upon the available soil water stored in the soil volume.

The TKWs have been identified for several crop and forage species (see references and temperature ranges in Table 8.2). Until recently, the TKW was determined by a rather tedious laboratory process in which the K_m values were measured in multiple assays of glyoxylate reductase from crop leaves at incremented temperatures (Burke *et al.* 1988). Burke (1990) and Fergusen and Burke (1991) presented an alternative approach based on the temperature dependence of reappearance of fluorescence following plant illumination. The latter approach may offer a procedure for rapid identification of a plant's temperature optimum, and may greatly expand the usefulness of the TKW approach.

In practice, the TKW provides a biological indicator of plant health that could potentially be used for irrigation management. In a patent application (SN 261510; filed June 17, 1994), the inventors, Drs Upchurch, Wanjura, Burke, and Mahan, describe an irrigation method in which the canopy temperature (T_c) of a target plant is repeatedly measured with an infrared thermometer at a regular time interval (t_i), and T_c is compared with the optimum plant temperature ($T_{c'}$). If $T_c \leq T_{c'}$ or if the VPD is restrictive to plant cooling, then irrigation is not indicated. If both $T_c > T_{c'}$ and VPD is not restrictive to plant cooling, then t_i is added to a time register. When the accumulated time in the time register exceeds the time threshold, transpirational cooling to reach the optimum is feasible, and a signal to the irrigation unit is generated.

8.3.8 Data fusion and model assimilation

This section would not be complete without some mention of the current trend toward fusion of information from many spectral bands into a single index and assimilation of TIR information into plant growth models. Though an in-depth discussion is beyond the scope of this chapter, a general discussion of the state-of-the-art is given here and references are provided for further investigation.

All the indices described in previous subsections rely on ancillary information to interpret the integrated soil-vegetation-atmosphere information in the TIR measurement. Menenti and Choudhury (1993) reviewed the correlations between surface temperature and other spectral measurements, and presented the implications for land surface studies. They suggested that z_o could be measured using laser sensors (Menenti and Ritchie 1992); G can be derived as a function of R_n, canopy LAI, and soil moisture using spectral reflectances and microwave observations (Clothier *et al.* 1986; Choudhury 1991); and soil moisture content could be estimated using microwave emissivity measurements (Schmugge 1985). In addition to these suggestions, there is evidence that incoming solar radiation, photosynthetically active radiation (PAR), and net shortwave radiation can be estimated from spectral reflectance in the visible and NIR spectral bands (Jackson 1984; Pinker and Ewing 1985; Frouin and Pinker 1995); f_c, and possibly E, can be measured with multi-frequency synthetic aperture radar (SAR) (Moran *et al.* 1997); and fractional absorbed PAR (*f*APAR) can be determined with spectral vegetation indices (Pinter 1993). There is also potential that such meteorological variables as rainfall, T_a, u, and VPD can be derived from satellite spectral images (e.g. Petty 1995; Toth *et al.* 1996). These are only a few of the studies that offer potential for spectral indices based on *data fusion,* a term for which the multispectral approach of the WDI is a good example.

The coupling of remote sensing (RS) and plant production models is a logical means for interpreting the above-mentioned information on atmospheric conditions, soil moisture, and crop activity derived from multispectral data. Plant production models provide a description of system status and an integration of biophysical processes at a time step shorter than the temporal frequency of RS information. Generally, such combined RS/modeling approaches utilize TIR measurements to determine the water status of plants and soils. For example, Moulin *et al.* (1998) suggested that TIR measurements would be useful in modeling the common model parameter, e_h, which is the efficiency of conversion of energy into dry matter (total or aboveground biomass). Maas *et al.* (1989) and Moran *et al.* (1995) incorporated measurements of T_s and T_a into a simple crop growth/soil moisture model to determine crop development and yield. Similarly, Guérif *et al.* (1993) derived a stress index from the TIR channel of NOAA-AVHRR satellite-based sensor to reduce daily carbon assimilation in a crop production model. Moulin *et al.* (1998) summarized the trend in modeling by noting that visible, NIR, and shortwave infrared information is generally used to retrieve some plant characteristics, whereas TIR information has been used to adjust some soil characteristics.

The state-of-the-art in RS/modeling approaches is exemplified by the Spatially Explicit Hydro-Ecological Model (SEHEM) developed by Nouvellon

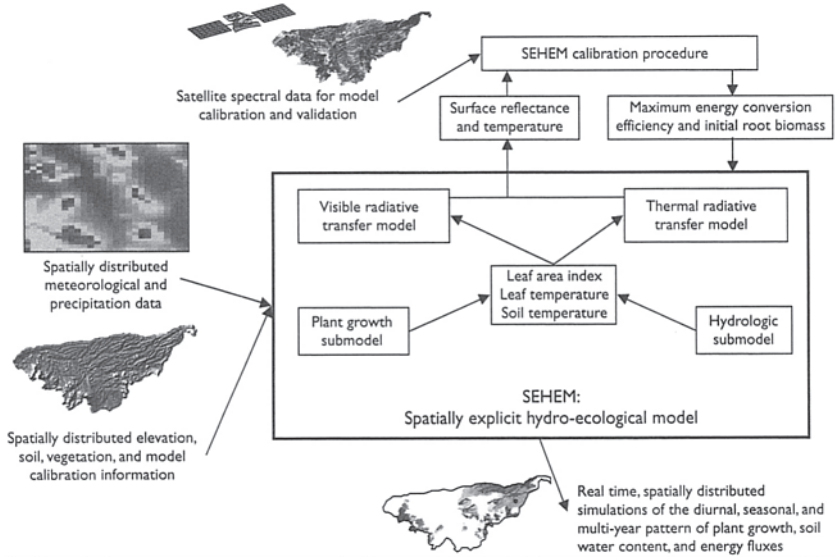


Figure 8.5 Diagrammatic representation of the SEHEM showing the inputs of surface and meteorological information.the SEHEM calibration procedure based on satellite spectral data, and the potential outputs of plant growth, soil water content, and energy fluxes. The images represent the USDAARSWalnut Gulch Experimental Watershed (WGEW) in Southeast Arizona (see Colour Plate XXVII).

et al. (2000) for grassland ecosystems. SEHEM is composed of a daily time-step plant growth model and an hourly time-step hydrological model, combined with visible and thermal radiative transfer models (illustrated in Figure 8.5). SEHEM requires inputs of spatially distributed information about soils, elevation, and vegetation, and is driven by hourly meteorological and precipitation data. By combining the plant growth, hydrological, and radiative transfer models, SEHEM can simulate such outputs as latent and sensible heat fluxes, soil water content, LAI, biomass of green and dead vegetation and roots, leaf and soil temperatures, surface temperature (T_s), and spectral reflectance (ρ_{red} and ρ_{NIR}). An iterative model calibration procedure is initiated to minimize the difference between modelsimulated T_s, ρ_{red}, and ρ_{NIR} and that measured by satellite-based sensors through adjustment of one-model parameter (maximum energy conversion efficiency) and one-model initial condition (root biomass). Once calibrated, the model becomes an accurate and powerful tool that can leverage intermittent TIR images (reportedly requiring only two to three images per year) to produce daily and hourly images of plant ecosystem health indicators. Using a 10-year time series of Landsat TM images over a semi-arid grassland watershed in Arizona, SEHEM has successfully produced multi-year simulations of plant growth, water and carbon budget (Figure 8.6), and hourly simulations of latent heat flux density (Figure 8.7) (Nouvellon *et al.* 2001). These results are just one example of the great potential of the RS/modeling approaches that are currently under development for multiple vegetation types and geographic locations.

8.4 Concluding remarks

Section 8.3 presented descriptions of eight TIR spectral indices for evaluation of plant ecosystem health. Three simple algorithms, SDD, CTV, and TSD, were the inspiration for subsequent, more complex stress indices such as $CWSI_t$, $CWSI_b$, and TKW, which are well documented and commercially accepted. Two relatively new approaches, WDI and $CWSI_q$, have potential for operational application, though further study and validation are warranted. One RS/modeling approach (SEHEM) was also described as an example of the potential use of TIR measurements as a source of model input, calibration, and validation.

Regardless of which index or model is chosen for application, the accuracy of the approach will depend largely on the proper implementation of the algorithms and the accurate measurement of the input variables. As noted by Norman *et al.* (1995), the limited success associated with the use of TIR remote sensing arises from the numerous variables exerting a strong effect on the relation between thermal emittance and surface conditions. As shown here, these variables include R_n, u, T_a, VPD, and the structure (leaf area, height, texture, etc.) and properties (albedo, emissivity, r_s, roughness, moisture)

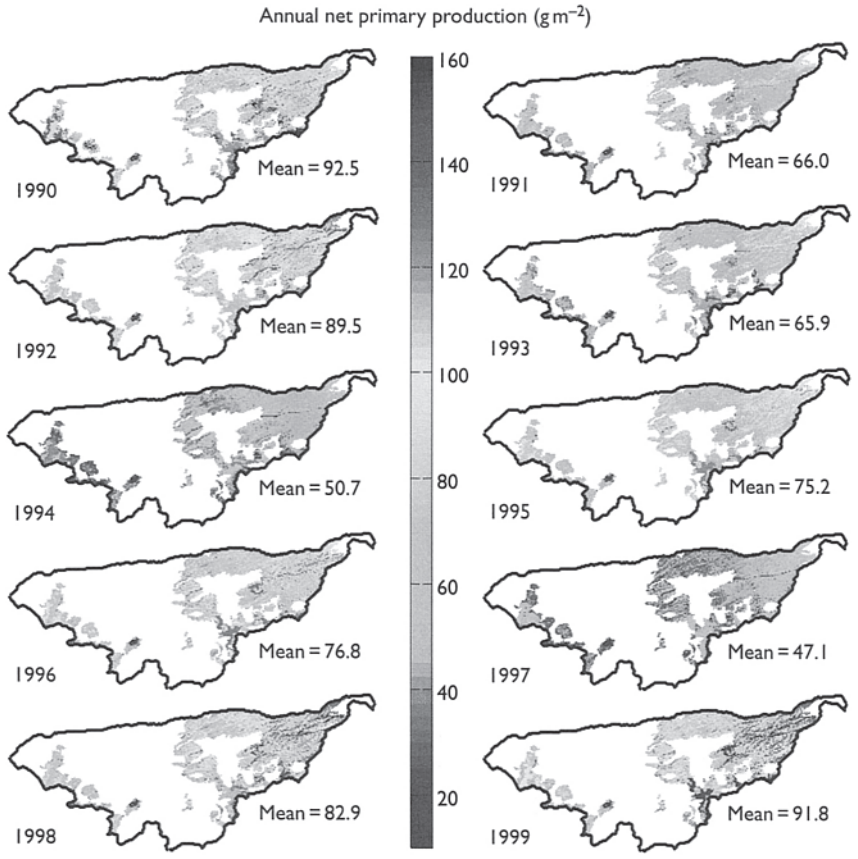


Figure 8.6 SEHEM simulations of annual net primary production over a 10-year period for grassland regions with the Walnut Gulch Experimental Watershed. Output validation was presented by Nouvellon et *al.* (2001) (see Colour Plate XXVIII).

of the vegetation and soil. Indices based on measurements of T_c rather than T_s require particular care with IR thermometry to avoid the complication of the soil background. Furthermore, with the use of handheld, airborne, and space-based IRTs, it is important to consider the sun/sensor/viewing geometry (e.g. Lagouarde *et al.* 1995) and the effects of the intervening atmosphere on temperature measurements (e.g. Perry and Moran 1994).

Despite these cautions, the future of IR thermometry as an indicator of plant ecosystem health is promising. TIR information fills a niche that has yet to be challenged with information from any other region in the spectrum. That is, remote sensing in the thermal spectrum is a proven means of detecting both early *and* late stages of plant water stress. Furthermore, plant

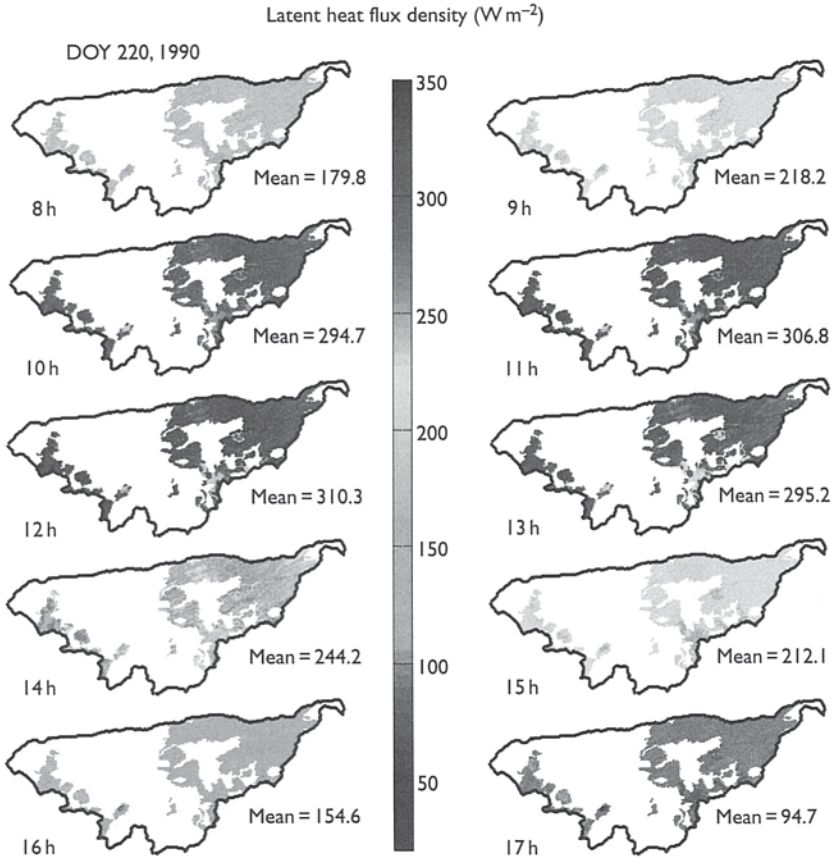


Figure 8.7 SEHEM simulations of hourly latent heat flux density for grassland regions within the Walnut Gulch Experimental Watershed. Output validation was presented by Nouvellon *et al.* (2001) (see Colour Plate XXIX).

temperature has been linked *directly* with plant water stress; whereas, surface reflectance and SAR backscatter have been associated with crop water stress indirectly, and in most cases, empirically. Considering the unique role of TIR and the importance of monitoring biosystem health, research and development should be continued, and serious consideration should be given to including high-resolution thermal sensors aboard upcoming satellites.

References

Allen, R.G. (1986) A Penman for all seasons. *J. Irrig. Drain. Eng.* 112:348–68.
Asari, A.Q. and W.E.Lommis (1959) Leaf temperatures. *Am. J. Bot.* 46:713–17.

Beljaars, A.C.M. and A.A.M.Holtslag (1991) Flux parameterization over land surfaces for atmospheric models. *J. Appl. Meteor.* 30:327–41.
Bowen, I.S. (1926) The ratio of heat losses by conduction and by evaporation from any water surface. *Phys. Rev.* 27:779–89.
Brutsaert, W.H. (1982) *Evaporation into the Atmosphere.* D.Reidel, London, 299 pp.
Burke, J.J. (1990) Variation among species in the temperature dependence of the reappearance of variable fluorescence following illumination. *Plant. Physiol.* 93: 652–6.
Burke, J.J., J.R.Mahan, and J.L.Hatfield (1988) Crop-specific thermal kinetic windows in relation to wheat and cotton biomass production. *Agron. J.* 80:553–6.
Burke, J.J. and D.R.Upchurch (1995) Cotton rooting patterns in relation to soil temperatures and the thermal kinetic window. *Agron. J.* 87:1210–16.
Clawson, K.L. and B.L.Blad (1982) Infrared thermometry for scheduling irrigation of corn. *Agron. J.* 74:311–16.
Choudhury, B.J. (1991) Multispectral satellite data in the context of land surface heat balance. *Rev. Geophys.* 29:217–36.
Clawson, K.L., R.D.Jackson, and P.J.Pinter Jr. (1989) Evaluating plant water stress with canopy temperature differences. *Agron.J.* 81:858–63.
Clothier, B.E., K.L.Clawson, P.J.Pinter Jr., M.S.Moran, R.J.Reginato, and R.D. Jackson (1986) Estimation of soil heat flux from net radiation during the growth of alfalfa. *Agric. For. Meteorol.* 37:319–29.
Curtis, O.F. (1936) Leaf temperatures and the cooling of leaves by radiation. *Plant. Physiol.* 11:343–64.
Curtis, O.F. (1938) Wallace and Clum, "Leaf temperatures": a critical analysis with additional data. *Am. J. Bot. 25:*761–71.
Eaton, F.M. and G.O.Belden (1929) Leaf temperatures of cotton and their relation to transpiration, varietal differences and yields. *US Dept. Agric. Tech. Bull* 91.
Ferguson, D.L. and J.J.Burke (1991) Influence of water and temperature stress on the temperature dependence of the reappearance of variable fluorescence following illumination. *Plant Physiol.* 97:188–92.
Frouin, R. and R.T.Pinker (1995) Estimating photosynthetically active radiation (PAR) at the earth's surface from satellite observations. *Remote Sens. Environ.* 51:98–107.
Fuchs, M. and C.B.Tanner (1966) Infrared thermometry of vegetation. *Agron.J.* 58:597–601.
Gardner, B.R., B.L.Blad, and D.G.Watts (1981a) Relationships between crop temperature, grain yield, evapotranspiration and phenological development in two hybrids of moisture stressed sorghum. *Irrig. Sci.* 2:213–24.
Gardner, B.R., B.L.Blad, R.E.Maurer, and D.G.Watts (1981b) Relationship between crop temperature and the physiological and phenological development of differentially irrigated corn. *Agron. J.* 73:743–7.
Gates, D.M. (1964) Leaf temperature and transpiration. *Agron. J.* 56:273–7.
Gay, L.W. and R.J.Greenberg (1985) The AZET battery-powered Bowen ratio system. In *Proc. 17th Conf. on Agric. and Forest. Meteorol.,* Scottsdale, AZ, Amer. Meteorol. Soc., Boston, MA, pp. 181–2.
Guérif, M., S.De Brisis, and B.Sequin (1993) Combined NOAA-AVHRR and SPOT-HRV data for assessing crop yields of semiarid environments. *EARSeL Adv. Remote Sens.* 2:110–23.
Hatfield, J.L. (1997) Plant-water interactions, chapter 3. *In Plants for Environmental Studies.* CRC Lewis Publishers, New York, pp. 81–103.
Hatfield, J.L. and J.J.Burke (1991) Energy exchange and leaf temperature behavior

of three plant species. *Env. Exp. Bot.* 31: 295–302.
Hatfield, J.L., J.J.Burke, J.R.Mahan, and D.F.Wanjura (1987) Foliage temperature measurements: a link between the biological and physical environment. In *Proc. Int. Conf. on Soil and Plant Water Status Utah State Univ.* July 6–10, Logan, UT, pp. 99–102.
Hiler, E.A. and R.N.Clark (1971) Stress day index to characterize effects of water stress on crop yields. *Trans. ASAE* 14: 757–61.
Huete, A.R. (1988) A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 27:47–57.
Idso, S.B. (1982) Non-water-stressed baselines: a key to measuring and interpreting plant water stress. *Agric. Meteorol.* 27:59–70.
Idso, S.B. and D.G.Baker (1967) Relative importance of reradiation, convection and transpiration in heat transfer from plants. *Plant Physiol.* 42:631–40.
Idso, S.B., R.D.Jackson, and R.J.Reginato (1977) Remote sensing of crop yields. *Science* 196:19–25.
Idso, S.B., R.D.Jackson, P.J.Pinter Jr., R.J.Reginato, and J.L.Hatfield (1981) Normalizing the stress-degree-day parameter for environmental variability. *Agric. Meteorol.* 24:45–55.
Jackson, R.D. (1982) Canopy temperature and crop water stress. In *Advances in Irrig.* 1:43–85.
Jackson, R.D. (1984) Total reflected solar radiation calculated from multiband sensor data. *Agric. For. Meteorol.* 33:162–74.
Jackson, R.D. (1987) The Crop Water Stress Index: a second look. In *Proc. Int. Conf. on Meas. of Soil and Plant Water Stress,* Utah State Univ., July 6–10, Logan, UT, pp. 87–92.
Jackson, R.D., R.J.Reginato, and S.B.Idso (1977) Wheat canopy temperature: a practical tool for evaluating water requirements. *Water Resour. Res.* 13:651–6.
Jackson, R.D., D.B.Idso, R.J.Reginato, and P.J.Pinter Jr. (1981) Canopy temperature as a crop water stress indicator. *Water Resour. Res.* 17:1133–8.
Kanemasu, E.T., M.L.Wesely, B.B.Hicks, and J.L.Heilman (1979) Techniques for calculating energy and mass fluxes. In B.J.Barfield and J.F.Gerber (eds) *Modification of the Aerial Environment of Crops.* Am. Soc. Agric. Eng., St Joseph, MI, pp. 156–82.
Kustas, W.P., B.J.Choudhury, Y.Inoue, P.J.Pinter Jr., M.S.Moran, R.D.Jackson, and R.J.Reginato (1990) Ground and aircraft infrared observations over a partially-vegetated area. *Int. J. Remote Sens.* 11:409–27.
Kustas, W.P. and J.M.Norman (1996) Use of remote sensing for evapotranspiration monitoring over land surfaces. *Hydrol. Sci. J.* 41:495–516.
Lagouarde, J.P., Y.H.Kerr, and Y.Brunet (1995) An experimental study of angular effects on surface temperature for various plant canopies and bare soils. *Agric. For. Meteorol.* 77:167–90.
Maas, S.J., R.D.Jackson, S.B.Idso, P.J.Pinter Jr., and R.J.Reginato (1989) Incorporation of remotely-sensed indicators of water stress in a crop growth simulation model. In *Proc.* 19th *Conf. on Agr. For. Meteorol.* March 7–10, Charleston, SC, AMS Boston, pp. 228–31.
Menenti, M. and B.J.Choudhury (1993) Synergistic use of thermal infrared and other spectral measurements for land surface studies. *In Proc. Thermal Remote Sensing Workshop,* La Londe les Maures, France, September 20–23, CEMAGREFENGREF, Montpellier, France, pp. 167–76.
Menenti, M. and J.C.Ritchie (1992) Estimation of effective aerodynamic roughness with altimeter measurements. *Proc. IGARSS'92,* IEEE, 1508–10.

Miller, E.C. and A.R.Saunders (1923) Some observations on the temperature of the leaves of crop plants. *J. Agric. Res.* 26:15–43.
Monteith, J.L. (1965) Evaporation and environment. *Symp. Soc. Exp. Biol.* XIX: 205–34.
Monteith, J.L. (1973) *Principles of Environmental Physics*. Arnold Press, London.
Monteith, J.L. (1981) Evaporation and surface temperature. *Q.J.R.Meteorol. Soc.* 107:1–27.
Monteith, J.L. and G.Szeicz (1962) Radiative temperature in the heat balance of natural surfaces. *Q.J.R.Meteorol.* 88: 496–507.
Moran, M.S., T.R.Clarke, Y.Inoue, and A.Vidal (1994) Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sens. Environ.* 49:246–63.
Moran, M.S., S.J.Maas, and P.J.Pinter Jr. (1995) Combining remote sensing and modeling for estimating surface evaporation and biomass production. *Remote Sens. Rev.* 12:335–53.
Moran, M.S., A.Vidal, D.Troufleau, Y.Inoue, and T.Mitchell (1997) Ku- and C-band SAR for discriminating agricultural crop and soil conditions. *IEEE Geosci. Remote Sens.* 36:265–72.
Moulin, S., A.Bondeau, and R.Delécolle (1998) Combining agricultural crop models and satellite observations: from field to regional scales. *Int. J. Remote Sens.* 19: 1021–36.
National Research Council (1994) *Rangeland Health*. National Academic Press, Washington DC, 180 pp.
Norman, J.M., M.Divakarla, and N.S.Goel (1995) Algorithms for extracting information from remote thermal-IR observations of the earth's surface. *Remote Sens. Environ.* 51:157–68.
Nouvellon, Y., S.Rambal, D.Lo Seen, M.S.Moran, J.P.Lhomme, A.Bégué, A.G. Chehbouni and Y.Kerr (2000) Modelling of daily fluxes of water and carbon from shortgrass steppes. *Agric For. Meteorol.* 100:137–53.
Nouvellon, Y., M.S.Moran, D.Lo Seen, R.B.Bryant, W.Ni, A.Bégué, A.G. Chehbouni, W.E.Emmerich, P.Heilman, and J.Qi (2001) Combining a grassland ecosystem model with Landsat TM imagery for a ten-year simulation of carbon and water budget. *Remote Sens. Environ.* 78:131–49.
Penman, H.L. (1948) Natural evaporation from open water, bare soil and grass. *Proc. R. Soc. Lond. A* 193:120–46.
Penman, H.L. (1953) The physical bases of irrigation control. In *Rep. 13th Int. Hort. Cong.* Vol. 2:913–23.
Perry, E.M. and M.S.Moran (1994) An evaluation of atmospheric corrections of radiometric surface temperatures for a semiarid rangeland watershed. *Water Resour. Res.* 30:1261–9.
Petty, G.W. (1995) The status of satellite-based rainfall estimation over land. *Remote Sens. Environ.* 51:125–37.
Pinker, R.T. and J.A.Ewing (1985) Modeling surface solar radiation: model formulation and validation. *J. Clim. Appl. Meteorol.* 24:389–401.
Pinter, Jr., P.J. (1993) Solar angle independence in the relationship between absorbed PAR and remotely sensed data for alfalfa. *Rem. Sens. Env.* 46:19–25.
Priestley, C.H.B. and Taylor, R.J. (1972) On the assessment of surface flux and evaporation using large-scale parameters. *Mon. Weather Rev.* 100:81–92.
Qiu, G.Y., K.Momii, and T.Yano (1996a) Estimation of plant transpiration by imitation leaf temperature I. Theoretical consideration and field verification. *Tran. Jpn. Soc. Irrig. Drainage Reclamation Eng.* 183:47–56.

Qiu, G.Y., K.Momii, and T.Yano (1996b) Estimation of plant transpiration by imitation leaf temperature II. Application of imitation leaf temperature for detection of crop water stress. *Tran. Jpn. Soc. Irrig. Drainage Reclamation Eng.* 185:43–9.

Reicosky, D.C. (1981) A research tool for evapotranspiration measurements for model validation and irrigation scheduling. In *Irrigation Scheduling for Water and Energy Conservation in the 80's, Proc. of the Am. Soc. of Agric. Engineers Irrig. Scheduling Conf.,* December, Chicago, II, pp. 18–26.

Rijtema, P.R. (1965) An analysis of actual evapotranspiration. *Agric. Res. Rep.* 659: 1–107.

Schmugge, T.J. (1985) Remote sensing of soil moisture. In M.G.Anderson and T.P. Burt (eds) *Hydrologic Forecasting,* John Wiley, New York, pp. 101–24.

Spittlehouse, D.L. and T.A.Black (1980) Evaluation of the Bowen ratio/energy balance method for determining forest evapotranspiration. *Atmosphere–Ocean* 18: 98–116.

Toth, J.J., D.C.Goodrich, M.S.Moran, R.Avissar, and A.F.Rahman (1996) Atmospheric modeling over southeast Arizona and northern Mexico as an input for field observation planning, *AGU Spring Meeting,* Baltimore, MD, May 20–24, p. 116.

van Bavel, C.H.M. and L.E.Myers (1962) An automatic weighing lysimeter. *Agric. Eng.* 43:580–3 (see also pp. 586–8).

Vidal, A. and C.Devaux-Ros (1995) Evaluating forest fire hazard with a Landsat TM derived water stress index. *Agric. For. Meteorol.* 77:207–24.

Wallace, R.H. and H.H.Clum (1938) Leaf temperatures. *Am. J. Bot.* 25:83–97.

Waggoner, P.E. and R.H.Shaw (1952) Temperature of potato and tomato leaves. *Plant. Physiol.* 27:710–24.

Wanjura, D.F. and D.R.Upchurch (1998) Evaluation of analytical methods for daily estimation of crop water stress. In *ASAE Ann. Intl. Meeting,* Orlando, FL, July 12–16, paper No. 982122, 2950 Niles Rd., St Joseph, MI 49085–9659, USA.

Wiegand, C.L. and L.N.Namken (1966) Influences of plant moisture stress, solar radiation and air temperature on cotton leaf temperature. *Agron. J.* 58:552–6.

Chapter 9

Exergy analysis of ecosystems
Establishing a role for thermal remote sensing

Roydon A.Fraser and James J.Kay

9.1 Introduction

Ecosystems are complex thermodynamic systems that evolve in time. Thermodynamics is the study of energy. Energy is characterized by *magnitude, form,* and *quality.* While the concept of energy magnitude (e.g. calorie, joule, watt, horsepower) and energy form (e.g. kinetic energy, potential energy, chemical energy, heat transfer, work transfer) are introduced in elementary or high school, few are familiar with the concept of energy quality, especially its quantification. Energy quality measures the capacity of energy, in its various forms, to do useful work. Interestingly, it is the quality of energy that provides an explanation for the continued existence of life on earth (Edgerton 1982; Kay 1984; Schneider and Kay 1994), and hence the existence of ecosystems. That is, the quality aspect of energy makes it possible to obtain and maintain organization in the form of life from a soup of disordered basic atomic elements. Of more immediate interest, the study of energy quality has the potential to provide a quantitative method to characterize the status, maturity, or stage of development of ecosystems, and to provide fundamental physical explanations, at least, in part, as to survival strategies and structures employed within ecosystems as they evolve.

The objective of this chapter is to establish the theoretical foundations required to quantitatively apply the energy quality concept to the study of ecosystems. The pseudo-property[1] of *maximum useful*[2] *to-the-dead-state*[3] work, commonly referred to by the specialized name *exergy,*[4] will be the tool employed to quantify the quality aspect of energy.[5]

In the process of establishing the foundations of the exergy concept, and consistent with the thermal remote sensing theme of this book, a role for ecosystem surface temperature measurements is identified (see section on "A possible role for ecosystem surface temperature measurements" and Sections 9.4.2 and 9.5). Although surface temperatures are necessary for determining the overall ecosystem exergy flows, they are not sufficient (see section on "A First look at the role of surface temperature"). Hence, it is beyond the scope of this chapter to provide complete calculation procedures for conducting an ecosystem exergy analysis.

As an introduction to the paradigm of energy quality, engineering applications are used to provide insight (Section 9.2). That is, since the application of an energy quality paradigm as applied to ecosystems is still in its infancy, advantage can be taken of insights gained from more mature, or at least less complex, applications, particularly in engineering. These engineering applications will lead to the conclusion that exergy is a measure of energy quality. Section 9.3 then formalizes the physics and mathematical foundations for quantifying energy quality, thus providing the theory necessary to admit quantification of an ecosystem's exergy content and flows. At this point, exergy destruction will be seen to be intimately linked with entropy production through the Gouy–Stodola theorem.[6] To strengthen the link between ecosystems and the exergy concept attention will focus on the dominant energy input to terrestrial ecosystems, solar energy (see section on "A first look at the role of surface temperature" and Section 9.4). An outcome of the solar exergy discussion is that surface temperature, and hence, remote sensing thermal imaging, may play a key role in characterizing an ecosystem's status, maturity, or stage of development (Section 9.5). This potential ability to characterize the state of an ecosystem will not be proven. It cannot at present. More data and analysis are needed. However, it is currently possible to establish the underlying physics of energy quality that may admit such ecosystem characterization in the future.

In essence, this chapter provides a detailed introduction to the exergy concept for those wishing to conduct exergy analyses of ecosystems, and in the process identifies thermal remote sensing as a necessary exergy analysis tool. This chapter does not provide a guide to performing a complete ecosystem exergy analysis, such efforts are for future work.

Please note that in this chapter the authors explicitly identify for the first time four new exergy classifications: *intrinsic exergy, transport exergy, restricted exergy,* and *accessible exergy.* Clarity of communication is critical. Distinguishing between these four classes of exergy will hopefully aid the exergy analyst in appreciating implicit assumptions behind a given exergy calculation.

9.2 The quality of energy paradigm

"The *first law* [*of thermodynamics*] deals with the quantity of energy in terms of a conservation rule. The *second law* [*of thermodynamics*] deals with the quality of energy. It is essentially a nonconservation rule" (Wark 1977). More precisely, the Second Law of Thermodynamics, in combination with the First Law of Thermodynamics and the Conservation of Mass, provide the rationale for defining, and the means for quantifying, energy quality. To speak of the quality of energy is to recognize that some forms of energy are more useful than others.

Before formalizing the concept of energy quality in Section 9.3, two outwardly simple, inwardly insightful, examples are given. These examples exploit the paradigm that energy is characterized not only by quantity, but also by quality. They are taken from the realm of engineering thermodynamics where the study of energy quality is reasonably well established and clear.

9.2.1 Engineering examples that use the quality of energy paradigm

The intent of the following two examples is to provide an incentive to the reader to learn more about the energy quality paradigm, to highlight the importance of temperature in thermodynamic system characterization, and to provide the foundations for the hypothesis that an ecosystem's surface temperature can provide a measure to quantitatively characterize an ecosystem's status, maturity, or stage of development.

Example 1: How good is the furnace in your home?

Consider the natural-gas furnace shown in Figure 9.1. The maximum combustion temperature, at constant pressure, that natural gas can attain is its adiabatic flame temperature ($T_H = T_{\text{Adiabatic Flame}} = T_{\text{Combustion}} \approx 2{,}000°C$) (Glassman 1987). The room temperature, T_R, is assumed to be constant at 20°C in this example while the outdoor environment temperature,[7] T_0, is assumed to be constant at 0°C. Heat transfer from the combustion gases to room air occurs across a heat exchanger, that is, $\dot{Q}_{\text{Combustion}}$ to $\dot{Q}_{\text{Room}}$.[8] This

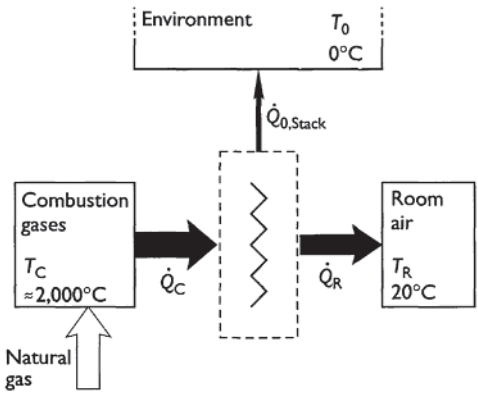


Figure 9.1 Schematic of a natural-gas home furnace.

heat exchanger is simply a sheet of metal that separates the combustion gases from the room air. Finally, a furnace must exhaust its combustion products (e.g. water, carbon dioxide, carbon monoxide); the energy lost to the environment via these combustion products is accounted for by stack losses, $\dot{Q}_{\text{Stack}}$, which contribute to the furnace's inefficiencies.

The performance of the home furnace shown in Figure 9.1 is quantified by its efficiency, η, which is defined and quantified as follows:

$$\dot{Q}\eta = \frac{\text{Benefit}}{\text{Cost}} = \frac{\dot{Q}_{\text{Room}}}{\dot{Q}_{\text{Combustion}}} = 85\% \quad (9.1)$$

An 85% efficient furnace is routinely referred to as a mid-efficiency furnace (Carson *et al.* 2000). High efficiency furnaces can achieve efficiencies of around 95%[9] (Lennox 2000) through the ingenious use of an additional heat exchanger in the stack that captures much of the stack losses.

Now imagine how you would respond to a salesperson who tried to sell you a revolutionary type of furnace with a claimed efficiency of 120%. Would you be suspicious? Hopefully yes given that a central expectation for an efficiency is that it be restricted to be less than or equal to 100%. For example, for the furnace system shown in Figure 9.1, conservation of energy[10,11] tells us that

$$\dot{Q}_{\text{Combustion}} = \dot{Q}_{\text{Room}} + \dot{Q}_{\text{Stack}} \quad (9.2)$$

or

$$\dot{Q}_{\text{Room}} \leq \dot{Q}_{\text{Combustion}} \quad (9.3)$$

hence, as expected, η has an upper bound of 100%, that is,

$$\eta = \frac{\dot{Q}_{\text{Room}}}{\dot{Q}_{\text{Combustion}}} \leq 100\% \quad (9.4)$$

Notice that as far as the calculation of a home furnace's efficiency is concerned its internal workings are irrelevant. That is, the exact design of the heat exchanger is of no concern. For example, is it a co-flow or a counter-flow heat exchanger,[12] or what is the heat exchanger's geometry? Answer: it does not matter. Only the energy flows across the furnace's system boundaries are needed to calculate its efficiency. This is not to say that the internal workings of the furnace are not important, they are if, for example, one wished to change the relative magnitudes of energy flows across system boundaries to improve efficiency, or if one wished to find a better (e.g. cheaper, more reliable) system that can maintain the same efficiency as a current system. The point is, *the internal workings of a thermodynamic system are irrelevant as far as*

calculating an efficiency is concerned, but not necessarily irrelevant with respect to how to optimize that system.

Now consider the much more complicated, more expensive, less conventional, exergy-conserving,[13] furnace shown in Figure 9.2. Concluding that the system shown in Figure 9.2 is still a furnace is based simply on its function (i.e. benefit) of providing room heating. For the home furnace shown in Figure 9.1, $\dot{Q}_{0,\text{Net}} = \dot{Q}_{0,\text{Stack}}$ while for the exergy-conserving furnace shown in Figure 9.2, $_{0,\text{Net}} = \dot{Q}_{0,\text{Exhaust}} - \dot{Q}_{0,\text{IN}}$.[14]

The *heat engine* and *heat pump* shown in Figure 9.2 are generic devices that convert thermal energy into work, and use work energy to pump thermal energy from cold to hot,[15] respectively. For sake of visualization, imagine the heat engine to be an internal combustion engine and the heat pump to be a refrigerator.[16] The internal combustion engine provides the work transfer, that runs the refrigerator. In turn, the heat pump cools (i.e. refrigerates) the environment while rejecting thermal energy to the room (just as the coils on the back of a refrigerator do).

The advantage of the exergy-conserving furnace shown in Figure 9.2 can be seen by answering the following question:

Question 1: For a fixed amount of fuel input (i.e. fixed $\dot{Q}_{\text{Combustion}}$) to the exergy-conserving furnace, what is the benefit received in the form of room heating (i.e. $\dot{Q}_{\text{Room}}$)?

First, the efficiency of a good diesel engine, η_{Diesel}, is greater than 40% (Heywood 1988), hence,

$$\dot{W} = \dot{W}_{\text{Diesel}} \geq 0.4\dot{Q}_{\text{Combustion}} \tag{9.5}$$

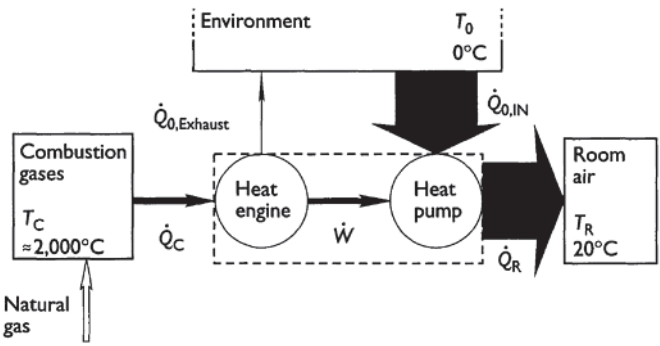


Figure 9.2 Schematic of an exergy-conserving natural-gas home furnace.

where $\dot{W}_{Diesel}$ is the work output of the diesel engine and $\dot{Q}_{Combustion}$ is the diesel fuel's heat of combustion.

Second, the coefficient of performance of a good heat pump, $COP_{Heat\ Pump}$, operating between 0 and 20°C can be greater than three (Reynolds and Perkins 1977; ASHRAE 1996), hence,

$$\dot{Q}_{Room} \geq 3\dot{W}_{Diesel} \tag{9.6}$$

Therefore, solving equations (9.5) and (9.6) for $\dot{Q}_{Room}$ in terms of $\dot{Q}_{Combustion}$ yields

$$\dot{Q}_{Room} \geq 1.2\dot{Q}_{Combustion} \tag{9.7}$$

or

$$\eta \equiv \frac{\text{Benefit}}{\text{Cost}} = \frac{\dot{Q}_{Room}}{\dot{Q}_{Combustion}} \geq 120\%! \tag{9.8}$$

Answer 1: The exergy-conserving furnace (Figure 9.2) can provide over 20% more room heating than a conventional furnace (Figure 9.1) for a given fuel input, $\dot{Q}_{Combustion}$.[17]

What happened? How is this possible?

It should not be possible to exceed an efficiency of 100% unless a calculation mistake was made or our efficiency definition is flawed. A flawed definition is, in fact, the case. The furnace efficiencies reported by the furnace industry, though intuitive in nature, are flawed.[18] Furthermore, what could have possibly led anyone to consider the more complicated furnace shown in Figure 9.2? The answer, as the caption to Figure 9.2 suggests: exergy considerations!

FIRST AND SECOND LAW EFFICIENCIES

One possible furnace efficiency definition based on the exergy concept is

$$\eta_{II,\ Furnace} \equiv \frac{\text{Benefit}}{\text{Cost}} = \frac{\dot{Q}_{room}}{\dot{Q}_{Room,\ Max}} \leq 100\% \tag{9.9}$$

which *is necessarily* less than or equal to 100% by definition. $\dot{Q}_{Room,\ Max}$ is calculated assuming advantage is taken of the useful work potential or exergy of the energy input, $\dot{Q}_{Combustion}$. The "II" subscript on $\eta_{II,\ Furnace}$ emphasizes that this efficiency invokes in $\dot{Q}_{Room,\ Max}$ a limit imposed by the Second Law of Thermodynamics, and hence, is called a *Second Law efficiency.* Correspondingly, the efficiency, $\dot{Q}$, given in equation (9.1) is referred to as a *First Law efficiency,*[19] and is characterized as simply a ratio of energies with no consideration given to the limits imposed by the Second Law of

Thermodynamics. Virtually all efficiencies reported outside the engineering literature, and even most within the engineering literature, are First Law efficiencies and are generally intuition based. Unfortunately, as the 120% efficiency result demonstrates, *thermodynamic intuition (or common sense) may not be so reliable*. For clarity purposes a subscript "I" will henceforth be added to all First Law efficiencies, i.e. $\eta = \eta_{\mathrm{I}}$.

Is the First Law efficiency defined in equation (9.1) wrong? No.

Is the First Law efficiency of equation (9.1) flawed? Yes.

Equation (9.1) is not wrong provided one recognizes the implicit constraint that it be restricted to use on "simple" furnaces; that is, those furnaces based only on heat exchanger technologies. Not surprisingly, few are aware of this implicit constraint and hence common acceptance of the flaw in equation (9.1) exists. This *implicit constraint severely restricts the paradigm under which one operates*. Second Law efficiencies, η_{II}, are not so constrained.

> It has been shown, by example (not proof), that the energy paradigm, based solely on the conservation of energy principle, is unnecessarily restrictive in the energy conversion system options it suggests, and that the exergy paradigm is much less restrictive. Since ecosystems are composed of an array of specialized energy conversion systems, this observation suggests that there may be value in investigating the ecosystem/exergy link more closely.

An excellent example of the paradigm breaking ability of an exergy analysis is given by Reistad (1980) who compares dominant US energy flows and exergy flows. In brief, electricity (i.e. power generation) and transportation First Law efficiencies are much less than residential, commercial, institutional, and industrial heating efficiencies, but electricity and transportation Second Law efficiencies, in stark contrast, are much higher than residential, commercial, institutional, and industrial heating efficiencies. Traditionally, to this day, national energy flows are reported on a First Law basis (Canada 1996); however, it is the Second Law viewpoint that correctly identifies those energy conversion technologies with the greatest potential for improvement.

Question 2: For a fixed amount of fuel input (i.e. fixed $\dot{Q}_{\mathrm{Combustion}}$) to the exergy-conserving furnace, what is the *maximum* benefit possible in the form of room heating (i.e. $\dot{Q}_{\mathrm{Room,\,Max}}$)?

$$\eta_{\mathrm{II,\,Carnot}} \equiv \frac{\mathrm{Benefit}}{\mathrm{Cost}} = \frac{\dot{W}_{\mathrm{Max}}}{\dot{Q}_{\mathrm{Combustion}}} = 1 - \frac{T_{\mathrm{L}}}{T_{\mathrm{H}}} \tag{9.10}$$

Returning to the exergy-conserving furnace of Figure 9.2, $\dot{Q}_{\text{Room, Max}}$ can be calculated using a reversible[20] heat engine and a reversible heat pump. To do so only requires knowledge of the Carnot efficiency[21] for a heat engine which is given by where T_L is the temperature (K) of a low-temperature reservoir (e.g. T_0) and TH is the temperature of a high-temperature reservoir (e.g. $T_{\text{Combustion}}$). Demonstrating that the Carnot efficiency is the maximum efficiency for a heat engine operating between two temperature reservoirs, and that it is a function of temperature only as given in equation (9.10), is left for the detailed presentations provided by virtually all first course in thermodynamics texts (e.g. Wark 1977; Reynolds and Perkins 1977; Black and Hartley 1991; Van Wylen *et al.* 1994; Cengel and Boles 1998).

Similarly, if one notes that a heat pump is simply a heat engine with all energy flow directions reversed, and that by the definition of *reversible,* the absolute magnitudes of these energy flows must be the same for a heat engine or heat pump operating between the same two temperature reservoir temperatures, then the reversible heat pump's coefficient of performance naturally follows to be

$$\text{COP}_{\text{II, Reversible Heat Pump}} \equiv \frac{\text{Benefit}}{\text{Cost}} = \frac{\dot{Q}_{\text{Room, Max}}}{\dot{W}} = \frac{1}{1-(T_L/T_H)} \tag{9.11}$$

Consequently, for an exergy-conserving furnace operating reversibly, a First Law efficiency based on equation (9.1) results as follows (see Appendix C):

$$\eta_{\text{I, Exergy Conserving Furnace}} = \frac{\dot{Q}_{\text{Room, Max}}}{\dot{Q}_{\text{Combustion}}} = 1,290\% \tag{9.12}$$

or a Second Law efficiency based on equation (9.9) results as follows:

$$\eta_{\text{II, Exergy Conserving Furnace}} = \frac{\dot{Q}_{\text{Room}}}{\dot{Q}_{\text{Room, Max}}} = \frac{\dot{Q}_{\text{Room, Max}}}{\dot{Q}_{\text{Room, Max}}} = 100\% \tag{9.13}$$

Answer 2: A reversible, exergy-conserving, furnace (Figure 9.2) can provide for a given fuel input, $\dot{Q}_{\text{Combustion}}$,[22] about 1,200% more room heating than the best (i.e. no stack losses) conventional furnace (Figure 9.1).

As for the energy-conserving furnace shown in Figure 9.1, its First Law efficiency is given by equation (9.1) as 85% while its corresponding Second Law efficiency is only a mere 6.6%. The disparity between these two efficiencies is a specific example of the observations of Reistad (1980) discussed earlier in this section.

The Second Law, or *exergy, viewpoint recognizes energy quality, not energy magnitude, considerations as the appropriate criteria for assessing the most effective use of an energy resource*. Such recognition directs, often in violation of intuition, one's analysis and efforts to those aspects of an energy conversion system that provide strategies for energy utilization improvements. Lessons learned from understanding the exergy viewpoint explain, for example, how to improve upon the conventional furnace (as demonstrated above), or why a combined cycle power plant[23] is inherently more efficient than a standard steam cycle power plant (Krenz 1984). These and other such exergy lessons currently exist in engineering. Future ecosystem exergy studies should reveal similar lessons.

Example 2: Believe it or not, it is easier to boil ice than water

While the home furnace example in section "Example 1: How good is the furnace in your home?" introduced the exergy paradigm, this example, the boiling of ice problem, aims to re-enforce the notion that the exergy paradigm offers a formal framework to characterize a thermodynamic system's departure from equilibrium. Ecosystems are thermodynamic systems that continually maintain out of equilibrium states; an ecosystem in thermodynamic equilibrium is dead. Therefore, it is reasonable to search for a thermodynamic parameter that measures a system's departure from equilibrium. In contrast, the conservation of energy paradigm (i.e. energy magnitude) says nothing about a system's departure from equilibrium.

Imagine that you have access to 1 kg of ice at – 20°C or 1 kg of water a 60°C, and that you have been contracted to provide 1 kg of boiling water at night.[24] Also imagine that water costs a million dollars per kilogram and that the only fuel available to heat the water is natural gas at 20°C and valued at a million dollars a gram. In order to maximize your profits you need to use as little natural gas as possible to boil either the water or the ice. Fortunately, you do have access to any piece of equipment you would like free of charge, including reversible heat engines and heat pumps. Let the environment temperature be 20°C thus positioning it 40°C above the temperature of the ice and 40°C below the temperature of the water.

The Question: Ideally, does it take less natural gas to bring the 1 kg of ice at -20°C, or 1 kg of water a 60°C, to a 100°C boil?

The Answer: It takes a factor of 3.0 less natural gas to bring the –20°C ice to a boil! That is, it is theoretically possible to bring the -20°C ice to 88°C[25] with no natural gas input. In fact, had the ice been at –45°C no natural gas would be needed!

The Answer is not surprising if one adopts an exergy perspective. Simply put, the 1 kg of -20°C ice has more exergy than the 1 kg of 60°C water. In effect,

the –20°C ice's departure from equilibrium with its 20°C environment is greater than that of the 60°C water; a fact reflected by the – 20°C ice's potential to reach 88°C with no natural gas input. In words, the work potential of the ice is first extracted as energy flows from the environment to the ice until the ice warms to the environment temperature, all the while this work potential is stored for later use. This stored work is then used to operate a device, that is, heat pump, to transfer additional energy from the environment until the water reaches 88°C. Since, in this case, a temperature of 100°C is not reached using the stored work, natural gas is then required, less natural gas though than would be needed by the 60°C water.

Ideal heating systems for both the –20°C ice and the 60°C water are shown in Figure 9.3. The major difference between the exergy-conserving systems shown in Figure 9.3 and the exergy-conserving furnace system shown in Figure 9.2 is that in Figure 9.3 the temperature of the system of interest does not remain constant, that is, the temperature of the ice or water increases while room temperature is constant in time. Appendix C details the exergy calculations behind The Answer given above.

An intuitive concern about the result, that it can take less natural gas to boil – 20°C ice than 60°C water, is that it appears to violate the First Law of Thermodynamics. The First Law demands that more energy must go into heating the ice than the water. There is, however, no conflict. The ice does require more energy than the water to heat to 100°C. Much of the energy needed to heat the ice, however, comes from the environment which is a vast resource of energy but not exergy. In effect, only the energy *and* exergy needed to heat the ice the last 12°C, from 88°C to 100°C, must come from the natural gas.

This boiling of ice example dramatically demonstrates *a key feature of exergy, it is positive or non-zero no matter in which direction a system is out of equilibrium with its environment.*[26] Correspondingly, the exergy of a system in equilibrium with its environment is zero. If this were not so, it would be possible to construct a car engine or furnace that requires no fuel, but only requires the air that surrounds it to operate; if nothing else, experience tells us that this is not possible. Therefore, we have the following two key observations:

> *Any System* out of equilibrium with its environment has the potential to do useful work. In other words, the intrinsic exergy[27] of any system is either positive or zero, it is never negative (assuming the work output from the system is defined as positive).
>
> *Corollary:* Any system in equilibrium with its environment has NO potential to do useful work, and therefore has zero exergy.
>
> *Intrinsic exergy* provides a quantifiable measure for how far out of equilibrium with the environment a system happens to be.

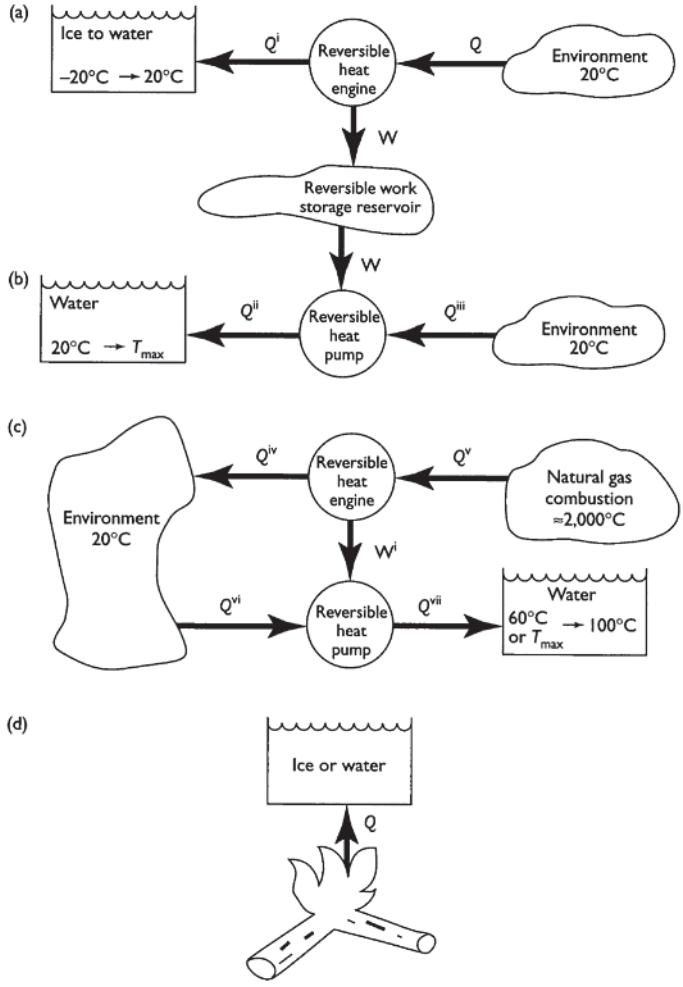


Figure 9.3 Exergy-conserving heating systems: together (a) and (b) bring the –20°C ice to the maximum temperature, T_{Max} (88ºC), possible without natural gas input (note that the work transfer into the reversible work storage reservoir equals the work transfer out), while (c) brings either the 60ºC water or the T_{Max} water to a boil. System (a) brings the –20°C ice to the environment temperature of 20°C, and system (b) brings the 20°C water to its maximum temperature, T_{Max}. Schematic (d) shows a conventional, non-exergy conserving, heating system.

A common feature of the exergy-conserving ice-boiling example given here, and the residential home furnace example given in section "Example 1: How good is the furnace in your home?" are that they make use of heat engines and heat pumps. This is no accident. In fact, it is important to note that *heat engines and heat pumps are the* only *means for heat transfer to take place* reversibly *between two thermal energy reservoirs at different temperatures* (Reynolds and Perkins 1977). In contrast, work transfer is inherently reversible provided that there are no frictional losses and that the process proceeds in a quasi-equilibrium[28] fashion. In short, designing an exergyconserving system is equivalent to designing a reversible system; *reversibility is key to preserving exergy.*

Finally, it is important to note that it is not being hypothesized that ecosystems strive to maximize the preservation of exergy. The authors do not hypothesize, and even *reject, the notion that exergy preservation strategies alone direct ecosystem design and evolution.* The authors *do hypothesize,* however, *that exergy transport and exergy destruction have key roles to play in ecosystem characterization* as discussed next.

A possible role for ecosystem surface temperature measurements

There are two key steps to appreciating the possible role to be played by ecosystem surface temperature measurements, at least with regards to working with the exergy paradigm. First, a possible role for exergy as a relevant ecosystem parameter must be established. Second, surface temperature must be established as a parameter for monitoring an ecosystem's exergy flows.

It is hypothesized that ecosystems strive to utilize exergy[29] *to their best advantage.* Ecosystems develop in a way which systematically increases their ability to degrade[30] the incoming (usually solar) exergy (Kay 1984; Schneider and Kay 1994).

EXERGY'S ROLE IN CHARACTERIZING ECOSYSTEMS

> *Fact 1:* Today's ecosystems are the result of an evolutionary process that has seen ecosystems composed of simple organisms (e.g. bacterium, algae) develop into ecosystems composed of complex, multicellular organisms (e.g. trees, humans) (Wicken 1987).

Therefore, it makes sense to look for a measure of system complexity, or order and organization, as one searches to discover how an ecosystem functions. Thermodynamic *entropy,*[31] an absolute measure of a system's thermodynamic disorder, is one such measure. Exergy is another possible measure of order and organization. Neither entropy nor exergy are claimed to measure complexity.

Beyond entropy and exergy there are many other measures of order and complexity proposed in the literature. Approaches adopted to quantify complexity include hierarchical approaches, geometric approaches, and algorithmic approaches (Cambel 1993). The quantification of complex behavior is discussed in the volume edited by Mayer-Kress (1986). Many scientists, coming from different fields, have offered tentative definitions of complexity and complexity measures (Margalef 1984; Berlinski 1986; Nicolis and Prigogine 1989; Gell-Mann 1994; Corbit and Garbary 1995; Kauffman 1995; Cillieres 1998; Ricard 1999). The volume by Peliti and Vulpiani (1988) brings together many different measures of complexity. There remains, however, no general theory of complexity (Hogan 1995). And, many complexity measures that can be quantified suffer, when applied to ecosystems, from practical measurement limitations making them effectively non-quantifiable. Those few indicators of ecosystem complexity that can be quantified (Odum 1995; Ulanowicz 1997) are beyond the scope of this chapter.

> *Fact 2:* The large exergy content of solar energy (see Section 9.4) provides ecosystems[32] with the high-quality energy input needed for an ecosystem's organized complexity to form from disorder (i.e. the raw materials of carbon, water, etc.) (Morowitz 1968; Odum and Odum 1976; Kay 1984; Ulanowicz and Hannon 1987; Edgerton 1982; Schneider and Kay 1994). For example, the phenomenon of a non-equilibrium system evolving to an ordered state as a result of fluctuations is referred to as "order through fluctuations" (Prigogine and Wiame 1946).

Ecosystems develop in a way which systematically increases their ability to degrade the incoming (usually solar) exergy. If this is the case, then measures of exergy degradation of an ecosystem can be used to characterize how well they are functioning, in a thermodynamic sense, and by implication, their degree of organization. This chapter focusses on exergy as a practical measure of ecosystem organization[33] and function, but not ecosystem complexity. At the outset, it must be recognized that the arguments given in this section represent facts and observations that appear to support the authors' hypothesis. However, validation of this hypothesis is the subject of active research (Allen 2000; Kay 2000a) and beyond the scope of this chapter.

Exergy analysis is favored for study because it is a well developed engineering tool that has demonstrated itself to be of great practical utility in the study of thermodynamic systems (Moran 1989; Li 1996; Bejan 1997).[34] Furthermore, the local environment affects the exergy utilization strategies pursued by organisms, and the authors hypothesize also by ecosystems. If an organism is to survive it must be able to *adapt* to its local environment.

Adaptation involves changing energy transfer approach(es), mechanism(s), or mode(s). For example, when temperature drops the human body invokes the mechanism of reducing blood flow to the skin and to the extremities. If temperature drops sufficiently, another mode, that of shivering, is induced. Temperature is a dominant parameter characterizing local environment; weather forecasts make this clearly evident. Another dominant parameter is the availability of water. Unlike entropy, exergy incorporates information about the local environment. Therefore, if, in characterizing an ecosystem, information about the local environment is desirable, then exergy, not entropy,[35] is favored as the thermodynamic parameter.

A FIRST LOOK AT THE ROLE OF SURFACE TEMPERATURE

> *Fact 3:* The high-exergy content of solar radiation is necessary, but not sufficient to support photosynthesis-dependent life[36] on earth. For example, the state of an organism's local environment is known to affect the ability of that organism to survive.

In its simplest form, a thermodynamic description of a system is in terms of the amount of energy and exergy entering, exiting, and stored within the system. For terrestrial ecosystems, such a model would naturally focus on the incoming solar energy and exergy, its primary source of energy and exergy. Therefore, parameters that characterize solar energy may be expected to play an important role in ecosystem characterization. Blackbody radiation is uniquely characterized by temperature alone (Incropera and DeWitt 1996). Solar energy can be well approximated as blackbody radiation originating from a thermal source at 5,762K (Weston 1992), while the earth's thermal radiation emissions can be well approximated as blackbody radiation originating from a thermal source at about 250 K (Krenz 1984). This hints at the importance of temperature in ecosystem characterization.

The fundamental importance of temperature reflected in equations (9.10) and (9.11) with regards to evaluating the exergy in the home furnace and boiling ice examples is a direct result of the fundamental role played by heat transfer. *Temperature gradients are the driving force behind heat transfer.*

Consider the ecosystem energy flows shown in Figure 9.4 where $\Phi_{T,\ Solar}$[37] is the solar radiation input (kJ s^{-1}), $\Phi_{T,\ Surface\ Emissions}$ is the thermal radiation emissions from surfaces within the ecosystem that cross the ecosystem boundary (kJ s^{-1}), $\Phi_{T,\ Background}$ is the atmospheric background thermal radiation (kJ s^{-1}), $\dot{Q}_{Convection}$ is the *convection*[38] heat transfer (kJ s^{-1}), and MINhIN and MQUThOUT represent mass flow transported energy in and out of the ecosystem (kJ kg^{-1} s^{-1}), respectively. is the mass flow rate (kg s^{-1}) and h is

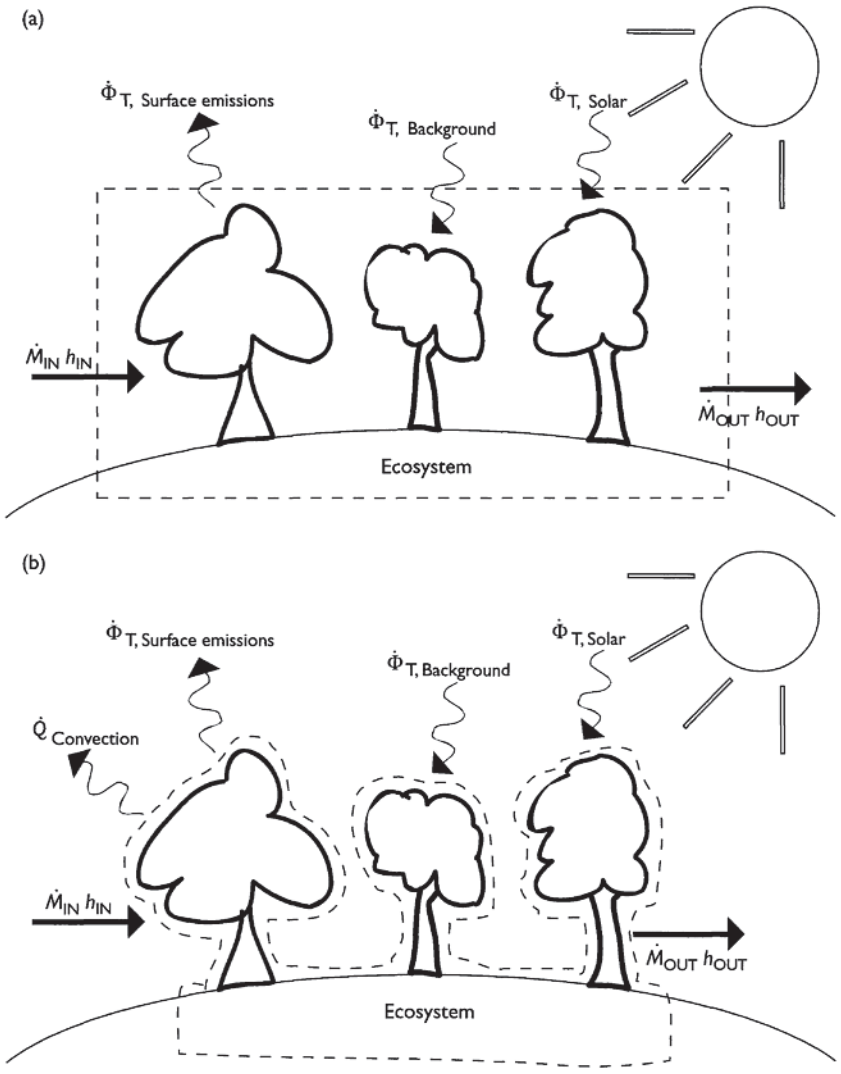


Figure 9.4 A terrestrial ecosystem viewed as a thermodynamic system. Two possible boundaries for this ecosystem are shown: (a) an atmospheric inclusive boundary, and (b) an atmosphere excluding boundary Dominant energy flows are solar radiation ($\Phi_{T,\ Solar}$), terrestrial thermal radiation from ecosystem surfaces ($\Phi_{T,\ Surface\ Emissions}$), atmospheric background radiation ($\Phi_{T,\ Background}$). and mass flow transported energy ($M_{IN}h_{IN}$ and $M_{OUT}h_{OUT}$). If the ecosystem boundary is selected adjacent to ecosystem surfaces as in (b) then convection heat transfer ($\dot{Q}_{Convection}$) can also be non-negligible.

specific enthalpy[39] (kJ kg^{-1}). No *shaft work*[40] transfer is indicted because it is zero for a fixed ecosystem boundary. Depending on where the ecosystem boundaries are located, there can be non-negligible *flow work*[41] transfer associated with transpiration and precipitation. Transpiration and precipitation, however, are not problematic for the ecosystem represented in Figure 9.4 since *flow work* is embedded within the mass flow energy-transfer terms. Also, depending on ecosystem boundary location, convection heat transfer can be negligible [e.g. Figure 9.4(a) given a system boundary that includes the atmospheric boundary layer and a large portion of the atmosphere, or given a system exposed to zero air velocity] or non-negligible [e.g. Figure 9.4(b) given a reasonable non-zero air velocity].

A simple global solar balance quickly demonstrates that solar radiation dominates other terrestrial energy inputs, including energy inputs from geothermal, tidal, and fossil fuel consumption sources. For an energy balance to exist there must be both inputs and outputs. It is surface and atmospheric reflectances and radiation emissions outputs that globally balance the solar radiation input, and dominantly contribute to the greenhouse effect that keeps our planet at a comfortable average surface temperature of about 13°C (Krenz 1984).

Locally, the dominant energy inputs and outputs for a terrestrial ecosystem, shown in Figure 9.4, are as follows: solar radiation, surface emissions (which in Figure 9.4(a) includes atmospheric radiation emissions), background radiation (e.g. atmospheric radiation from outside the system, or from other surfaces such as mountains or buildings), mass flow related energy fluxes, and possibly convection.

For the ecosystem shown in Figure 9.4(b), two of the three energy flux outputs are directly controlled by surface temperature, T_{Surface}. Specifically, the rate of surface radiation emissions is controlled by T_{Surface} according to the Stefan-Boltzmann Law,[42] while the rate of convection heat transfer is controlled by the temperature gradient between the environment temperature, T_0, and T_{Surface} according to Newton's Law of Cooling.[43] Only the mass flux energy output is, in general, T_{Surface} independent. The disconnect between ecosystem mass flux energy outflow (e.g. transpiration, precipitation, river flow) and surface temperature is the primary reason a quantification of overall ecosystem exergy destruction remains beyond the scope of this chapter. A secondary reason is a shortage of appropriate experimental measurements from ecosystems.

Conclusion: *Ecosystem surface temperature controls major ecosystem energy flux outputs, and hence exergy flux outputs.* That is, ecosystem surface temperature is a central thermodynamic variable needed to be measured. Determining ecosystem exergy fluxes is a precursor to determining exergy destruction by ecosystems.

FREE ENERGY AND EXERGY

> *Fact 4:* The exergy, not the energy content, of a chemical *determines* the minimum amount of energy needed to construct that chemical from a soup of stable elements[44] found in the environment.[45]

Edgerton (1982) describes the minimum energy required to produce a biochemical product as the *biological free energy.* He then goes on to identify exergy[46] as a reasonable measure of this biological free energy. Therefore, if the creation of chemicals is the name of the game for ecosystems to *function,* then energy quality as measured by exergy, and not simply energy magnitude, is a relevant thermodynamic parameter of interest (Jørgensen 1977; Jørgensen and Mejer 1981; Jørgensen and Müller 2000).

An important caveat to discussions of exergy in ecological systems concerns the relationship between exergy and Gibb's free energy. Edgerton's (1982) reference to a *biological free energy* is no doubt due to an attempt to draw an analogy with *Gibbs free energy.* Gibbs free energy is used extensively in the analysis of biological processes. Edgerton, however, makes it clear that one cannot simply use Gibbs free energy to describe the energies involved in biological chemical reactions because, as he correctly observes, solar energy is of high quality yet if approximated as blackbody radiation has a Gibbs free energy of zero![47] It is the biological free energy of solar energy that drives photosynthesis. A clear distinction between biological free energy and Gibbs free energy is, therefore, needed because (a) Gibbs free energy is often loosely referred to simply as *free energy,*[48] thus leaving the door open for confusion with biological free energy, and (b) chemical exergy is quantitatively often well approximated by a chemical's Gibbs free energy, thus confusion is again possible if one begins thinking of Gibb's free energy as simply another term for chemical exergy (Krenz 1984).

9.3 Theoretical foundations for the exergy paradigm

Having argued that an exergy analysis of ecosystems has the potential to provide important insights into ecosystem organization and function, the exergy concept is now developed in more detail.

This section formally introduces the reader to the definition of *exergy* as the *maximum useful to-the-dead-state work*. Since this chapter is intended to be an introduction to the exergy concept, with the purpose of establishing a link between ecosystem exergy analysis and thermal remote sensing, a minimum of mathematics is used by focussing on a non-reacting, control mass,[49] system. Nevertheless, the essential elements of the exergy concept are captured by the non-reacting, control mass systems analyzed here. The

reader is referred to Section 9.3.4 for a brief discussion on how the equations derived here can be generalized.

In the course of writing this chapter the authors realized that a number of refinements to the exergy concept, as it is currently used in engineering systems analysis (Bejan 1997; Li 1996; Tsatsaronis 1999; Moran 1999), are necessary for its application to ecological systems. This chapter introduces these refinements to current exergy terminology including a clear distinction between surroundings and environment (Section 9.3.1), a less restrictive concept of dead state with subsequent identification of a stable-equilibrium dead state (Section 9.3.2), and a less restrictive concept of maximum work (Section 9.3.2). More significantly, this chapter also introduces the notion of different classifications of exergy, and specifically the classifications of intrinsic exergy, transport exergy, restricted exergy, and accessible exergy (Section 9.3.5[50]).

9.3.1 Thermodynamic systems

The term *system* will be defined in a very broad sense. *A system simply identifies the subject of discussion or analysis. The system must be defined by the analyst for the particular problem at hand.* As stated by Reynolds and Perkins (1977),

> A system might be a particular collection of matter, such as the gas in a bottle. Or it might be a region in space, such as the bottle and whatever happens to be in it at the moment. Sometimes we include fields in our definition of the system; for example, the gas in the bottle and the electric field in the bottle might be defined as the system. At other times fields are defined to be outside the system; thus the gas may be the system, but the fields that occupy the same space are considered external to the system. Another situation in which two systems share the same space occurs in the analysis of ionized gases; the ions are often treated as one system, and the electrons as another. Interacting systems are often of quite different types; for example, in the study of liquid droplets the liquid interior to the surface is sometimes treated as one system, and the surface molecules as another. A system might be very simple, such as a piece of matter, or very complex, such as a nuclear power generation plant. Matter may flow through a system, such as a jet engine, or the system may be completely devoid of matter, such as the system of radiation in an enclosed volume.

In order to write down thermodynamic equations describing a system it is necessary for that system to be well defined by the analyst. This leads to the very important conclusion that a corresponding *system diagram* (or system schematic) is required every time a thermodynamic equation is written!

> It is *imperative*[51] that a *System Diagram* accompanies every First Law[52] equation and every Second Law equation in order to establish the assumed positive direction for mass transfers, work transfers, heat transfers, and radiation transfers.[53] It is not required that the assumed transfer direction match the actual transfer direction except in the case of radiation transfer.[54] If the assumed direction for mass, work, or heat transfer is incorrect a negative value simply results.

Remembering that exergy measures the work potential between a system and its environment (Section "Example 2: Believe it or not, it is easier to boil ice than water"), specification of the environment is as important as specification of the system itself. Generally, everything that is not included in the system is called the *surroundings* or the *environment* of the system (Gyftopoulos and Beretta 1991; Bejan 1997; Wark and Richards 1999). Dictionaries also identify surroundings and environment as synonyms. Nevertheless, *it is necessary for the purposes of clearly defining exergy that a system's surroundings and environment not be synonyms.*

With reference to Figure 9.5, the authors offer the following terminology: *Surroundings* continue to be defined as everything not included in the system. As such, surroundings can be divided into two components: the *immediate surroundings*[55] is that portion of the surroundings affected by, or affecting, system processes; and the *non-immediate surroundings* is that portion of the surroundings unaffected by, and that do not affect, system processes. The boundary between the immediate and non-immediate surroundings can, therefore, be modeled as an *isolated system boundary.* In turn, the immediate surroundings can be divided into two components: the *immediate environment* is that portion of the immediate surroundings across which property gradients exist driving heat transfer, work transfer, mass transfer, and net radiation transfer processes that interact with the system; and the *non-immediate environment* is that portion of the immediate surroundings which is free of irreversibilities. By the definition of irreversibility this implies that the entropy production, P_S, *within* the non-immediate environment is zero, that is, $P_{S,\,\text{Non-Immediate Environment}} = 0$. Of special interest for exergy purposes is the identification of the *reference environment,* or simply *environment.*[56] The atmosphere is often modeled as a reference environment with fixed temperature, pressure, and chemical composition (Ahrendts 1980; Szargut *et al.* 1988; Moran 1999). In the case of a power plant, river or lake water may be considered the reference environment (Moran and Shapiro 2000). The exergy reference environment is that system in the nonimmediate environment used to define the zero reference state, or dead state, for the system. Section 9.3.2 discusses the dead state which in turn identifies the selection criteria for the reference environment. Figure 9.5 provides a schematic representation of the preceding definitions.

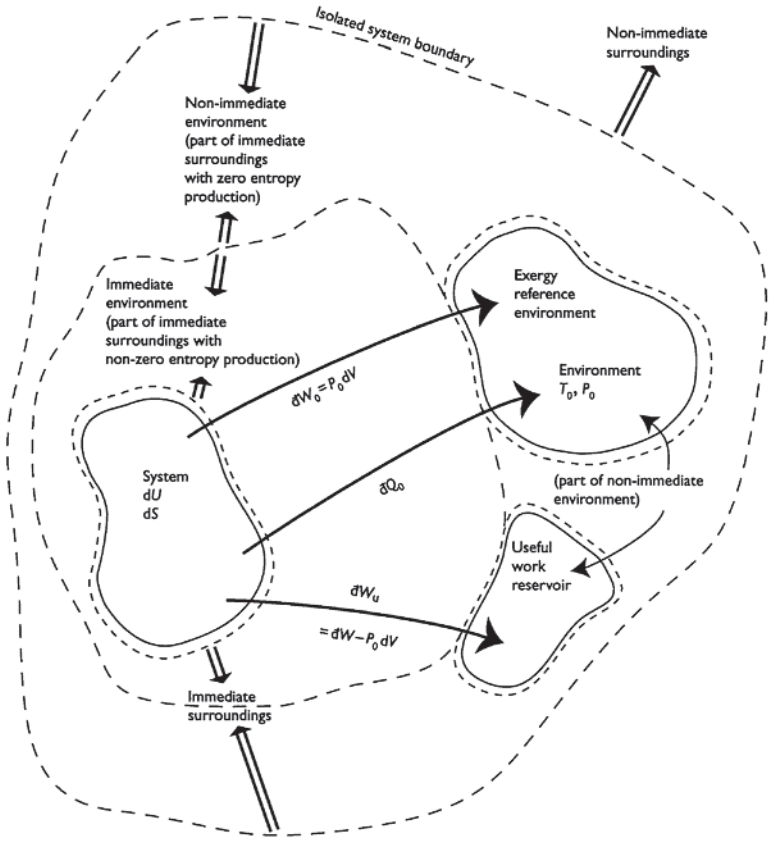


Figure 9.5 A control mass system *diagram* of a system is one that interacts with the environment via work transfer and heat transfer, and that is capable of doing useful work. The terms system, environment, immediate environment, non-immediate environment, reference environment, immediate surroundings, isolated system boundary, and non-immediate surroundings are also indicated. Entropy production is zero in the non-immediate environment.

9.3.2 Combining the First and Second laws: exergy

The exergy concept consists of the First and Second Laws of Thermodynamics plus three additional concepts: *useful work*, *maximum work*, and *dead state*. *Useful work* is that fraction of work that can go into lifting a weight in the non-immediate surroundings. Pressure-volume work, also referred to as *P* d*V work*, *done on the environment* (i.e. P_0 *d*V) is not useful as it cannot be used to lift a weight.[57] *Maximum work* is literally the maximum work transfer that can be obtained by a system. For the system shown in Figure 9.5, the

maximum work is obtained if all processes are reversible, that is, if the entropy production, in both the system and immediate environment, is zero ($P_{S,}=P_{S,\ System} + P_{S,Immediate\ Environment} = 0$; $P_{S,Non\text{-}Immediate\ Environment} = 0$ by definition).[58] A system is said to be in the *dead state* when it is in *thermodynamic equilibrium* with its *reference environment,* where the reference environment is identified as that non-immediate environment system that *maximizes a system's exergy calculation*.[59] For example, if the atmosphere is selected as the reference environment, then an ideal gas mixture is at the dead state when it is at the temperature and pressure of the atmosphere (i.e. in thermal and mechanical equilibrium), has no kinetic or potential energy relative to the atmosphere (i.e. zero velocity and zero elevation above a reference level); does not react with the atmosphere (i.e. chemically inert), and has no concentration gradients with respect to the atmosphere (i.e. no net diffusion).

> A system in thermodynamic equilibrium is macroscopically identifiable by the absence of all force and thermodynamic property gradients.

Of particular interest is the *stable-equilibrium reference environment* used to define the stable-equilibrium dead state. A *stable-equilibrium environment* is a non-immediate environment system that may interact with the system, whose intensive properties are unaffected by the system (e.g. temperature, pressure, specific internal energy, chemical concentrations), whose extensive properties may change (e.g. internal energy, volume, entropy, mass), and which is in thermodynamic equilibrium. The atmosphere, river water, and lake water are often modeled as stable-equilibrium environments. A *stableequilibrium reference environment* is that stable-equilibrium environment that maximizes a system's exergy calculation. Therefore, a system is at its *stable-equilibrium dead state*[60] *when it is in* thermodynamic equilibrium with a stable-equilibrium reference environment.

A special, but important, dead state is the *thermal-mechanical stableequilibrium dead state*[61] that exists when a system is only in thermal and mechanical equilibrium with its environment. It is not necessarily in, for example, chemical equilibrium with its environment. The dead state used in most exergy analyses found in the engineering literature is, for practical reasons, the thermal-mechanical stable-equilibrium dead state (e.g. Van Wylen *et al.* 1994; Bejan 1997; Cengel and Boles 1998; Wark and Richards 1999). One practical reason being that cost-effective mechanisms for extracting the work potential, or exergy, from post-combustion chemical species gradients do not, in general, exist. From an ecosystem perspective, a thermal-mechanical stable-equilibrium dead state may prove useful when performing an exergy analysis of a cold blooded animal, but determining such utility is beyond the scope of this chapter.

It is valuable to note that engineers define only one type of dead state, the stable-equilibrium dead state. This is an implicit restriction that has served the engineering profession well, but in considering ecosystems, *the authors have found it necessary to generalize the definition of dead state* to include the possibility that the reference environment's intensive properties may change as it interacts with the system. This would be the case when the system is comparable in size to the reference environment. For example, consider the case of the Amazonian rainforest (the system) and its local atmosphere (the reference environment). The Amazonian rainforest is so massive that it can control/change its local climate. That is, the Amazonian rainforest system is so large that it is actually capable of changing the thermodynamic state of its reference environment, a situation that led the authors to identify the engineering concept of dead state as a special case, that is, the special case of a stable-equilibrium dead state.

Engineers have not had the need to consider any dead state other than a stable-equilibrium dead state because the engineering systems analyzed have not approached the scale of a rainforest. This may change if, for example, cities ever begin to be analyzed as complex energy-conversion systems.

Control mass exergy balance equation

For the non-reactive, control mass system shown in Figure 9.5, whose species and phase concentrations equal that of the environment,[62] the First and Second Law equations for the system are as follows (neglecting kinetic and potential energy):

First Law equation (or *energy balance equation*) for Figure 9.5 system:

$$\mathrm{d}U = -đW - đQ_0 \tag{9.14}$$

Second Law equation (or *entropy balance equation*) for Figure 9.5 system:

$$đ\mathcal{P}_S = \mathrm{d}S_{\text{System}} + \mathrm{d}S_{\text{Immediate Environment}} = \mathrm{d}S + \frac{đQ_0}{T_0} \tag{9.15}$$

where U is the system's internal energy (kJ), W is the total work output of the system (kJ), Q_0 is the convection heat transfer to the environment (kJ), T0 is the environment temperature (K), S is the system's entropy (kJ K^{-1}), and is the entropy production in both the system and the immediate environment (kJ K^{-1}). The two different differential symbols, d and d, are used to distinguish between *exact* and *inexact differentials*,[63] respectively. An inexact differential is one that is path or process dependent. Thermodynamic properties such as entropy, internal energy, temperature, mass, etc., are state dependent, not path dependent, and hence, are described by exact differentials.

Equations (9.14) and (9.15) are combined by eliminating the convection heat transfer to the environment term. The resulting combined equation is then rearranged to isolate the work output from the system to yield

$$đW = -\mathrm{d}U + T_0\,\mathrm{d}S - T_0\,đ\mathcal{P}_S \tag{9.16}$$

The work represented by dW is the actual work output from the system. The useful work is determined by subtracting the pressure-volume work (P0 dV work) done by the system on the environment. It will be noticed that the negative sign in front of the T_0 $\mathrm{dP_S}$ term shows that entropy production results in a loss in ability to do work since T_0 and $\mathrm{dP_S}$ can never be negative. Therefore, to maximize useful work in this example is to minimize entropy production; that is, to zero the entropy production in equation (9.16). Therefore, the resulting *maximum useful work* equation is

$$đW_{\text{Useful, Maximum}} = -(\mathrm{d}U + P_0\,\mathrm{d}V - T_0\,\mathrm{d}S) \tag{9.17}$$

where $W_{\text{Useful, Maximum}}$ is the maximum useful work (kJ), P_0 is the environment pressure (kPa), and V is the system's volume (m^3).

Finally, the system's exergy can be calculated by integrating equation (9.17) to the dead state. Therefore, the system's *exergy, X*,[64] is defined by

$$\begin{aligned}\text{Exergy (kJ)} = X &\equiv \int_{\text{initial state}}^{\text{dead state}} đW_{\text{Useful, Maximum}} \\ &= W_{\text{Useful, Maximum, to-the-dead-state}}\end{aligned} \tag{9.18}$$

Equation (9.18) reveals why *exergy* is referred to in Section 9.1 as a system's *maximum, useful, to-the-dead-state, work*.

> By convention, unless stated otherwise, it is understood that exergy is defined by integrating to the stable-equilibrium dead state.

One may be inclined to claim that by the definition of dead state that T_0, P_0, and environment species and phase compositions, must be constant; however, recall that the environment is not defined as an external system whose intensive properties are unaffected by the system, but as *that* external system that *maximizes a system's exergy* (Section 9.3.2). Furthermore, recall the Amazonian rainforest discussion of Section 9.3.2 where the system is so massive that it is actually capable of changing the state of the environment. Therefore, the *stable-equilibrium assumption is essential to recognizing exergy as a pseudo-property;* that is, the resulting calculated exergy will depend only on the state of the system if T_0 and P_0 are fixed. Note, however, that although T_0 and P_0 may be assumed constant for a specific calculation of exergy, they need not be constant between exergy calculations. For example,

the exergy of natural gas in summer is less than in winter. For those engineers and scientists who are still inclined to always require T0 and P_0 to be constant (i.e. always require a stable-equilibrium dead state) for a given exergy calculation, the authors must object. The exergy concept must be capable of reflecting reality, and therefore must be capable of admitting the possibility that ecosystems exist, such as the Amazonian rainforest, which can affect the state of their reference environment.[65]

To obtain an expression for X is straightforward having assumed a stableequilibrium dead state. With T_0 and P_0 constant, the exergy for the control mass system shown in Figure 9.5 can be expressed as follows by integrating equation (9.17) as per equation (9.18):

$$X_{CM} = (U + P_0V - T_0S) - (U_0 + P_0V_0 - T_0S_0) \tag{9.19}$$

where X_{CM} is the control mass (CM) exergy, U, V, and S are the initial internal energy, volume, and entropy of the system, respectively, and U_0, V_0, and S_0 are the internal energy, volume, and entropy of the system when it is in equilibrium with the environment, respectively.

Just as internal energy cannot be quantified in an absolute sense until a frame of reference is chosen, an absolute measure of exergy cannot be established until a dead state is chosen as a reference frame. Nevertheless, exergy is still only a pseudo-property because exergy does not adhere to the *stable-equilibrium state postulate*[66] (Gyftopoulos and Beretta 1991) for thermodynamic properties. In particular, the dead state aspect of exergy's definition introduces information about the environment which is independent of the state of the system.

Equation (9.19) is the classic control mass exergy[67] expression as can be found in almost any modern introductory thermodynamics text (Black and Hartley 1991; Van Wylen *et al.* 1994; Cengel and Boles 1998; Wark and Richards 1999; Moran and Shapiro 2000).[68] Remember that work out of the system has been defined in Figure 9.5 as positive, if it had been defined as negative, the negative of equation (9.19) would result.

Do not forget the immediate environment

What is the implication of the zero entropy production assumption in the section on "Control mass exergy balance equation"? Simply put, it implies that all processes in the system and immediate environment must be reversible. That is, simply making all processes inside the system reversible is not necessarily sufficient to maximize the useful work!

> Communication with the environment is a central component of the exergy concept.

The central point of this section is to identify the *need to consider* the immediate environment in an exergy analysis, a viewpoint or paradigm not readily apparent when a *system-centric viewpoint* is adopted. A systemcentric viewpoint ignores thermodynamic entropy producing processes in the surroundings.[69] For example, it is very common for engineers to fail to see how the reversible furnace given in Figure 9.2 can be improved, even when you tell them it can be improved. Why is it so difficult to see the possible improvement? The answer lies in the fact that the system-centric viewpoint directs their thinking. How can one see the possible improvement? The answer lies in the *isolated-system viewpoint* observation that *there is always another immediate environment to be considered*. In the case of the exergy-conserving furnace of Figure 9.2, one must consider the immediate environment of the combustion system which includes the natural gas input. It is then discovered that the combustion process from natural gas to reaction products is irreversible. This, in turn, suggests looking for a reversible method of extracting the chemical potential energy from the natural gas, which, in turn, leads one to consider replacing the combustion chamber and heat engine with a fuel cell; a change that indeed improves the furnace.

The distinction between a system-centric viewpoint and an isolated-system viewpoint (see Appendix E) hints at the non-uniqueness of the system model used to calculate exergy. Consider, for example, that although a Carnot cycle engine can be used in Figures 9.3(a) or 9.El(b) to transfer thermal energy reversibility between the system and the environment, a Stirling cycle engine or an Ericsson cycle engine (Reynolds and Perkins 1977) could also have been used since all three reversible engines have the same efficiency, namely the Carnot efficiency as given in equation (9.10). *The exergy of a system does not depend on how the system and local environment are made reversible, only that they are made reversible.*

That is, even for an ideal, reversible, system there is no unique structure to that system. An interesting implication of this observation is that it suggests that "life is a tradeoff" (Kay 2000b), or that as ecosystem evolution has the option to migrate to many different thermodynamically equivalent systems. Since there is no unique reversible system for doing reversible work, then

> one cannot expect a unique irreversible ecosystem to exist for a given set of environment conditions; history[70] must also play a role in ecosystem development.

In conclusion, do not forget the immediate environment, and strive to understand what assumptions, both implicit and explicit, have been made for a selected immediate environment. In the language of Keenan (Hatsopoulos and Keenan 1965), the father of the exergy concept, the *communication channels* must be considered.

The exergy transfer (balance) equation

The exergy transfer equation provides a means to quantify exergy destruction and provides the means to trace exergy as it progresses through an ecosystem. Recall the hypothesis that ecosystems strive to utilize exergy to their best advantage (see section "A possible role for ecosystem surface temperature measurements"). Specifically, *ecosystems strive to control exergy destruction.* Hence, having a means to quantify exergy destruction is very important.

The control mass exergy equation, equation (9.19), developed in the section on "Control mass exergy balance equation" establishes the maximum useful work output for the system shown in Figure 9.5. Work output is shown going to a *useful work reservoir.* The useful work reservoir can be thought of simply as *a system where all the work transfer goes into reversibly lifting a weight.* Once the system has lifted the weight in the useful work reservoir, this stored gravitational energy can be used to lift another weight in a second useful work reservoir. In fact, any energy stored in the useful work reservoir is also the exergy of that useful work reservoir. With this understanding, the exergy transferred, $dX_{Transferred}$, into the useful work reservoir, for the system shown in Figure 9.5, is as follows:

$$\begin{aligned} đX_{\text{Transferred}} \equiv đW_{\text{Useful}} &= đW_{\text{Useful, Maximum}} - T_0\, đ\mathcal{P}_S \\ &= -(dU + P_0\, dV - T_0\, dS) - T_0\, đ\mathcal{P}_S \\ &= dX_{\text{System}} - T_0\, đ\mathcal{P}_S \\ &= dX_{\text{System}} - đX_{\text{Destroyed}} \end{aligned} \tag{9.20}$$

where W_{Useful} is the actual useful work obtained from the system, X_{System} is the system's exergy, and $X_{Destroyed}$ is the exergy destroyed during the work transfer process where

$$đX_{\text{Destroyed}} \equiv đX_{\text{Destroyed, System}} + đX_{\text{Destroyed, Immediate Environment}} \tag{9.21}$$

Please observe that system exergy has been expressed as an exact differential in equation (9.20). This is a subtle but very important observation. To use an exact differential implies that exergy is path or process independent. Any changes in a stable-equilibrium thermodynamic property (e.g. internal energy) is path independent because a reversible path can always be found in property state space to link any two states, and that thermodynamic property is uniquely defined at each point along this reversible path. Maximum useful work is not uniquely defined in state space, and hence remains path dependent, because information about the environment is introduced, that is, information that is independent of the state of the system. In particular, T_0 and P_0 remain

free to vary, which leaves the maximum useful work free to vary depending on the path followed by T_0 and P_0 during the integration of $dW_{\text{Useful, Maximum}}$. *By imposing a dead state not only is the end state of the system fixed, but so is the T_0 and P_0 path to that end state fixed.* In particular, T_0 and P_0 are no longer free to vary arbitrarily, and hence system exergy becomes a function of state space only. In effect, by defining a dead state not only is the end state for an exergy calculation fixed, but the integration path followed by equation (9.18) is also constrained. Therefore, system exergy, X_{System}, is a system property for a given dead state, nevertheless, the dead state is itself still system independent, thus we speak of X_{System} as being a *pseudo-property.*

Equation (9.20) expresses the idea that exergy can be transferred from system to system. Therefore, a *general exergy balance equation* can be written as follows:

$$\begin{bmatrix}\text{Exergy change}\\ \text{in system}\end{bmatrix} = \begin{bmatrix}\text{Exergy flow}\\ \text{IN}\end{bmatrix} - \begin{bmatrix}\text{Exergy flow}\\ \text{OUT}\end{bmatrix} - \begin{bmatrix}\text{Exergy destruction}\\ \text{in system}\end{bmatrix} \tag{9.22}$$

which yields the *exergy transfer rate equation*

$$\frac{dX_{\text{System}}}{dt} = X_{\text{Transferred, IN}} - \dot{X}_{\text{Transferred, OUT}} - \dot{X}_{\text{Destroyed}} \tag{9.23}$$

where $\dot{X} \equiv dX/dt$. Equation (9.23) does not refer to a specific system diagram because it applies to any system.

9.3.3 Lost work and the Gouy-Stodola theorem

Observe that equations (9.22) and (9.23) subtract an exergy destruction term instead of adding an exergy production term. This is because entropy production leads to a loss in ability to do work, that is, exergy destruction. This link between exergy destruction and entropy production will now be formalized.

Lost work,[71] W_{Lost}, is defined as

$$đW_{\text{Lost}} \equiv đW_{\text{Reversible}} - đW \tag{9.24}$$

where $W_{\text{Reversible}}$ is the work transfer from the system if all processes are reversible, and W is the actual work output from the system. Consider the system shown in Figure 9.5, if the work done on the environment is subtracted from both $W_{\text{Reversible}}$ and W in equation (9.24) then

$$đW_{\text{Lost}} = đW_{\text{Useful, Maximum}} - đW_{\text{Useful}} \tag{9.25}$$

or, using equation (9.20), equation (9.25) becomes

$$đW_{\text{Lost}} = T_0\, đ\mathcal{P}_S \tag{9.26}$$

which is known as the *Gouy-Stodola theorem,* "in memory of the first two thermodynamicists who were at least partially successful in convincing their contemporaries that (equation (9.26)) deserves attention" (Bejan 1997). Remember that dP_S is the entropy production in *both* the system and the immediate environment (see section on "Control mass exergy balance equation" and Appendix E). Finally, again using equation (9.20), one obtains

$$đX_{\text{Destroyed}} = T_0\, đ\mathcal{P}_S \tag{9.27}$$

or

$$đX_{\text{Destroyed}} \geq 0 \tag{9.28}$$

from which one can answer the question can exergy be produced? The answer is no.

Exergy can only be destroyed, never produced.[72]

Three *words of caution* when using the Gouy-Stodola theorem to calculate exergy destruction indirectly from a previous entropy production calculation.

1 First, do not forget that P_S includes entropy production in both the system and the immediate environment. Recall the system-centric and isolated-system viewpoints discussed in section "Do not forget the immediate environment." Chambadal's power plant (Chambadal 1957; Bejan 1997) is a classic example of the need to consider both the system and the immediate environment when calculating entropy production. Unfortunately, the authors have observed, in the ecological literature, instances of P_S being determined for only the system while P_S in the immediate environment is inappropriately ignored. Omitting the immediate environment from consideration has lead to a misunderstanding of the relationship between exergy, energy, and entropy.
2 Second, notice that, in general, when integrating equation (9.26) that T_0 is not necessarily constant. For example, if the time frame for integration spans winter and summer, then it would be reasonable to consider a dead state whose temperature varies. This said, in the special case of a stable-equilibrium dead state, T_0 is constant, thus equation (9.26) can be integrated to yield $W_{\text{Lost}} = T_0 P_S = X_{\text{Destroyed}}$.

3 Third, if entropy production is calculated locally within the interior of the system and immediate environment on the basis of identified entropy production mechanisms, it is very easy to unintentionally miss a specific mechanism. Adding to this difficulty is the problem of correctly modeling the local entropy production in a complex system. Whether there is value in identifying and quantifying local entropy production in ecosystems at a mechanistic level is a matter for future research. It certainly has value in engineering as exampled by the field of entropy production minimization[73] (Bejan 1995). Fortunately, a local understanding of entropy production mechanisms is unnecessary to measure overall exergy destruction. Measurement of ecosystem surface temperatures are one component of the measurements needed to determine overall ecosystem exergy destruction.

9.3.4 Exergy of a generalized system

The five most common generalizations to the system shown in Figure 9.5, and equations (9.14) and (9.15) are (1) to add kinetic and potential energy, (2) to add mass flows in and out of the system thus making the control mass a control volume, (3) to add additional paths for heat transfer, (4) to add chemical species considerations, and (5) to add radiation energy transfer. Table 9.1 summarizes these five generalizations.

Given the importance to ecosystems of solar radiation energy transfer, it is discussed in further detail in Section 9.4.

Although useful work, and hence exergy, is defined with respect to an ability to lift a weight it must be recognized that

all reversible work modes are equivalent from an exergy perspective.

For example, when all processes are reversible, 100% of the work input into electrostatic potential energy can be transferred into surface tension energy storage, which in turn can be transferred into gravitational potential energy in the lifting of a weight. Hence, the lifting of the weight simply serves as a model to aid understanding, it does not limit the number of type of reversible work modes that can operate, nor does it detract from the fact that *work transfer is 100% exergy transfer.* Some have claimed that exergy refers only to mechanical work; this is not true, exergy refers to all reversible work modes.

When discussing generalizations to a model some limit to these generalizations must be drawn. For example, quantum and relativistic effects have been ignored, gradients in field forces that act interior to the system have been excluded (e.g. gravitation), etc. However, there is one

Table 9.1 Generalizations commonlya applied to thermodynamic systems

	Figure 9.5, generalization	*Form of term added to an exergy calculation*	*Comment*
1	Kinetic energy, potential energy	$E = U + \mathrm{KE} + \mathrm{PE}$; where E is the total energy (kJ), U the internal energy (kJ), KE the kinetic energy (kJ), and PE the potential energy (kJ)	Total energy E replaces internal energy U in the First Law equation
2	Cross boundary mass flow	$\dot{M}(e + Pv) = \dot{M}(h + \frac{1}{2}v^2 + gz) = \dot{M}h^\circ$; where $\dot{M}$ is the mass flow rate ($\mathrm{kg\,s^{-1}}$), e the specific total energy ($\mathrm{kJ\,kg^{-1}}$), P the pressure (kPa), v the specific volume ($\mathrm{m^3\,kg^{-1}}$), h the specific enthalpy ($\mathrm{kJ\,kg^{-1}}$), v the velocity ($\mathrm{m\,s^{-1}}$), g the gravitational acceleration ($\mathrm{m\,s^{-1}}$), z the elevation (m), and h° is specific total enthalpy ($\mathrm{kJ\,kg^{-1}}$)	Enthalpy incorporates both the transport of total energy with mass flow, and the work required to transport this mass across the system boundary. This work is referred to as flow work; it is quantified by the Pv term
3	Convection heat transfer	$\dot{Q}_i(1 - T_0/T_i)$; where $\dot{Q}_i$ is the convection heat transfer rate from any system in the non-immediate environment except the reference environment ($\mathrm{kJ\,s^{-1}}$), T_0 the reference environment temperature (K), and T_i the non-immediate environment, non-reference environment, system temperature corresponding to $\dot{Q}_i$ (K)	Recall, conduction heat transfer is that limiting case of convection heat transfer when advection is zero
4	Chemical reactions, chemical diffusion	$\dot{M}_i\mu_i$ or $\dot{N}_i\hat{\mu}_i$; where $\dot{M}_i$ is the mass flow rate of chemical species i ($\mathrm{kg\,s^{-1}}$), $\dot{N}_i$ the molar flow rate of chemical species i ($\mathrm{kmole\,s^{-1}}$), μ_i the mass chemical potential of species i, and $\hat{\mu}_i$ the molar chemical potential of species i	Chemical potentials govern both the exergy of chemical reactions and the exergy of diffusion. When working with chemical reactions it is generally more convenient to work on a molar basis, whereas in the absence of chemical reactions a mass basis is often preferred
5	Radiation energy transfer	$\dot{Q}_{q,r}(1 - (4/3)(T_0/T_q))$ and $\dot{Q}_r(1 - (4/3)(T_0/T_r))$; where $\dot{Q}_{q,r}$ is the radiation energy transfer to system surface r from radiation source q ($\mathrm{kJ\,s^{-1}}$), $\dot{Q}_r$ the radiation energy transfer from system surface r ($\mathrm{kJ\,s^{-1}}$), T_q the temperature of radiation source q (K), and T_r the temperature of system surface r (K)	Blackbody or graybody radiation energy transfer is assumed. See Appendix F for a discussion on the $\frac{4}{3}$ factor

Note

a Common generalizations are based on those generalizations that tend to be dominant or relevant from an energy magnitude perspective. For example, although the earth's magnetic field may impact ecosystem processes, it does not contribute significantly to energy magnitude considerations, and is therefore not considered.

additional generalization that should be mentioned that applies to all the common generalizations presented in Table 9.1. This is the generalization that thermodynamic properties can, in general, vary continuously with location. Therefore, it sometimes becomes necessary to consider the integral form of the terms presented in Table 9.1, and equations (9.14) and (9.15). For example, the total energy can be expressed as

$$E \equiv U + \mathrm{KE} + \mathrm{PE} = \int_{\mathrm{System}} \left[\mathrm{u} + \frac{1}{2}\mathrm{v}^2 + \mathrm{gz} \right] \mathrm{d}m \tag{9.29}$$

where m is the mass. Although it is straightforward to write the First and Second Law equations in integral form, it is, in general, not convenient and often distracts from the overall physics being presented. But, ignoring the integral form also hides some of the physics. Consider again the integral equation for the total energy given in equation (9.29), but this time in a little more detail:

$$\begin{aligned} E &= \sum_{i=1}^{\mathrm{I}} \int_{\mathrm{System}} \left[\mathrm{u}_i + \frac{1}{2}\mathrm{v}_i^2 + \mathrm{gz}_i \right] \mathrm{d}m_i \\ &= \int_{\mathrm{System}} \left[\langle \mathrm{u} \rangle_i + \frac{1}{2}\langle \mathrm{v}^2 \rangle_i + \mathrm{g}\langle \mathrm{z} \rangle_i \right] \mathrm{d}m \\ &= M\langle \mathrm{u} \rangle_{i,m} + \frac{1}{2}M\langle \mathrm{v}^2 \rangle_{i,m} + Mg\langle \mathrm{z} \rangle_{i,m} \end{aligned} \tag{9.30}$$

where m_i is the mass of species i, and M is the total system mass. Looking at equation (9.30) one discovers that the specific internal energy, velocity, and elevation presented earlier in this chapter are actually mass $()_m$ and species $()_i$ weighted averages. The mass weighted character of elevation should come as no surprise since technically center of mass is a mass weighted quantity. Nevertheless, it does often surprise. Just remember, there is *always* another level of generalization that can be made.

9.3.5 Intrinsic, transport, restricted, and accessible exergy

A glance through almost any recent engineering thermodynamics text will reveal that the control mass exergy, X_{CM}, given in equation (9.19), is commonly referred to as *non-flow exergy* (Bejan 1997; Cengel and Boles 1998). If kinetic and potential energies are not neglected in the system shown in Figure 9.5, then the general non-flow exergy equation results as follows:

$$\chi^{\circ}_{\text{Non-Flow}} = (e - e_0) + P_0(v - v_0) - T_0(s - s_0) \tag{9.31}$$

where $\chi^{\circ}_{\text{Non-Flow}}$ represents specific total non-flow exergy (kJ kg^{-1}), e is the specific total energy (kJ s^{-1}), v is the specific volume (m^3 kg^{-1}), P is the pressure

(kPa), T is the temperature (K), s is the specific entropy (kJ kg^{-1}), and the subscript 0 refers to reference environment properties. Engineering systems admissible to a non-flow exergy analysis include automotive engines, heat pumps, pressure vessels, and power plants.

When the system under study also includes mass flows the exergy of the fluid streams entering or leaving the system are commonly referred to as a *flow exergy*. If kinetic and potential energy are not neglected in the steady-state steady-flow (SSSF) system shown in Figure 9.6, then the general flow exergy equation results as follows (Bejan 1997; Cengel and Boles 1998):

$$\chi^{\circ}_{\text{Flow}} = (h^{\circ} - h^{\circ}_{0}) - T_0(s - s_0) \tag{9.32}$$

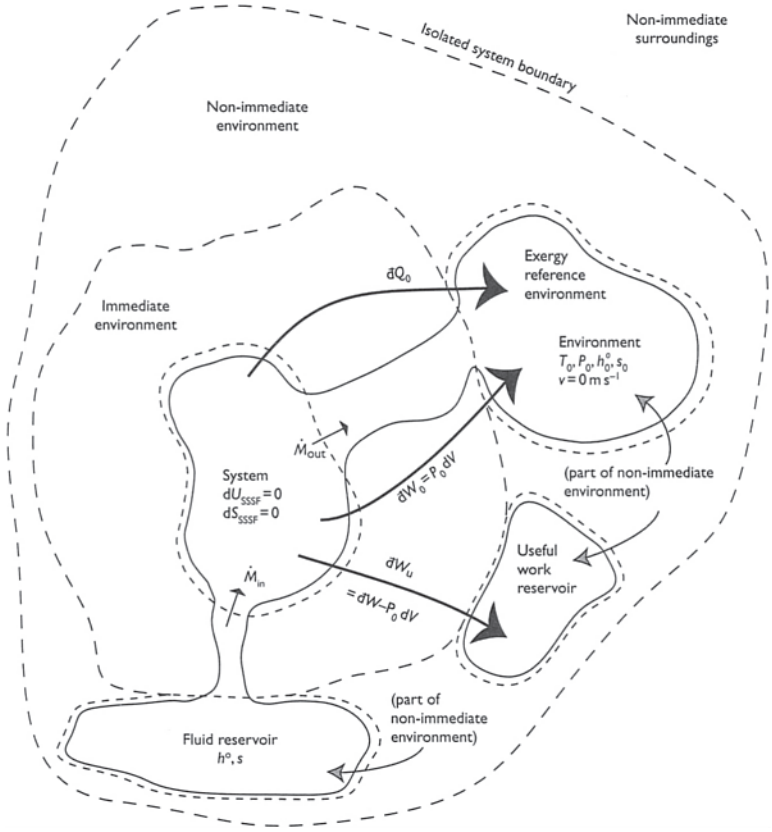


Figure 9.6 System diagram used to calculate flow exergy. Assumes an SSSF, non-reacting, and a fluid reservoir[74] species composition equal to that of the environment.

where χ^{o}_{Flow} represents specific total flow exergy (kJ kg^{-1}), and h^{o} is the specific total enthalpy (kJ kg^{-1}). Engineering systems admissible to a flow exergy analysis include jet engines, steam turbines, pumps, compressors, furnaces, and power plants.

Besides the visual differences between equations (9.31) and (9.32), there is a subtle but critical conceptual difference between non-flow and flow exergy. *Non-flow exergy* measures the useful work potential of *system content,* it is not concerned with the useful work potential that may exist in the surroundings. In contrast, *flow exergy* measures the useful work potential of that portion of the system's *surroundings that enters or exits the system;* it is not concerned with the useful work potential stored within the system itself.[75]

It is important to note that flow exergy, χ_{Flow}, is composed of two parts: the specific non-flow exergy, $\chi_{\text{Non-Flow}}$, of the inflow (or outflow), and the flow work exergy,[76] $\chi_{\text{Flow Work}}$, associated with the inflow (or outflow). This can be demonstrated by adding the specific non-flow exergy to the specific flow work exergy as follows:

$$\begin{aligned}\chi^{\circ}_{\text{Flow}} &= \chi_{\text{Non-Flow}} + \chi_{\text{Flow Work}} \\ &= [(e - e_0) + P_0(v - v_0) - T_0(s - s_0)] + [(P - P_0)v] \\ &= (e + Pv) - (e_0 + P_0 v_0) - T_0(s - s_0) \\ &= (h^{\circ} - h^{\circ}_0) - T_0(s - s_0)\end{aligned} \tag{9.33}$$

By noting that flow exergy includes flow work exergy, and by noting that flow work exergy can be negative for pressures less than environment pressures, one concludes that it is possible for flow exergy to be negative. In effect, a net work input to the system shown in Figure 9.6 may be required in order to maintain SSSF flow. Furthermore, because flow exergy can be negative it cannot be used as a measure of departure from equilibrium with the environment in the same way intrinsic exergy can, as discussed in section "Example 2: Believe it or not, it is easier to boil ice than water." Nevertheless, flow exergy is still useful as an exergy analysis tool since it can be used to determine the exergy destruction within a system (Wark and Richards 1999). For example, exergy destruction can be calculated from the exergy balance given in Figure 9.7.

Understanding that flow exergy and non-flow exergy are different at a fundamental concept level led the authors to rethink the concept of exergy and its interpretation. For example, the exergy of a can of gasoline from the perspective of someone who has no access to engines, fuel cells, matches, or other devices that can access the exergy content of the gasoline is inaccessible exergy. As another example, wood can be burned or left to decompose. In either case, access to the wood's exergy exists, but the ability to utilize all the exergy is restricted by the particular process that consumes the wood and

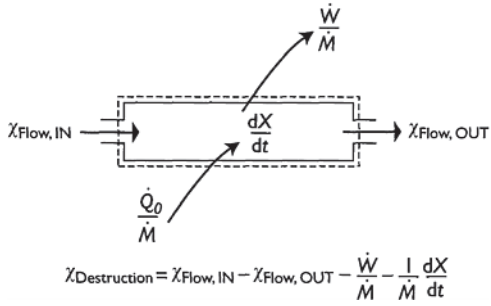


Figure 9.7 Exergy destruction, $\chi_{Destruction}$, as determined from a system's shaft work output, W, from a system's heat transfer with the environment, Q_0, from flow exergies, $\chi_{FLOW,OUT}$, and $\chi_{FLOW,IN}$, and from a change in system exergy, χ_{System}.

leads to the idea of restricted exergy. These and other conceptual refinements of the notion of exergy allowed the authors to suggest a resolution to the controversy surrounding the calculation of solar exergy (Section 9.4). In the end, the authors have classified exergy into seven types.

The authors identify *four fundamental classes of exergy:*

1 Intrinsic exergy, X^{i} (control mass or control volume exergy),
2 Transport exergy, X^{t} (SSSF energy in-flow exergy),
3 Restricted exergy, X^{r},
4 Accessible exergy, X^{a} (Corollary: inaccessible energy X^{ia}).

One *exergy hybrid:*

5 Restricted-access exergy, X^{ra}.

One *performance-based exergy:*

6 Extracted exergy, X^{ex} (actual useful work).

One *theoretical limitation* oriented *exergy:*

7 Hidden exergy, X^{h}.

Intrinsic exergy and *transport exergy* are introduced to differentiate between a work potential calculation of system content, and that of the energy of the surroundings that enter or exit the system, respectively. *Restricted exergy*[77]

is introduced to address the finite time, finite size, and structural constraints of ecosystems that unavoidably impose inherent irreversibilities. *Accessible exergy* is introduced to address structural constraints that inherently limit an ecosystems access to one or more modes of work potential and which may or may not introduce inherent irreversibilities. The hybrid of *restrictedaccess exergy* is introduced to characterize situations that simultaneously involve both restricted exergy and accessible exergy constraints. *Extracted exergy* is introduced as an exergy analogy to the First Law based concept of actual work. *Hidden exergy* is introduced to remind exergy users that an exergy calculation assumes all relevant[78] reversible work mode[79] gradients and thermal gradients are known and have been considered. Although developed for ecosystems, these seven exergy concepts can be applied to any thermodynamic system.

The major characteristics of intrinsic exergy, transport exergy, restricted exergy, accessible exergy, restricted-access exergy, extracted exergy, and hidden exergy are summarized in Table 9.2 with examples. Establishing the relative importance of these seven types of exergy on the development and function of ecosystems is a subject for future research. The authors do speculate, however, that in the future the concepts of restricted, accessible, and restricted-access exergy will prove useful in the analysis of ecosystems.

It is left to the reader to compliment the examples of Table 9.3 with additional engineering, biological, and ecosystem examples.

9.4 The exergy of solar energy

The sun is responsible for maintaining Earth's ecosystems (Schrodinger 1944; Kay 1984; Ulanowicz and Hannon 1987; Edgerton 1982; Schneider and Kay 1994, 1995). Virtually all life on Earth depends on photosynthesis, either directly or indirectly, for life-sustaining energy. Only in extremely rare circumstances, such as found around deep ocean thermal vents (Lutz and Kristof 2000; Stover 2000), may ecosystems exist in the absence of solar energy inputs.

Experience as expressed by the Second Law of Thermodynamics and the Spontaneous Equilibrium Principle[80] reveals that the natural direction for ecosystems to follow is towards a state of maximum disorder, that is, an equilibrium state. Ecosystems, on the other hand, continue to maintain a high degree of order and organization.[81] One question arises, how is ecosystem organization maintained? The answer, by utilizing the exergy content of solar radiation. Please note that *access to the non-zero exergy of solar radiation is necessary, but not sufficient, to maintain an ecosystem*. This important point is simply a reflection of the fact that *exergy content characterizes thermodynamic order,* it does not characterize organization, rather it is hypothesized that *exergy destruction characterizes organization*.

Table 9.2 Seven types of exergy: intrinsic, transport, restricted, accessible, restrictedaccess, extracted, and hidden[a] exergy

	Assumptions	Intrinsic	Transport	Restricted	Accessible	Restricted - Access	Extracted	Hidden
1	Useful work.	■	■	■	■	■	■	■
2	Maximum work.	■	■	■	■	■		■
3	To-the-dead-state work.	■	■	■	■	■		■
4	No friction.	■	■	■	■	■		■
5	All relevant reversible work modes[b, c] are accessible[d] *reversibly*.	■	■					
6	Thermal (temperature) gradients are accessible *reversibly*[e].	■	■					
7	Steady-state, steady-flow (SSSF).		■	▨	▨	▨		
8	All relevant reversible work modes are accessible.			■				
9	Relevant thermal gradients are accessible.			■				
10	Access to relevant reversible work mode and thermal gradients is constrained by finite size, finite time, and/or system structure[f].			■		■		
11	One or more relevant reversible work modes and/or thermal gradients are not accessible[g].				■	■		
12	All actual irreversibilities exist.						■	
13	All unknown and/or non-relevant[h] reversible work mode gradients, and/or thermal gradients, are accessible reversibly.							■

Notes

■ Indicates this assumption applies.

▨ Indicates that this assumption need not apply, but it may.

a All other exergy types have a corresponding hidden exergy, for example, hidden intrinsic exergy compliments intrinsic exergy.

b "The reference 'relevant reversible work modes' means that we count only important work modes for the system in question and do not count irreversible work modes" (Reynolds and Perkins 1977).

c Each thermodynamic property gradient corresponds to one reversible work mode and one exergy destruction (entropy production) mechanism. The only exception being a temperature gradient that corresponds to heat transfer, not a reversible work mode.

d Accessible means that a work mode mechanism exists to extract exergy from a given property gradient.

e That is, a reversible heat engine is accessible. Section 9.4 discusses complications associated with thermal radiation.

f Excludes system structures that prevent access to the exergy stored in a given property gradient.

g No assumption is made concerning the fate of the exergy stored in the inaccessible property gradient. It may, however, be referred to as *inaccessible exergy*.

h Implicit in the selection of the dead state is the selection of which property gradients are to be considered relevant.

Table 9.3 Distinguishing feature and examples: intrinsic, transport, restricted, accessible, restricted-access, and extracted exergy

Exergy type	*Distinguishing feature and examples (and notes)*
1 Intrinsic exergy, X^{i}	Intrinsic exergy is the theoretical maximum capacity to do useful work because of thermodynamic gradients between a system and its environment; it is the exergy *within* a system Examples: i Exergy of a liter of gasoline, a pound of food, a log of wood, or an entire ecosystem ii Non-flow exergy provided the system's thermal–mechanical stable-equilibrium dead-state composition equals that of the stable-equilibrium reference environment[a] iii Exergy of the energy within a turbine at any given instant in time Notes: A Intrinsic exergy is necessarily greater than or equal to zero ($X^{i} \geq 0$). This observation leads to the conclusion presented in section "Example 2: Believe it or not, it is easier to boil ice than water" that, "intrinsic exergy provides a quantifiable measure for how far out of equilibrium with the environment a system happens to be" B In determining intrinsic exergy all processes are considered reversible, hence, exergy destruction (and correspondingly entropy production) is necessarily zero ($X_{\text{Destroyed}} = 0$) C For a given stable-equilibrium dead state, intrinsic exergy (kJ) behaves as an extensive thermodynamic property. Correspondingly, specific intrinsic exergy (kJ kg^{-1}) is intensive D The system is analyzed as a control mass with one exception: mass flow with the reference environment is allowed as necessary in order to permit chemical equilibrium with the reference environment to be reached
2 Transport exergy, X^{t}	Transport exergy is the theoretical maximum capacity to do useful work of energy that *enters* a SSSF system Examples: i Exergy of steam flowing into a turbine, natural gas flowing into a furnace, food entering an organism, or water vapor flowing out of an ecosystem into the environment ii Flow exergy provided the system's thermal–mechanical stable-equilibrium dead-state composition equals that of the stable-equilibrium reference environment Notes: A Transport exergy may be negative because it incorporates flow work exergy. Negative transport exergy simply implies that work input to the system is required in order to maintain the imposed SSSF constraint B For a given stable-equilibrium dead state, mass flow transport exergy behaves as an intensive thermodynamic property *only* (kJ kg^{-1}). That is, transport exergy and specific transport exergy are equivalent C The SSSF assumption makes the intrinsic exergy of the system used to calculate transport exergy irrelevant

(Continued)

Table 9.3 (Continued)

Exergy type	*Distinguishing feature and examples (and notes)*
3 Restricted exergy, X^r	Restricted exergy is the theoretical maximum capacity of energy to do useful work when *all reversible*[b] *work mode*[c,d] *and thermal gradients are accessible*, but the *method for accessing these gradients is restricted* by time, space, or structure Examples: i Work output from a Diesel engine following an ideal Diesel cycle, as distinct from the exergy output from an internal combustion engine following an ideal Stirling cycle ii Solar exergy that can be captured by reptiles through absorptive thermal heating iii Exergy delivered by an ideal[e] co-generation[f] facility combusting natural gas and providing process steam at 200°C; the exergy delivered equals the work output plus the exergy of the process steam iv Non-flow or flow exergy[g] as given by equations (9.31) and (9.32), respectively Notes: A In determining restricted exergy, exergy destruction is minimized given the governing restriction(s) ($X_{Destroyed}$ is minimized, $X_{Destroyed} \geq 0$) B Entropy production minimization is a tool that can be used to calculate restricted exergy (Bejan 1995)
4 Accessible exergy, X^a (corollary: inaccessible exergy[h], X^{ia})	Accessible exergy is the theoretical maximum capacity of energy to do useful work when *direct access to some, but not all, reversible work mode or thermal gradients exists*[i]. (Corollary: Inaccessible exergy is the theoretical maximum capacity to do useful work that is *unobtainable due to an inability to access some, perhaps all, reversible work mode or thermal gradients*) Accessible exergy examples: i Exergy of coal from the perspective of a power plant which cannot make use of the exergy stored within the concentration gradient between the coal ash and the reference environment ii Exergy of corn from the perspective of a human who can digest the starch but not the cellulose iii Exergy of champagne from the perspective of a human who can digest the alcohol but cannot access the exergy attributable to the pressure difference between the dissolved CO_2 and the atmospheric dead state Inaccessible exergy examples: i For a coal fired power plant, the exergy content of coal ash ii For a human, the exergy content of cellulose iii For a tree, the exergy of the voltage difference that generates lightening[j] Notes: A Accessible work mode and thermal gradient processes are considered reversible ($X_{Destroyed} = 0$) with the possible exception of solar energy (see Section 9.4 and Appendix F) B It is not relevant if inaccessible work mode(s) and/or thermal gradient(s) exergy destruction mechanisms are or are not active. For example, it is irrelevant that the exergy of cellulose is unchanged by human consumption, while the voltage difference exergy that drives lightning is destroyed during the lightening process

(Continued)

Table 9.3 (Continued)

C An accessible exergy calculation is always less in magnitude than the corresponding intrinsic or transport exergy calculation except in the special case when the exergy content in all inaccessible work modes and/or thermal gradients is zero

D Although both restricted and accessible exergy are technically controlled by system structure, restricted exergy is concerned with structure(s) that *inhibits* access to a particular work mode while accessible exergy is concerned with structure(s) that *denies* access to a particular work mode

E Accessible exergy plus inaccessible exergy for a given system is equal to intrinsic or transport exergy, whichever is applicable

5 Restricted-access exergy, X^{ra}

Restricted-access exergy is the theoretical maximum capacity of energy to do useful work when *access to some, but not all, work mode or thermal gradients exists*, and the *method of access to one or more of these accessible gradients is restricted* by time, space, or structure

Examples:

i Exergy delivered by an ideal co-generation facility combusting coal and providing process steam at 200°C, but where the facility cannot make use of the exergy within the concentration gradient between the ash and the reference environment

ii Exergy of corn from the perspective of a human who cannot digest cellulose, and calculated assuming ideal human digestion with its space and time limitations that inherently impose irreversibilities

iii Solar exergy captured using photosynthesis which is constrained by the wavelengths that can be accessed, and restricted by the photosynthetic processes

Notes:

A Exergy destruction is in general non-zero ($X_{Destroyed} \geq 0$)

B Restricted-access exergy combines the independent constraints of restrict ed exergy and accessible exergy

6 Extracted exergy, X^{ex}

Extracted exergy is the *actual useful work* output or captured input

Example:

i The potential energy increase in a weight lifted to a higher height

ii The actual work delivered by an engine to propel an automobile from point A to point B

iii The exergy required to run the process of photosynthesis added to the exergy stored in chemical bonds created by photosynthesis

Note:

A Intrinsic exergy, transport exergy, restricted exergy, accessible exergy, or restricted-access exergy minus extracted exergy is the intrinsic exergy destruction, $X^{i}_{Destroyed}$, transport exergy destruction, $X^{t}_{Destroyed}$, restricted exergy destruction, $X^{r}_{Destroyed}$, accessible exergy destruction, $X^{a}_{Destroyed}$, or restricted-access exergy destruction, $X^{ra}_{Destroyed}$, respectively

(Continued)

Table 9.3 (Continued)

Exergy type	*Distinguishing feature and examples (and notes)*
7 Hidden exergy, X^h	Hidden exergy is the theoretical maximum capacity of energy, in *non-relevant or unknown thermodynamic property gradients*, to do useful work Example: i For biological systems the exergy of nuclear energy would constitute a portion of the systems' hidden exergy Note: A Implicit in the selection of the dead state is the selection of which property gradients are to be considered relevant and non-relevant B Since hidden exergy includes the exergy of unknown property gradients, it cannot, in general, be explicitly quantified

Notes

a This statement assumes that the only relevant reversible work mode is pressure-volume (P dV) work.

b Each reversible work mode corresponds to a thermodynamic property gradient, e.g. *P* d*V* work corresponds to a gradient in pressure. A non-reversible work mode is friction.

c Identifying the dead state implicitly identifies which work mode gradients are considered relevant. For example, if the dead state's reference environment is considered to be composed of specific percentages of O2, N2, CO2, and H2O only, then, for an engine combustion system, the exergy analyst is either implicitly identifying the exergy contribution from nitric oxides (NO and NO2) as non-relevant. Without an explicit comment by the exergy analyst as to the reason for the implicity non-relevance of NO and NO2, it is not clear if NO and NO2 are non-relevant because their exergy contribution is negligible, because of practical limitations in knowledge, or because the analyst is unaware of their existence. Whatever the reason, the exergy of the NO and NO2 is considered hidden exergy.

d For biological systems, where chemical reactions dominate, nuclear energy work modes are generally considered non-relevant.

e The adjective *ideal* means *minimum entropy production*. The tool of entropy production minimization (Bejan 1995) provides one means for determining a system's ideal operating point.

f An example of a co-generation system is a natural gas turbine that produces electricity and whose exhaust gases are used for space heating.

g In special cases a restricted exergy may also be an intrinsic or transport exergy. For example, this is the case for non-flow and flow exergy in the special case when the system's thermal-mechanical stable-equilibrium dead-state composition equals that of the stable-equilibrium reference environment.

h *Inaccessible* exergy applies only to known relevant work modes to which access has been denied. In the case of non-relevant work modes, or of unknown work modes, this inaccessible useful work potential is referred to as *hidden inaccessible exergy*. For example, nuclear reactions are generally considered non-relevant and consequently do not contribute to inaccessible exergy; rather, nuclear energy exergy is generally considered hidden inaccessible exergy. For known work modes, it is context dependent as regards to what is to be considered hidden inaccessible exergy.

i When there are multiple gradients of one type, for exmaple, multiple temperature differences creating multiple thermal gradients, not all of these same type gradients need to be accessible. Accessible exergy only requires access to at least one of the multiple, same type, gradients.

j Although a pine tree may partially, and indirectly, access the exergy made available by lighting, when a lightening ignited fire initiates pine cone activity, this is not considered accessible exergy since it is not directly accessed.

Without exergy an energy source cannot sustain the irreversible processes in ecosystems. Therefore, the non-zero exergy content of solar radiation is necessary for ecosystems to battle inherent irreversibilities that would otherwise sentence that ecosystem to death by thermodynamic equilibrium with its environment.

> Ecosystems maintain order by utilizing the exergy content of solar radiation that exists due to the sustained temperature gradient between the sun and the ecosystem's environment.
>
> *Corollary 1:* Ecosystems control the method and rate of solar radiation exergy destruction.
>
> *Corollary 2:* An ecosystem can only exists in the presence of exergy destruction.

In order to track exergy destruction through an ecosystem, and hence characterize an ecosystem's usage of solar exergy, it is necessary to know how much exergy that ecosystem starts with, that is, it is necessary to determine the exergy of solar radiation.

9.4.1 The solar exergy controversy

In the engineering and physics communities the exergy of solar radiation is of great interest because it represents an upper limit to the useful work producing ability of solar collectors. For example, the maximum theoretical electrical output possible from a solar collector. This great interest is reflected by the numerous publications concerned with the maximum work potential of solar energy (Petela 1964; Spanner 1964; Castans 1976, 1983; Landsberg and Mallinson 1976; Press 1976; De Vos and Pauwels 1981, 1986; Jeter 1981; Gribik and Osterle 1984; Haught 1984; Landsberg 1986; Bejan 1987; Bošnjakovic and Knoche 1988; Kabelac 1991). However, even given all this interest and work, a controversy still exists with regard to how to correctly quantify solar exergy.

The first step to resolving this controversy is to recognize that the radiation exergy literature may be divided into two groups, which differ by the initial question posed. The first group considers a closed system while the second group considers a SSSF open system (Kabelac 1991). The essence of the difference is that the closed system group has been trying to calculate an intrinsic exergy, while the open system group has been trying to calculate a transport exergy. As identified in Section 9.3.5 by the authors, intrinsic and transport exergy are fundamentally different, hence, any controversy between groups that may have once existed is now resolved. Nevertheless, controversies still exist within each group.

The group one controversy surrounding the exergy of closed system blackbody radiation is thoroughly discussed by Bejan (1987, 1997).

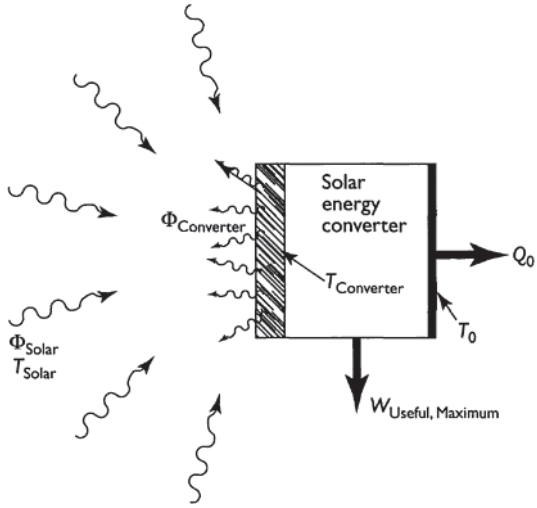


Figure 9.8 A system that continuously converts blackbody solar radiation into work (adapted from Kabelac (1991)).

Since ecosystems are open systems to solar radiation, this section focuses on the group two controversy concerning solar exergy from the open system perspective.

Kabelac (1991) discusses in some detail the three competing expressions found in the literature for maximizing the work output from solar energy for an open system. All three expressions are derived based on the model shown in Figure 9.8. Each expression and their corresponding assumptions are summarized in Table 9.4.

It is not the intent of this section, or this chapter, to resolve the controversy over which solar exergy expression given in Table 9.4 is correct; such an effort is a subject for future research. All three solar exergy expressions in Table 9.4 are, however, discussed in turn in Section 9.4.2 where a bias towards favoring equation (9.34) is proposed.

9.4.2 A role for surface temperature measurements

As identified in section "A first look at the role of surface temperature," an ecosystem's surface temperature governs two major modes of energy transfer from an ecosystem: thermal radiation surface emissions energy transfer and convection heat transfer. These two energy transfer considerations alone are sufficient to justify the need for ecosystem surface temperature measurements for an ecosystem *energy analysis*.

Table 9.4 Competing expressions for the maximum work output from solar radiation. Based on Figure 9.8 and Kabelac (1991)

Model	*Proposed solar exergy expression, assumptions,*[a] *and comments*
1, 2, and 3	*Implicit assumptions* i Useful work ii Maximum work iii To-the-dead-state work iv No friction *Explicit assumptions* v Steady-state steady-flow vi Ideally concentrated solar radiation[b] vii Diffuse blackbody radiation[c] viii Conduction or convection to the environment at temperature T_0
1	$W_{\text{Maximum}} = \Phi_{\text{T, Solar}} \left(1 - \frac{4}{3} \frac{T_0}{T_{\text{Solar}}} + \frac{1}{3} \frac{T_0^4}{T_{\text{Solar}}^4} \right)$ (9.34) *Assumptions* ix Zero entropy production[d] x Finite area *Comment* A T_{surface} equals T_0 B Φ_T is given by the Stefan–Boltzmann law, $\dot{\Phi}_T = \sigma A T^4$
2	$W_{\text{Maximum}} = \Phi_{\text{T, Solar}} \left(1 - \frac{T_{\text{Optimum}}^4}{T_{\text{Solar}}^4} \right) \left(1 - \frac{T_0}{T_{\text{Optimum}}} \right)$ $= \Phi_{\text{T, Net}} \left(1 - \frac{T_0}{T_{\text{Optimum}}} \right)$ (9.35) where T_{Optimum} is the surface temperature, T_{Surface}, that satisfies the expression $0 = 4T_{\text{Surface}}^5 - 3T_0 T_{\text{Surface}}^4 - T_{\text{Solar}}^4 T_0$ (9.36) *Assumptions* x Finite area xi Solar radiation is converted to thermal energy[e] *Comment* C T_{Surface} is explicitly represented by T_{Optimum}

(Continued)

Table 9.4 (Continued)

Model	*Proposed solar exergy expression, assumptions,*[a] *and comments*
3	$W_{Maximum} = \Phi_{T,\,Net}\left(1 - \frac{T_0}{T_{Solar}}\right)$ (9.37)
	Assumptions xi Solar radiation is converted to thermal energy xii $T_{Surface}$ approaches T^{f}_{Solar} xiii Finite $\Phi_{T,\,Net}$[g]
	Comment D Area approaches infinity in order to maintain finite $\Phi_{T,\,Net}$ as $T_{Surface}$ approaches T_{Solar}

Notes

a Assumptions are discussed in more detail in Appendix F.

b Solar radiation is highly directional given the small solid angle (θsun = 6.8×10^{-5} steradians) intercepted by the sun in the hemisphere of the sky. Ideally, solar energy can be concentrated using non-imaging optics to increase the incident solar radiation intensity by a factor of $1/\sin^2 \theta$sun ≈ 46,000 while maximizing the solid angle distribution of the incident solid angle to 2π steradians (Welford and Winston 1989).

c The blackbody assumption is introduced to simplify the equations governing radiation energy transfer. Refer to Appendix F for an explanation of why radiation energy transfer *is* not to be considered heat transfer, a fact lost on most engineers and scientists. The diffuse radiation assumption implies that both incident and emitted radiation is distributed equally in all directions. Consequently, the assumption of diffuse solar radiation implies use of an ideal solar concentrator (Bejan 1997).

d The zero entropy production assumption is not valid if the incoming radiation is converted to thermal energy and if the absorbing surface is at a different temperature than that which produced the incoming radiation.

e Non-zero entropy production exists at the absorbing surface since it is at a different temperature than that which produced the incoming radiation, that is, $T_{surface}$ does not equal *T*Solar.

f Zero entropy production exists since the conversion to thermal energy at the absorbing surface can take place reversibly when $T_{Surface}$ approaches $T_{Solar.}$

g As $T_{surface}$ approaches $T_{Solar,}$ $\Phi_{T,Solar}$ and $\Phi_{T,Converter}$ approach infinity. It is assumed, however, that $\Phi_{T,Solar} - \Phi_{T,Converter}$ remains constant at $\Phi_{T,Net.}$

Parameter definitions

Referring to Figure 9.8 and Kabelac (1991), the terms used above are as follows: $W_{Maximum}$ is the maximum work and also the maximum useful work output for each model (kJ); $\Phi_{T,\,Solar}$ is the incoming solar radiation (kJ); *T*0 is the environment temperature (K); $T_{surface}$ is the effective blackbody temperature of the sun (5762 K (Weston 1992)); *T*Surface is the system's surface temperature (K); $T_{Optimum}$ is the optimum surface temperature that maximizes work output (K); and $\Phi_{T,Net}$ is the net radiation energy transfer to the system and is equal to the incoming solar radiation energy minus the emitted converter surface radiation (kJ); σ is the Stefan-Boltzmann constant ($\sigma = 5.66961 \times 10^{-8}$W m^{-2} (kJ);A is surface area, and ΦT is the radiation energy per unit time emitted from a surface of area A at temperature T (kJ s^{-1}).

From an *exergy analysis* perspective surface temperature is also essential in order to enable calculation of exergy destruction resulting from the finite-temperature-difference convection heat transfer between the ecosystem surface and the atmosphere (refer to Figure 9.4(b)). The purpose of this section is to further demonstrate a possible need for ecosystem surface temperature measurements if one wishes to perform an ecosystem *exergy analysis*. This will be accomplished by considering, in turn, the applicability of each of the three proposed solar exergy expressions given in Table 9.4.

It is informative to note that a *solar exergy* calculation also reveals the *maximum exergy destruction* to be found in an SSSF solar energy balance.

Model 1: zero entropy production, finite area system, solar exergy

Petela (1964), Press (1976), and Landsberg and Mallinson (1976) independently derived the solar exergy expression given by equation (9.34).

The surface temperature, T_{Surface}, does not explicitly appear in equation (9.3) because it has been set equal to the environment or dead state temperature, T_0, which it necessarily equals in order to maximize the work output from the system shown in Figure 9.8. Furthermore, since entropy production is zero and no consideration has been given to system size, process time, or system structure, equation (9.34) represents a measure of *transport exergy*. By recognizing that in the derivation of equation (9.34) that T_{Surface} is equal to T_0, then equation (9.34) can be re-written as follows:

$$X^{\text{t}}_{\text{Solar, Model 1}} = \Phi_{\text{T, Solar}} \left(1 - \frac{4}{3} \frac{T_{\text{Surface}}}{T_{\text{Solar}}} + \frac{1}{3} \frac{T^4_{\text{Surface}}}{T^4_{\text{Solar}}} \right) \tag{9.38}$$

in which case the *need for a surface temperature measurement becomes explicitly clear.*

If one objective for an ecosystem is to maximize the available exergy it can do so by controlling its surface temperature. As shown in Table 9.5, a lower surface temperature results in more available solar exergy. This is an interesting observation because it is consistent with the surface temperature measurements of Luvall and Holbo (1991) and Akbari (1995) who observe that mature natural ecosystems (e.g. forests) are cooler than manmade ecosystems (e.g. farmland, lawns), and therefore, offers support to the author's hypothesis (see section "Example 2: Believe it or not, it is easier to boil ice than water") that ecosystem development may entail an attempt by the ecosystem to control exergy destruction.

It is valuable to keep in mind that all proposed solar exergy models given in Table 9.4 assume use of an ideal solar concentrator (Explicit assumption vi).

Table 9.5 Solar exergies for selected surface temperatures (model 1)

Surface temperature, T_0		*Relative solar exergy:* $(X^t_{\text{Solar, Model 1}} / \Phi_{T,\text{Solar}})$ $(kJ_{\text{Solar Exergy}}/kJ_{\text{Solar Energy}})$
13°C (286 K)	Earth's average surface temperature (Krenz 1984)	0.9338
20°C (293 K)	Commonly assumed environment temperatures	0.9322
27°C (300 K)		0.9306
60°C (333 K)	Observed asphalt temperature (Luvall 2000)	0.9229

Note
$T_{\text{Solar}} = 5,762$ K (Weston 1992).

In effect, background radiation from the atmosphere and deep space are ignored. Background radiation effects on calculated solar exergy introduce restricted exergy and accessible exergy considerations. Such considerations remain a topic for future solar energy research. The primary point of Table 9.5 is to demonstrate that solar exergy increases with decreasing surface temperature. Admittedly, the effect is not large, however, the authors suspect that once background radiation effects are considered, the effect will be larger.

Model 2: non-zero entropy production, finite area system, solar exergy

Castans (1976), Haught (1984), De Vos and Pauwels (1981), and Bejan (1997) favor the solar exergy expression given by equation (9.35).

The surface temperature, T_{surface}, explicitly appears in equation (9.35) as the optimum surface temperature, T_{Optimum}. For a solar temperature of $T_{\text{Solar}} =$ 5,762K and an environment temperature of $T_0 =$ 300K, the optimum surface temperature is $T_{\text{Optimum}} =$ 2,465 K, high even for a fossil fuel combustion temperature (Glassman 1987). Consequently, T_{Optimum} is an absurd surface temperature for an ecosystem to attempt to attain. In contrast, it is not quite so absurd for an engineer to consider designing a solar collector that conforms to the constraints of Model 2 and therefore would be optimized in the limit represented by equation (9.35). The finite area assumption of Model 2 places a structural restriction on the system. Furthermore, the requirement that the solar radiation must be converted to thermal energy is a second structural constraint on the system. Therefore, equation (9.35) describes a measure of *restricted exergy.* Given the non-zeroentropy production in Model 2, the ratio of incoming solar exergy to incoming solar energy is less than that

found for Model 1. For example, at an environment temperature of 300 K, this ratio for the restricted-access exergy of Model 2 is 0.8489 $kJ_{Solar\ Exergy}/KJ_{solar\ Energy}$, while as seen in Table 9.5 for the transport exergy of Model 1 it is 0.9306 $KJ_{Solar\ Exergy}/KJ_{Solar\ Energy}$.

Since ecosystems have access to photosynthesis it is not true that incoming solar radiation is necessarily converted to thermal energy before further utilization, hence, *the restricted exergy described by equation (9.35) cannot be argued to apply to ecosystems because it is unnecessarily restrictive*. Therefore, the absurdly high surface temperature required by equation (9.35) is of no concern to the analysis of solar exergy from the perspective of an ecosystem.

Model 3: zero entropy production, infinite area system, solar exergy

Kabelac (1991) proposes the solar exergy expression given by equation (9.36). Equation (9.36) is simply the Carnot efficiency for a heat engine operating between two high- and low-temperature thermal energy reservoirs at T_{Solar} and T_0, respectively. This is an appealing result as it would mean that the exergy of radiation energy transfer is described in a similar fashion to the exergy of conduction or convection heat transfer.

The surface temperature, $T_{surface}$, is implicitly present in equation (9.37) through F_{Net} where $T_{Surface}$ appears explicitly in the definition of F_{Net} as follows:

$$\Phi_{Net} \equiv \lim_{T_{Surface} \to T_{Solar}} A(T)\sigma(T_{Solar}^4 - T_{Surface}^4) \tag{9.39}$$

As with Model 2 (section "Model 2: non-zero entropy production, finite area system, solar exergy"), $T_{surface}$ is absurdly large, being equal to the sun's effective blackbody surface temperature of 5,762 K. Furthermore, $T_{surface}$ cannot be varied as it necessarily approaches T_{Solar}.

Kabelac (1991) argues that the infinite area implication of Model 3 is consistent with the infinite area implication of reversible conduction or convection heat transfer for the Carnot engine, and therefore should not present a conceptual problem.

A conceptual problem, however, does seem to exist. Three observations concerning the infinite area requirement of Model 3 suggests equation (9.36) is unacceptable as a measure of solar exergy. These three observations are as follows: (a) The infinite area requirement for reversible heat transfer is not an exact parallel to the infinite area limit imposed on radiation energy transfer. For example, it is possible for a finite amount of reversible heat transfer to take place in a finite amount of time over a finite area! This fact is poorly understood even among engineers who routinely state that reversible heat

transfer requires either infinite area or infinite time. The fact is, for conduction, for example, an infinite thermal conductivity also works to provide finite heat transfer rate across a finite area and an infinitesimally small temperature difference. Given the common use and poor understanding in the thermodynamic community of the infinite area conclusion for reversible heat transfer, it is understandable why Kabelac (1991) endorsed the infinite area parallel between radiation and heat transfer without considering the differences. (b) Embedded in the exergy concept is the idea of terrestrial usefulness and practicality. Therefore, a terrestrial-based solar collector cannot be expected to have an infinite area that would extend the system outside its immediate surroundings. Unfortunately, that physics of radiation energy transfer demands that an infinite area solar collector necessarily[82] extends outside a system's immediate environment. In contrast, it is possible for conduction heat transfer to take place over an infinite area in a finite volume of space if the surface is fractal in nature. And (c), an infinite area solar collector requires an infinite amount of solar radiation to strike its surface at any given instant, however, the sun is of finite area and hence can emit only a finite amount of solar radiation at any given instant.

Therefore, there are serious concerns about the appropriateness of equation (9.36) as a measure of solar exergy. Consequently, the authors take the position that *equation (9.36) is meaningless from an exergy perspective,* a conclusion consistent with Bejan's (1997:476) statement that the infinite area requirement is "totally unrealistic."

The correct solar radiation exergy equation

The authors do not suggest that they have definitively concluded equation (9.34) to be the correct measure of solar exergy, but the arguments presented in earlier do suggest that equation (9.34) is currently the best expression for calculating solar radiation exergy from an ecosystem's perspective, and therefore is cautiously accepted for the time being.

Although it need not have been so, it is nice to see that equation (9.34) further justifies the need for surface temperature measurements.

9.4.3 Importance of air temperature

The authors implicitly argued in section "Model 1: zero entropy production, finite area system solar exergy" using equation (9.34) that by controlling surface temperature an ecosystem can control the solar exergy available for destruction. The difficulty is that it is actually the dead state, or the temperature of the reference environment, that determines the solar exergy available for destruction. It just so happens in the derivation of equation

(9.34) that $T_{Surface}$ approaches T_0 in the exergy calculation limit. Therefore, in section "Model 1: zero entropy production, finite area system, solar exergy," it is implicitly assumed that ecosystem surface temperature correlates with T_0, where T_0 will often be appropriately taken as the local atmospheric air temperature. The validity and implications of this assumption need to be investigated in future work and therefore are beyond the scope of this chapter. This assumption needs to be investigated in order to better understand how ecosystems may control the solar exergy available for destruction, and the recent ecosystem air temperature measurement work of Akbari (1995) may assist in this investigation.

Recall that the authors hypothesized that *ecosystems strive to utilize exergy,* and that this utilization includes controlling exergy destruction (section "A possible role for ecosystem surface temperature measurements"). What has been argued here is that the utilization of exergy may also include controlling the amount of exergy available for destruction.

9.5 Linking thermal remote sensing to ecosystem organization

9.5.1 Life as on exergy destroyer

Section "Exergy's role in characterizing ecosystems" introduces the idea that exergy provides one indicator of order and organization. Following this and associated lines of reasoning has lead to the postulation of the exergy destruction principle (Kay 2000b): The *exergy destruction principle*[83] can be stated as follows:

> A system exposed to an inflow of exergy will be displaced from equilibrium. The response of the system will be to organize itself so as to destroy the exergy as thoroughly as circumstances permit, thus limiting the degree to which the system has moved from thermodynamic equilibrium. Furthermore, the further the system is moved from equilibrium, the larger the number of organizational (i.e. dissipative) opportunities which will become accessible to it, and consequently, the more effective it will become at exergy destruction.

Schrödinger's work (Schrödinger 1944) suggests[84] it appropriate to consider living systems as non-equilibrium dissipative systems (Schneider and Kay 1995). Given the exergy destruction principle and the right conditions, the emergence of living systems should be expected as a means of furthering the mandate of exergy destruction. In particular, the earth is an open

thermodynamic system with a large exergy flow impressed upon it by the sun. Consequently, physical and chemical processes will emerge to destroy the incoming exergy. For example, energy shifts (conversion of shortwave radiation to longer-wave infrared), absorption, and meteorological and oceanographic circulation will degrade much of the incoming solar exergy. And, as argued elsewhere (Kay 1984; Kay and Schneider 1992; Schneider and Kay 1994), life is simply another means of destroying solar exergy.

The origin of prebiotic life is the development of yet another route for the destruction of exergy. A stepwise progression of stages can be recognized in the emergence of prebiotic organized structures: the formation of simple molecules, the formation of biomonomers (amino acids, sugars), the formation of bipolymers (polypeptides, nucleic acids), the aggregation of bipolymers onto microspheres, and the emergence of protocells as functional relationships develop among microspheres (Wicken 1987). Therefore, life should be viewed as the sophisticated end in the continuum of development of natural dissipative structures from physical, to chemical autocatalytic, to living systems. Then, life, with its requisite ability to reproduce, insures that these dissipative (exergy destroying) pathways continue, and has evolved strategies to maintain these dissipative structures in the face of a fluctuating physical environment.

Wicken (1987) notes that living systems are a unique example of dissipative structures because they are self-creating, rather than a product of only impressed forces. In conclusion, the origin of life should not be seen as an isolated event but as a holistic process that represents the emergence of yet another class of processes whose goal is the dissipation of thermodynamic gradients through exergy destruction.

9.5.2 Ecosystems as exergy degraders

Following from Section 9.5.1, ecosystems can be viewed as the biotic, physical, and chemical components of nature acting together as nonequilibrium dissipative processes (Kay 2000b). As ecosystems develop or mature they should therefore develop more complex structures and processes with greater diversity, more cycling, and more hierarchical levels all to abet exergy destruction (Kay 1984; Kay and Schneider 1992; Schneider and Kay 1994).

Surface temperature as a possible indicator of ecosystem maturity

The energetics of terrestrial ecosystems provides an encouraging example of the thesis that ecosystems will develop so as to destroy exergy more effectively. The exergy destruction across an ecosystemis a function of the difference

between incoming and outgoing exergy flows (see section "A first look at the role of surface temperature"). Assuming that the exergy of incoming solar energy and the exergy of outgoing surface radiation emissions dominate ecosystem exergy flows, and assuming ecosystems are bathed by the same amount of incoming solar energy, then the most mature ecosystem should re-radiate its energy at the lowest exergy level, that is, the ecosystem should have the lowest surface temperature.

As previously reported (Schneider and Kay 1994) and previously documented (Luvall and Holbo 1991), the surface temperature of tropical forest, mid-latitude varied, and semiarid terrestrial ecosystems were measured using a Thermal Infrared Multispectral Scanner (TIMS). In keeping with the assumptions of the previous paragraph measurements are made within minutes of solar noon on a clear day. The one unmistakable trend is that the more developed ecosystem is cooler. For example, a grassland is warmer than an adjacent forest, or a young forest is warmer than an adjacent, more mature, forest. This is an encouraging result consistent with the exergy destruction principle.

A similar ecosystem surface temperature versus ecosystem maturity trend is seen in the work of Akbari (1995) where a more tightly controlled set of experiments than those of Luvall and Holbo (1991) were conducted. Control included making the measurements at the same solar time of day in the same area under no wind, clear sky, and consistent soil moisture conditions. A lawn (single species of grass) had the warmest temperature, an undisturbed hay field was cooler, and a field which has been naturally regenerating for 20 years was coldest. Also, another field that had been regenerating for 20 years was disturbed by mowing; after mowing, its surface temperature rose significantly, but very quickly returned to its cooler pre-disturbance temperature. These surface temperature observations are again consistent with the exergy destruction principle.

In conclusion, ecosystem surface temperature appears to correlate with ecosystem maturity as evidenced by some preliminary experiments, and thus appears to support the exergy destruction principle. However, the ultimate strength of this correlation, and the hypothesis that ecosystems develop in a way which systematically increases their ability to destroy incoming exergy, remains to be determined in future works.

9.6 Conclusions

A Second Law perspective in the engineering sciences has proven extremely useful to understanding and designing energy efficient, effective, and optimized thermodynamic systems. Fundamentally, ecosystems are complex thermodynamic systems concerned with the transformation of mass and

energy into usable forms, for example, work. As such, the Second Law of Thermodynamics has an important role to play in understanding ecosystem phenomena.

A central concept to a Second Law analysis is *exergy,* where exergy is defined as the *maximum useful to-the-dead-state work*. Through the GouyStodola theorem exergy destruction is seen to be a practical manifestation of entropy production. Exergy destruction, however, is more than simply a different way to represent entropy production because it inherently incorporates additional information about the system's environment.

Lessons learned from the exergy viewpoint explain, for example, how to greatly improve upon the conventional furnace, why a combined cycle power plant or cogeneration facility is inherently more efficient than a standard steam cycle power plant, or why an electric heat pump is superior to a radiant electric heater. These and other such lessons exist in engineering. Future ecosystem exergy studies should reveal similar lessons about how ecosystems are structured. In effect, an exergy analysis of ecosystems provides a new perspective from which to understand and characterize ecosystems, correspondingly, the *authors argue that exergy transport and exergy destruction have key roles to play in ecosystem characterization.*

In this chapter, it is demonstrated that *ecosystem surface temperature controls energy and exergy fluxes in and out of ecosystems*. Measuring an ecosystem's surface temperature is fundamental to the quantification of both energy and exergy fluxes. Therefore, thermal remote sensing has a major role to play in quantitatively characterizing ecosystems.

A dynamic relationship exists between an ecosystem and its environment that does not exist in engineering systems. By recognizing this, the authors discovered that the engineer's exergy concept is too restrictive or insufficient to be entirely satisfactory when used to characterize a complex thermodynamic system such as an ecosystem. This has lead the authors to identify, for the first time, six new exergy classifications: *intrinsic exergy, transport exergy, restricted exergy, accessible exergy (corollary: inaccessible exergy), restricted-access exergy,* and *extracted exergy.* Intrinsic exergy corresponds to the useful work potential of the energy stored within a system. *Intrinsic exergy provides a quantifiable measure for how far out of equilibrium with the environment a system happens to be.* Transport exergy corresponds to the SSSF useful work potential of the energy that *enters* a system; there is no exergy contribution from the energy stored within the system. Solar exergy is an example of transport exergy. Restricted exergy considers a system's structure and operational restrictions, for example, the maximum theoretical photosynthetic capacity of a plant restricts the exergy a plant can extract from the wavelengths of solar energy it has access to. Accessible exergy considers the system's capability of coupling to an exergy source, for example,

a plant can access solar exergy via photosynthesis whereas an animal does not have the same access to solar exergy. A system's reference environment defines the relevant work mode gradients and thermal gradients; all other gradients contribute to *hidden exergy.* The authors were pleased to discover that by clearly separating out and identifying these six exergy classifications they were apparently able to resolve a long standing controversy over the correct relationship to use to quantify solar exergy.

As a foundation to these exergy classifications, the authors have also extended the engineering concept of dead state and clarified the definition of environment by introducing seven new or refined definitions: *immediate surroundings, non-immediate surroundings, immediateenvironment, non-immediate environment, stable-equilibrium reference environment, stable-equilibrium dead state,* and *thermal-mechanical stableequilibrium dead state.* This clarification makes it clear that *communication with the environment is a central component of the exergy concept.* That is, do not forget exergy destruction in the immediate environment. Unfortunately, the authors have observed in the ecological literature instances where the immediate environment has been inappropriately ignored, and consequently, the exergy has been incorrectly defined.

Appendix A: exergy terminology conventions

Today, the term *exergy,* originally introduced in Europe in the 1950s, is being favored globally over the term *availability,* a term made popular in the United States by the MIT School of Engineering in the 1940s. It is being favored globally partly because it is shorter, it rhymes with energy and entropy, and it can be adapted without requiring translation (Cengel and Boles 1998).

The purpose of Tables 9.A1 and 9.A2 is to convey to the reader an appreciation for the need to clearly define the exergy terminology one is using. To assume a colleague implicitly defines a given exergy term in the same fashion as oneself is unwise as evidenced by the terminology variations identified in Tables 9.A1 and 9.A2. Table 9.A3 summarizes new, refined, or clarified exergy terminology used in this chapter.

Table 9.A1 Examples of variations in exergy terminology found in the literature

	Exergy	*Non-flow exergy*	*Flow exergy*	*Available energy*	*Non-flow availability*	*Flow availability*
This chapter	Exergy, or work potential	Non-flow exergy	Flow exergy	N/A	N/A	N/A
Bejan (1997)	Exergy (p. 108)	Non-flow exergy (p. 124)	Flow exergy (p. 128)	N/A	Non-flow availability (p. 123)	Flow availability (p. 127)
Black and Hartley (1991)	Availability (p. 303)	Closed-system availability (p. 305)	Open-system availability (p. 305), or Stream availability (p. 303)	N/A	N/A	N/A
Cengel and Boles (1998)	Exergy, or Useful work potential (p. 421)	Non-flow exergy (p. 434)	Flow exergy (p. 435)	N/A	N/A	N/A
Edgerton (1982)	Available energy (p.161)	Closed-system available energy, or Availability (p. 134)	Maximum work (p. 144)	N/A	N/A	Exergy or essergy (p. 131)
Li (1996)	Availability (p. 47)	Non-flow availability (p. 49)	Availability (p. 48)	N/A	N/A	N/A
Van Wylen *et al.* (1994)	Availability (p. 313)	Non-flow availability (p. 315)	Flow availability, or Exergy (p. 314)	N/A	N/A	N/A
Wark (1977)	Availability (p. 345)	Availability (p. 358)	Stream-availability (p. 364)	Available energy (p. 346)	N/A	N/A

Table 9.A2 Examples of variations in exergy concept terminology found in the literature

Source	*Useful work*	*Maximum work*	*Maximum useful work*	*Maximum useful to-the-dead-state work (i.e. exergy)*	Irreversibility
This chapter	Useful work	Maximum work	Maximum useful work	Maximum useful to-the-dead-state work	Exergy destruction, or lost work
Bejan (1997)	Available work (p. 113)	Reversible work (p. 111)	Reversible available work (p. 114), or maximum delivery of useful (available) mechanical power (p. 114)	Non-flow-exergy (p. 124), or flow exergy (p. 128)	Lost available work or lost exergy (p. 114)
Black and Hartley (1991)	Useful work (p. 305)	Reversible work (p. 296)	Reversible useful power (p. 298)	Closed-system availability (p. 304) or Open-system availability (p. 304)	Irreversibility Rate: Actual work − maximum work (p. 298), or equivalently Actual useful work − maximum useful work (p. 298)
Cengel and Boles (1998)	Useful work (p. 423)	N/A	Reversible work (p. 242)	Maximum reversible work (p. 420)	Irreversibility (p. 424)
Edgerton (1982)	Useful work (p. 2)	N/A	Available energy (p. 2)	Maximum useful work (p. 131)	Available-energy loss (p. 134)
Li (1996)	Useful work (p. 48)	Available work (p. 48)	Available useful work (p. 48)	Work potential (p. 49)	Irreversibility (p. 53)
Van Wylen *et al.* (1994)	Maximum reversible work − work on surroundings	Reversible work (p. 315)	N/A	Maximum available work (p. 315)	Irreversibility (p. 315)
Wark (1977)	Useful work (p. 356)	Maximum work (p. 366)	N/A	Maximum useful work (p. 356)	Irreversility: Actual Work − Maximum work (p. 366), or equivalently Actual useful work − Maximum useful work (p. 367)

Table 9.A3 New, refined, or clarified exergy terminology introduced in this chapter

	Exergy terminology	*Section*	*Status*
1	Maximum useful to-the-dead-state work	9.1 and 9.3.5[a]	New
2	Intrinsic exergy	9.3.5[b]	New
3	Transport exergy	9.3.5	New
4	Restricted exergy	9.3.5	New
5	Accessible exergy	9.3.5	New
6	Inaccessible exergy	9.3.5	New
7	Restricted-access exergy	9.3.5	New
8	Extracted exergy	9.3.5	New
9	Hidden exergy	9.3.5	New
10	Immediate surroundings (synonym: Local environment)	9.3.1	Clarified (clarified)
11	Non-immediate surroundings	9.3.1	Clarified
12	Immediate environment	9.3.1	Refined
13	Non-immediate environment	9.3.1	Refined
14	Reference environment	9.3.1	New
15	Stable-equilibrium environment	9.3.2	New
16	Stable-equilibrium reference environment	9.3.2	New
17	Stable-equilibrium dead state	9.3.2	New
18	Thermal–mechanical stable-equilibrium dead state	9.3.2	New
19	Maximum work	9.3.2, 9.3.5, and 9.4.1	Refined
20	System-centric viewpoint	Appendix E[c]	New
21	Isolated-system viewpoint	Appendix E[c]	New
22	Environment	9.3.1	Context dependent
23	Restricted dead state	9.3.2	Discontinued

Notes
a Also described in the section on "Control mass exergy balance equation."
b Also described in the section on "Example 2: Believe it or not, it is easier to boil ice than water."
c Also described in the section on "Do not forget the immediate environment."

Appendix B: nomenclature

a Blackbody radiation constant (7.565×10^{-16} J m^{-3} K^{-4})
c Speed of light (2.998×10^{8} m s^{-1}) or specific heat
d Exact differential
d Inexact differential
g Gravitational acceleration
h Plank's constant (6.626×10^{-34} J s)
k_B Boltzmann constant (1.380×10^{-23} JK^{-1})
m_i Mass of species i (kg)
n System constituents (kmol)
v Velocity (m s^{-1})

z Elevation (m)
A Area (m^2)
E, e Total energy, specific total energy (kJ, kJ kg^{-1})
H, h Enthalpy, specific enthalpy (kJ, kJ kg^{-1})
h^o Total enthalpy (kJ)
I Total number of species
KE Kinetic energy (kJ)
M Total system mass (kg)
M Mass flow rate (kg s^{-1})
M Mass flow rate of species i (kg s^{-1})
N_i Molar flow rate of species i (kmol s^{-1})
P Pressure (kPa)
p Any thermodynamic property
PE Potential energy (kJ)
Q Heat transfer (kJ)
Q Heat transfer rate (W)
Q_i Heat transfer rate from non-immediate environment, non-reference environment, system I (W)
S, s Entropy, specific entropy (kJ kg^{-1} K^{-1})
T Temperature (K)
T_i Temperature of non-immediate environment, non-reference environment, system i (K)
T_H High temperature (K)
T_R Room temperature (K)
U Internal energy (kJ)
V, v Volume, specific volume (m^3, m^3 kg^{-1})
v Velocity (m s^{-1})
W Work transfer (kJ)
W Work transfer rate (W)
P_S Entropy production (kJ K^{-1})
β Relevant work mode parameters
η Efficiency
η_I First Law efficiency
η_{II} Second Law efficiency
μ_i Molar chemical potential of species i (kJ kg^{-1})
μ_i Mass chemical potential of species i (kmol kg^{-1})
σ Stefan-Boltzmann constant (5.667×10^{-8} W m^{-3} K^{-4})
Φ_T Radiation energy transfer (kJ)
ψ_T Radiation entropy transfer (kJ K^{-1})
X, x Exergy, specific exergy (kJ, kJ kg^{-1})
$X°$ Total exergy (kJ)
X^i Intrinsic exergy (kJ)
X^t Transport exergy (kJ)
X^r Restricted exergy (kJ)

X^{a} Accessible exergy (kJ)
X^{ra} Restricted-access exergy (kJ)
X^{ex} Extracted exergy (kJ)
$\langle\ \rangle_{i}$ Species-weighted average
$\langle\ \rangle_{m}$ Mass-weighted average

Additional subscripts:
0 Reference environment
$1 \rightarrow 2$ From state 1 to state 2
1,2,… State point
q Radiation source
r System surface
CM Control mass
IN Transport into a system
Max Maximum
OUT Transport out of a system
words Words in a subscript are self-explanatory

Appendix C: exergy-conserving calculations

9.C1 Exergy-conserving furnace

In section "First and second Law efficiencies" a calculation of $\dot{Q}_{\text{Room, Max}}$ is needed. It is a two-step process to solve for $\dot{Q}_{\text{Room, Max}}$. Referring to Figure 9.2 for the temperatures used, first, equation (9.10) is used to calculate $\dot{W}_{\text{Max}}$ as follows:

$$\begin{aligned}\dot{W}_{\text{Max}} &= \left(1 - \frac{T_0}{T_{\text{Combustion}}}\right)\dot{Q}_{\text{Combustion}} \\ &= \left(1 - \frac{273\,\text{K}}{2{,}273\,\text{K}}\right)\dot{Q}_{\text{Combustion}} = 0.88\,\dot{Q}_{\text{Combustion}}\end{aligned} \tag{9.C1}$$

Second, $\dot{Q}_{\text{Room, Max}}$ is determined by equation (9.11) to be

$$\begin{aligned}\dot{Q}_{\text{Room, Max}} &= \frac{\dot{W}_{\text{Max}}}{(1 - (T_0/T_{\text{Room}}))} = \frac{0.88\,\dot{Q}_{\text{Combustion}}}{(1 - (273\,\text{K}/293\,\text{K}))} \\ &= 14.6(0.88\,\dot{Q}_{\text{Combustion}}) = 12.9\,\dot{Q}_{\text{Combustion}}\end{aligned} \tag{9.C2}$$

Observe that absolute temperatures must be used; do not use Celsius or Fahrenheit.

9.C2 Bringing ice to a boil

In section "Example 2: Believe it or not, it is easier to boil ice than water", a calculation of the maximum temperature, T_{Max}, that –20°C ice can achieve without the addition of the natural gas is needed and given below. Both ice and liquid water are assumed incompressible with constant specific heats. Referring to Figure 9.3(a), the differential work output from the reversible (Carnot efficiency) heat engine, as the ice is heated from T_1 = –20°C to T_2 = 0°C, is given by

$$\begin{aligned} đW_{1\to 2} &= \left(1-\frac{T}{T_0}\right) đQ_{0,1\to 2} \\ &= \left(1-\frac{T}{T_0}\right)(đQ_{1\to 2}+đW_{1\to 2}) \end{aligned} \tag{9.C3}$$

or by rearrangement and noting that $dQ_{1\to 2} = dU = Mc_{Ice}\,dT$ by the First Law of Thermodynamics

$$\begin{aligned} W_{1\to 2} &= \int_{T_1}^{T_2}\left(\frac{T_0}{T}-1\right)Mc_{Ice}\,dT \\ &= Mc_{Ice}\left[T_0\ln\left(\frac{T_2}{T_1}\right)-(T_2-T_1)\right] \end{aligned} \tag{9.C4}$$

where $W_{1\to 2}$ is the work output from state 1 to state 2, T is the instantaneous ice temperature, T_1 is the temperature at state 1, T_2 is the temperature at state 2, T_0 = 20° C is the environment temperature, Q_0 is the heat transfer from the environment, Q is the heat transfer to the ice, U is the internal energy of the ice, M is the mass of ice, and c_{Ice} is the specific heat of ice.

At 0°C the ice undergoes a constant temperature phase change from ice to liquid water. During this phase change the work output from the reversible heat engine is given by

$$W_{2\to 3} = W_{\text{Ice}\to\text{Liquid Water}} = Mh_{\text{Fusion}}\left(1-\frac{T_2}{T_0}\right) \tag{9.C5}$$

where $W_{2\to 3}$ is the work output from state 2 to state 3 and h_{Fusion} is the specific heat of fusion.

Observe that the pressure-volume work associated with the decrease in volume as the ice melts occurs at atmospheric pressure and hence is not useful work. Any useful work potential associated with this change in volume is accounted for through the enthalpy of fusion.

The remaining useful work potential is obtained as the liquid water is heated from 0°C to the environment temperature of 20°C, and is given by

$$W_{3\to4} = Mc_{\text{Liquid Water}}\left[T_0 \ln\left(\frac{T_0}{T_2}\right) - (T_0 - T_2)\right] \tag{9.C6}$$

where $W_{3\to4}$ is the work output from state 3 to state 4, and $c_{\text{Liquid Water}}$ is the specific heat of water. Equation (9.C6) is derived in the same fashion as equation (9.C3).

The specific heat of ice is c_{Ice} = 2.0 KJ kg^{-1} K^{-1} and of water is $c_{\text{Liquid Water}}$ = 4.2 kJ kg^{-1} K^{-1} (Reynolds and Perkins 1977), and the specific heat of fusion for water is h_{Fusion} = 334 kJ kg^{-1} (CRC 1978). Therefore, the total exergy[85] content of the -20°C ice is

$$\chi_{\text{Ice}} = W_{1\to2} + W_{2\to3} + W_{3\to4} = 4.6 + 22.8 + 1.4 = 28.8\,\text{kJ kg}^{-1} \tag{9.C7}$$

Once the ice's exergy has been extracted and stored, it can be used by the reversible heat pump system shown in Figure 9.3(b) to heat the now 20°C water to a maximum temperature of T_{Max}. The equation governing T_{Max} is given by

$$\begin{aligned} W_{4\to5} &= \int_{T_4}^{T_5}\left(1 - \frac{T_0}{T}\right) đQ \\ &= Mc_{\text{Liquid Water}}\left[(T_{\text{Max}} - T_0) - T_0 \ln\left(\frac{T_{\text{Max}}}{T_0}\right)\right] \end{aligned} \tag{9-C8}$$

where $W_{4\to5}$ is the work output from state 4 to state 5, T is the instantaneous water temperature, $T_4 = T_0$ is the temperature at state 4, $T_5 = T_{\text{Max}}$ is the temperature at state 5, Q is the heat transfer to the water. Furthermore, $W_{4\to5}$ is also the exergy of the hot water. Therefore,

$$W_{4\to5} = M\chi_{\text{Ice}} \tag{9.C9}$$

Equation (9.C9) explains why equation (9.C8) can be and is written as a work output calculation instead of a heat input calculation. That is, the energy flow magnitudes in operating a reversible heat pump are equal the energy flow magnitudes for a reversible heat engine except all flow directions are reversed.

Iterating equations (9.C8) and (9.C9) for T_{Max} yields a T_{Max} of about 88°C or 361K.

It is instructive to note that in calculating T_{Max} the mass, M, of the ice is not needed. This is reflective of the pseudo-property character of exergy.

Appendix D: Gibbs free energy of blackbody radiation

The internal energy, U, for a system consisting only of blackbody radiation of temperature T occupying a volume V is given by

$$U_{\text{Blackbody Radiation}} = aVT^4 \tag{9.D1}$$

where $a = (8p^5 / 15)(k_B / h^3c) = 7.565 \times 10^{-16}$ J m^{-3} K^{-4}, k_B is Boltzmann's constant, h is Plank's constant, and c is the speed of light in a vacuum.

The radiation pressure, P, corresponding to equation (9.D1) is given by

$$P_{\text{Blackbody Radiation}} = \frac{a}{3}T^4 \tag{9.D2}$$

and entropy, S, in volume, V, by

$$S_{\text{Blackbody Radiation}} = \frac{4}{3}aVT^3 \tag{9.D3}$$

Finally, Gibbs free energy, G, is defined by

$$G \equiv U + PV - TS \tag{9.D4}$$

Therefore, substitution of equations (9.D1)-(9.D3) into equation (9.D4) yields

$$G_{\text{Blackbody Radiation}} = 0 \tag{9.D5}$$

Equations (9.D1)-(9.D3) were taken from Bejan (1997).

Appendix E: distinguishing between the system-centric viewpoint and the isolated-system viewpoint

Figure 9.E1 schematically shows two approaches to making all processes reversible. Figure 9.El(a) represents the *system-centric viewpoint* while Figure 9.El(b) represents the *isolated-system viewpoint*. The importance of the isolated-system viewpoint is that it clearly reveals the *need* to consider the environment. For those familiar with the dS_i entropy production viewpoint[86,87] of irreversible thermodynamics (Prigogine 1955), Figure 9.El(a) is the dS_i viewpoint ($dS_i = dP_S = dP_{S,\text{System}}$; $dP_{S,\text{Immediate Environment}} = 0$). Figure 9.El(b) is the viewpoint that recognizes that something can possibly be done with energy flows that exits a system, that is, $dP_S = dP_{S,\text{System}} + dP_{S,\text{Immediate Environment}}$. For example, the exiting energy may have the potential to lift a weight. Energy magnitude is not the relevant factor for determining how high the weight can be lifted, energy quality is the relevant factor. The isolated-system viewpoint (Figure 9.Elb) represents an important paradigm shift from the system-centric viewpoint (Figure 9.Ela).

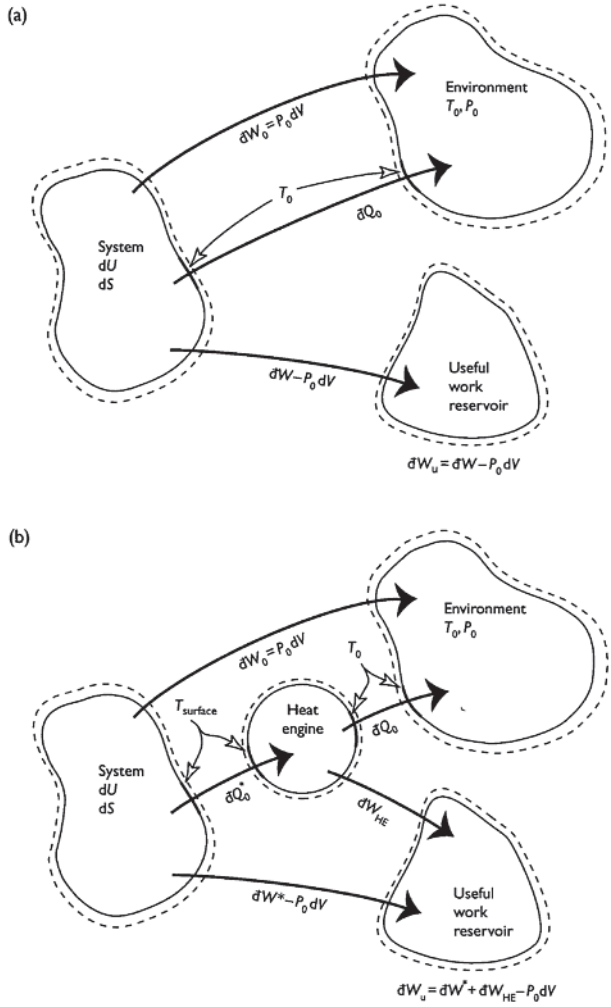


Figure 9.EI Two viewpoints for calculating system exergy: (a) the system-centric viewpoint; and (b) the isolated-system viewpoint.

One way to think of the difference between the system-centric viewpoint and the isolated-system viewpoint of exergy is as follows: the system-centric viewpoint internalizes all entropy production or exergy destruction processes; the isolated-system viewpoint maintains system boundaries. The isolated-system viewpoint then proceeds to make processes in the local environment reversible by the addition of a reversible heat engine.[88]

Does one expect these two viewpoints to yield the same exergy for the same initial system? Yes. Therefore, as stated in section "Do not forget the immediate environment", the exergy of a system does not depend on how the system and local environment are made reversible, only that they are made reversible.

The isolated-system viewpoint is the viewpoint adopted in this chapter.

Appendix F: solar exergy assumptions

The purpose of this section is to elaborate on the assumptions summarized in Table 9.4.

The following assumption comments of Table 9.4 are directed to the reader who wishes to be exposed to some of the finer details of exergy analysis in the presence of radiation energy transfer, and in turn acquire a flavor for one area of potential future research. Assumption and model numbering used below are consistent with those used in Table 9.4.

iii *To-the-dead-state work:* The appropriate dead state for radiation energy is not clear. Bejan (1987) managed to unify the various proposed equations in the literature for calculating the exergy of enclosed radiation by noting that any differences were the result of different initial states and different dead states for the radiation. The question then arrises, should the dead state for enclosed radiation be a volume of radiation at the environment temperature, or should it be an empty enclosure with no volume. There is as yet no consensus on the appropriate dead state for enclosed radiation. As for solar radiation, not only does it also suffer from an uncertainty on the correct dead state, there continues to be no unifying theory for open system radiation energy.

v *Steady-state, steady-flow:* Over a limited period of time this may be a good approximation, but the intensity ($W\ m^{-2}$) of solar radiation varies throughout the day and is zero at night. It would not be unexpected in the future to have this assumption/restriction modified in an alternative solar exergy calculation in order to account for the non-stationary nature of solar radiation.

vi *Ideally concentrated solar radiation:* Notice that the solar energy conversion system shown in Figure 9.8 does not include atmospheric background radiation, and that the solar radiation input is over an entire half-hemisphere solid angle even though the sun only subtends a solid angle of 6.8×10^{-5} steradians in the sky. These observations reflect an implicit maximization in the concentration of solar radiation. Authors such as Haught (1984) have considered the work potential of non-concentrated solar radiation, or diffuse solar radiation; however, the work output from non-concentrated solar radiation systems is

necessarily less than that for corresponding concentrated solar radiation systems.

vii *Diffuse blackbody radiation:* From an ecosystem's perspective the major problem with this assumption is that even if the spectrum for solar radiation striking the outer atmosphere is blackbody like, the spectrum of the solar radiation that actually reaches the ground is distorted by atmospheric gases such as water vapour (H_2O) and carbon dioxide (CO_2). The necessity of a blackbody (or graybody) assumption is that it enables one to conveniently express the entropy content of solar radiation as

$$\Psi_{\mathrm{T,\ Radiation}} = \frac{4}{3}\frac{\Phi_{\mathrm{T,\ Radiation}}}{T_{\mathrm{Radiation}}}$$

(Planck 1966).[89]

viii *Conduction or convection to the environment at temperature* T_0*:* This assumption simply takes advantage of the accepted thermodynamic notion of reversible heat transfer, for example, as assumed by a Carnot engine (Reynold and Perkins 1977). Reversible energy transfer to the environment is needed to maximize the work output from the system shown in Figure 9.8. The difficulties with reversible radiation energy transfer are discussed in assumptions (ix), (xi), and (xiii); it is these concerns that lead to the controversy over which maximum work output expression in Table 9.4 may be considered a correct expression for solar radiation exergy. In short, this assumption is a convenience that attempts to avoid additional controversy.

ix *Zero entropy production* (= 0): Two objections to this assumption can be made.

I One objection that may be raised concerning the expression in Model 1, if it is taken to represent solar exergy, is that it can be negative. In truth this is not a problem just as negative flow exergy is not a problem (Cengel and Boles 1998). Only intrinsic exergy is necessarily greater than or equal to zero. Model 1, however, is a measure of restricted exergy given the SSSF assumption (Assumption I), and hence, is not restricted from negative values; it is possible for work input to be required in order to maintain the SSSF operation.

II A second, much more serious objection, is that, in order for the entropy production to be zero, the incoming radiation cannot be absorbed reversibly by the system's radiation emitting surface at temperature T_{Surface} unless T_{Surface} equals T_{Solar}, in which case the device's efficiency is zero for a finite area system (T_{surface} equals $T_{\mathrm{Converter}}$ in Figure 9.8). It is well known that if the temperatures T_{Solar} and T_{surface}

are not equal then thermal radiation energy transfer is inherently irreversible (Wurfel 1982; De Vos and Pauwels 1983; Kabelac 1991). The problem then becomes how is the work potential of the incoming radiation to be extracted? In particular, the problem pertains to the difficulty one has in conceiving of a sub-system located between the incoming solar radiation and the outgoing surface radiation that captures all the incoming radiation through a reversible direct energy conversion phenomenon. The specific difficulty is that if *all* incoming solar radiation is *directly converted* to a work mode, then this surface must have an absorptivity of one. An absorptivity of one implies an emissivity of one, and an emissivity of one implies the presence of surface thermal emissions at $T_{\text{Converter}}$. Unfortunately, the original necessity of looking to a direct energy conversion device was to bypass any irreversible conversion to thermal energy by the direct energy conversion device. Photosynthesis and the photoelectric effect are examples of irreversible direct energy conversion phenomena. In short, the concern is that reversible radiation energy transfer to a finite size surface not at T_{Solar} may not be conceptually achievable.

x *Finite area:* The implications of this assumption for Model 1 are incorporated into the discussion of Assumption xiii. It is valuable to note that the Model 3 maximum work output expression derivation avoids the entropy production concerns of Assumption xiii by admitting an infinite area collector. Admitting an infinite area collector, however, simply substitutes one problem for another. For example, as noted by Bejan (1997: p. 476), the infinite area case is "totally unrealistic." The finite area assumption in Model 2 is not a problem since non-zero entropy production has been admitted through Assumption xi.

xi *Solar radiation is converted to thermal energy:* In Model 2, this assumption is realistic as a model, but from an exergy perspective it a priori admits defeat in the attempt to maximize the work output from the device shown in Figure 9.8. Defeat in the sense that it is a priori accepted that there must be non-zero entropy production due to the inherent irreversibility of converting radiation to thermal energy, as discussed in Assumption ix(II). Therefore, the maximum work output determined under the assumptions of Model 2 is necessarily a form of *restricted exergy.*

Unlike in Model 2, Assumption xi does not a priori admit defeat in Model 3. In particular, Model 3 represents that special case where radiation conversion to thermal energy is reversible, that is, when the

temperature of the solar radiation and surface are equal (see Assumption xii).

xii $T_{surface}$ *approaches* T_{Solar}: This assumption is identical to the assumption used to establish the special case situation when convection or conduction heat transfer is reversible (Reynolds and Perkins 1977). That is, convection or conduction heat transfer is only reversible when the temperature difference across which heat transfer takes place approaches zero. Given the acceptance of the concept of reversible heat transfer,[90] the $T_{surface}$ approaches T_{Solar} assumption is consistent with past practice, and hence appears reasonable.

xiii *Finite* F_{Net}: This assumption has *an analogy* in the requirement that a finite amount of convection or conduction heat transfer must continue to exist even as the temperature difference across which the heat transfer takes place necessarily approaches zero for reversible heat transfer (see Assumption xii). The assumption is not an identical requirement, however, for two reasons:

First, when Assumptions xii and xiii are combined one concludes that a finite amount of reversible thermal radiation energy transfer is only possible if the collector area approaches infinity. In contrast, reversible heat transfer is possible in two different limits, that of infinite area or of infinite time. The infinite time limitation is not admissible for reversible thermal radiation energy transfer because the rate of thermal energy transfer *to* the system is not set by a temperature difference (which can be made infinitely small in the case of heat transfer), but by the Stefan–Boltzmann law (Incropera and DeWitt 1996) that governs blackbody radiation emissions intensity.

Second, the infinite area limit implies the need for an infinite amount of thermal radiation input within any given time interval. As the area increases so does the magnitude of the incoming radiation intercepted by this area. One may argue that this second implication justifies Bejan's (1997: p.476) conclusion that the infinite area limit is "totally unrealistic," even conceptually. In contrast, Kabelac (1991) argues that the infinite area limit is conceptually acceptable. In short, the reversible radiation energy transfer limit imposed in Model 3 may not be conceptually realizable.

In closing this section it is informative to note that none of the three possible solar radiation exergy models summarized in Table 9.4 consider the possibility of a time lag between radiation energy input and radiation energy output. A time lag, however, does exist for ecosystems which emit thermal radiation day and night. This observation simply reflects the possibility that a more appropriate model quantifying the exergy of solar radiation may still need to be developed.

Notes

1 The "pseudo-property" descriptor applies to the concept of *useful work* as defined in Section 9.3. In brief, a true thermodynamic property is fixed by the thermodynamic state of the system. Useful work, on the other hand, is also a function of the environment. However, it behaves like a thermodynamic property for a *given reference environment,* hence the adjective pseudo-property.
2 The adjective *useful* is used to indicate that work done on the environment is not considered useful. For example, useful work can be used to turn the wheels of a car while work done on the environment cannot.
3 *Dead State* is defined in Section 9.3.2.
4 It is extremely important to note that the opportunity for terminology confusion abounds. For example, Wark (1977) and Van Wylen *et al.* (1994) refer to maximum useful to-the-dead-state work as *availability,* not as *exergy.* Furthermore, Wark (1977) refers to non-flow exergy as *availability* while Li (1996) refers to *flow exergy* as *availability;* Van Wylen *et al.* (1994) refers to *flow exergy* simply as *exergy.* Refer to Appendix A for a brief review of exergy terminology found in the literature. Appendix A is extremely important to understanding the need for clarity of communication. *In this chapter, exergy will refer to the idea of maximum useful to-the-dead-state work for* any *system.* When the exergy concept is necessarily applied in a more restricted sense a suitable adjective will be added. For example, the exergy that can be delivered by a steady-state, steady-flow, control volume system is termed *flow exergy.* In addition to Appendix A, refer to Table 9.2 for a list of exergy adjectives used.
5 By definition, exergy is measured in the same units as work transfer, namely the same units as used to measure energy magnitude.
6 It is easy to misinterpret results obtained using the Gouy-Stodola theorem because it hides much of the underlying physics. Refer to Section 9.3.3 for a discussion of the Gouy-Stodola theorem.
7 Symbol subscripts are summarized in Appendix B: Nomenclature. A subscript "0" is always used to refer to the environment or surroundings.
8 A dot above a variable indicates a rate term. That is, Q has units of power, for example, Joules/second or Watts.
9 This is an Annual Fuel Utilization Efficiency (AFUE). The US government's minimum AFUE rating for furnaces is 78%.
10 Steady-state is assumed in equation (9.2).
11 For clarity reasons a conservation of energy equation cannot be written without a corresponding system diagram (e.g. Figure 9.1) indicating work transfer and heat transfer energy flow directions. A negative work or heat transfer rate then simply means that energy transfer takes place in a direction opposite to that indicated in the system diagram.
12 A co-flow heat exchanger would have the room air and combustion gases flow in the same direction on opposite sides of the heat exchanger. A counter-flow system would have the room air and combustion gases flow in opposite directions (Incropera and DeWitt 1996).
13 This furnace is *exergy-conserving* in its *design philosophy.* Similarly, the furnace shown in Figure 9.1 is *energy-conserving* in its design philosophy since it attempts to transport 100% of $Q_{Combustion}$ to Q_{Room}. The exergy-conserving nature of the furnace shown in Figure 9.2 is established by noting that if all processes are reversible there is no loss in ability to do work, or in particular, no loss in *exergy.*
14 As will be seen, $Q_{0,Net} = Q_{0,\,Exhaust} - Q_{0,IN} < 0$. That is, on an energy quantity basis there is a net transfer of energy from the environment to the furnace as reflected

by the thicker arrow exiting the environment. Or put another way, the exergy-conserving furnace of Figure 9.2 refrigerates the environment.

15 A heat pump provides a means to actively transport thermal energy in a direction opposite to the natural heat transfer direction of hot to cold. However, the Second Law of Thermodynamics requires that heat transfer must always take place from hot to cold (Reynolds and Perkins 1977). The heat pump does not violate the Second Law of Thermodynamics because it simply creates *locally* the appropriate temperature gradient.

16 Conceptually, a heat pump is equivalent to a refrigerator in that they both pump thermal energy from a lower temperature to a higher temperature. They differ by purpose only; a heat pump utilizes the high-temperature thermal energy output from the system to heat while a refrigerator utilizes the low-temperature thermal energy input to the system to cool.

17 Had Question 1 asked what fuel savings would result for a given Q_{Room}, then the exergy-conserving furnace in the given example provides at least a 17% fuel savings over a conventional furnace.

18 A flawed efficiency may still be useful. If all furnaces whose efficiencies are being compared are similarly constructed, for example, conform to the system model of Figure 9.1, then the flawed efficiency given by equation (9.1) can provide a measure of the relative performance of the various furnaces.

19 First Law efficiencies are also referred to as *thermal efficiencies* (Wark and Richards 1999).

20 A reversible process is defined as a process that can be reversed without leaving any trace on the *surroundings* (Cengel and Boles 1998). That is, both the system and the surroundings can be returned to their initial states at the end of the reversible process.

21 A Carnot engine is not the only engine with a Carnot efficiency (Reynolds and Perkins 1977). All reversible heat engines operating between two temperature reservoirs possess the Carnot efficiency. For example, an ideal Stirling engine or an ideal Ericsson engine. This implies that there are many different engine design options available, even for maximum engine efficiency engines.

22 Had Question 2 asked what fuel savings or additional cost would result for a given QRoom then the reversible exergy-conserving furnace provides about a 93% fuel savings over the conventional furnace.

23 A *combined cycle power plant* is simply a power plant that combines both a gas turbine with a steam or Rankine cycle.

24 The "at night" constraint is given to emphasize that the only sources of energy are (a) the liquid water at 60°C, (b) the ice at –20°C, (c) the environment at 20°C, and (d) the natural gas at 20°C. For example, solar radiation energy is not available in this example.

25 See Appendix C for the calculation of this 88°C result.

26 Technically it is *intrinsic exergy* that is being discussed that is necessarily positive or zero. The "intrinsic" adjective technicality is discussed in Section 9.3.5.

27 It is important to understand that this observation is not referring to the exergy, or useful work potential, of material flowing through a system, but rather speaks only of the exergy within the system itself. *Intrinsic exergy* is formally defined in Section 9.3.5. In brief, it is the work potential of a system, not the work potential of material flowing into or out of a system, and not the work potential of a system constrained by size, time, and/or structure.

28 A *quasi-equilibrium* process, also referred to as *quasi-static* process, is an idealized process during which a system is internally infinitesimally close to a state of equilibrium at all times (Wark and Richardson 1999).

29 To *utilize* exergy well includes the ability to destroy exergy thoroughly. It does not mean that the only goal is to preserve exergy, that is, be efficient.
30 Energy *degradation* is distinctly different from energy *dissipation*. Energy dissipation is the loss in ability to do work due to entropy production. It represents only one of several methods to degrade energy. Energy degradation is a loss in ability to do work from the *perspective of the system utilizing the energy*. It represents a loss in ability to do work which may result due to dissipation, due to energy changing to a useless form of energy, or due to exergy that moves beyond the reach of the system. Energy degradation is also concerned with exergy density and time. Elaborating on the authors concept of energy degradation is beyond the scope of this chapter, but will be the subject of a future paper.
31 Entropy is a thermodynamic property just as temperature and pressure are thermodynamic properties. Whereas temperature is a macroscopic measure of random molecular kinetic energy and pressure is a macroscopic measure of the force per unit area generated by molecular collisions, entropy is a measure of molecular randomness or uncertainty (Reynolds and Perkins 1977).
32 Some ecosystems should possibly be excluded from this generalization. For example, those ecosystems found around deep ocean thermal vents (Stover 2000; Lutz and Kristof 2000).
33 For a discussion on the sometime subtle but important differences between system order, organization, and complexity, see Corning and Stephen (1998a, 1998b), Kay (1984, 2000b), and Schneider and Kay (1994).
34 Exergy analysis related engineering tools such as entropy generation minimization (Bejan 1995) remain beyond the scope of this chapter.
35 Exergy destruction and entropy production are related in the sense that they both correspond to a loss in ability to do work as evidence by the Gouy-Stodola theorem presented in Section 9.3.3
36 Photosynthesis-dependent life includes not only organisms that can photosynthesize, but also all those organisms dependent on photosynthesizing organisms through the food chain. For example, cattle, humans, lions, and butterflies.
37 Radiation energy transfer is distinctly different from convection heat transfer as clearly demonstrated by the entropy flux associated with radiation being 4/3 larger than the entropy flux associated with convection (see Appendix F, item vii) (Planck 1966). Consequently, the symbol Φ_T is used to distinguish radiation energy transfer from convection heat transfer, which is commonly represented by the symbol $\dot{Q}$. The T subscript on Φ_T indicates that radiation energy and temperature are associated through Planck's distribution.
38 By definition the authors consider *conduction* a special case of convection where fluid flow (advection) is zero. Convection is composed of both advection and thermal diffusion effects. Conduction is driven by thermal diffusion only.
39 Enthalpy (kJ) is a convenient thermodynamic property which by definition is equal to the fundamental thermodynamic properties of internal energy plus pressure times volume (Reynolds and Perkins 1977). Specific enthalpy is simply the enthalpy scaled by mass (kJ kg^{-1}).
40 *Shaft work* is all work transfer that crosses a system's boundary except *flow work* (Reynolds and Perkins 1977). Shaft work is represented by W, flow work is not.
41 Work is associated with a force acting over a distance. *Flow work* is associated with the pressure force acting over the distance traveled by the flow as it crosses the system boundary (Reynolds and Perkins 1977; Moran and Shapiro 2000).
42 The *Stefan-Boltzmann Law* for blackbody radiation is given by $\dot{\Phi}_{T_{Surface}} =$

$\sigma A T^4_{Surface}$, where $\dot{\Phi}_{T_{Surface}}$ is the radiation energy transfer rate, s is the StefanBoltzmann constant (W m^{-2} K^{-4}), and A is surface area. The spectral emissions from α real surface are far more varied than provided for by the Stefan-Boltzmann Law, however, for practical purposes, real surfaces are still governed by $T^4_{Surface}$ a dependence through the graybody assumption and the empirical measurement of a surface specific emissivity, ε (Incropera and DeWitt 1996).

43 The expression $q" = h(T_{Surface} - T_0)$ is known as *Newton's Law of Cooling* where $q"$ is the convection heat flux (W m^{-2}) and h is the convection heat transfer coefficient (W m^{-2} K^{-1}). h is frequently empirically determined (Incropera and DeWitt 1966).

44 *Stable elements* include O_2, N_2, H_2, and C. The stable form of an element is simply the chemically stable form of that element at the environment temperature and pressure. If an element exists in more than one stable form at the environment temperature and pressure then one of the forms should be specified as the stable form. For example, the stable form for carbon is assumed to be graphite, not diamond. For purposes of tabulating a chemical's ethalpy of formation, Gibb's free energy of formation, or exergy, a standard reference state for the environment of 25°C and 1 atm is generally selected (Rossini *et al.* 1952; Moran 1989; Cengel and Boles 1998).

45 The stable elements that constitute the chemical composition of the environment varies. One popular model for the composition of atmospheric air at 25°C and 1 atm sets the mole fractions for N_2 at 0.7567, O_2 at 0.2035, H_2O at 0.0303, CO_2 at 0.0003, and other gases at 0.0092 (Moran 1989). This is clearly only a model as the humidity is fixed.

46 Edgerton (1982) actually uses the term available energy in place of exergy; see Table A1.

47 See Appendix D.

48 *Helmholtz free energy* is also often loosely referred to simply as free energy.

49 A control mass system is a thermodynamic system with no mass flow across its boundary. In contrast, a *control volume* system is a system with mass flow across its boundary. A control mass system should not be confused with a closed system, nor a control volume system with an open system. A *closed system* has no energy flow across its boundary (and therefore no mass flow too), while an *open system* has energy flow across its boundary (which may or may not be associated with mass flow).

50 The authors are currently preparing a more detailed work on these new exergy classifications.

51 On this point the authors are adamant. Without an explicit system diagram people must implicitly assume system diagrams, but if assumed diagrams differ, confusion, misinterpretations, and errors result.

52 The First Law refers to the First Law of Thermodynamics; the Second Law refers to the Second Law of Thermodynamics. The First Law equation is also called the energy balance equation or Conservation of Energy equation; the Second Law equation is also called the entropy balance equation.

53 Energy transferred by radiation is referred to as *radiation energy transfer* or simply *radiation transfer.* It is not to be referred to as radiation *heat* transfer because doing so is one major reason people do not recognize that radiation transfer actually transfers a third more entropy with it than convection or conduction heat transfer (see Appendix F). This recognition problem exists because the vast majority of heat transfer texts deal only with First Law concepts.

54 Fortunately, a priori determination of the direction of radiation transfer is not difficult. The net radiation transfer to a surface is simply the sum of incoming radiation absorbed by that surface minus the radiation emitted by that surface. Therefore, all system diagram surfaces must be drawn indicating both incoming and outgoing radiation transfer unless radiation transfer is considered negligible, and hence not included in the system diagram.

55 A synonym for immediate surroundings is *local environment*. The term local environment is not used in this chapter because it merely distracts from clearly separating the idea of surroundings from the idea of environment. However, as a matter of technical clarity, there is no problem in using the term local environment.

56 Care is necessary when using the term *environment*. In addition to it being used to specifically refer to the *exergy reference environment,* it can also be used as a synonym for the *immediate surroundings*. The distinction must be made based on context.

57 Any additional $P\,\mathrm{d}V$ work (i.e. $(P - P_0)\,\mathrm{d}V$) can go into lifting a weight and is therefore useful.

58 Please note that *maximum work* is not necessarily *reversible work*. Reversible work is defined as the work transfer during a reversible process, that is, a zero entropy production process (Ps = 0). Maximum work may correspond to minimum theoretical entropy production, not zero entropy production; see Sections 9.3.5 and 9.4.1.

59 It is possible that there may be more than one system to choose from in the non-immediate environment as the reference environment. For example, a choice between atmospheric air and river water. A central concept of exergy is that it *maximizes* the work potential, therefore there exists the requirement that the appropriate non-immediate environment system to be selected to act as the reference environment be one that maximizes the exergy calculation.

60 The term *stable-equilibrium dead state* introduced here by the authors reflects a new refinement to the definition of dead state. The adjective "stableequilibrium" is based on the *stable-equilibrium state principle* (Gyftopoulos and Beretta 1991). In particular, the stable-equilibrium dead state yields an exergy expression in which the calculated exergy becomes a pseudo-property, that is, it is dependent only on the state of the system.

61 The *thermal-mechanical stable-equilibrium dead state* is actually referred to in the engineering literature as the *restricted dead state* (Bejan 1997). This is unfortunate given the need to introduce restricted exergy in Section 9.3.5, unfortunate because the restriction placed on the dead state is concept imposed while the restrictions placed on the exergy calculation in Section 9.3.5 are system imposed. The adjective "thermal-mechanical" was chosen for clarity reasons as it explicitly states the type of equilibrium to be reached. Therefore, in the exergy terminology introduced in this chapter, *restricted dead state is left undefined, never to be used*.

62 Without this assumption that species and phase concentrations be equal to that of the environment, it would be impossible to bring the system into complete thermodynamic equilibrium with its environment since, in this example, mass transfer is defined to be zero. A system in thermal and mechanical equilibrium with the environment, but not chemical equilibrium, is said to be in the *thermalmechanical equilibrium dead state*.

6 3 Exact differentials are dependent only on the end states of the process. For example, the change in a system's internal energy as that system proceeds from state 1

to state 2 is given by

$$\int_1^2 \mathrm{d}U = U_2 - U_1 \equiv \Delta U$$

Inexact differentials are dependent on both the end states of the process and the actual path or process followed. For example, one does not speak of the change in work transfer between states 1 and 2, but rather one speaks of the work transfer *from* states 1 to 2. Symbolically this is represented by

$$\int_1^2 đ\mathrm{W} = \mathrm{W}_{1\to 2} = \mathrm{W}$$

For convenience, the $1 \to 2$ subscript can be dropped if it does not create confusion. For example, when the system only proceeds from state 1 to state 2; it does not proceed onto a state 3. Of note, some practitioners use $d(\delta \mathrm{W})$ (Nicolis and Prigogine 1977) in place of dW. The authors discourage this as the δ does not indicate as clearly as the $1 \to 2$ subscript that the quantity in question is path dependent; however, it is recognized that the *d* notation is useful for perturbation studies.

64 The Greek letter chi *(X,χ)* will be used to represent exergy. *X* is total exergy in kJ and χ is specific exergy in kJ kg^{-1}. The letter chi is used since it resembles the Roman letter *x* whose pronunciation is the first syllable in the word exergy. The Roman letter *x* is not used due to its heavy use as a general mathematical variable.

65 A subtle extension of the non-constant T_0 and P_0 idea is the idea that the exergy concept must also admit the possible need to switch between non-immediate environment systems as the reference environment, even during a single exergy calculation. The ultimate goal in an exergy calculation is to maximize work, not to necessarily fix the reference environment.

66 The *stable-equilibrium state postulate* states that, "any property *P* can be written as a function of the form

$$P = P(E, n_1, n_2, \ldots, n_r, \beta_1, \beta_2, \ldots, \beta_s) = P(E, \boldsymbol{n}, \boldsymbol{\beta})$$

where the explicit dependencies of *P* on *E*, *n*, and *ß* are determined by the system, that is the constituents, the internal forces, the external forces, and the constraints (Gyftopoulos and Beretta 1991). *E* is energy, *n* is a constituent, *ß* is a relevant parameter such as volume, *r* is the number of different constituents, and *s* is the number of different relevant parameters.

67 Also known as non-flow exergy.

68 Many of these texts refer to exergy as availability.

69 While a *system-centric viewpoint* ignores entropy production processes in the surroundings, an *isolated-system viewpoint* considers the entropy production in the immediate surroundings. See Appendix E for further discussion on the systemcentric and isolate-system viewpoints and their relation to the $\mathrm{d}S_i$ viewpoint of irreversible thermodynamics.

70 Darwin (1859) was among the first to introduce a role for history into science.

71 Some people have taken issue with the term *lost work* as they believe the term is inaccurate or confusing since it may implicitly communicate to some that something "lost" may also be "found." The authors suggest those who dislike the term lost work use the term *permanently lost work*.

72 By equation (9.27), the conclusion that exergy can never be produced is as strong

and valid as the Second Law of Thermodynamics statement that entropy can never be destroyed.

73 Also referred to as entropy generation minimization.

74 A *fluid reservoir* is characterized as a non-immediate environment system that undergoes no shaft work or heat transfer, but can reversibly supply or accept a mass flow with no change to its specific properties.

75 Consider the system shown in Figure 9.6 used to calculate flow exergy. In calculating flow exergy the size of this system is irrelevant. That is, the stored exergy content of the system is irrelevant, and the SSSF assumption is how information about system content is removed from the calculation.

76 *Flow work exergy* is defined by $\chi_{\text{Flow Work}} = (P - P_0)v$ where P is pressure (kPa), v is the specific volume (m^3 kg^{-1}), and the subscript 0 refers to the reference environment (Cengel and Boles 1998).

77 As discussed in Section 9.3.2, the concept of *restricted exergy does not invoke the concept of restricted dead state*. For clarity of communication the terminology of restricted dead state must never be used, rather one speaks of thermal-mechanical stable-equilibrium dead state.

78 Relevant in relationship to what is important to the system in question. In the context of exergy, identification of the dead state implicitly identifies which work mode gradients are considered relevant. For example, if the dead state's reference environment is defined as the atmosphere composed of specific percentages of oxygen, nitrogen, carbon dioxide, water, etc., then implicitly nuclear reactions that could change the composition are neglected or considered non-relevant.

79 Each reversible work mode corresponds to a thermodynamic property gradient, e.g. $P\,dV$ work corresponds to a gradient in pressure. A non-reversible work mode is friction.

80 The classical Second Law of Thermodynamics distinguishes between allowable and impossible processes. In effect, it identifies which process directions are possible and which are not. It does not, however, require that a system spontaneously move towards a maximum entropy or equilibrium state. The *Spontaneous Equilibrium Principle* is additionally required. The Spontaneous Equilibrium Principle simply states that an isolated system *naturally and spontaneously* proceeds to a state of equilibrium. This separation of concepts is consistent with the Law of Stable Equilibrium presented by Hatsopoulos and Keenan (1965) and Kestin (1968), which combines the two concepts into one law.

81 Recall, order and organization are different (see section on "Exergy's role in characterizing ecosystems") (Corning and Stephen 1998a,b).

82 The solar radiation energy flux striking the earth is proportional to the earth's projected area normal to the direction of solar radiation, that is, that of a circle, and not the earth's actual area, that is, that of a sphere. Therefore, even if the earth's surface was fractal in nature and of infinite area, the solar radiation energy flux reaching the earth would be unchanged, remaining proportional to projected area. Therefore, an infinite area solar collector must necessarily extend beyond the earth into outer space.

83 In Kay (2000b), the *exergy destruction principle* was called the *exergy degradation principle*. The change of name to exergy destruction principle is necessary for the purpose of clarity since exergy destruction is clearly defined both in this chapter and in the literature, while exergy degradation is not.

84 First, Schrödinger used the concept of negentropy to effectively describe living systems as non-equilibrium dissipative systems. Second, and very importantly, *negentropy is not exergy destruction*! Third, today, negentropy is a non-

recommendable thermodynamic term. For example, negentropy has created considerable confusion suggesting that information processes negate the Second Law of Thermodynamics. The authors suspect that if Schrödinger were working today he might use exergy destruction instead of negentropy to describe his thoughts on life; unfortunately, the concept of exergy destruction had not yet been developed in 1944.

85 This is an intrinsic exergy calculation to the thermal-mechanical stableequilibrium dead state.

86 Specifically, dS_i is used to represent entropy production internal to the system only. Although the dS_i viewpoint does lead to the important conclusion that entropy production is greater than or equal to zero locally, it does not possess the practical, or engineering, viewpoint incorporated into exergy. A classic engineering example that demonstrates the linkage between system and environment entropy production is Chambadal's power plant (Chambadal 1957). In particular, if entropy production in the immediate environment is ignored an incorrect optimum exhaust temperature for the power plant can be obtained (Bejan 1995).

87 Technically dS_i is an inexact differential, however, it is written as dS_i instead of dS_i to be consistent with the irreversible thermodynamic literature.

88 A reversible heat engine, such as the Carnot heat engine, extracts the maximum work output from a heat transfer process. Its use is also the only way to reversibly transfer energy by heat transfer across a finite temperature difference, and hence, its introduction is unavoidable for any reversible process involving heat transfer across a finite temperature difference.

89 The $\frac{4}{3}$ factor often comes as a surprise to most engineers who have studied radiation heat transfer for years. This is because virtually all courses and texts on radiation heat transfer, some quite voluminous (Siegel and Howell 1992), tackle radiation heat transfer from a First Law of Thermodynamics perspective only! That is, the entropy of solar radiation and the Second Law of Thermodynamics is completely ignored. The Second Law or exergy perspective of radiation energy transfer is still in its infancy, even in engineering where the exergy concept is otherwise reasonably well developed.

90 Remember, *heat transfer* refers to convection or conduction heat transfer only, it does not refer to radiation energy transfer. Radiation energy transfer has unique considerations that differentiate it from convection or conduction heat transfer. For example, the $\frac{4}{3}$ factor discussed in Assumption vii.

References

Ahrendts, J. (1980) Reference states. *Energy Int. J. 5:667–77.*

Akbari, M.H. (1995) Energy-based indicators of ecosystem health. Master of Science, Department of Crop Sciences, University of Guelph, Guelph, Ontario.

Allen, T. (2000) Can we use energy based indicators to characterize and measure the status of ecosystems, human, disturbed and natural? Presentation at *Advances in Energy Studies Workshop: Exploring Supplies, Constraints, and Strategies,* Porto Venere, Italy, May 25.

ASHRAE (1996) 1996 *ASHRAE Handbook: HVAC Systems and Equipment.* American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), Inc., Atlanta, GA.

Bejan, A. (1987) Unification of three different theories concerning the ideal conversion

of enclosed radiation. *J. Sol. Energy Eng.* 109:46–51.
Bejan, A. (1995) *Entropy Generation Minimization: The Method of Thermodynamic Optimization of Finite-size Systems and Finite-Time Processes.* CRC Press, New York.
Bejan, A. (1997) *Advanced Engineering Thermodynamics,* 2nd edn. John Wiley, Toronto, Ontario.
Berlinski, D. (1986) The language of life. In J.L. Dasti and A. Karlquist (ed.) *Complexity Language and Life: Mathematical Approaches.* Springer-Verlag, Berlin.
Black, W.Z. and J.G.Hartley (1991) *Thermodynamics,* 2nd edn. HarperCollins, New York.
Bošnjakovic, F. and K.F.Knoche (1988) *Technische Thermodynamik,* Teil 1, 7th edn. Steinkopff-Verlag, Darmstadt.
Cambel, A.B. (1993) *Applied Chaos Theory: A Paradigm for Complexity.* Academic Press, New York.
Carson, Dunlop and Associates (2000) http://www.carsondunlop.com/reports/hieff.html, Toronto, Ontario, April.
Castans,M.(1976)Bases Fisicas del Aprovechamiento de la Energia Solar.*Rev. Geofis.* 35:227–39.
Castans,M.and S.M.Jeter (1983) Comments on "Maximum conversion efficiency for the utilization of direct solar radiation." *Sol. Energy.* 30: 293.
Cengel, Y.A. and M.A.Boles (1998) *Thermodynamics: An Engineering Approach,* 3rd edn. McGraw Hill, Toronto, Ontario.
Chambadal, P. (1957) *Les Centrales Nucleaires.* Armand Colin, Paris, pp. 41–58.
Cillieres, P. (1998) *Complexity and Postmodernism: Understanding Complex Systems.* Routledge, London.
Corbit, J.D. and D.J.Garbary (1995) Fractal dimension as a quantitative measure of complexity in plant development. *Proc. R. Soc. Lond. B.* 262:1–6.
Corning, P.A. and S.J.Kline (1998a) Thermodynamics, information and life revisited, Part I: "To Be or Entropy." *Syst. Res. Behav. Sci.* 15:273–95.
Corning, P.A. and S.J.Kline (1998b) Thermodynamics, information and life revisited, Part II: "Thermoeconomics" and "Control Information." *Syst. Res. Behav. Sci.* 15:453–82.
Darwin, C. (1859) *The Origin of Species.* John Murray, London.
De Vos, A. and H.Pauwels (1981) On the thermodynamic limit of photovoltaic energy conversion. *Appl. Phys.* 25:199–225.
De Vos, A. and H.Pauwels (1983) Comment on a thermodynamic paradox presented by P. Würfel. *J. Phys. C. Solid State Phys.* 16:6897–909.
De Vos, A. and H.Pauwels (1986) Discussion of "The second law efficiency of solar energy conversion." *J. Solar Energy Eng.* 108:80–3.
Edgerton, R.H. (1982) *Available Energy and Environmental Economics.* Lexington Books, D.C. Heath and Co., Toronto, Ontario.
Gell-Mann, M. (1994) *The Quark and the Jaguar: Adventures in the Simple and the Complex.* W.H.Freeman and Company, New York.
Glassman, I. (1987) *Combustion,* 2nd edn. Academic Press, New York.
Government of Canada (1996) *The State of Canada's Environment.* Ministry of the Environment, Ottawa, Ontario, Canada.
Gribik, J.A. and J.F.Osterle (1984) The second law efficiency of solar energy conversion. *J. Sol. Energy Eng.* 106:16–21.
Gyftopoulos, E.P. and G.P.Beretta (1991) *Thermodynamics: Foundations and Applications.* Collier Macmillan Canada, Toronto, Ontario.
Hatsopoulos, G.N. and J.H.Keenan (1965) *Principles of General Thermodynamics.* Robert E.Krieger Publishing Company, Malabar, FL.

Haught, A.F. (1984) Physics considerations of solar energy conversion. *J. Sol. Energy Eng.* 106:3–15.
Heywood, J.B. (1988) *Internal Combustion Engine Fundamentals*. McGraw-Hill Book Company, Toronto, Ontario.
Hogan, J. (1995) From complexity to perplexity. *Sci. Am.* June: 104–9.
Incropera, F.P. and D.P.DeWitt (1996) *Fundamentals of Heat and Mass Transfer,* 4th edn. John Wiley, Toronto, Ontario.
Jeter, S.J. (1981) Maximum conversion efficiency for the utilization of direct solar radiation. *Sol. Energy.* 26:231–6.
Jørgensen, S.E. and H.Mejer (1977) Ecological buffer capacity. *Ecol. Model.* 3: 39–61.
Jørgensen, S.E. and H.F.Mejer (1981) Exergy as key function in ecological models, In *International Symposium Energy and Ecological Modelling.* US International Society for Ecological Modelling, Louisville, Kentucky.
Jørgensen, S.E. and F.Müller (eds) (2000) *Handbook of Ecosystem Theories and Management,* CRC Press-Lewis Publishers, Boca Raton, FL.
Kabelac, S. (1991) A new look at the maximum conversion efficiency of black-body radiation. *Sol. Energy.* 46 (4): 231–6.
Kauffman, S.A. (1995) *At Home in the Universe: The Search for Laws of Complexity.* Viking Press, London.
Kay, J. (1984) Self-Organization in Living Systems. PhD Thesis, Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada.
Kay, J. (2000a) Can we use energy based indicators to characterize and measure the status of ecosystems, human, disturbed and natural? Presentation at *Advances in Energy Studies Workshop: Exploring Supplies, Constraints, and Strategies,* Porto Venere, Italy, May 25.
Kay, J. (2000b) Ecosystems as self-organizing holarchic open systems: narratives and the second law of thermodynamics. In S.E.Jørgensen and F.Müller (eds) *Handbook of Ecosystem Theories and Management.* CRC Press-Lewis Publishers, Boca Raton, FL, pp. 135–60.
Kay, J.J. and E.D.Schneider (1992) Thermodynamics and measures of ecosystem integrity. In D.H.McKenzie, D.E.Hyatt, and V.J.Mc Donald (eds) *Ecological Indicators, Volume 1*, *Proceedings of the International Symposium on Ecological Indicators,* Fort Lauderdale, Florida, Elsevier, pp. 159–82.
Kestin, J. (1968) *A Course in Thermodynamics*. Hemisphere Press, New York.
Krenz, J.H. (1984) *Energy Conversion and Utilization,* 2nd edn. Allyn and Bacon, Inc., Toronto, Ontario.
Landsberg, P.T. and J.R.Mallinson (1976) Thermodynamic constraints, effective temperatures and solar cells. *Int. Colloquium on Solar Electricity,* CNES, Toulouse, pp. 27–46.
Landsberg, P.T. (1986) An introduction to nonequilibrium problems involving electromagnetic radiation. In J.Casas-Vázquez *et al.* (ed.) *Recent Developments in Nonequilibrium Thermodynamics.* Springer-Verlag, Berlin, pp. 224–67.
Lennox Dimension Two-Stage Gas Furnace (2000) http://www.lennox.com/, Lennox Industries Inc., Richardson, TX, April.
Li, K.W. (1996) *Applied Thermodynamics: Availability Method and Energy Conversion, Combustion: An International Series*. Taylor and Francis, Washington, DC.
Lutz, R.A. and E.Kristof (2000) New eyes on the oceans: deep sea vents—science at the extreme. *National Geographic,* National Geographic Society, Washington, DC, pp. 116–27, October.
Luvall, J.C. (2000) personal communication. Jeffrey C. Luvall, GHCC/MSFC/NASA,

Marshall Space Flight Center, Huntsville, AL, April.
Luvall, J.C. and H.R.Holbo (1991) Thermal remote sensing methods in landscape ecology. In M.Turner, and R.H.Gardner (eds), *Quantitative Methods in Landscape Ecology.* Chapter 6. Springer-Verlag, Berlin.
Margalef, R. (1984) Ecosystems: diversity and connectivity as measurable components of their compliation. In *The Science and Praxis of Complexity.* United Nations University, Tokyo.
Mayer-Kress, G. (1986) *Dimensions and Entropies in Chaotic Systems.* SpringerVerlag, Berlin.
Moran, M.J. (1989) *Availability Analysis: A Guide to Efficient Energy Use.* ASME Press, New York.
Moran, M.J. (1999) Fundamentals of exergy analysis and exergy-aided thermal system design. In A.Bejan and E.Mamut (eds) *Thermodynamic Optimization of Complex Energy Systems.* Kluwer: Academic Publishers, London.
Moran, M.J. and H.N.Shapiro (2000) *Fundamentals of Engineering Thermodynamics,* 4th edn. John Wiley, Toronto, Ontario.
Morowitz, H.J. (1968) *Energy Flow in Biology: Biological Organization as a Problem in Thermal Physics.* Academic Press, New York.
Nicolis, G. and I.Prigogine (1977) *Self-Organization in Non-equilibrium Systems.* John Wiley, New York.
Nicolis, G. and I.Prigogine (1989) *Exploring Complexity: An Introduction.* W.H.Freeman and Company, New York.
Odum, H.T. and E.C.Odum (1976) *Energy Basis for Man and Nature.* McGraw-Hill Book Company, Toronto, Ontario.
Peliti, L. and A.Vulpiani (eds) (1988) *Measures of Complexity.* Springer-Verlag, Berlin.
Petela, R. (1964) Exergy of heat radiation. J. *Heat Transfer* 86:187–92.
Planck, M. (1966) *Theorie der Wärmestrahlung,* 6th edn. Barth Verlag, Leipzig.
Press, W.H. (1976) Theoretical maximum for energy from direct and diffuse sunlight. *Nature London* 264:734–5.
Prigogine, I. and J.M.Wiame (1946) Biologie et Thermodynamique des Phenomenes Irreversible. *Experientia* II: 451–3.
Prigogine (1955) *Introduction to Thermodynamics of Irreversible Processes,* 2nd edn. John Wiley, New York.
Reistad, G.M. (1980) Available energy utilization in the United States. In R.A. Gaggioli (ed.) *Thermodynamics: Second Law Analysis.* American Chemical Society, Washington, DC, pp. 93–109.
Reynolds, W.C. and H.C.Perkins (1977) *Engineering Thermodynamics.* McGraw-Hill Inc., Toronto, Ontario.
Ricard, J. (1999) *Biological Complexity and the Dynamics of Life Processes.* Elsevier Science, New York, NY.
Rossini, F.D., D.D.Wagman, W.H.Evans, S.Levine, and I.Jaffe (1952) *Selected Values of Chemical Thermodynamic Properties,* NBS Cir. 500, National Bureau of Standards, Washington, DC.
Schneider, E.D. and J.J.Kay (1994) Life as a manifestation of the second law of thermodynamics. *Math. Comput. Model.* 19 (6–8): 25–48.
Schneider, E.D. and J.J.Kay (1995) Order from disorder: the thermodynamics of complexity in Biology. In M.P.Murphy and L.A.J.O'Neill (eds) *What is Life: The Next Fifty Years. Reflections on the Future of Biology.* Cambridge University Press, Cambridge pp. 161–72.
Schrodinger, E. (1944) *What is Life?: The Physical Aspect of the Living Cell.* Cambridge University Press, Cambridge.
Siegel, R. and J.R.Howell (1992) *Thermal Radiation Heat Transfer,* 3rd edn. Taylor

and Francis, Philadelphia, PA.
Spanner, D.C. (1964) *Introduction to Thermodynamics*. Academic Press, London.
Stover, D. (2000) Creatures of the thermal vents. Ocean Planet Exhibition, Smithsonian Institute, Washington, DC.
Szargut, J., D.R.Morris, and F.R.Stward (1988) *Exergy Analysis of Thermal, Chemical and Metallurgical Processes*. Hemisphere, New York.
Tsatsaronis, G. (1999) Strengths and limitations of exergy analysis. In A.Bejan and E.Mamut (eds) *Thermodynamic Optimization of Complex Energy Systems*. Kluwer Academic Publishers, London.
Ulanowicz, R.E. (1997) *Ecology, the Ascendent Perspective*. Columbia University Press, New York.
Ulanowicz, R.E. and B.M.Hannon (1987) Life and the production of entropy. *Proc. R. Soc. Lond. B*. 232:181–92.
Van Wylen, G., R.Sonntag, and C.Borgnakke (1994) *Fundamentals of Classical Thermodynamics,* 4th edn. John Wiley, Toronto, Ontario.
Wark, K. (1977) *Thermodynamics*. McGraw-Hill, New York.
Wark, K., Jr. and D.E.Richards (1999) *Thermodynamics,* 6th edn. McGraw-Hill, Toronto, Ontario.
Weast, R.C. and M.J.Astle (eds) (1978) CRC Handbook of Chemistry and Physics, 59th edn. CRC Press, Boca Raton, FL.
Welford, W.T. and R.Winston (1989) *High Collection Nonimaging Optics*. Academic Press, Toronto, Ontario.
Weston, K.C. (1992) *Energy Conversion*. West Publishing Company, New York.
Wicken, J.S. (1987) *Evolution, Thermodynamics, and Information: Extending the Darwinian Program*. Oxford University Press, New York.
Würfel, P. (1982) The chemical potential of radiation. *J. Phys. C. Solid State Phys.* 15:3967–85.

Part III

Thermal infrared instruments and calibration

Chapter 10

Calibration of thermal infrared sensors

John R.Schott, Scott D.Brown and Julia A.Barsi

10.1 Overview and scope

This chapter deals with the radiometric calibration of thermal infrared (TIR) sensors from an end-to-end systems perspective. Our intention is to provide the basis for calibration of laboratory, field, and flight instruments. This is of obvious use to the operators of these instruments, but even if you are only using TIR image data from a satellite, it will be important in understanding how to convert that data to surface temperature values. Because of the increasing availability and use of many band systems, we will include manychannel sensors or spectrometers throughout our discussion; however, the approach is also valid for single-band instruments.

Our initial goal in most TIR remote sensing studies can often be simply stated as the need to identify the spectral emissivity and the kinetic temperature of each object (pixel) in the scene. Achieving this goal involves careful calibration of laboratory, field, and flight instrumentation, ongoing procedures to monitor this instrumentation, and algorithms to convert sensed data (i.e. digital counts) to the radiometric domain where we have established our calibration references.

Regrettably, calibration to the sensor reaching radiance using onboard blackbodies as illustrated in Figure 10.1(a) is only the first step in quantitative image analysis. The other three fundamental steps are conceptually illustrated in Figure 10.1(b)–(d). These steps consist of conversion of the sensorreaching radiance to the surface-leaving radiance (Figure 10.1(b)), separation of the surface-leaving radiance into an emitted and reflected component [calculation of the background component (Figure 10.1(c))], and finally separation of the emitted component into emissivity and temperature-driven components [i.e. solving for temperature and emissivity (Figure 10.1(d))]. In most cases these steps are not as easily separable as we have described them here, and we shall resort to a number of tricks to achieve our goal of measuring the temperature and spectral emission structure of the earth (cf. Gillespie *et al.* 1996). However, in all cases one common component prevails, that is the need for good radiometric calibration of laboratory field and flight instruments (cf. Guenther 1991).

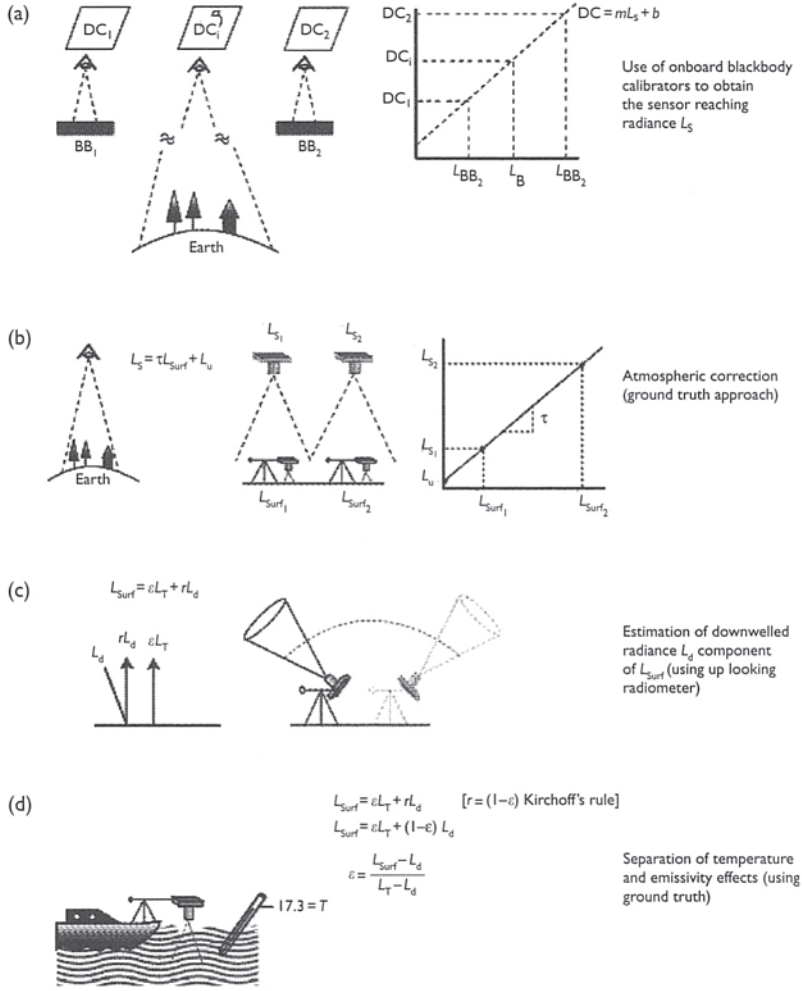


Figure 10.1 Steps in end-to-end system calibration.

10.1.1 Radiometric terms

We begin with a discussion of temperature. The *true* or *kinetic temperature* of an object is a result of the vibrational and translational motion of the atoms and molecules that make up the object. The kinetic temperature can be measured by direct contact with a chemical thermometer or electro-mechanical detector such as a thermopile. This approach allows the instrument to measure the temperature via conduction of the heat from the contact surface of the object. However, theoretically there exists a temperature

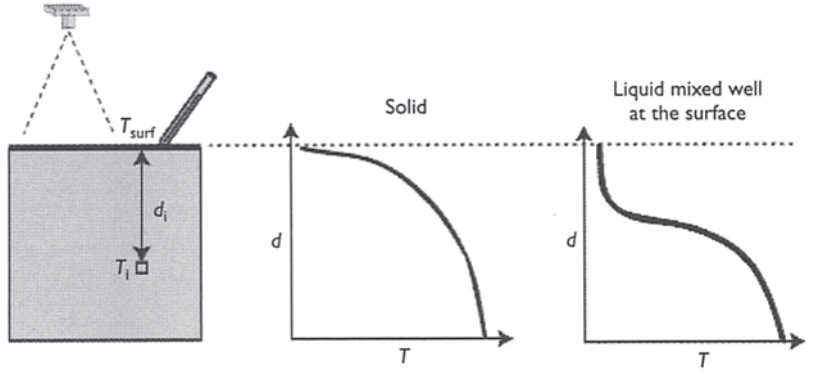


Figure 10.2 Temperature gradients with depth (d) exist within solids and liquids which vary depending on thermodynamic properties. The surface or skin temperature (T_{surf}) may not reflect the temperature of the bulk (T_i).

gradient within the object that is a function of the material's thermal conductivity. (Figure 10.2).

We must, therefore, ask which temperature we wish to measure. Typically, we are interested in the bulk or average temperature of the object. However, for materials with lower thermal conductivities the temperature gradient through the bulk will be greater, and the surface or the skin temperature will not be indicative of the bulk temperature. This issue regarding the actual temperature being measured will be very important in our discussions pertaining to calibration standards and standard monitoring.

In addition to contact or conductive measurements, the temperature of an object can also be remotely sensed by measuring the radiation emitted by the object. Recall that the radiance from a perfect radiator or blackbody is described by the Planck equation, and is expressed as

$$L_{BB\lambda}(T) = \frac{2\hbar c^2}{\lambda^5}(e^{\hbar c/\lambda kT} - 1)^{-1} \tag{10.1}$$

where L_λ is the spectral radiance (W m^{-2} μm^{-1} sr^{-1}), h is Planck's constant (6.6256×10^{-34} Js), c is the speed of light (3×10^8 m s^{-1}), λ. is wavelength (m, nm, or μm), k is the Boltzmann gas constant (1.38×10^{-23} K^{-1}), and T is the surface temperature (K). However, the perfect radiator is an idealized concept, and radiance measured from a material at a known temperature is usually less than the blackbody radiance. This observation gives rise to the measured radiance equation, which is expressed as

$$L_\lambda(T) = \varepsilon(\lambda)L_{BB\lambda}(T) \tag{10.2}$$

where $L_{BB\lambda}(T)$ is the Planckian radiance from a blackbody at the temperature$T$ of the object observed. The *spectral emissivity* ($\varepsilon(\lambda)$) is a materialdependent radiation property that indicates how efficiently the surface emitscompared to an ideal radiator. Because the emissivity is a material-dependentproperty, it is often more important than the temperature for materialmapping and identification studies.

At this point, we can define another commonly used temperature metric called the *apparent temperature, brightness temperature,* or *radiometric temperature*. The apparent temperature of an object is the kinetic temperature which a perfect radiator would be required to maintain, to generate the radiometric signal measured from the object.

10.1.2 Justifying calibration

The basic goal of instrument calibration is to relate instrument measurements to the instrument reaching radiance. If this can be accomplished to a high degree of certainty, then other techniques can be applied to transform these measurements to physical properties of the object being sensed (primarily, temperature, and emissivity). We will achieve these goals by discussing the use of lab (primary) and field (secondary) source standards to inject known radiances into the instrument so that the corresponding measurements can be calibrated. The calibration of these instruments can be broken down into two processes: the *radiometric calibration* which verifies the instrument's ability to correctly measure the magnitude of incident radiation and the *spectral calibration,* which verifies the ability to discern the spectral distribution of the incident radiation. In operation, if we look regularly at a pair of sources with known radiance and record the image level (digital count) they produce, then we have an end-to-end system calibration (assuming linearity). With these data, we can convert any digital count in an image to an observed radiance level over that spectral channel (cf. Figure 10.1(a)). The process can be repeated for each spectral channel. The spectral bandpass must also be determined as discussed in Section 10.2.2.

10.2 Lab calibration

Calibration in the TIR relies almost exclusively on the use of radiational source standards. In the visible and near-infrared (VNIR) spectral regions, there is an ongoing migration in the standards community toward the use of detector-based standards. This is driven by the inherent stability of modern VNIR detectors. It is the lack of a similar temporal stability in thermal imaging detectors that forces the use of source-based standards and also drives much of our calibration strategy. Because all field and flight instruments rely on the use of reference standards, we will begin our discussion with a treatment

of calibration source standards. Finally, in closing this section, we should point out that, while we will emphasize source standards, there is a growing use of detector standards in the form of electrical substitution radiometers for very precise work in standards laboratories (cf. Wolfe 1998).

10.2.1 Radiometric standards

The type of source we will be most concerned with in TIR calibration is the blackbody. This is a source that approximates a perfect radiator (i.e. $\varepsilon = 1$) and, as a result, the spectral radiance is described by the Planck function (equation 10.1). In principle, our standardization process is simplified (at least conceptually) to a temperature standard (i.e. if we know the temperature of the blackbody, we know its spectral radiance). In fact, we can only approximate a blackbody (and there are many ways to do so) and only approximately know its surface kinetic temperature. The following paragraphs discuss various blackbody designs (cf. Figure 10.3) and their respective performance attributes for our applications.

For the most precise work done in the laboratory, *melt-point blackbody* standards (Figure 10.3(a)) are used. These blackbodies are typically cylindrical or conical cavities open at the end to allow observation into the cavity. The cavity walls are made of low reflecting material (i.e. highly emissive) and since no flux can leave the cavity without bouncing from the walls several times the effective emissivity is very close to 1 (emissivities of 0.9999 are common for National Institute of Standards and Technology (NIST) traceable melt point blackbodies). The cavities are made of a thin-walled thermally conductive cone surrounded by a very pure elemental material (e.g. cesium). The standard material is maintained at its melting point by a separate set of thermal controllers and thermal monitors. Because of the heat of fusion, this is a very stable temperature location and our knowledge of the cavity temperature is largely limited by the purity of the material used as the transition material. The radiance from these sources can be known very accurately, and they can be used as *primary standards*. They have several limitations, three of which make them impractical for day to day use in most laboratories. They are expensive, limited to one temperature (radiance level), and have a small useable size (i.e. aperture), making them difficult to use directly with large aperture, low resolution systems. They also tend to be quite large, which limits their use in some applications.

In order to achieve a range of temperatures, multiple blackbodies are required with the cavities controlled by the melting point or boiling point of different materials. An alternative approach is to utilize a *thermally controlled blackbody* that can be adjusted through a range of temperatures (Figure 10.3(b)). This can be done by controlling the boiling point with the pressure of an inert gas over the fluid. The cavity will be very stable at the liquid to gas transition temperature. By carefully monitoring and controlling the vapor

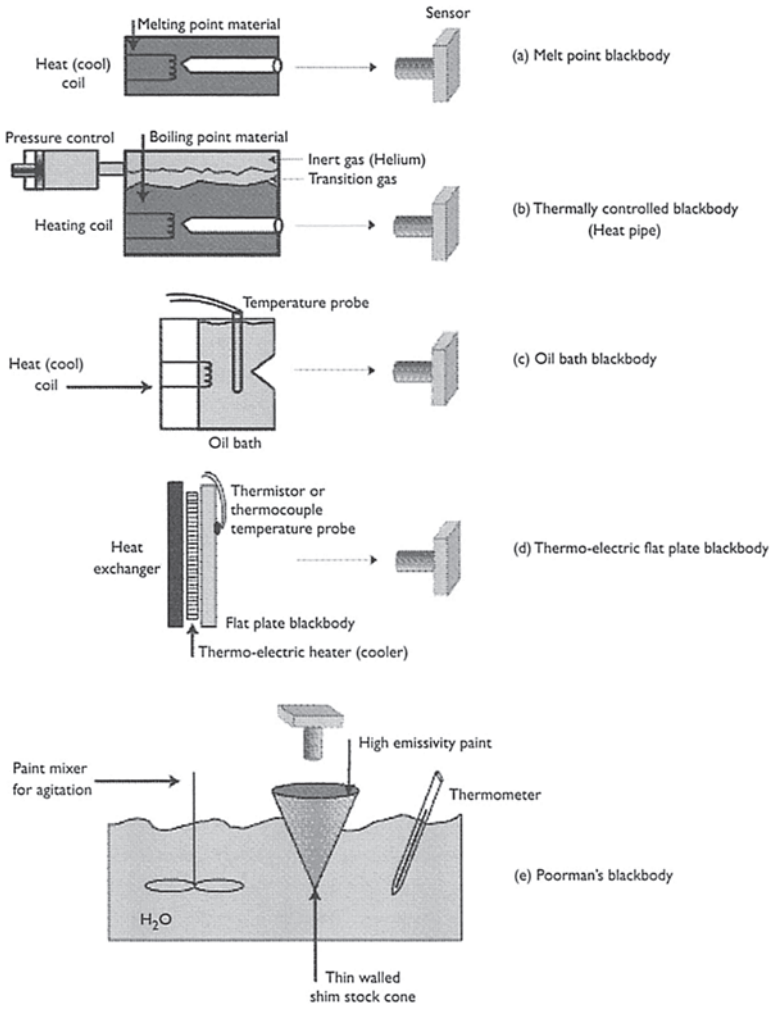


Figure 10.3 Illustration of common types of blackbodies.

pressure, the boiling point temperature can be controlled over a wide range and still be known very accurately. Typically, the controlled temperature cavities have emissivites of (0.999) and the temperature uncertainty is of the order of 0.1 K or better. These sources still suffer the limitations of high cost, large physical size, and small useful source size (e.g. 1–2 cm aperture).

A more cost-effective alternative for common use in the laboratory is the *liquid bath blackbody.* These use a temperature controlled insulated bath filled with a circulating fluid (usually oil, hence the common name *oil bath*

blackbody (Figure 10.3(c))). The fluid is in thermal contact with a thin walled cone, the outside of which is coated with a highly emissive material (typically a special paint). The bath temperature is carefully monitored with a bridgetype thermometer immersed in the circulating liquid. This type of blackbody is reasonably affordable, can have a larger surface area (although very large sources are difficult to build because of thermal uniformity and space logistics), and can cover the range of temperatures needed for most earth observation work. They are still somewhat large and the fluid circulation systems make them impractical for many field and most flight operations. The instruments in daily use at the Rochester Institute of Technology (RIT) have emissivities of about 0.995 and temperatures uncertainties of approximately 0.05. They have the marked advantage of reasonable cost, ease of use, and source sizes that are sufficiently large enough to eliminate lengthy and costly alignment time during calibration setup. As a result, they are commonly used for many day to day operations with the more exotic sources only used periodically to update the oil baths.

In standards jargon, the melt point blackbodies are used as *primary standards* and the oil baths as *secondary standards*. Rigorously speaking, even the melt-point blackbodies are secondary standards since they are typically calibrated to the primary melt-point blackbody at NIST.

For field or in-flight calibration of instruments, a *thermo-electric flat plate blackbody,* is commonly used (Figure 10.3(d)). These standards utilize thermo-electric heating/cooling devices to control the transfer of heat between a high conductivity flat plate and a heat exchanger. The plate is typically coated with a special paint to increase the emissivity. To increase further the effective emissivity, the plate surface may be grooved (pyramidal) or covered with a honeycomb (waffle). To monitor the surface temperature of the radiation surface, thermistors or thermocouples are placed directly into and/or on the surface. Flat plate blackbodies are widely used because they do not utilize liquids that may be spilled in the rough environment of a field collection or in an aircraft. Additionally, these devices can be made very compact and can be oriented at various angles (which liquid-type blackbodies cannot) making them more appropriate as internal calibration sources for field and flight instruments.

The more impoverished reader may want to consider the *poorman's blackbody* (Figure 10.3(e)). It consists of a simple thin walled metallic cone (we make them out of shim stock) painted with a high emissivity paint submerged in a water bath. If the water bath is well circulated, then the blackbody cone should be at the temperature of the water. The limitations of this approach are that in its simplest form, the blackbody can only be viewed vertically, the temperature range is limited (though it is acceptable for most earth observation) and the emissivity of the blackbody may deviate significantly from one. An even simpler approach involves just using a well-mixed water bath and taking advantage of the high spectrally flat emissivity of water

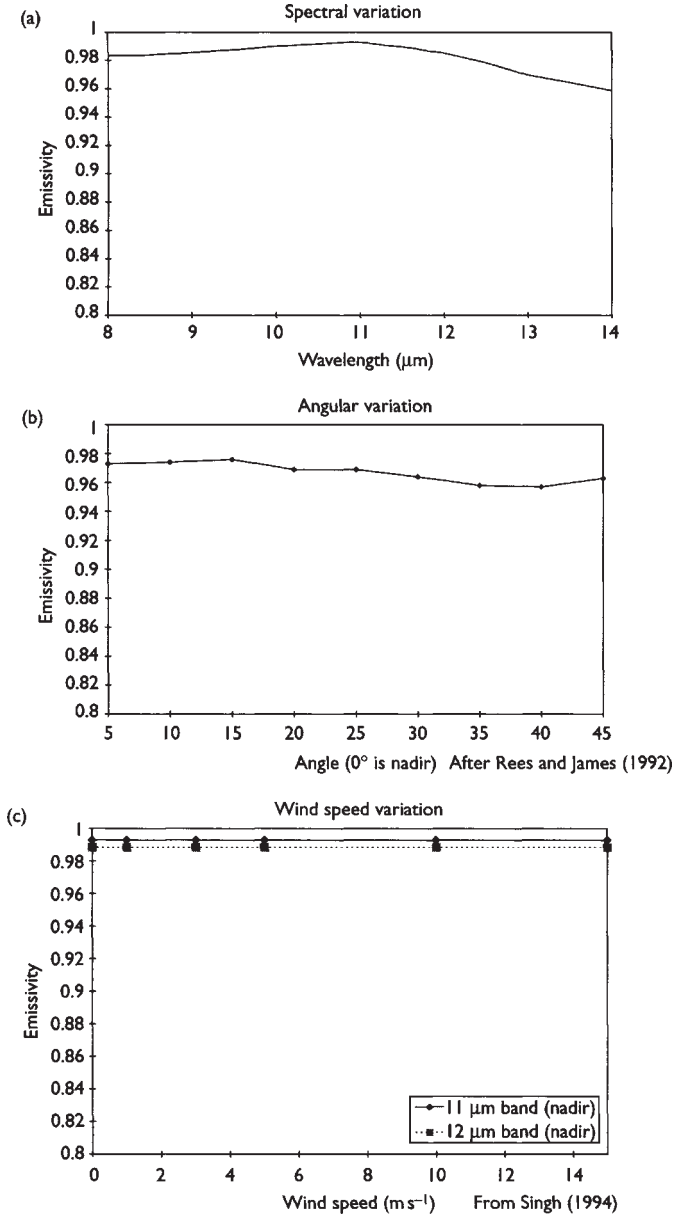


Figure 10.4 Plots showing the emissivity of natural water as a function of (a) wavelength, (b) view angle, and (c) wind speed. The data in (a) are for normal viewing. The data in (b) are for the 8–14 μm spectral range. The data in (c) are for 1 μm wide bands.

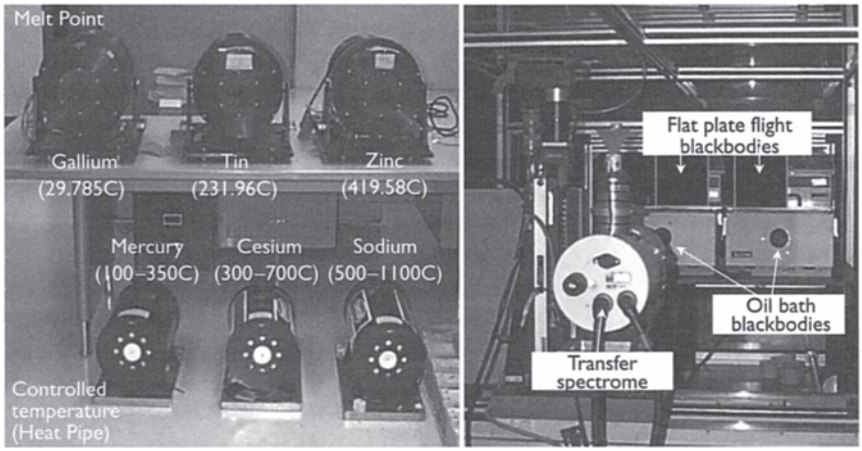


Figure 10.5 Photos showing various types of blackbodies. Images courtesy of Rochester Institute of Technology's Digital Imaging and Remote Sensing Laboratory.

across most of the electromagnetic spectrum (cf. Figure 10.4). This approach eliminates any decoupling of the skin temperature of the blackbody from the water temperature. Clearly, the water bath approach is not very attractive for flight instruments, but it can be very useful in the field, particularly as a backup if other equipment fails. Figure 10.5 shows photographs of several types of blackbodies.

Table 10.1 gives a quick summary of the expected errors in calibration of various types of blackbody sources. The errors associated with the use of a blackbody are very much a function of the environment in which the measurements are taken. This is because the largest unknown or unaccounted error is typically the reflected-radiance from the surround. Let us consider several ways to calculate the "known" radiance from a blackbody. In the simplest case, re-expressing equation (10.1), we would assume the blackbody was truly black and the temperature was known. In this case, the spectral radiance would be

$$L_\lambda = L_{\mathrm{BB}\lambda}(T) \quad (\mathrm{W\,m^{-2}\,sr^{-1}\,\mu m^{-1}}) \tag{10.3}$$

The effective radiance in a particular bandpass would be

$$L_i = \int R_i' L_{\mathrm{BB}\lambda}(T)\,\mathrm{d}\lambda \quad (\mathrm{W\,m^{-2}\,sr^{-1}}) \tag{10.4}$$

where R_i' is the peak normalized spectral response over the bandpass of interest (i.e. for the ith spectral band). Many times the effective spectral

Table 10.1 Errors associated with various blackbody sources and calibration equations.[a] All radiance values are expressed in terms of apparent temperature (K)

Type of instrument	*Temperature uncertainty (K)*	*Emissivity*	*Emissivity uncertainty*	*Radiance error (s_i) using equation (10.7′)*	*Radiance error using equation (10.3′) (s_{x-3})*	*Radiance error using equation (10.3) and cooler background (s_{x-3})*
Melt-point blackbody	0.01	0.9999	0.0001	0.01	0.01	0.01
Controlled temperature precision cavity radiator	0.10	0.9981	0.001	0.10	0.11	0.14
Oil bath blackbody	0.05	0.995	0.005	0.10	0.14	0.33
Flat plate blackbody	0.10	0.96	0.008	0.18	0.76	1.89
Poorman's blackbody	0.10	0.98	0.005	0.14	0.39	0.96
Water as a blackbody	0.10	0.985	0.01	0.21	0.35	0.84

Notes

$$L_{X-3} = L_{BB}(T_{BB}) \tag{10.3′}$$

$$L_i = \varepsilon L_{BB}(T_{BB}) + (1-\varepsilon)L_{BB}(T_B) \tag{10.7′}$$

$$s_i = \left(\left(\frac{\delta L_i}{\delta \varepsilon}\right)^2 s_\varepsilon^2 + \left(\frac{\delta L_i}{\delta L(T_{BB})}\right)^2 s_{L(T_{BB})}{}^2 + \left(\frac{\delta L_i}{\delta L(T_B)}\right)^2 s_{L(T_B)}{}^2\right)^{1/2}$$

bias errors between equations (10.3′) and (10.7′):

$$s'_{x-3} = L_i - L_{x-3}$$

$$s_{x-3} = (s_i^2 + s'_{x-3}{}^2)^{1/2}$$

a *Assumptions:* TBB = 320 K, $\Delta\lambda$ = 8–14 μm, and background temperature T_B = 300 ± 1 K or T_B = 260 ± 1 K for the last column. All radiance errors are expressed as apparent temperature with $\Delta L/\Delta T$ computed for a 300 K source. s_ε, $s_{L(T_{BB})}$, and $s_{L(T_B)}$, and are the uncertainties in emissivity, radiance due to uncertainty on blackbody temperature, and radiance due to uncertainty in background temperature. The last two columns incorporate the bias errors associated with using equation (10.3′) to approximate equation (10.7′).

radiance for a particular bandpass is more useful. It can be expressed as

$$L_{i\lambda} = \frac{\int R'_i L_{BB\lambda}(T)\,d\lambda}{\int R'_i\,d\lambda} \quad (\mathrm{W\,m^{-2}\,sr^{-1}\,\mu m^{-1}}) \tag{10.5}$$

For convenience, we will express most of our calibration equations in terms of spectral radiance, but recognize that any radiometric expression can be converted to effective radiance or effective spectral radiance by weighting by the appropriate responsivity expression. Because most of our blackbodies are not truly black, we need to modify equation (10.3) to account for emissivity, that is, re-expressing equation (10.2),

$$L_\lambda = \varepsilon(\lambda) L_{BB\lambda}(T) \tag{10.6}$$

Using equation (10.3), we would calculate too much radiance coming directly from the blackbody. Using equation (10.6) is more complete, but it is also an approximation in that it neglects the reflected radiance term. Thus, the most appropriate expression for the spectral radiance from a calibration source can be expressed as

$$L_\lambda = \varepsilon(\lambda)L_{BB\lambda}(T) + [1 - \varepsilon(\lambda)]L_{BB\lambda}(T_b) = \varepsilon(\lambda)L_{BB\lambda}(T) + r(\lambda)L_{BB\lambda}(T_b) \tag{10.7}$$

where we have used Kirchoff s rule to express the reflectance of the blackbody as $r(\lambda) = 1 - \varepsilon(\lambda)$ and we have assumed that the background radiance can be approximated by the spectral radiance from a blackbody having the temperature of the background (T_b). Simple examination of equation (10.7) shows that if this is the more correct expression, then equation (10.6) will always underestimate the radiance and equation (10.3) may over or under the estimate radiance depending on the temperature of the blackbody relative to the background. In fact, if the blackbody and the background are at the same temperature, then equation (10.7) and (10.3) yield the same results. The importance of these approximations is shown in Table 10.1. The first column labeled radiance error describes the error in the knowledge of radiance from the blackbody if equation (10.7) is used. It reflects errors due only to uncertainties in the input parameters (T, T_b, and ε). The last two columns include the bias errors due to the common practice of approximating the radiance using equation (10.3) instead of rigorously using equation (10.7) (cf. Moeller *et al.* 1996). Two cases, one with a background relatively close to the blackbody temperature, and one with a background with quite a different temperature, simulating a cold sky are presented. Because most of us cannot think in radiance units, it is often more convenient to work in apparent temperature. This is the temperature a perfect blackbody would have to be at to generate the radiance observed. The errors in Table 10.1 are expressed for convenience in units of apparent temperature or more rigorously the change in temperature needed to generate the corresponding change in radiance. Since the change in radiance per unit change in temperature varies with temperature, we use changes relative to a 300 K source for these illustrations. It is clear that "blacker" blackbodies and those with surround temperatures close to the target temperature simplify the problem and reduce errors. It is also clear that in many cases we need to use the full rigor of equation (10.7).

Finally, it is important to consider, as we proceed, what degree of calibration is necessary for a particular task. The cost in terms of instrumentation, manpower, and time increases significantly if very small temperature errors are required. Most studies need to evaluate what temperature/emissivity knowledge is required for the particular application. Then, an error propagation study can predict the level of instrument calibration required

and from there the laboratory and flight calibration errors that can be tolerated.

10.2.2 Spectral standards

To this point, we have emphasized only radiance levels and the use of source standards. We should point out that there is also a need to perform wavelength calibration of most instruments. The spectral calibration consists of characterizing the relative spectral response of each channel in the imaging sensor as a function of wavelength. Typically, this is done by placing a continuous source like a hot blackbody at the entrance aperture of a monochrometer. The monochromatic energy exiting the monochrometer is used to irradiate the sensor, usually through an optical collimator. By scanning the monochromator through a range of wavelengths, the relative response of the imager as a function of wavelength can be determined. This assumes that the relative source radiance (i.e. source temperature) is known, along with the relative throughput of the monochromater-collimator combination. In order to verify the wavelength calibration of the monochromator, sources with a well-known narrow line structure are required. One way to do this is to use a line source (e.g. a CO_2 laser). Another approach is to use a filter to selectively pass or absorb only a narrow wavelength range from a broadband source. Because of their narrow absorption features, transparent cells filled with a gas with very well-defined spectral transmission can be used for this purpose.

10.2.3 Use of transfer standards to calibrate field or flight blackbody sources

You will typically need to transfer information about your laboratory source calibration to field or flight blackbodies for more operational use. Often, size, space, weight, and electrical power requirements drive us toward some form of flat plate thermo-electrically controlled blackbody for operational instruments. In order to calibrate these field units, we will use our hopefully well-characterized laboratory sources and a transfer radiometer to transfer the calibration to the field unit. This is done using the procedures illustrated in Figure 10.6. First, a radiometer is used to look at a standard blackbody at two temperatures (Figure 10.6(a)). Ideally, to reduce temperature drift in the radiometer, two standard blackbodies would be used. These blackbodies are set at temperatures that are slightly above and below the temperature of the field blackbody(s). Then the field blackbody is measured. The spectral or bandpass radiance from the standard blackbodies can be calculated using the procedures described in the previous section. The radiometer can then be calibrated by assuming the relationship between radiance and counts is linear, at least over the small range represented by the temperature difference in the standard blackbodies. The radiance for the field instrument can then be

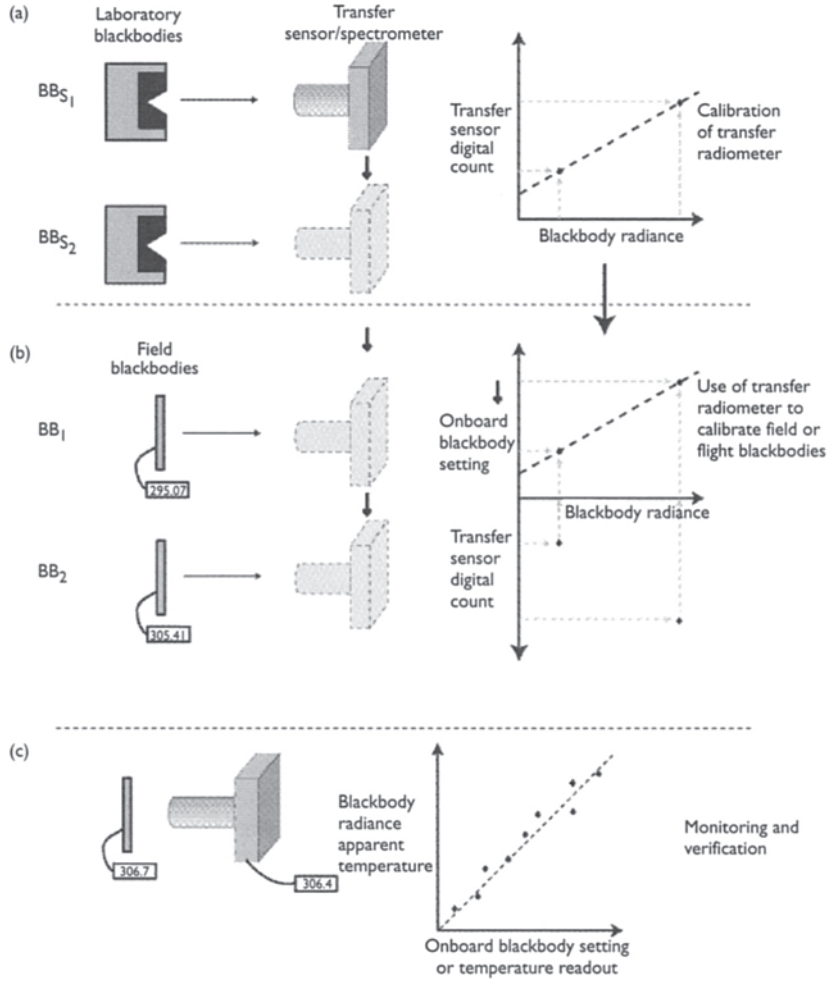


Figure 10.6 Illustration of steps involved in initial calibration of reference blackbodies: (a) use of blackbodies to calibrate a transfer radiometer; (b) use of the transfer radiometer to calibrate a point in a field blackbody readout; and (c) combination of many readout point using steps (a) and (b) to generate an overall calibration of a field blackbody.

interpolated using the two-point calibration and the observed signal from the radiometer when observing the field blackbody (Figure 10.6(b)).

The field blackbody will also have a setting or readout usually proportional to or approximately equal to its kinetic temperature. Ideally, this is the signal from a thermistor in direct contact with the surface of the blackbody. If we then plot the blackbody readout versus the interpolated radiance (often expressed in apparent temperature for convenience), we have the first point in our calibration curve. This entire procedure is repeated over the entire

operating range of the field blackbody (Figure 10.6(c)). It is important to recognize that most infrared radiometers suffer from long-term drift so that for accurate work, the localized piecewise linear recalibration of the radiometer should be repeated for each measurement. If a spectral calibration is required, then this procedure needs to be repeated at each wavelength range of interest. However, because the main variable being monitored is the radiometric temperature of the field blackbody, spectral interpolation should not introduce significant error.

Based on the resulting calibration, we should be able to predict the radiance from the field blackbody quite accurately assuming three critical assumptions hold. First, that the field blackbody is stable (i.e. the radiance is always the same for any given readout value). Second, that the readout sensor closely tracks the surface kinetic temperature (a common flaw in blackbodies is a sensor that is imbedded into or is partially insulated from the skin temperature of the blackbody). Third, that the background radiance in the field is comparable to the laboratory background. The stability can be easily checked with repeated measurements, the readout tracking can be tested by running the blackbody at a high or low temperature relative to ambient and then circulating ambient air over the surface. The surface temperature may change (depending on the temperature control circuit), but the readout and radiance should still generate points on the calibration curve indicating that the temperature probe is accurately tracking the skin temperature. The background radiance may be significantly different in the field than in the laboratory. To correct for this, we would need to know the effective emissivity of the blackbody, as well as the effective background temperatures in the field and during calibration. We could then use equation (10.7) and subtract out the reflected laboratory background radiance and add in the reflected field background radiance for each measurement. Clearly these corrections may be unnecessary if the blackbody is sufficiently black, the backgrounds have similar temperatures or our error tolerances are high compared to the errors introduced by background effects (cf. Table 10.1). The emissivity of a surface can be measured using specialized instrumentation as described by Salisbury and D'Arian (1992) or using a simplified though less precise approach described by Schott (1986).

10.2.4 Calibration of field sensors and in-flight calibration

The calibration of field and flight sensors would ideally be a simple extension of the calibration of the laboratory transfer radiometer as described in the previous section. For many field instruments and applications, this is indeed the case. If the blackbody fully fills the entrance aperture of the field or flight instrument, then we can easily perform a full up sensor calibration. In the simplest case, the instrument observes two blackbodies (or, if necessary, sequentially observes a single blackbody at different temperatures) at

temperatures that approximately span the temperature range to be measured. Note, that once an instrument is involved, we should always use the effective radiance terms as described in equations (10.4) and (10.5). The output voltage or digital count of the sensor is then plotted against blackbody radiance to generate a two point calibration curve. This process can be repeated for each channel in a multichannel instrument and each detector in an imager with multiple detectors. It assumes that the response of the instrument is linear with radiance over the temperature range of interest. This should be carefully verified in the laboratory by generating a detailed plot of radiance versus signal out for many blackbody temperature levels over the entire operating range of interest. If the instrument response is found to be non-linear, several options exist. The first is to treat the response as piecewise linear over several sub-regions of the total operating range. This, of course, means that several calibration points (i.e. several blackbody levels need to be measured in the field each time an instrument is calibrated). For many flight instruments this is impractical and more than two points may not be available. In this case, the functional form of the non-linearity of the system response (or more typically its deviation from linearity) can be calculated and the function forced to fit through the two known calibration points.

Because of the inherent drift in many infrared instruments, it is often necessary to regularly perform calibration in the field. On the other hand, many instruments have some type of internal blackbody to which they frequently normalize the response (i.e. perform a bias adjustment). This process minimizes the effect of drift in the instrument and can reduce or eliminate the need for regular recalibration in the field. However, the reader should be cautioned that many instruments, even with internal references, will have a change in their response if the ambient temperature changes. Again, this should be carefully evaluated in the laboratory so that the need for field calibration is known in advance.

Flight instruments can be calibrated using the same two-point approach as field instruments if full aperture blackbodies can be located ahead of the first optical element (or window). This is commonly done for line scanner type instruments using the back scan time as shown in Figure 10.7. Each revolution of the mirror generates one or more line(s) of image data and allows the sensor to see the known radiance from two blackbodies. This allows a full two-point recalibration of the instrument with each rotation of the mirror. The radiance from each blackbody is known (or can be calculated, if necessary, using equations (10.7) and (10.4)) and a count versus radiance calibration can be performed for each detector in each band. Then, every count in the line(s) associated with that rotation of the mirror can be converted to radiance. The entire process is repeated for each rotation of the mirror. This full aperture approach is very attractive because the black-bodies are viewed through the entire optical system in exactly the same way the earth is viewed, as a result, we get a complete end-to-end calibration on a regular

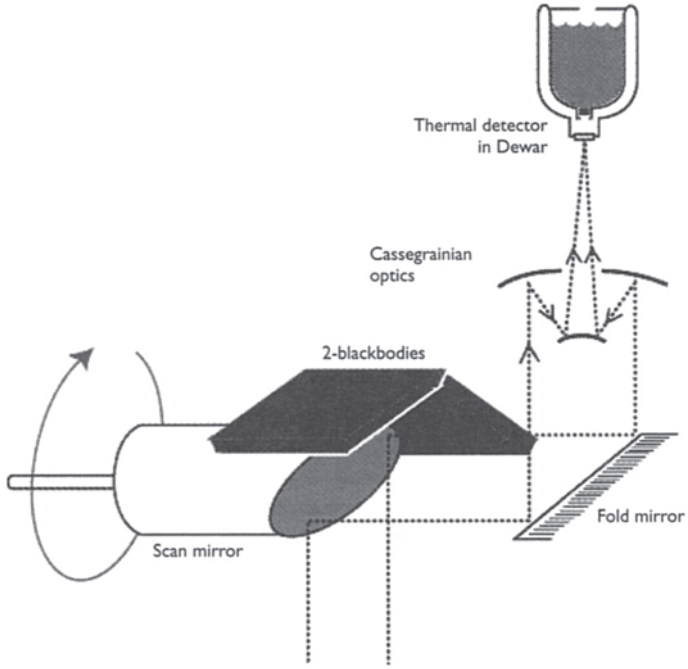


Figure 10.7 Illustration of blackbodies used for calibration during the backscan of a TIR line scanner.

basis so that any drift in the instrument response should be completely removed.

Regrettably, this approach is often not possible with whisk or push broom imagers or where the primary optic is large. Whisk broom scanners often do not scan far enough off the image area to fully image a full aperture blackbody. Push broom scanners have essentially no comparable dead time during an acquisition to view the calibrator and the cost, weight, power, and non-uniformity problems associated with large blackbodies make them impractical for many large-aperture systems. An alternative approach used with some systems is to use full-aperture calibrators only periodically during image acquisition. For example, full-aperture blackbodies may be moved in front of the imager before or after each image acquisition. A pair of images of the blackbody at different temperatures can then be used to calibrate the entire image assuming the system is stable over the period of image acquisition. In many cases, the detectors will have been at least bias restored on a line by line basis using a reference closer to the detectors (i.e. behind the telescope) that is somehow chopped into the field of view of the detectors. This line by

line restoration accounts for short-term drift with the full-aperture blackbodies used to define an absolute end-to-end calibration and to account for long-term drift.

Unfortunately, full-aperture calibration is often not available in many flight systems. In these cases, the regular calibration is done using black-body sources that are introduced somewhere along the optical train (usually after the telescope). For example, in the case of the ETM+, a calibration wand is flipped into the optical path during the dead time when the scan mirror is turning around (cf. Figure 10.8). In the TM case, the wand consists of a high emissivity background surface at constant temperature and a mirror that reflects the radiance from a small blackbody into the optical path and onto the detectors. The wand blocks any radiation coming through the telescope and becomes the source for radiance reaching the detectors. As the wand moves across the detector's field of view, the background is used as a flat plate blackbody whose temperature and, therefore, radiance is known. Then the mirror fills the detector's field of view and reflects a known blackbody radiance onto the detectors (cf. Barker *et al.* 1985). Since this occurs with every mirror oscillation, each line of data has a complete two-point linear calibration update. In the simple linear case, we can write an expression for each detector in each band of the form

$$\mathrm{DC}_{ij} = m'_{ij} L_{\mathrm{BB}i}(T) + b'_{ij} \tag{10.8}$$

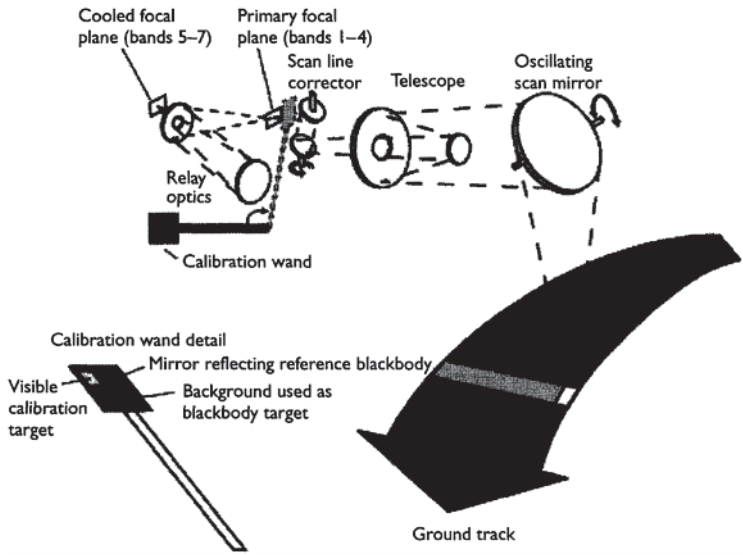


Figure 10.8 Optical illustration showing how the calibration wand is introduced to calibrate the latter stages of the Landsat Enhanced Thematic Mapper +.

where DC_{ij} is the digital count in the *i*th band from the *j*th detector (e.g. on TM 4 and 5 there are four thermal lines acquired per oscillation requiring four detectors), $L_{BBi}(T)$ is the radiance from the blackbody in the *i*th band *due* to its temperature *T*, and m'_{ij} and b'_{ij} are the detector linear gain and bias terms for the *i*th band and the *j*th detector. If the detector exhibits non-linear characteristics, they can be included as corrections to the linear fit using preflight characterization data. The problem with equation (10.8) is that it neglects the transmissive losses and additive radiance from the optical elements ahead of where the calibrator is inserted. Because it is most convenient to place the calibrators in a region where the optical beam is narrow, they are usually behind at least the telescope and possibly some conditioning optics. As a result, several mirrored surfaces are neglected, in the wand-type calibration, which collectively have a significant transmissive loss. In addition, all of these surfaces will have an emissivity equal to one minus their reflectance and as a result, they are radiation sources. The structures that support the mirrors also acts as radiation sources (e.g. the spider web that supports the secondary mirror in the Thematic Mapper telescope) that contribute a significant radiation load (bias level) that is also neglected by the wand. These effects must be taken into account if we are to have an accurate calibration of the instrument. In most cases, the bias correction and possibly the gain associated with the forward optics will be a function of the temperature of the optical elements and the telescope optical cavity. If these surfaces change temperature in flight (which they commonly do unless the cavity temperature is actively controlled) then the fore optics correction must include adjustments based on the temperature of the optical surfaces and background. This can be accomplished using radiometric models, empirical fits, or, more typically, a combination of the two where a radiometric model is adjusted to fit empirical observation.

The empirical fit is accomplished pre-flight using known radiance sources ahead of all of the optical elements. This is essentially the procedure we described for calibration of field instruments. In this case, a collimator may be used with a small blackbody rather than a full-aperture blackbody to fill the entrance aperture with a known radiance level. The instrument's overall linear calibration response can be expressed as

$$DC_{ij} = m_{ij}(T_o)L_i + b_{ij}(T_o) \tag{10.9}$$

where L_i is the entrance aperture radiance in band *i* and $m_{ij}(T_o)$ and $b_{ij}(T_o)$ are the end-to-end instrument gain and bias. The functional dependence of the gain and bias on the temperature(s) (T_o) of the forward optical elements are explicitly noted. However, we should recognize that the form of this functional dependency is usually a complex radiometric model including the temperatures of the optical elements and their background, the emissivity of the elements, and the geometric form factors for each element. As a result,

the values of $m_{ij}(T_o)$ and $b_{ij}(T_o)$ will change with changes in the operating condition of the instrument. Thus, we need to have a solution for all the possible operating conditions of the instrument. To simplify this somewhat, we can express the radiance relationship more explicitly in terms of the dependence on the forward optics as

$$L_{BBi} = L_i g_i(T_o) + c_i(T_o) \tag{10.10}$$

where L_{BBi} is the radiance reaching the location of the internal calibrator (i.e. where the blackbodies on the wand are located), L_i is the radiance reaching the sensor in the ith band, and $g_i(T_o)$ and $c_i(T_o)$ are the band dependent multiplicative (gain) and additive (bias) effects due to the propagation of the image radiance from the front of the sensor to the onboard calibrator. Substituting equation (10.10) into equation (10.8) yields

$$DC_{ij} = m'_{ij} g_i(T_o) L_i + b'_{ij} + m'_{ij} c_i(T_o) \tag{10.11}$$

By comparison with equation (10.9), we see that

$$m_{ij} = m'_{ij} g_i(T_o) \quad \text{or} \quad g_i(T_o) = \frac{m_{ij}}{m'_{ij}} \tag{10.12}$$

and

$$b_{ij} = b'_{ij} + m'_{ij} c_i(T_o) \quad \text{or} \quad c_i(T_o) = \frac{b_{ij} - b'_{ij}}{m'_{ij}} \tag{10.13}$$

This means that by running both an internal calibrator (equation 10.8) and an end-to-end calibrator (equation 10.9), we can isolate the unknown effects due to the forward optics (equations 10.12 and 10.13). By repeating this evaluation over the range of operating conditions of the instrument (e.g. heating and cooling the telescope or individual optical elements) the functional dependency of the fore optics gain and bias on monitored surface temperatures can be established. In flight, the internal calibrator is used in conjunction with the fore optics gain and bias (obtained from lookup tables or models based on the monitored temperature(s) of the telescope) to generate the overall calibration coefficients (m_{ij}, b_{ij}) (cf. equations 10.12 and 10.13).

We should point out that this is just one of the many procedures that can be used to attempt to account for the effects of optical components ahead of an internal calibrator. Another approach might assume that the gain term (g_i) was a constant and the bias term alone varies with instrument conditions. If we look to space (i.e. essentially zero radiance) just before an image acquisition then the observed signal is equal to the overall system bias (b_{ij}) and the effect of the forward optics can be computed using equation (10.13). Regrettably, many imagers cannot regularly point to deep space and even if

they could, the radiance levels may be so small that the bias level may be on an extremely non-linear portion of the response curve or even below the signal threshold for the instrument. Other options include closing a shutter over the entrance aperture of the telescope and using the shutter as an end-to-end calibrated radiance source. By changing the temperature of the shutter, a full two-point end-to-end calibration assessment is available in space. Clearly, this can only be done periodically and an internal calibrator would still be necessary to remove short term variations in detector response.

10.2.5 Onboard calibrator monitoring

No matter what form of blackbody calibration is used, some additional form of periodic end-to-end testing is highly desirable because of potential changes in an instrument over its lifetime. This is particularly true of space-based instruments with long lifetimes. Over time, the optical surfaces in the telescope may change affecting the fore optics calibration. Without some periodic way to do an end-to-end assessment, we would never know if changes took place and might, for example, continue to use incorrect terms for the gain and bias correction for the telescope. One way to assess the endto-end performance of satellite systems is with ground truth or underflight assessments. These approaches are discussed in Section 10.3.

One of the fundamental questions with any calibration procedures concerns the long-term performance of the calibration reference itself. This is particularly true of space programs where it is very difficult to do a detailed periodic reassessment of the calibrator against other well-known reference standards. As a result, most systems try to employ some form of onboard, often redundant, monitoring. The first monitor for most systems is the thermistor (thermocouple) imbedded or attached to the surface of each blackbody. These, rather than any pre-calibrated control signal, are used to estimate the true kinetic temperature and, therefore, the radiance from the instrument. In many systems, multiple monitoring probes are used. This not only provides a redundant check but also, on large blackbodies, can provide a check of thermal uniformity. Regrettably, the temperature monitoring probes are only of use if they are truly monitoring the skin temperature of the blackbody, which is what is observed radiometrically. For cavity-type radiators, this is typically not a problem (the surface is usually close to radiational and convective equilibrium). However, for many flat plate radiators used in full-aperture calibrators and even some internal calibrators, the surface may not be close to a thermal equilibrium with the surroundings. In these cases, the surface, temperature must be maintained by conducting heat to or away from the surface. This inevitably generates gradients near the surface, which can be difficult to measure. Imbedded thermistors may be slightly below the surface or be slightly insulated from the surface by the coating used to blacken the surface. Surface-mounted probes may change

the local radiational field where the measurement is taken by forming a radiational shield. Well-designed blackbodies attempt to reduce these effects by using highly conductive materials that make it hard to maintain local thermal gradients. However, the potential is always there and the user should always check carefully for any decoupling of the radiometric and kinetic temperatures. One way that has been proposed to do this in flight is to use a simple radiometer to monitor the temperature of the blackbody. Very stable broadband low-frequency bolometer-style detectors can be used for this purpose. These detectors essentially use an internal thermal reference to provide long-term stability. Their slow response times and low sensitivity make them unacceptable for imaging purposes, but do not limit their utility for radiometric monitoring of onboard calibrators. The radiometric temperature can then be compared to the kinetic temperature measured by the thermistor to see if any systematic thermal decoupling is taking place.

10.3 Post-launch verification

One of the most critical concerns after the launch of a new TIR satellite imager is whether the pre-launch calibration is still valid. This same concern also applies to new airborne systems (i.e. can we trust the onboard calibrator). Indeed, for many satellite systems employing internal calibrators, there is a periodic need (e.g. yearly) to verify the calibration of the instrument. In this section, we will emphasize procedures for post-launch verification (or recalibration) of satellite-based imagers but the approach could also be applied to airborne sensors. There are two basic approaches to this process. The first involves assuming that the onboard sensor is calibrated, inverting to surface radiance or temperature, comparing the results with ground truth, and determining whether any residual error is within the compounded errors of the procedures used. If the error between inversion, and truth is larger than the propagated error due to sensor calibration, atmospheric inversion, and ground truth measurement, then the instrument calibration needs to be updated. This ground truth based approach is discussed in Section 10.3.1. An alternative, though similar, approach is to predict the radiance values an imaging sensor should see and compare these values to observed values. Only if the error between prediction and observation exceeds the propagated errors associated with the prediction process is an updated calibration required. This approach typically employs underflight measurements and is discussed in Section 10.3.2.

10.3.1 Selection and use of ground truth targets

In order to evaluate or update the calibration of an imaging system, it is critical that we have two or more known radiance values. Ideally, these would

be two full-aperture blackbodies at different temperatures located in front of the imaging system. We could then compare the blackbody radiance values to the values measured by the imager. If the errors exceeded expectations, then equation (10.9) could be used to perform an updated calibration of the gain and bias values for each detector. Regrettably, we do not yet have a blackbody calibration lab in space. As a result, this test of system performance is most commonly done using radiance values from the imaging sensor propagated to earth and compared to surface radiance values (cf. Schott and Schimminger 1981; Schott 1993). This comparison could also be done by propagating the surface radiance to space and comparing the measured and predicted values at the spacecraft. However, because we most often want to evaluate how well we can invert image values to surface radiance or temperature, it is more common to measure errors in terms of surface radiance or temperature. In either case, we have a critical need for surfaces whose radiance is well known. If we express the surface-leaving radiance as

$$L(0) = \varepsilon L_{\mathrm{BB}}(T) + rL_{\mathrm{d}} \tag{10.14}$$

then it is clear that ground truth may consist of directly measured values of surface radiance $L(0)$ or well-known values of emissivity, surface temperature, and downwelled radiance (L_{d}), which can be used to compute the surface radiance. We would need these values for fully resolvable targets in each spectral band of interest. Furthermore, to effectively evaluate the instrument gain and bias, we would need to have at least two targets at significantly different radiance levels (ideally covering the operating range of the imager). Because of the dependence of radiance on temperature, emissivity, and downwelled radiance, it is important to insure that these values will not change between the measurement time and the imaging time. The emissivity of most surfaces can be assumed constant for long periods of time (soiling, moisture content, and phenological changes being some obvious exceptions) and selecting stable air masses can reduce the temporal dependence on L_{d}. The importance of L_{d} can also be significantly reduced by selecting high emissivity (low reflectivity) targets. The time dependency of temperature remains a critical problem. The diurnal heating and cooling cycle keeps the temperature of most objects in at least a slow state of flux. This change is minimized late at night. However, although temperatures are most stable, the scene contrast is also often lowest late at night, making it difficult to obtain the appropriate temperature ranges. To minimize errors due to thermal changes with time, it is best to choose objects with high thermal inertia and, if possible, times of slow change in diurnal temperature values. There is also an issue of whether to directly measure radiance or to compute it from measurements or estimates of temperature, emissivity, and downwelled radiance. It is simplest to measure the radiance of two or more objects at the time of the overflight. There are several practical problems with this approach

that may often force us to choose the indirect solution. The first is the need to match the spectral response of the surface face radiometer to the response of the imaging sensor. The match is never identical, which means that we need to work in some units other than raw radiance. If the spectral response of the two sensors is nominally the same (e.g. 10–12 μm) then if we carefully convert each sensor's radiance to apparent temperature, we can do a comparison of radiance values expressed as apparent temperature. If the target's emissivity values are approximately flat, this approach introduces only small errors (cf. Section 10.4). If the sensor's bandpass values differ significantly in bandwidth or spectral shape, then we must go through a more involved process to predict the radiance the imaging sensor should see (i.e. to predict truth). In this case, we need to convert the radiance in one spectral band to the radiance we would expect in another. Without spectral data, this can only be an approximation. The errors in the process are minimized if the spectral response functions are similar, the emissivities are approximately constant over the whole spectral range, and downwelled radiance effects are minimized (i.e. sky temperature or L_d when expressed as apparent temperature is approximately constant with wavelength). The transfer process involves estimating the unknown terms in equation (10.14) using the observed surface radiance value and thus predicting what the groundleaving radiance would be in the spacecraft bandpass. For example, if we measured the temperature and estimated the emissivity from lab data or a lookup table based on material type, we could solve for L_d and express it as the apparent temperature of the sky (T_{sky}). Then the effective radiance in the sensor spectral bandpass could be estimated as

$$L(0) = \int [\varepsilon L_{BB}(T_\lambda) + rL_{BB\lambda}(T_{sky})]R'_S(\lambda)\,d\lambda \tag{10.15}$$

where R'_S (λ) is the spacecraft sensor's peak normalized spectral response. The errors in equation (10.15) are almost as large as if we just used indirect measurements to begin with, so, in general, surface radiance measurements are only useful if they are a close spectral match to the satellite system or the target is very nearly a blackbody. Even under these conditions, the spectral radiance measurements can pose a serious logistical problem. This is because the measurements need to be made essentially simultaneously with the overflight of two or more surfaces that may be very large. For example, the ground sample sizes for Landsat 7, ASTER, Landsat 5, MODIS, and AVHRR are approximately 60, 90, 120, 1,000, and 1,000 m, respectively. To account for sampling error, we would need to characterize an area approximately three times the ground sample size on each side. This can be a tall order if the surface has any significant spatial variability. This drives us to seek large targets with constant temperature and emissivity over areas spanning several pixels. Any variability in radiance would need to be recorded so that we could effectively predict the aggregate radiance observed from space. As a

result of these restrictions, we may need one or more radiometers per sample for low-resolution systems. On the other hand, when the resolution is of the order of a meter (e.g. for airborne systems), then the logistics are far more tractable and the use of a field radiometer becomes very attractive.

All of these constraints have pushed us towards the use of water as one of the most convenient ground truth targets. It has a number of very useful characteristics, not the least of which is its ubiquitous nature on the surface of the planet. In addition, the high thermal inertia of water and the tendency of fluids to mix leaves us with large surfaces that often have temporally and spatially constant temperature. Also, if the water is well mixed, then a kinetic measurement is very indicative of the skin temperature that will be radiometrically observed. Finally, water as shown in Figure 10.4 has a very high spectrally flat emissivity that does not vary appreciably for angles near nadir. As a result, both kinetic and surface radiometric measurements of water can be effectively used as ground truth. One important limitation becomes important in the case of calm water, particularly during high solar loading conditions (i.e. good remote sensing days). If the water is very calm (i.e. unmixed pools, ponds, even lakes if there is little wind) then solar heating (or radiational and evaporative cooling) can induce a sharp thermal gradient in the surface water. This has two negative effects. First, it means that simple kinetic measurements will not accurately reflect the skin temperature. Second, it may set up a condition where there is substantial non-uniformity in the surface temperature if there is a source of disturbance (e.g. boat wakes).

In order to utilize our ground truth measurements, we must first convert the radiance measured with the overhead system to comparable values. Typically, this involves solving for the surface-leaving radiance, apparent surface temperature, or kinetic temperature of the ground truth target(s). This requires correction for the influence of the atmosphere. In the simplest case, effects of atmospheric transmission and path radiance can be expressed as

$$L(h) = \tau L(0) + L_{\mathbf{u}} \tag{10.16}$$

where $L(h)$ is the measured sensor-reaching radiance, $L(0)$ is the surfaceleaving radiance, and τ and L_u are the effective transmission and upwelled radiance (path radiance), respectively. There are a number of methods to estimate τ and L_u for each band (cf. Schott 1997). One of the most straightforward is to use a radiative transfer code such as MODTRAN (cf. Berk *et al.* 1989). The critical inputs to the MODTRAN code are the temperature and relative humidity of the atmosphere as a function of altitude at the time of the overflight and the spectral response of the sensor in each band. Ideally, the atmospheric data would be obtained from a radiosonde balloon launched at the time and location of the ground truth campaign. In practice, data from regional airports must often be extrapolated in space and time to generate

estimates of the atmospheric profile at the point of interest. Once $L(0)$ values are obtained, they can be directly compared to surface radiance values (assuming common spectral response in the sensors). Often to minimize cross-calibration errors, the comparisons are done using effective spectral radiance values over the bandpass (cf. equation 10.5 or apparent temperature values). If the truth consists of kinetic temperature values, then the image derived surface radiance needs to be converted to temperature or vice versa using equation (10.14) and a band-dependent radiance to temperature lookup table (i.e. an inversion of the Planck equation).

At this point, we can compare the truth to predicted values and estimate the residual error. If we are calibrating a newly launched sensor, we would have performed a parallel study of the end-to-end error propagated to surface radiance. This would include sensor noise, sensor calibration error, and errors in whatever atmospheric correction was used (cf. Section 10.4). If the measured errors are within the tolerance errors from the error propagation model, we would assume the sensor calibration is still valid. If not, we could update the sensor calibration by propagating the ground truth values to sensor-reaching radiance values using equation (10.16). These become the equivalent of the blackbody values used in equation (10.9) for laboratory calibration of the instrument. Clearly this step of calibration update is not done casually and is usually the result of repeated tests that show a consistent pattern of error in the overall instrument calibration.

10.3.2 Use of ground truth for end-to-end sensor calibration

It is also possible at this point to perform a ground truth based calibration of the imaging sensor. This would commonly be used for two purposes. The first would be to perform an end-to-end calibration of the sensing system incorporating atmospheric and sensor effects. In this case, the sensor need not be calibrated (only stable) and no external atmospheric correction is required. If the sensor response is linear, then we can combine equations (10.9) and (10.16) to yield

$$\mathrm{DC}_i = \begin{cases} m_i L_i(h) + b_i \\ m_i(\tau_i L_i(0) + L_{ui}) + b_i \\ a_i L_i(0) + C_i \end{cases} \tag{10.17}$$

where $a_i = m_i\tau_i$ and $C_i = m_i L_{ui} + b_i$ are the end-to-end linear calibration coefficients for the ith spectral channel which can be found by simple linear regression of image counts versus ground truth radiance values. This approach is applicable to both aerial and satellite systems and provides a complete systems calibration and tends to reduce any systematic errors by minimizing them simultaneously. Regrettably, the solution is only valid if the atmosphere

and sensor response are constant. Thus, another portion of the same satellite image 50 miles away may have a different atmosphere or another flight line 30 minutes later may have both a different atmosphere and different sensor gain and bias. Because of this, the ground truth based calibration is usually only used for research studies where the cost of a field campaign for calibration of each image is warranted. Scarpace *et al.* (1974) demonstrated the effectiveness of this approach using two small pools at different temperatures for end-to-end system calibration of an aerial TIR sensor.

This brings us to the second use of ground truth for calibration. In this case, we assume we have a well-calibrated sensor and wish to evaluate the effects of the atmosphere (cf. Palluconi 1996). For example, we may be comparing atmospheric correction techniques or attempting to assess the error associated with using a particular technique (e.g. the MODTRAN radiative transfer model described above). In this case, if we regress the ground truth derived surface radiance ($L_i(0)$) against the sensor-reaching radiance ($L_i(h)$), then the slope and intercept yield the atmospheric transmission (τ_i) and the path radiance (L_{ui}), respectively (cf. equation 10.16). These values can then be compared to the corresponding values predicted by an atmospheric correction technique.

10.3.3 Use of underflights for post-launch sensor calibration verification

The other common approach to verification of the calibration of a satellitebased TIR imager is to use an underflight with a well-calibrated airborne TIR imager. Because airborne systems can be well calibrated in a laboratory immediately before a flight and often employ multiple full-aperture blackbody calibrators, they can be quite well calibrated. To take advantage of this, we fly the airborne system under the satellite, but above most of the earth's atmosphere. If we acquire data at essentially the same time as the satellite, we can select uniform targets with a range of temperatures whose groundleaving radiance can be assumed identical and whose radiance is measured at aircraft and satellite altitudes. For these assumptions to be valid, four things must hold true. First, the temperature of the targets must be constant over the period between when the airborne and satellite data are acquired. This can be optimized by selecting targets with high thermal inertia, acquiring data near diurnal extremes, and minimizing the time difference between the underflight and satellite acquisitions. Because we often have little control over satellite acquisition time, it is even more critical that we carefully select our targets and control acquisition times. The second thing we must be certain of is that the airborne and satellite radiances are observed from the same target. This is greatly facilitated by selecting large radiometrically uniform targets where image registration and sampling issues are minimized. Selection of sites can often be facilitated by using the higher spatial resolution sensor (usually the airborne sensor) to select a region of

uniform radiance that is substantially larger than the area sampled by the pixel(s) used to characterize the target in the low-resolution sensor. The third requirement or implicit assumption is that the emissivity of the targets relative to the sensors be the same. Since the two sensors are often not along the same line of site, we want to choose targets whose emissivity is approximately Lambertian (i.e. constant) over the relevant angles. The fourth requirement is that the airborne and satellite sensors have the same response. Since this is never completely accomplished, we seek to minimize spectral response differences by closely matching the sensor responses, choosing targets with constant spectral emissivities over the bandpass and using spectral radiance or apparent temperature values to perform any comparison.

In reviewing all these requirements, we find that large water bodies meet many of the requirements needed to minimize differences between the aerial and satellite radiances. The thermal inertia is very large. The emissivity is well known, spectrally flat, and approximately constant over a range of view angles and water conditions (cf. Figure 10.4). For a well mixed water body, the temperature can be uniform over very large areas. One potential limitation with water as the primary target is that only a limited radiance range may be available. This limitation can often be overcome by selecting a study site where different types of water bodies are available (e.g. open water and bays, lakes and rivers, large river discharge into ocean, etc.).

If we were truly above all of the atmosphere, we could directly compare the airborne and satellite radiance values and directly determine if there were any calibration error in the satellite system. In fact, the airborne radiance values must be corrected for the effects of the upper atmosphere (cf. Figure 10.9). This correction can be estimated using radiative transfer codes based on radiosonde estimates of the temperature and water vapor in the upper atmosphere. The expected satellite-reaching radiance can be predicted from the observed aircraft radiance measurements according to

$$L_i(s) = L_i(h)\tau_i(s-h) + L_{ui}(s-h) \tag{10.18}$$

where $L_i(s)$ is the predicted radiance in the ith band at the satellite, $L_i(h)$ is the observed radiance at aircraft altitude (h), and $\tau_i(s-h)$ and $L_{ui}(s-h)$ are, respectively, the transmission and path radiance from altitude h to space. The atmospheric transmission and path radiance are predicted for each band by the radiative transfer process and are a potential source of error. By flying over most of the atmosphere, we can minimize the magnitude of the atmospheric correction. This means that we will only introduce small errors in small terms resulting in small final errors in the predicted sensor-reaching radiance. The final assessment involves comparing the difference between the observed and predicted satellite-reaching radiance

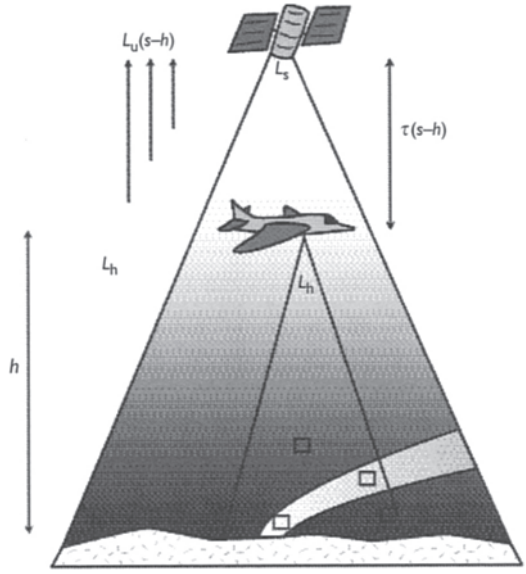


Figure 10.9 Illustration of how the observed radiance at the aircraft (L_h) can be propagated through a thin atmospheric layer (s–h) to yield the radiance at the spacecraft, L_s.

values relative to the expected errors in the two measurements. If the observed differences exceed measurement errors, they can be used to update the satellite sensor calibration using equation (10.9). This final step is essentially the same approach as that of the ground truth method described in Section 10.3.1. The only difference is that the errors in the predicted radiance can often be substantially reduced by eliminating the errors associated with using the radiative transfer codes through the lower atmosphere. In addition, the aerial image data allows for a wide choice of potential targets to optimize the radiance range and spatial uniformity. Schott and Schimminger (1981) and Schott (1993) describe the use of this underflight approach to verify the post-launch calibration of the Heat Capacity Mapping Radiometer and the Landsat Thematic Mapper Thermal sensors.

Underflight techniques have also been used to evaluate how well satellite sensors can predict surface-leaving radiance. By flying a calibrated airborne sensor over the same targets at a series of altitudes, a radiance versus altitude relationship can be established and extrapolated to zero altitude. If the lowest flight altitudes are quite low (hundreds of feet over water is common) then the aircraft imaging radiometer is essentially collecting ground truth (cf. Figure 10.10). These few extrapolated ground truth points are used with the corresponding data at any flight altitude to compute the atmospheric

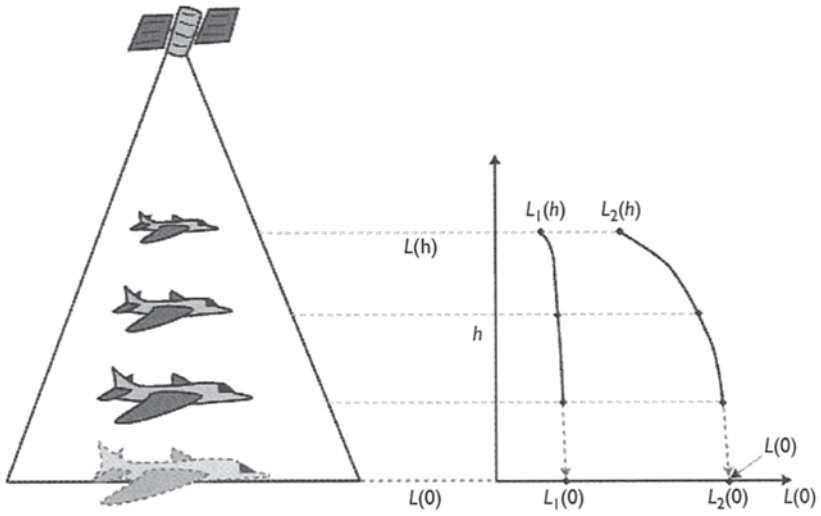


Figure 10.10 Illustration showing how underflight data can be used for comparison (ground truth) with satellite estimates of surface-leaving radiance. $L_1(h)$ and $L_2(h)$ represent the measured radiance at altitude for targets 1 and 2, respectively.

parameters using equation (10.16). If this multialtitude profile technique is used under a satellite overpass, then the atmospherically calibrated aircraft data at altitude can be used to predict surface radiance values that can be directly compared to the satellite estimates of surface radiance. This approach has been used to evaluate various methods for atmospheric inversion of calibrated satellite data (cf. Schott 1993).

10.4 Error sources and effects

In this section, we want to briefly look at some of the sources of error in radiometric calibration and temperature measurement. This treatment is intended to provide the reader with some insights into the relative magnitudes of error sources for some simple cases. It is important to recognize that, for any particular sensor, target set and atmospheric correction method, a sensitivity and error propagation study should be undertaken. As a result, the terms included in the error analysis presented here are more important than the actual values that are often case specific or may have been selected simply to illustrate a point. The reader is referred to Beers (1957) and Rubenstein (1981) for a thorough treatment of error propagation.

We can begin by considering our calibration problem as comprising three components. These three components are the surface radiance, the satellite

radiance, and the atmospheric correction that relates one to the other. We can look at the surface and satellite radiance measurements separately and consider the sources of error that contribute to the total error in each measurement. Then we can look at the error in the atmospheric correction terms to see how they influence the error in propagating surface radiance to space or more typically, space-based radiance to surface radiance.

10.4.1 Error in sensor-reaching radiance

Table 10.2 contains examples showing error sources associated with the measurement of sensor-reaching radiance for two types of satellite sensors and one airborne sensor. To be consistent, we are using 10.5–12.5 μm for the spectral bandpass for all examples and we will express all radiometric errors in terms of the equivalent temperature change in a 300 K blackbody necessary to produce that change in radiance. In the case of the first sensor, we are assuming it is a system similar to the Landsat TM shown in Figure 10.8. We have estimated the error in absolute radiance for the onboard blackbodies to be 0.15 K in apparent temperature using the approach described in the discussion of Table 10.1. The uncertainty introduced by the fore optics radiance model generates an independent source of error also at the 0.15 K level. The system noise for this sensor (detector, preamplifier, conditioning electronics) when expressed as apparent temperature is 0.1K. Finally, the uncertainty due to the quantization of the analog signal into digital counts generates an additional uncertainty of 0.3 K for this example. Since these error sources are independent, we can generate an overall estimate of the error, in measurement of the radiance associated with a digital count value, to be the square root of the sum of the squared values from the individual sources. The errors are expressed in apparent temperature for convenience. We see that in this case the overall error in measured radiance would be about 0.4 K dominated by quantization noise. The second sensor example is essentially the same as the first with a smaller quantization step and, therefore, less uncertainty introduced by quantization which in this case (i.e. because we were limited by quantization error) substantially

Table 10.2 Errors in measurement of sensor-reaching radiance due only to sensor effects. All values expressed in terms of apparent temperature (K). Sensor 1 is similar to Landsat TM, sensor 2 is similar to Landsat ETM+, and sensor 3 is indicative of a range of airborne instruments

	On-board calibrator	*Fore-optics model*	*Sensor noise*	*Quantization noise*	*Overall error*
Sensor 1	0.15	0.15	0.10	0.30	0.3808
Sensor 2	0.15	0.15	0.10	0.10	0.2550
Sensor 3	0.20	0.00	0.10	0.05	0.2291

improves overall performance. This is the reason the Landsat 7 thermal data are available at two gain levels. The low gain level ensures no clipping of the data will occur but limits the radiometric performance. The higher gain level has smaller step sizes and, therefore, less quantization error. The third sensor has characteristics we might expect from an airborne instrument. The absolute radiance error due to the use of full aperture flat plate blackbodies is estimated at 0.2 K (cf. Table 10.1). Since end-to-end calibration is used, there is no fore optics error and in this example (assuming a 10-bit A–D) very low quantization error. These errors represent the expected error in the sensor-reaching radiance for three realistic sensor examples. Regrettably, these are not the only errors in the calculation of apparent surface temperature or in verification of in-flight sensor calibration.

10.4.2 Error in surface-leaving radiance

A second source of error in many cases is the error in surface-leaving radiance. Table 10.3 shows several examples of error calculations associated with surface-leaving radiance measurement or estimation. These errors represent our uncertainty in "ground truth" measurements. The first two radiance measurements assume the ground truth data are measured with a field radiometer. The two temperature measurement lines assume we predict radiance from knowledge of the temperature and emissivity. The final radiance measurement assumes the ground truth is the result of extrapolated radiance measurements from an aircraft using the method illustrated in Figure 10.10. For the radiance measurements we have an initial error associated with the field radiometer and usually dominated by the absolute accuracy of the instrument calibration (cf. the error sources given in Table 10.1). Because the radiance we observe from an airborne or satellite system covers a larger area than the field radiometer and because there is some misregistration of the image and field data, we have to recognize that the spatial variability in radiance will introduce error in our estimate of the surface radiance that the overhead imager will observe. There are three major sources for this variation. They are the spatial variation in temperature [$\Delta T(x,y)$], the variation in emissivity (due to spatial variations [$\Delta\varepsilon(x, y)$] and angular variations [$\Delta\varepsilon(\theta)$]), and variation in temperature with time [$\Delta T(t)$]. In our example, we assume that the field radiance measurement(s) (usually an average of several samples over a target site) attempts to characterize the surface radiance seen by one or more pixels in the overhead imaging system. The spatial variation in temperature adds uncertainty to this estimate. This uncertainty will be minimized for regions of near constant temperature (e.g. water in our example). The spatial variation in emissivity is a similar source of error. In many cases, this error can be minimized by selecting uniform material surfaces for ground truth. An additional source of error can be introduced if the emissivity of the surface has any variation with view angle. For example, if

Table 10.3 Errors in surface-leaving radiance associated with ground truth measurements. All radiance values are expressed as apparent temperature (K). Calculations are for a 300 K target with emissivities of 0.90 (land) and 0.986 (water). The effective background temperature associated with the downwelled radiance is 243.15 K

	Target radiance-temperature error sources (K)					
	1. Radiance measurement error	*2. Temperature measurement error*	*3. Spatial $\Delta T(x,y)$*	*4. Temporal $\Delta T(t)$*	*5. RSSΔT (s_{LTBB})[a]*	*6. $s(L_T)$[b]*
Radiance measurements, land	0.150	0	0.150	0.15	0.26	0.23
Radiance measurements, water	0.150	0	0.100	0.05	0.19	0.18
Temperature measurements, land	0	0.250	0.150	0.15	0.33	0.30
Temperature measurements, water	0	0.100	0.100	0.05	0.15	0.15
Aircraft-based water radiance	0.250	0	0.050	0.05	0.26	0.26

	Emissivity errors (emissivity units)			
	Measurement error ε	*Angular, spatial variation ε*	*RSSε (s_ε)*	*$s(L_\varepsilon)$[c]*
Radiance measurements, land	0	0.01	0.01	0.43
Radiance measurements, water	0	0.005	0.01	0.21
Temperature measurements, land	0.015	0.01	0.02	0.77
Temperature measurements, water	0.005	0.005	0.01	0.30
Aircraft-based water radiance	0	0.004	0.00	0.17

	Downwelled radiance errors (K)			
	Measured error L_d	*Estimated error L_d*	*$s(L_d)$*	*Overall radiance error $s_{L(0)}$*
Radiance measurements, land	0.5		0.05	0.49
Radiance measurements, water	0.5		0.01	0.28
Temperature measurements, land		4.4	0.44	0.87
Temperature measurements, water		4.4	0.06	0.34
Aircraft-based water radiance		4.4	0.06	0.31

Notes
Relevant equations:

$$L(0) = \varepsilon L_{TBB} + (1-\varepsilon)L_d$$

$$s_{L(0)} = \left(\left(\frac{\delta L(0)}{\delta L_{TBB}}\right)^2 s_{L_{TBB}}^2 + \left(\frac{\delta L(0)}{\delta \varepsilon}\right)^2 s_\varepsilon^2 + \left(\frac{\delta L(0)}{\delta L_d}\right)^2 s_{L_d}^2\right)^{1/2}$$

$$= ((s(L_{TBB}))^2 + (s(L_\varepsilon))^2 + (s(L_d))^2)^{1/2}$$

where

$$s(L_{TBB}) = \left|\frac{\delta L(0)}{\delta L_{TBB}} s_{L_{TBB}}\right|, \quad s(L_\varepsilon) = \left|\frac{\delta L(0)}{\delta \varepsilon} s_\varepsilon\right|, \quad s(L_d) = \left|\frac{\delta L(0)}{\delta L_d} s_{L_d}\right|$$

$s_{L(0)}, s_{L_{TBB}}, s_{L_d}$, and s_ε are the uncertainties in surface-leaving radiance, radiance due to target temperature, downwelled radiance, and emissivity.

a Root sum square error from columns 1–4.
b Error associated with target temperature and radiance.
c Error associated with target emissivity (K).

the ground truth radiance is measured vertically and the imager is 15° off nadir, the surface radiance values will differ if there is any non-lambertian behavior in the target emissivity. Finally, we need to recognize that the field radiance and the overhead measurements are usually not exactly coincident in time. Thus, we will have an additional error depending on the rate of change in temperature with time and the time between the relevant measurements. The radiance measurement rows in Table 10.3 reflect typical errors in these terms propagated to an overall radiance error. The contributions to overall error are grouped into uncertainties due to the target temperature, the target emissivity, and the downwelled radiance. This example demonstrates how we can take advantage of the spatial and temporal stability of water to improve our estimates of surface radiance. Note that we have assumed that the sky radiance (L_d) is nearly constant for all measurements and, therefore, contributes very little to error to the radiance estimates. If the surfaces are specular or the sky variable over the measurement time, we would need to include the variation in downwelled radiance as an error source.

If we are relying on temperature estimates to predict surface radiance, we need to contend with some additional sources of error. The first error is in the individual temperature measurement. The recording devices themselves (usually thermistors) can be readily calibrated to 0.1 K or better. The problem comes from getting the thermistor to effectively record the skin temperature of the surface being measured. For solids this is a problem of getting the thermistor in thermal contact with the surface (potentially changing the surface by the measurement process) or imbedding the thermistor and dealing with surface gradients. For water, we have the surface gradient issue if the water is not well mixed. Snyder *et al.* (1997) point out a number of problems associated with trying to use even very uniform land surface features for ground truth. Their estimates of individual measurement errors, as well as spatial and temporal variations in radiance over land, have influenced our selection of reasonable error estimates for this sensitivity assessment. We have somewhat optimistically set the land temperature error at 0.2, but caution that errors much higher than this are common depending on target and instrument placement. For reasonable well mixed water (Hudson River), Schott (1979) sites examples where thermistor and surface radiometer values have essentially the same error (within 0.1 K) when compared to calibrated surface radiances derived from over flights. We have, therefore, chosen (0.1 K) as our temperature measurement error for water. The spatial variability in temperature and emissivity, the angular variation in emissivity, and the temporal variation in temperature all have the same impact whether we measure radiance or temperature so these values are the same as for our radiometric examples.

In the case of direct measurement of temperature, we must also estimate the emissivity and the downwelled radiance in order to solve equation (10.14)

for the surface-leaving radiance. We have assumed that emissivity is estimated based on material type and tabulated laboratory measurements. These errors could be reduced somewhat by making numerous measurements of emissivity in the field for the target of interest. Again, because of the inherent variability in land surface feature, these error sources are larger over land. Finally, we need to estimate the downwelled radiance. One way to do this is to directly measure the irradiance onto the target in the bandpass of interest. This will yield a good estimate of L_d with small errors due to slight temporal variation in L_d and instrument errors. However, if we are directly measuring temperature, it is usually because we do not have a field radiometer and we will need an alternative way to estimate the downwelled term. This can be done using radiosonde data and a radiative transfer model such as MODTRAN. The errors in this estimate can be rather large, but their effect can be mitigated by choosing targets with high emissivity values to reduce the reflected radiance contribution. In our examples, we see that water with its high emissivity largely negates the error from L_d, while land surfaces with emissivities typically around 0.9–0.95 will have a non-negligible error associated with errors in L_d.

For the direct temperature measurement approach, we can attribute the total surface-leaving radiance error to uncertainties in temperature, emissivity, and downwelled radiance. For our example, we have estimated each of the component errors and expressed them in terms of apparent temperature in order to compare their relative importance to the overall error in the surface radiance value (which is also expressed in terms of apparent temperature). From this assessment, we clearly see that measuring radiance is more attractive than using contact temperature methods. However, if 10s or 100s of readings are required to characterize the spatial variability and the readings must be near coincident in time to reduce temporal variations, then cost may drive us to the thermistor approach. In addition, we see that water is a significantly better target than land, if well-mixed water bodies are available as targets. Finally, we see that by using an imaging radiometer located in an aircraft over water, we can take dual advantage of water as a target, and a radiometer as a sensor to minimize spatial variability effects by directly sampling and averaging out most spatial variation. Figure 10.11 shows an example under flight image and the size of Landsat 7 (60 m) pixels projected onto the "ground truth image." With careful target selection, the thermal uniformity within the pixels can reduce sampling and spatial variability effects with this approach to negligible levels compared to instrument errors.

It is important to note in comparing Tables 10.2 and 10.3 that in many cases the error in radiance from a large sample as measured by the overhead sensor may be smaller and in some cases much smaller than the error in the estimate of the ground-leaving radiance for that same large sample. The reason for this is due in part to the effects of sampling, but we should also recognize that we have placed all the spatial and temporal registration errors

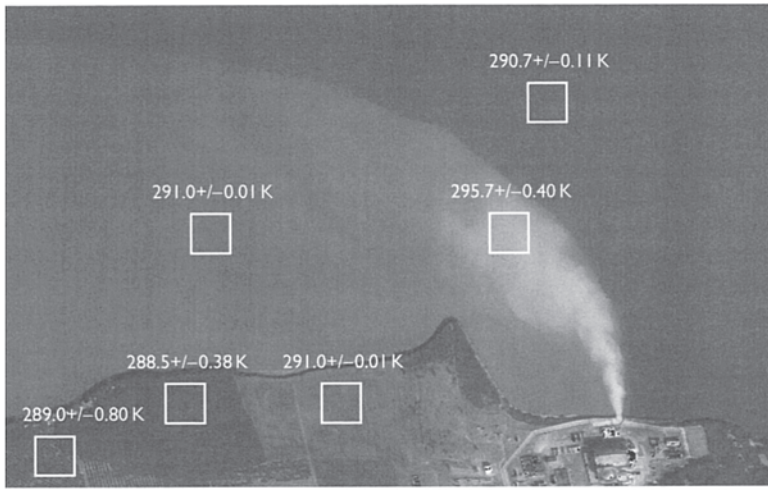


Figure 10.11 Thermal infrared image of a power plant cooling water discharge into Lake Ontario. The overlaid boxes on this aerial image represent the approximate size of 60 m pixels (e.g. Landsat ETM+) that might be ground truthed by this underflight image. The values shown with each box are the mean and standard deviation of the temperature associated with the 2-m pixels within the box. The image is from one of the 70+channels of RIT's Modular Imaging Spectrometer Instrument.

on the ground truth. In fact, these errors can be assigned to either the overhead measurement or the ground measurement or split between them. We also need to recognize that the measured radiance values in Tables 10.2 and 10.3 are not directly comparable. Table 10.2 represents the error in sensor-reaching radiance for a single pixel (i.e. the error in $L(h)$ in e.g. equation 10.16). Table 10.3 represents the error associated with trying to measure (using ground truth) the radiance leaving the ground over the often very large area represented by that exact same pixel (i.e. the error in $L(0)$ in e.g. equation 10.16). In the long run, our goal is to predict the value (and the error in the value) of one of these measurements from the other. For example, we want to operationally predict surface radiance values from image-derived measurements of the sensor-reaching radiance. However, during calibration experiments, we want to predict the sensor-reaching radiance (and error) from surface radiance measurements. In both cases, we need to investigate the errors associated with atmospheric correction using equation (10.16).

10.4.3 Atmospheric sources of error

Our purpose here is not to try to describe atmospheric correction or inversion methods, but merely to illustrate the nature of error propagation and the

magnitudes associated with simple methods as they relate to instrument calibration. The errors in this case are illustrated in Table 10.4. The only sources of error are uncertainties in the atmospheric transmission (τ) and upwelled radiance (L_u). However, in this case we have a new twist in that the

Table 10.4 Error due to uncertainties in atmospheric correction in the 10–12μm range. Sensor-reaching and ground-leaving radiance errors are expressed in terms of apparent temperature (K)

	T (K)	τ	τ *error*	L_u ($W\,m^{-2}\,sr^{-1}$)	L_u *error* ($W\,m^{-2}\,sr^{-1}$)	*Error in sensor reaching radiance*	*Error in target leaving radiance*
MODTRAN, local radiosonde (to space)	285	0.671	0.00518	6.265	0.02821	0.2990	0.4454
	305	0.671	0.00518	6.265	0.02821	0.3456	0.5148
MODTRAN, interpolated radiosonde (to space)	285	0.671	0.01025	6.265	0.05823	0.5910	0.8805
	305	0.671	0.01025	6.265	0.05823	0.6830	1.0176
Aerial profile technique (to 7 km)	285	0.679	0.00512	6.126	0.02500	0.2974	0.4380
	305	0.679	0.00512	6.126	0.02500	0.3438	0.5064
MODTRAN interpolated radiosonde (3 km to space)	285	0.929	0.00046	0.932	0.01316	0.0408	0.0440
	305	0.929	0.00046	0.932	0.01316	0.0366	0.0394
MODTRAN interpolated radiosonde (7 km to space)	285	0.979	0.00004	0.139	0.00350	0.0108	0.0110
	305	0.979	0.00004	0.139	0.00350	0.0090	0.0092

Notes
Relevant equations:

$$L(s) = \tau L(0) + L_u$$

$$s_{L(s)} = \left(\left(\frac{\delta L(s)}{\delta \tau} \right)^2 s_\tau^2 + \left(\frac{\delta L(s)}{\delta L(0)} \right)^2 s_{L(0)}^2 + \left(\frac{\delta L(s)}{\delta L_u} \right)^2 s_{L_u}^2 + \rho \frac{\delta L(s)}{\delta \tau} \frac{\delta L(s)}{\delta L_u} s_\tau s_{L_u} \right)$$

$$L(0) = \frac{L(s) - L_u}{\tau}$$

$$s_{L(0)} = \left(\left(\frac{\delta L(0)}{\delta \tau} \right)^2 s_\tau^2 + \left(\frac{\delta L(0)}{\delta L(s)} \right)^2 s_{L(s)}^2 + \left(\frac{\delta L(0)}{\delta L_u} \right)^2 s_{L_u}^2 + \rho \frac{\delta L(s)}{\delta \tau} \frac{\delta L(s)}{\delta L_u} s_\tau s_{L_u} \right)$$

Assumptions: τ and L_u are correlated at $\rho = -0.85$ and measured radiance (i.e. $L(s)$ and $L(0)$) values are error free (i.e. $s_{L(s)} = s_{L(0)} = 0$).

values of τ and L_u are typically highly correlated (at least when they are predicted by a radiative transfer code). This means that errors in one tend to compensate in part for errors in the other. This is important because our knowledge of the atmospheric parameters is not particularly good in most operational cases. We have illustrated five cases. Four of them assume that the atmospheric variables are estimated using radiative transfer models and estimates of the atmospheric profile. Many alternative methods for atmospheric compensation exist (cf. Schott 1993). However, the only method available operationally for all sensors is the use of radiosonde data with radiative transfer codes so we have emphasized it here. How well the radiative transfer models work is very much a function of the accuracy of the atmospheric profile data, the clarity of the atmosphere and the spectral region being sensed. For these examples, we have consistently chosen a reasonable clear atmosphere and the 10.5–12.5 μm spectral region that is reasonably free of major atmospheric absorption features. As a result, we are looking at some of the best conditions available for remote sensing. In two cases, we use the radiative transfer model to propagate to space. In the first case, the radiosonde is assumed to be launched over the ground truth site at the time of overpass (i.e. a very well characterized atmosphere such as might be available during an intensive calibration effort). In the second case, we assume the atmospheric temperature and water vapor profile must be interpolated over space and time from the nearest radiosonde release site (i.e. most major airports in the world release balloons at 12:00 and 24:00 h GMT).

We are assuming that surface temperature and relative humidity data are available to force the interpolation to actual values for the lowest atmospheric layer. This greatly improves the radiometric calibration and can usually be done operationally. In propagating the error for these cases, we have assumed the surface radiance is known perfectly and estimated the error in sensor-reaching radiance due only to the atmosphere. We have also inverted the equation and computed the error in retrieved surface radiance due only to error in the atmospheric correction (i.e. assuming no error in sensed radiance). As we see the errors due to the atmosphere can be sizeable, particularly for the operational situation if only radiative transfer models are used.

The third case shows the effect of using the multialtitude technique (cf. Figure 10.10) to effectively collect ground truth and remove the atmosphere by a linear fit to equation (10.16). This approach yields more satisfying results but is only available for aircraft collections or in the rare case when an aerial system is flown under a satellite as part of a calibration campaign. The fourth and fifth cases are the same as the second except that we are only propagating from the aircraft altitude to space. These cases represent the cases used in underflight evaluation of a satellite sensor where we know the radiance reaching the aircraft and want to estimate the error introduced by the atmosphere in extrapolating to space (cf. Figure 10.9). These cases dramatically illustrate the value of using the aircraft data to generate "truth"

measurements above most of the atmosphere. Because the upper atmosphere is so thin, it has a small effect on the final signal. In addition, the upper atmosphere is less variable and, therefore, less error is introduced due to uncertainties in the temperature, water vapor profile. The net result is that the atmosphere introduces more tolerable errors. For all the cases studied, we have included errors associated with two target temperatures. This is because the compensating atmospheric errors tend to reduce errors in measurement of targets with temperatures close to the surface air temperature (i.e. about 290 K in this case). Thus, we see that the errors for the near ambient target are less than the hot target. It is important that the user carefully evaluate errors for the range of targets, atmospheric conditions, spectral bands, etc., of interest for each study.

For evaluating or updating the calibration of a satellite system, we would combine the ground truth errors computed using one of the approaches illustrated in Table 10.3 with the atmospheric error from an approach illustrated in Table 10.4 to generate the sensor-reaching radiance error. If the difference between the observed and predicted values exceeds the combination (root mean square) of the error from the estimation of sensor-reaching radiance and the instrument error (Table 10.2), we should question the calibration and perform an update using the best estimates of sensor-reaching radiance as inputs to equation (10.9). If the sensor is determined to be in calibration, we can predict the surface-leaving radiance using an atmospheric correction method. The error can be approximated as the root mean square of the sensor measurement error (Table 10.2) and the atmospheric correction error (Table 10.4). However, as in all these cases, a full error propagation should be performed to obtain precise estimates.

10.4.4 Errors due to spectral mismatch

If the ground truth sensors and the overflight sensor have the same bandpass our analysis is complete at this point. In practice, this is rarely the case. The ground truth or underflight radiometer will usually have bands that range from a nominal match to being considerably different. If the bands are a nominal match and both sensors avoid strong absorption features, then we may introduce little error if we use spectral radiance or apparent temperature values and assume the bands are identical. If the spectral mismatch is larger or there is strong spectral structure, then significant error can be introduced by spectral mismatch. This is a particularly significant problem in the 3–5 μm window where the steep slope of the Planck function introduces serious problems when spectral mismatch occurs. For our example cases, we have picked two degrees of spectral mismatch. Our space imaging sensor has a bandpass of 10.5–12.5 μm (we used the actual response of Landsat 5). In one case, we will look at using a ground truth radiometer with a bandpass of

10.4–12.4 μm. It is shifted slightly and has small changes in the shape of the spectral response such as we would expect from an instrument with different detectors and filters. In the second case, the ground truth instrument is substantially different with a spectral bandpass of 9.5–11.5 μm. For both comparisons, we look at targets at two temperatures and with two different spectral shapes. The temperature impacts the spectral shape of the Planck function (though this effect is minimal for targets near 300 K at ≈10 μm). For spectral shapes, we use water as one case, which is quite flat spectrally (cf. Figure 10.4), and a target whose emissivity increases linearly with wavelength at a rate of 0.02 emissivity units per micrometer starting at 0.88 at 8 μm. This target is characteristic of real objects exhibiting a moderate spectral structure.

To approximate the errors introduced by spectral mismatch, we will simulate the steps a smart user might apply to minimize mismatch artifacts. First, we use a radiative transfer model to generate an estimate for the atmospheric terms for one atmospheric condition, then we can solve for the surface-leaving spectral radiance observed by the ground truth instrument and the spectral radiance reaching the overhead instrument using the following expressions

$$L_{\Delta\lambda 1}(0) = \frac{\int [\varepsilon(\lambda) L_{\mathrm{BB}\lambda}(T) + (1 - \varepsilon(\lambda)) L_{\mathrm{d}\lambda}] R'_1(\lambda)\, \mathrm{d}\lambda}{\int R'_1(\lambda)\, \mathrm{d}\lambda} \tag{10.19}$$

and

$$L_{\Delta\lambda 2}(s) = \frac{\int ([\varepsilon(\lambda) L_{\mathrm{BB}\lambda}(T) + (1 - \varepsilon(\lambda)) L_{\mathrm{d}\lambda}] \tau(\lambda) + L_{\mathrm{u}\lambda}) R'_2(\lambda) \mathrm{d}\lambda}{\int R'_2(\lambda)\, \mathrm{d}\lambda} \tag{10.20}$$

where $L_{\Delta\lambda 1}(0)$ and $L_{\Delta\lambda 2}(s)$ are the effective spectral radiance observed by the ground truth and space instruments, respectively, and $R'_1(\lambda)$ and $R'_2(\lambda)$ are the spectral response functions for each instrument. The smart user recognizing a potential for spectral mismatch would take the effective surface spectral radiance and convert it to an apparent temperature $T_{\mathrm{app1}}(0)$ and then propagate this apparent temperature to the expected spectral radiance at the space instrument according to

$$L'_{\Delta\lambda 2}(s) = \frac{\int (L_{\mathrm{BB}\lambda}[T_{\mathrm{app1}}(0)] \tau(\lambda) + L_{\mathrm{u}\lambda}) R'_2(\lambda)\, \mathrm{d}\lambda}{\int R'_2(\lambda)\, \mathrm{d}\lambda} \tag{10.21}$$

The difference between $L'_{\Delta\lambda 2}(s)$ and $L_{\Delta\lambda 2}(s)$ represents the error due just to spectral mismatch. This is expressed as an apparent temperature error for the cases illustrated in Table 10.5. Consideration of Table 10.5 indicates that when the spectral mismatch is small, targets are spectrally flat and we are in

Table 10.5 Effects of spectral mismatch when comparing to a spacecraft spectral band from 10.5–12.5 μm radiance band. The subscript I refers to the ground truth and aircraft spectral band and 2 refers to the satellite spectral band

	$L_{\Delta\lambda 1}(0)$	$T_{app}(0)$	$L_{\Delta\lambda 2}(s)$	$T_{app}(s)$	$L'_{\Delta\lambda 2}(s)$	$T'_{app}(s)$	$L'_{\Delta\lambda 2}(s) - L_{\Delta\lambda 2}(s)$ (T_{app})
Case A: $\Delta\lambda = 10.4$–$12.4\,\mu m$							
Water							
$T = 285$	7.378	284.490	7.335	284.305	7.336	284.316	0.011
$T = 305$	9.886	304.113	8.988	297.777	8.974	297.665	−0.112
Spectral target							
$T = 285$	7.224	283.174	7.238	283.436	7.236	283.416	−0.019
$T = 305$	9.635	302.290	8.829	296.567	8.810	296.425	−0.142
Case B: $\Delta\lambda = 9.5$–$11.5\,\mu m$							
Water							
$T = 285$	7.535	284.545	7.335	284.305	7.340	284.353	0.049
$T = 305$	10.331	304.349	8.988	297.777	8.995	297.827	0.050
Spectral target							
$T = 285$	7.312	282.792	7.238	283.436	7.207	283.158	−0.278
$T = 305$	9.955	301.879	8.829	296.567	8.774	296.148	−0.419

Note
See equations (10.19)–(10.21) for expressions $L_{\Delta\lambda 1}(0)$, $L_{\Delta\lambda 2}(s)$, and $L'_{\Delta\lambda 2}(s)$.

a benign spectral region, the spectral mismatch error will be very small. However, even in the benign spectral region we are considering, the errors can become significant if the targets have spectral structure and the spectral mismatch is significant. In cases where significant error occurs due to spectral mismatch, there may be a partial remedy. This is because much of the error may be systematic and, therefore, predictable (this is especially true if it is caused primarily by the Planck function). While the details are beyond the scope of our treatment here, the conceptual approach is simple. We generate an error function to systematically correct the predicted radiance values $L'_{\Delta\lambda 2}(s)$ for spectral mismatch induced error. We then compute the non-systematic or random error using *Monte Carlo* methods. This random component is the error we would associate with a systematically corrected spectrally mismatched radiance estimate. This error must be combined with the error sources illustrated in Table 10.4 to generate the overall error.

10.4.5 Summary

In closing this section on error assessment and this chapter on calibration issues, the reader is reminded that accurate calibration of TIR systems is not a trivial task. Further, the example error estimates are very case-specific and should not be considered indicative of any particular instrument or scenario. Thus, in order to achieve proper calibration, careful laboratory and field

measurements such as those suggested here should be followed. In parallel with these procedures, a case-specific error assessment should be conducted so that we know what confidence to place on any measurement.

Appendix: definition of symbols

$L_{BB\lambda}$ Spectral radiance from a blackbody [W m^{-2} sr^{-1} μm^{-1}]
T Temperature (K)
h Planck's constant (6.6256×10^{-34} J s)
c Speed of light (3×10^{8} m s^{-1})
λ Wavelength (m)
k Boltzmann gas constant (1.38×10^{-23} K^{-1})
ε Emissivity
L_{λ} Spectral radiance (Wm^{-2} sr^{-1} μm^{-1})
R' Peak normalized spectral response
r Reflectivity
m', b' Gain (counts/radiance unit) and bias (radiance units) obtained using an internal onboard calibrator
m, b Instrument gain (counts/unit radiance) and bias (radiance units)
g, c The multiplicative gain (loss) and additive bias (radiance units) of the forward optics
L_{d} Downwelled radiance from the sky onto the earth's surface (W m^{-2} sr^{-1})
τ Atmospheric transmission
L_{u} Path radiance along the target sensor line of sight (W m^{-2} sr^{-1})
a, C Linear gain (counts/unit radiance) and bias (radiance units) incorporating instrument and atmospheric effects
h Aircraft altitude (km)
s Spacecraft altitude (km)
t Time (s)
i Subscript indicating bandpass number
b Subscript indicating background
j Subscript indicating detector number
o subscript indicating forward optical elements

References

Barker, J.L., R.B.Abrams, D.L.Ball, and K.C.Leung (1985) Radiometric calibration and processing procedure for reflective bands on Landsat-4 protoflight thematic mapper. Landsat-4 Science Characterization Early Results, NASA conference publication 2355, Vol. II, part 1, pp. 47–86.

Beers, Y. (1957) *Introduction to the Theory of Errors*. Addison-Wesley, Reading, MA.

Berk, A., L.S.Bernstein, and D.C.Robertson (1989) MODTRAN: a moderate

resolution model for LOWTRAN 7. GL-TR-89–0122, Spectral Sciences Inc., Burlington, MA.
Gillespie, A., T.Matsunaga, S.Rokugawa, and S.Hook (1996) Temperature and emissivity separation from advanced spaceborne thermal emission and reflection radiometer (ASTER) images. *Proc. SPIE* 2817:82–97.
Guenther, B.W. (ed.) (1991) Calibration of passive remote observing optical and microwave instrumentation. *Proc. SPIE* 1493:314.
Moeller, C.C., P.S.Grant, D.D.Laporte, L.E.Grumley, P.Hajek, W.P.Mansel, J.S.Myers, and S.White (1996) Blackbody emissivity considerations for radiometric calibration of the MODIS airborne simulator (MAS) thermal channels. *Proc. SPIE* 2820:44–55.
Palluconi, F.D. (1996) Validation of the ASTER thermal infrared surface radiance data product. *Proc. SPIE* 2820:97–104.
Rees, W.G. and S.P.James (1992) Angular variation of the infrared emissivity of ice and water surfaces. *Int.J. Remote Sens.* 13 (15): 2873–86.
Rubinstein, R.Y. (1981) *Simulation and the Monte Carlo Method.* Wiley, New York.
Salisbury, J.W. and D.M.D'Arian (1992) Emissivity of terrestrial material in the 8–14 μm atmospheric window. *Remote Sens. Environ.* 42:83–106.
Scarpace, F.L., R.P.Madding, and T.Green III (1974) Scanning thermal plumes. In *Ninth International Symposium on Remote Sensing of Environment.* Ann Arbor, Michigan.
Schott, J.R. (1979) Temperature measurement of cooling water discharged from power plants. *Photogrammetric Eng. Remote Sens.* 45 (6): 753–61.
Schott, J.R. (1986) Incorporation of angular emissivity effects in long wave infrared image models. *Proc. SPIE* 685:44–52.
Schott, J.R. (1993) Methods for estimation of and correction for atmospheric effects on remotely sensed data. *Proc. SPIE* 1968 (51): 448–82.
Schott, J.R. (1997) *Remote Sensing: The Image Chain Approach*. Oxford University Press, Oxford.
Schott, J.R. and E.W.Schimminger (1981) Data use investigations for applications explorer mission A (Heat Capacity Mapping Mission). Calspan Report No. 6175M-l, NASA Accession #E81–10079.
Singh, S.M. (1994) Effect of surface wind speed and sensor view zenith angle dependence of emissivity on SST retrieval from thermal infrared data: ATSR. *Int.J. Remote Sens.* 15 (13): 2615–25.
Snyder, W.C., Z.Wan, Y.Zhang, and Y.Feng (1997) Requirements for satellite land surface temperature validation using a silt playa. *Remote Sens. Environ.* 61 (2): 279–89.
Wolfe, W.L. (1998) *Introduction to Radiometry,* Vol. TT29. SPIE, Optical Engineering Press.

Chapter 11

MUST—a medium scale surface temperature mission dedicated to environment and agriculture

Alain Vidal, Philippe Duthil, Catherine Ottlé, Vicente Caselles, Antonio Yagüe and John Murtagh

11.1 Introduction

The Medium Scale Surface Temperature (MUST) study was carried out in the framework of the European Commission (DG XII) fourth "Research and Development Work Programme." The objective of this study was the definition and demonstration of interest of a large swath, medium resolution thermal infrared imager mission, named MUST. More precisely speaking the specific objectives were:

- to demonstrate the relevance and efficiency of the products of the MUST mission in the relevant application fields and to assess the economical benefits of the mission;
- to further develop methodologies for retrieving thermal- and waterrelated surface parameters from the sensor data;
- to design a medium-resolution, large-swath thermal imager, that is, compact and affordable;
- to analyze the operational implementation of the ground segment.

The study was co-ordinated by Matra Marconi Space (MMS) and their partners Cemagref (France), CNRS/CETP (France), the Universitat de Valencia (Spain), INFOCARTO (Spain), and the NRSCL (UK). It included the whole Mission and System definition process, starting with the definition of the user requirements, including the space and ground segments, the cost estimates, and ending with the evaluation of the MUST mission benefits versus costs and the final recommendations on the potential continuation of the programme. A development and implementation of the MUST sensor was then proposed in the framework of the European Space Agency Coastal Zone Earth Watch mission.

II.2 The MUST mission and related applications

II.2.1 Applications

The application of thermal infrared measurements from space are based on the relation existing between surface temperature and the soil and vegetation hydric state as introduced later. They can be classified into three main classes: (a) the assessment of the vegetation hydric state, important for applications such as agriculture (crop yield forecasts, potential stress due to drought, illness, or other pests), irrigation management, and forest fires risks assessment; (b) the assessment of surface (soil and vegetation) evapotranspiration, and thereby the evaluation of water consumption, useful for irrigation management and the evaluation of soil moisture that is helpful in hydrology applications; (c) the assessment of surface temperature itself or the air temperature as a by-product of surface temperature. The related applications are mapping frosts on agricultural surfaces or heat islands on urban surfaces. In addition, the MUST thermal infrared data are expected to be useful for the global monitoring of the biosphere and as a contribution to the Global Circulation Models providing data on the water fluxes at the global scale. The different fields of operational applications for the thermal infrared data are listed in Table 11.1.

II.2.2 The MUST information products

The MUST information products can be classified into three types, based on equation (11.1):

$$T_s = T_a + (T_s - T_a) \tag{11.1}$$

where T_s is the surface temperature measured by MUST and T_a the air temperature. This simple equation explains the double dependence of T_s on: (a) the climatic conditions, expressed through T_a; (b) the energy balance of

Table 11.1 Main land applications identified for a thermal imager

Domain	*Parameter of interest*
Agriculture	Hydric state of vegetation for crop yield forecasts and irrigation management
	Areas of frost risks
Irrigation	Irrigation water consumption assessment
Forests	Areas of fire risks
Hydrology	Hydric state of soil
Environment	Heat islands in urban centres
Scientific biosphere global monitoring, Global Circulation Models	Complement to VEGETATION data: water fluxes, hydric state of vegetation and soils

the considered surface, where equilibrium is the difference between surface and air temperatures ($T_s - T_a$).

Product type 1: vegetation stress index product

Measured through $T_s - T_a$, this product mainly concerns crop yield estimation in agriculture, irrigation monitoring, and risk assessment of forest fires. The evaluation of vegetation stress is derived from the analysis of the surface energy balance terms. The energy balance is usually expressed with the following equation:

$$R_n = G + H + LE \tag{11.2}$$

where R_n is the net radiation flux, G the soil heat flux, H the sensible heat flux, and LE the latent heat flux or evapotranspiration. This partition depends on the availability of water in soil (for soil evaporation) or in canopy (for canopy transpiration). As shown by many authors (Perrier 1975; Jackson *et al.* 1981), a reduction of soil/plant surface evapotranspiration results in an increase of $T_s - T_a$, whereas an increase of evapotranspiration results in a decrease of $T_s - T_a$.

Physically, T_s ranges from a maximum value of $T_{s\,max}$ when evapotranspiration is null ($LE = 0$) to a minimum value of $T_{s\,min}$ when evapotranspiration reaches its maximal (or potential) value $LE = LE_p$ (Moran *et al.* 1994; Vidal *et al.* 1997). LE_p depends on the atmospheric conditions (air temperature and moisture) and on the plant characteristics (resistance to heat exchange with air and resistance to evapotranspiration).

The ratio of actual LE to LE_p (LE/LE_p) provides a precise assessment of the vegetation stress, which is minimal when $LE/LE_p = 1$, and maximal when $LE/LE_p = 0$. Several indices have been developed to estimate this ratio, LE/LE_p, using remote sensing measurements. The more classical ones are based on the CWSI (Crop Water stress Index) approach where (Jackson *et al.* 1981):

$$\frac{LE}{LE_p} \approx 1 - \mathrm{CWSI} = \frac{T_s - T_{s\max}}{T_{s\min} - T_{s\max}} \tag{11.3}$$

Product type 2: daily/weekly surface evapotranspiration product

Estimated also through $T_s - T_a$, this product mainly concerns irrigation monitoring and water resources management. A generic expression has been derived by many authors (Jackson *et al.* 1977; Seguin and Itier 1983; Vidal and Perrier 1988) from the surface energy balance for estimating the daily evapotranspiration from an instantaneous midday remote sensing measurement of $T_s - T_a$:

$$LE_{\mathrm{d}} = R_{\mathrm{nd}} + A - B\,(T_{\mathrm{s}} - T_{\mathrm{a}}) \tag{11.4}$$

where LE_d and R_{nd} are the daily evapotranspiration and net radiation, A and B are constants depending on the canopy, and $T_s - T_a$ is the instantaneous difference between surface and air temperatures measured near midday.

Product type 3: interpolated air temperature, T_a

This is derived by correlating surface and air temperature, assuming air temperature to be known at some meteorological station point. Some of the primary applications include frosts prediction and detection of urban heat islands. A strong correlation is found between surface and air temperatures, when low air temperatures occur, which are the usual conditions when frosts maps or urban heat island maps, are required.

11.2.3 Methodology followed for assessing the user requirements and benefits

The User Requirements phase has been a major step in the definition of the MUST Mission and System, as no structured user community exists. The scientific community has not necessarily evaluated all the issues related to end-user requirements for information products using land surface temperature. The user requirements and benefit assessments have therefore been established with three National user groups in United Kingdom, Spain, and France (Table 11.2).

The user groups were involved in two main steps of the process. First, they expressed their requirements in terms of products and services. Second, after the products had been simulated, they indicated more precisely their interest for the products. This provided an assessment of the benefits derived from MUST products by the user community.

11.2.4 The information products' requirements and simulations

The main applications in agriculture, water resources, and forest fires will be presented henceforth. In all the cases, MUST surface temperatures were simulated from Landsat TM thermal IR data (120-m resolution). Since 250-m resolution was envisaged for MUST, Landsat thermal data were resampled at 250-m resolution using bicubic convolution. The maps presented in this chapter derived from such resampled thermal IR data.

Table 11.2 Composition of user groups in the three partner countries of the MUST project

France	Agriculture	EC MARS Project, Agricultural College/Research Institute (remote sensing department), Cereals Trader, Sugar Beet Technical Institute
	Irrigation	Irrigation companies (South West, South East)
	Forest fires	Forest administration (Haute Corse), Services Provider in Forest Fire mitigation
	Frosts prevention	Fruits Production Technical Institute, Forestry Producers Association
	Water resources	Water distribution by large companies
Spain	Agriculture, irrigation	Agronomic Research Institute
	Frosts risk and damage	Meteorological Institute
	Forest fires	Administration of Andalucia
	Heat island	Urban environment administration
UK	Agriculture	Agricultural Advisory Service (ADAS), Horticulture research, Farming online, Value added company
	Irrigation	ADAS, School of Agriculture (Silsoe)
	Frosts prevention	UK Met Office, British Sugar

MUST information products for agriculture

INPUTS TO YIELD PREDICTION MODELS

Users described that yield prediction models do not sufficiently take into account the actual vegetation stress. In this field, remote sensing is already used (e.g. by the EU MARS project), but it primarily involves the estimation of biomass using reflected solar wavelengths. Following the present tendencies in the use of EO data for yield prediction, it was suggested to use MUST data as a direct input in "efficiency" models, for example, the Monteith model (Monteith 1972), or the 3M "Modified Monteith Model" recently developed by the MARS project with Cemagref (Laguette *et al.* 1995, 1997). In these models, the dry matter (DM) is estimated as a cumulative product of efficiencies and global radiation (R_g), then transformed into crop yield using harvest indexes (HI). In this case, a MUST-derived water requirement satisfaction index SI can be used in the expression of the conversion efficiency, which is usually considered as a constant:

$$\text{Yield} = \text{HI} \cdot \text{DM} = \text{HI} \cdot \int \varepsilon_s \, (\varepsilon_{i0}\text{NDVI}_n) \, (\varepsilon_{c0} \, (t) \, \text{SI}) \, R_g \, dt \qquad (11.5)$$

where NDVI_n is the NDVI (normalized difference vegetation index) normalized between its maximal and minimal values during the crop season, $\text{NDVI}_n = (\text{NDVI} - \text{NDVI}_{min})/\text{CNDVI}_{max} - \text{NDVI}_{min})$, SI is a linear function of CWSI, ε_s is the climatic efficiency, ε_{i0} is the interception efficiency for maximal NDVI, and ε_{c0} is the conversion efficiency for maximal SI. The product of R_g

with efficiencies is integrated from the beginning of the cropping season to the date of the cycle where yield is estimated/predicted. The aforementioned authors have shown that, when the "3M model" is used with a continuous series of NOAA-AVHRR images, the final yield of wheat can be retrieved with a precision of 1.2 tons ha^{-1} instead of 2.4 tons ha^{-1} obtained when not accounting for water stress effects on yield.

SIMULATED PRODUCTS

The 3M model was applied on maize fields in the Orthez region (South West of France). Yield prediction figures obtained with remote sensing data have been compared to actual yield figures derived from *in situ* measurements in sample plots. The ideal process would have been to acquire remotely sensed data along the whole crop season with a sampling interval of typically 10 days and integrate them. Unfortunately, this was not possible because Landsat TM images were available in cloud-free conditions on a single date (20 July, 1996). Consequently, it was decided to compare this single date remote sensing result (which is actually the DM accumulation derivative) with the *in-situ* DM variation measurement averaged on the period around the available date.

The results, sketched in Figure 11.1, are not conclusive on the capability of IR-derived water stress information to improve the crop DM and yield prediction. Since this result is not coherent with the aforementioned MARS project research results, it is believed that it is a consequence of the single-date available acquisition.

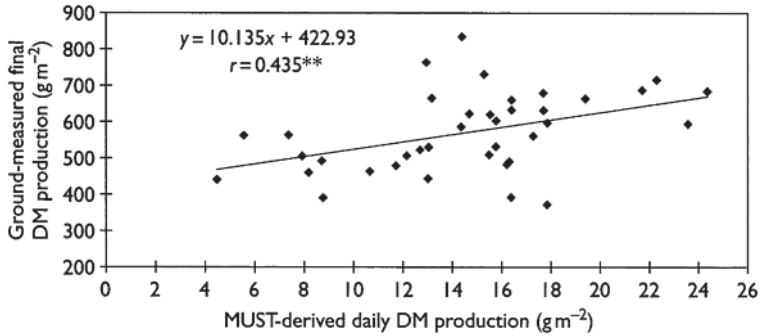


Figure 11.1 Comparison of the daily dry matter (DM) production estimated from onedate MUST-simulated thermal IR data with the ground measured final DM production on maize (Orthez—France).

MUST information products for irrigation and water resources

The users involved in irrigation, from both agricultural and water management points-of-view, identified three information products. In order of priority, these are: the spatial distribution of water consumption (derived from the evapotranspiration LE), maps of irrigated surfaces, and maps of crop water stress for monitoring water application and irrigation scheduling.

The users involved in water quality management (the domestic water distribution companies) were interested in soil moisture maps at the scale of small to medium watershed area. This information provides the means for identifying and assessing the importance of water contributing areas, as input for water quality models. They were also considering the crop water consumption (*LE* estimation) to derive infiltration/runoff as input for water quality models.

SIMULATED PRODUCTS

The objective of the simulations was mostly to show the users spatially distributed evapotranspiration information at 250-m resolution to demonstrate its advantage in comparison to sampled information and to 1-km resolution information. The simulated products are therefore daily evapotranspiration maps on the sites of Orthez (France) (Figure 11.2), the

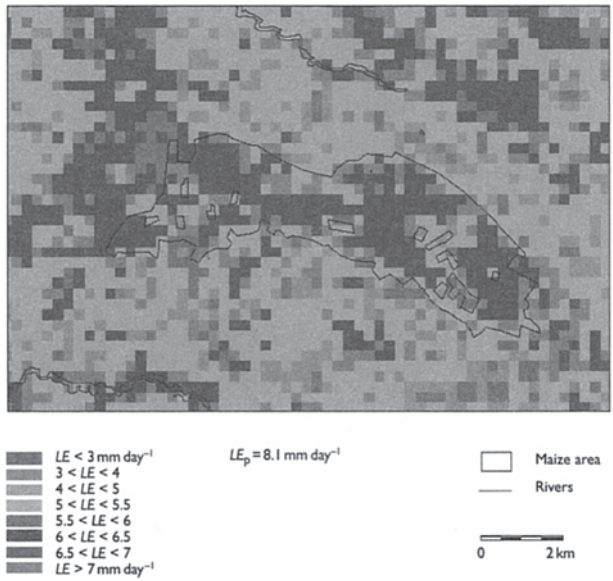


Figure 11.2 Daily evapotranspiration map obtained from MUST-simulated thermal data using equation (11.4) (Orthez—France) (see Colour Plate XXX).

Orgeval river basin (France, part of the Seine river basin), and of Barrax (Albacete–La Mancha-Spain), using the approach in equation (11.4).

Forest fires

Fire-fighting authorities have been using short-term fire risk indexes for a long time. These indexes are usually based on actual and predicted meteorological parameters, such as wind speed, air moisture, and temperature. Vegetation stress is usually represented by a simple budget between rainfall and potential evapotranspiration, which is difficult to transpose to forest areas, mainly due to spatial variations in the terms of this budget, and on how this budget is exploited by soil and tree root zones. It has recently been shown that using surface temperature measurements to derive the vegetation stress improved the fire risk prediction on both a short-term (daily forecast) and mid-term (weekly-monthly) range (Vidal *et al.* 1994; Vidal and DevauxRos 1995). Based on this rationale and on the operational way to fight fires in Corsica, two types of requirements were expressed by the fire fighting users:

- a real-time, daily-risk index integrating climatic and vegetation stress, at the scale of large forested areas (typically larger that 50,000 ha) useful for a better positioning of the fire fighting teams put in alert during summer months;
- a weekly risk index at a more local scale, usually for areas ranging from 5,000 to 20,000 ha, needed in order to support decisions on concentrating or moving means (staff and material) of fire watch patrols.

In addition, the forests officials were interested in two types of products:

- long-term risk maps on usually stressed areas to be used for the establishment of risk prevention plans at a 1/50,000 scale;
- *fire damage maps*: the thermal infrared data to be used in combination with visible, near-infrared (NIR), and short wave infrared (SWIR) data are expected to significantly enhance the accuracy of the damage maps established with visible, NIR, SWIR data only.

SIMULATED PRODUCTS

The different types of products have been simulated for Corsica (Figure 11.4) and Spain (Figure 11.5), assuming that MUST would enable an observation every day or 2–3 days. In the case of Corsica, an extension of CWSI (see equation 11.3) to sparse vegetation, called Water Deficit Index (WDI), has been used. This index, introduced by Moran *et al.* (1994) and applied to forests by Vidal and Devaux-Ros (1995), is based on the representation

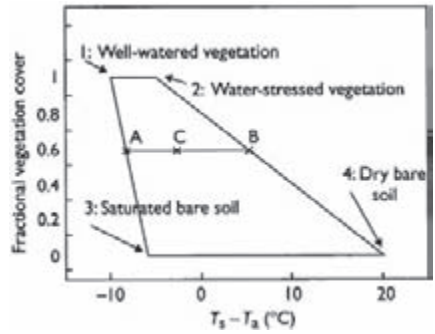


Figure 11.3 The theoretical trapezoidal shape showing the different biomass versus water stress conditions of the canopy-soil continuum (from Moran *et al.* 1994). The WDI of point C is given by AC/AB as shown in equation (11.6).

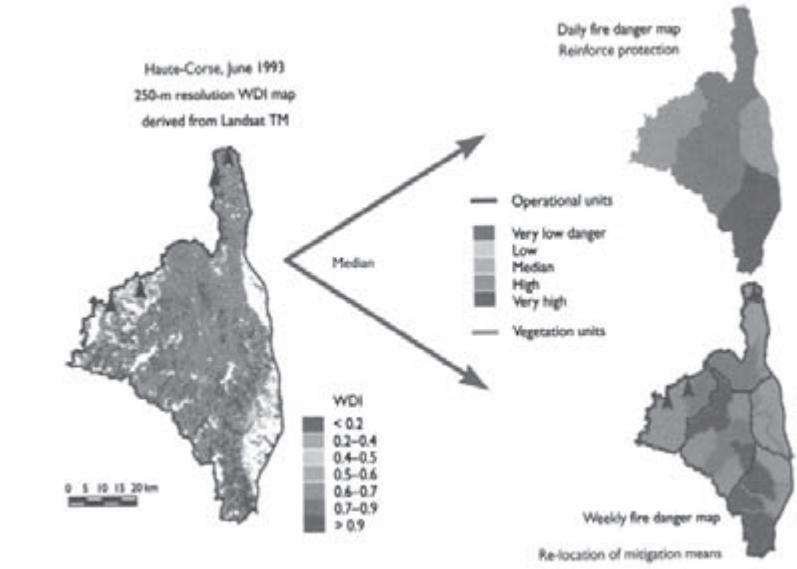


Figure 11.4 Daily and weekly fire risk index on the right part are the results of sub-sampling a full scale risk index obtained from MUST-simulated thermal data (on the left), useful for the establishment of 1/50,000 long-term risk maps (see Colour Plate XXXI).

of the soil-canopy continuum conditions in a fractional vegetation cover versus the difference between surface and air temperature ($T_s - T_a$) diagram. Actually, its position is theoretically comprised within a trapezoidal pattern: Figure 11.3 presents such a pattern and the definition of its limits.

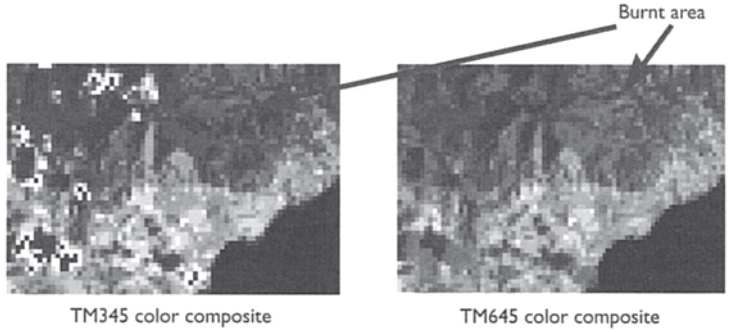


Figure 11.5 Classification of fire damaged areas using different bands of a Landsat TM image. Respectively, red, NIR, SWIR (on the left), and thermal infrared, NIR and SWIR (on the right). The latter provides a higher accuracy (see Colour Plate XXXII).

These authors have proposed both a theoretical and a graphic simple estimation of the soil-canopy evaporation for a given fractional vegetation cover, knowing its potential evaporation LE_p:

$$\mathrm{SAVI} = (\rho_{\mathrm{NIR}} - \rho_{\mathrm{R}})/(\rho_{\mathrm{NIR}} + \rho_{\mathrm{R}} + L)(1 + L) \tag{11.6}$$

where T_s is the composite surface temperature of the soil-canopy continuum as estimated from thermal infrared measurements, BC and AB are the distances represented in Figure 11.3, and the wet and dry indices correspond to the left and right limits of the trapezoid.

The main interest of this approach is the possibility of estimating both $T_s - T_a$ and fractional vegetation cover from remote sensing measurements. In the WDI approach, both NDVI and Soil Adjusted Vegetation Index (SAVI) have been used to estimate fractional vegetation cover:

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{R}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{R}}} \tag{11.7}$$

$$\frac{LE}{LE_p} = \frac{(T_s - T_a) - (T_s - T_a)_{\mathrm{dry}}}{(T_s - T_a)_{\mathrm{wet}} - (T_s - T_a)_{\mathrm{dry}}} = \frac{BC}{AB} = 1 - WDI \tag{11.8}$$

where ρ_{NIR} and ρ_{R} are the reflectances in the sensor's near-infrared and red wavebands, and L is a unitless constant assumed to be 0.5 for a wide variety of leaf area index values (Huete 1988).

11.2.5 System requirements derived from user requirements

From the above step of identification of MUST applications and information products, a synthetic table (Table 11.3) was prepared and validated by users during each user meeting and after national interviews.

Table 11.3 Synthesis of user requirements

	Agriculture	*Irrigation*	*Frosts heat islands*	*Forest fires*	*Water resources*
Information product	Yield forecast maps	Evapotransp. map, water stress maps	Air temperature maps	Water stress maps	Soil moisture evapo-transpiration map
Frequency	Daily to weekly	Weekly	Daily	Daily	Daily
Overpass time	Midday	Midday	Late night	Midday	Midday
Spatial resolution	250 m	250–100 m	250 m	250–1,000 m	250 m
Precision	1–2 K	0.5–1.5 K	<1 K orchards; 1–2 K trees	0.5–1 K	1 (SVAT)–2 K
Delivery	1–2 days	Week	Month	Real-time	Real-time

MUST frequency and overpass

A 1-day revisit was requested by most but actually only some applications require a real daily revisit (frosts, heat islands, forest fires) and this can be achieved quite easily since these applications concern atmospheric conditions or areas where the cloud coverage is low to null. The overpass time might be a critical issue. Most applications request data within 2 h before or after the time of maximal surface temperature (between 12.30 and 13.00 local solar time). This would be satisfied by a late morning overpass. However, frost monitoring requires night-time acquisition, which should preferably correspond to a late night overpass.

MUST spatial resolution

This critical issue was finally solved with 250-m resolution, thought to be sufficient for most applications. However, 100-m resolution was recommended only for irrigation. For this specific area, the expected information products are compatible with 250-m resolution, but might be perturbed by a succession of irrigated and non-irrigated fields.

MUST temperature precision

This issue was the most discussed. A simple look at bibliography and classical figures issued from other thermal IR missions (Landsat TM, AVHRR, ASTER, PRISM) actually showed that users required a 1 K precision whereas ground calibration cannot achieve a precision better than 1 K. Therefore, to be as user-oriented as possible, our assessment of user requirements on MUST

temperature precision was based on the analysis of information products precision, as it is easy to figure that the final temperature precision strongly depends on the expected information product.

APPLICATIONS BASED ON VEGETATION STRESS ESTIMATION

For all these applications, the users require that data from instrument be capable of discerning five stress classes for stress indexes ranging from 0 to 1. This means that a 10–20% precision on the stress index is fully acceptable, which, as shown in equation (11.3), corresponds to a 10–20% precision on the $T_{s\,min}$ - $T_{s\,max}$ range. The expected precision on the surface temperature measurement depends on meteorological conditions and was simulated for

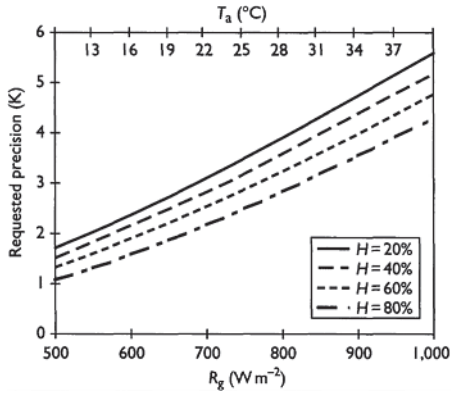


Figure 11.6 MUST requested precision for vegetation stress estimation on wheat (r_a = 40 s m^{-1}).

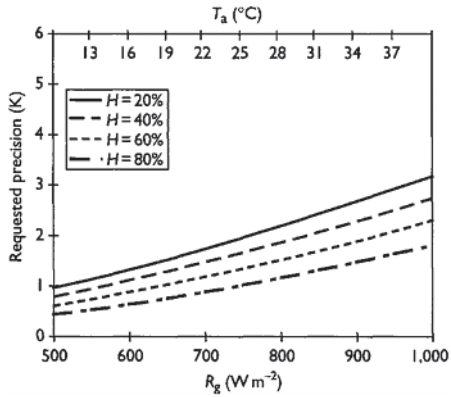


Figure 11.7 MUST requested precision for vegetationstressestimationonforest(r_a = 16 s m^{-1}).

wheat (Figure 11.6) and forest (Figure 11.7) as follows. The $T_{s\,min} - T_{s\,max}$ range mostly depends on canopy aerodynamic resistance r_a, air temperature T_a, global radiation R_g, and relative air moisture $H\%$. To simplify the simulation conditions, T_a was supposed to be related with R_g, and (T_a, R_g) couples were combined with different values of $H\%$ to derive values of T_s precisions corresponding to 20% of T_{smin}–T_{smax}.

These simulations show that for crop stress assessment a precision of 1–2 K is always acceptable, even in cold and humid conditions typical in northern Europe. Furthermore, a precision of 2–3 K is still acceptable for conditions with higher temperature and radiation values (e.g. for almost all irrigated regions). For forests on the opposite, simulations show that a precision of 0.5–1 K may be required for certain conditions (lower temperature and radiation values), which are rarely met during forest fire periods.

APPLICATIONS BASED ON SURFACE FLUX ESTIMATION

For these applications, two precision levels were requested. "Irrigation" users requested a precision of better than 10% on the estimation of the daily evapotranspiration, essentially to improve existing estimations. "Water quality" users required a precision of about 50%, but insisted on having a good description of the spatial distribution of such a typical three-class information to improve their hydrological models. If we consider the most constraining requirement and equation (11.4) with a maximal value of parameter B of 0.6 mm K^{-1} (Brasa *et al.* 1996), the required precision on the surface temperature can be estimated through: $\delta LE(\text{mm}) \approx 0.6\delta T_s$, with an expected δLE ranging from 0.3 (temperate regions) to 0.6 mm (Mediterranean and tropical regions), which yields a δT_s of 0.5–1.0 K for the worst case, that is, a final precision of 0.5–1.5 K.

APPLICATIONS BASED ON AIR TEMPERATURE INTERPOLATION

For these applications, interpolation of air temperature restricts the requirement to a relative error to be added to the precision obtained on T_a ground-measurements, usually considered to be about 0.5 K. The best precision was required for frost monitoring on orchards, as a difference of 0.5–1.0 K may be very important for certain fruit varieties.

II.3 The MUST system derived main characteristics

II.3.1 Main mission characteristics

The MUST instrument is to be accommodated on a low earth orbit space platform. The most likely candidate is an Earth Watch satellite of the European Space Agency to be launched around 2004 (Coastal Zone Earth Watch).

Table 11.4 Main characteristics of the MUST mission

Ground sampling	250 m at nadir, 460 m at swath edge
Swath width	1,420 km
Spectral channels	10.3–11.3 μm, 11.3–12.3 μm
Temperature resolution	0.24 K (lower band), 0.43 K (upper band)
Absolute temperature accuracy	1 K
Operations	Continuously operating and transmitting images in L band to local stations

The typical orbit of this satellite is sunsynchronous, with an altitude of 825 km and an overpass local time of 9.00 or 10.30 local solar time. An important issue extensively discussed during the MUST study was the combination of thermal IR data with visible and near-IR data, which is required by many of the applications described above. To simplify the instrument and keep it affordable, it was assumed that the candidate platform would provide simultaneous data in these wavelenghts. Therefore, this issue is not discussed here.

The main mission characteristics are provided in Table 11.4 and discussed subsequently.

INSTRUMENT SWATH

Most of the MUST candidate applications require a frequent revisit at the same location of a few days to 1 week, but considering cloud cover conditions, a daily to 2-day "orbital" revisit time appears to be mandatory. Considering a 820 km orbit, this requirement translates in a swath larger than 1500km (±40°) to ensure 2-day revisit in identical conditions (daylight or eclipse). For coherence with instrument design issues, the baseline for MUST is an instrument with a ±38°, 1400 km on the Ground, ensuring 1–3 days of revisit at latitudes around 40° as illustrated in Figure 11.8.

SPATIAL RESOLUTION

The drawback of the wide swath option is that it implies a medium spatial resolution. A 250-m resolution has been elected as a compromise between the user requirements and the different constraints for technological implementation.

SELECTION OF THE SPECTRAL BANDS

The issues of the atmospheric and emissivity corrections required to retrieve surface temperature from thermal infrared remote sensing measurements

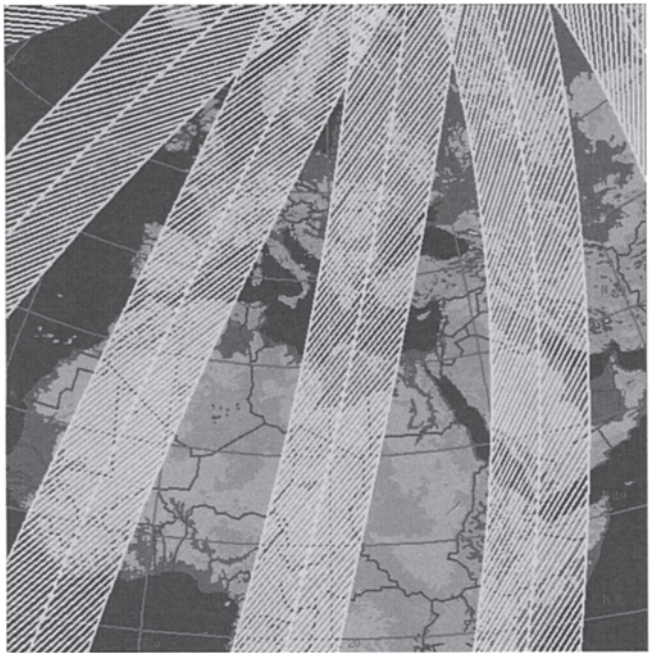


Figure 11.8 The MUST swath, 1400 km, with a ± 38° field of view on a 820 km orbit, ensures a 1–3 days of revisit time at latitudes around 40°. This figure shows orbit tracks and instrument swath on three successive orbits, from the right to the left, each separated by the 100-min orbital period.

have been largely addressed by many scientists. In the case of MUST, it was decided to adopt the approach proposed by Valor and Caselles (1996). The atmospheric correction is performed by the simultaneous use of two spectral channels around 10.5 and 11.5 μm and the application of the split window algorithm, with deduces the actual temperature from the measured temperatures according to the following equation:

$$T = T_1 + A(T_1 - T_2) + B(T_1 - T_2)^2 + C(1 - \varepsilon) + D\Delta\varepsilon + E \quad (11.9)$$

T_1 and T_2 are the equivalent blackbody temperatures measured by spectral channels 1 and 2. ε is the mean emissivity of the land target in channels 1 and 2 ($\varepsilon = \varepsilon_1 + \varepsilon_2/2$), $\Delta\varepsilon = \varepsilon_1 - \varepsilon_2$. The coefficients A, B, C, D, and E have been calculated by a regression analysis based on a very large range of different types of atmosphere. Depending on the data available, the emissivity figures ε_1 and ε_2 can either be known or need to be estimated. In this last case, a mean emissivity value is considered for vegetation (0.985) and soil (0.96). Then the emissivity of the land covered by both vegetation and soil is derived from their respective percentages of land cover and emissivity:

$$\varepsilon = \varepsilon_V P_{V+} \varepsilon_S (1 - P_V) + 4 < d\varepsilon > P_V(1 - P_V) \quad (11.10)$$

This approach has a significant operational advantage over some other methods in that it does not require a knowledge of the land characteristics (such as emissivity), or the atmosphere conditions. It also does not require several images (in day and night conditions).

OPERATIONS

The MUST sensor would operate continuously at 250-m resolution and the resulting data will be continuously transmitted in L-band to local users potentially present in the satellite transmission visibility circle. The advantage of the local ground stations is the immediate reception of the data, which is a significant advantage for these applications that only have a value if the information obtained is very recent. This is the case for some applications of MUST including, for instance, risk assessment of forest fires.

11.3.2 Overall system description

The overall system view is provided in Figure 11.9.

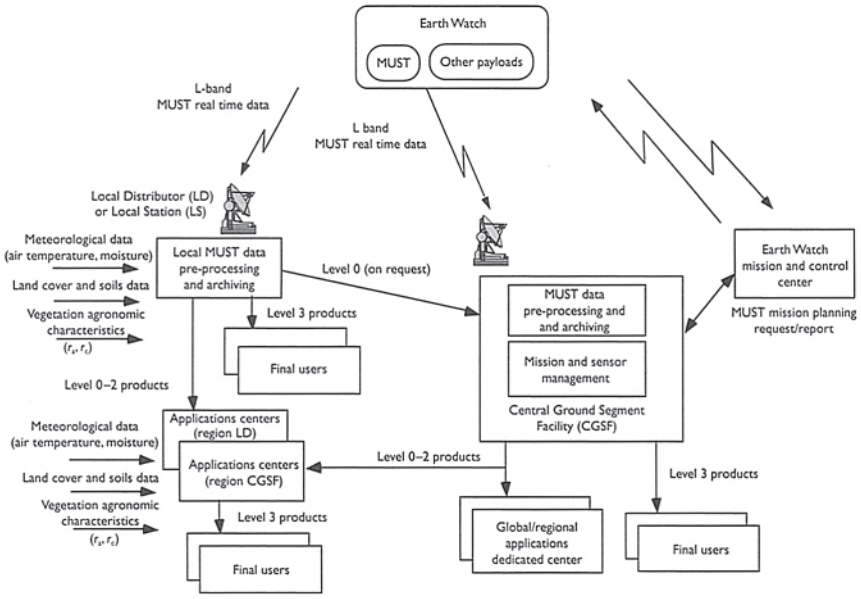


Figure 11.9 MUST overall system.

Space segment main features

In order to be independent from the host satellite (Earth Watch or any other), the MUST on-board data handling and communications (L-band) facilities are stand-alone and independent from the host satellite payload.

The MUST sensor data are not on-board recorded (no recorder accommodated on board). This is for cost-effectiveness and because it is believed that there will be a rapid network of local stations covering all the land areas, as such local stations can be procured at low cost and even through the upgrading of existing NOAA HRPT stations.

11.3.3 MUST products definition

The MUST products definition is sketched in Table 11.5.

11.3.4 Application ground segment (GS)

The preliminary proposed definition of the Application GS is based on a distributed architecture, where most of the value added products are generated in local distributed facilities. It consists of:

Table 11.5 The different levels of MUST products

Level	*Name*	*Processing definition*
0	Raw data, Channels 1 and 2	MUST raw numerical counts
1A	Brightness temperature at instrument input, Channels 1 and 2	Numerical counts conversion to radiance (instrument calibration coefficients) and then to brightness temperature (Planck's law inversion)
1B	Geo-corrected brightness temperature, at instrument input, Channels 1 and 2	The same as 1A for temperature retrieval. In addition, application of geometric correction and resampling in an earth reference, identically for both spectral channels
2A and 2B	Land Surface Temperature (LST) (2B geo-corrected with Ground control points)	Application of the split window algorithm to level 1B brightness temperatures (LST are therefore geometrically earth referenced)
3A and 3B	Geophysical value added products (3B geo-corrected with GCPs) • Water deficit index (WDI) • Evapotranspiration (LE) • Soil moisture (SM)	Application of specific algorithms to retrieve geophysical value added products, from level 2A LST

- A Central Ground Segment Facility (CGSF) that ensures the minimal required mission and sensor management, archives and sells data received at the main Ground Station (acting like an LD), performs routinely low-level processing of the MUST data up to level 1B, maintains an electronic data distribution and cataloging system covering the whole earth, and acts as a central point for commercial and technical information about the MUST mission and sensor.
- Local Distributors (LD) and associated Application Centers (AC): formed by L-band receiving antennas and associated processing facilities, geographically distributed world-wide to cover systematically the land areas, working under special favorable licensing and royalty agreements of mutual co-operation with the CGSF to build a global coverage, produce high-level value added products for their reception area and distribute on-time raw data and derived information products to end users (EU).
- Local Stations (LS) and Application Centers (AC): these are receiving (L band) and processing facilities freely distributed around the world, licensed to process their own received raw data. They could perform the basic processing of MUST data under the appropriate standard commercial licensing contract and generate free value added products for their own needs or local customers.
- Application Centers: free world-wide processing facilities without their own L band reception antenna that receive raw commercially licensed MUST data from either the central ground segment (CGSF) or an LD station to generate their own free value added products (i.e. levels 2 and 3) based on MUST data and usually other local sources of ancillary data.
- End Users: final users of MUST level 2 and 3 products distributed from CGSF or any AC of the world.

11.3.5 The MUST payload description

Overall payload presentation

The instrument is a pushbroom concept made of three optical modules of 26° field (400 km swath) each. The lines of sight of the modules are biased by ±26° in order to provide a total field of 78° and the swath width of 1400 km. A calibration system made of two blackbodies at about 250 and 350 K is fixed in front of the optical modules. A calibration mechanism allows the three modules to be fed successively by the blackbody. In rest position, the cold blackbody faces the cold sky. Heaters on both blackbodies are used to adjust the temperature. Electronic boxes are fixed on the opposite side of the bracket supporting the optical modules (Figure 11.10).

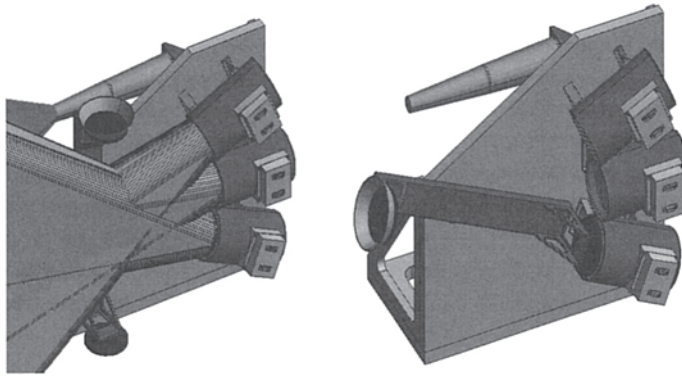


Figure 11.10 Overall concept of MUST instrument in observation and calibration modes.

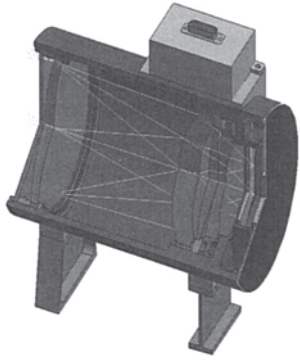


Figure 11.11 Optical module concept.

Optical modules

Each optical module consists of a four-lens dioptric system. A ZnSe entrance window limits the flux entering the telescope and ensures a quiet environment to the optics (Figures 11.11 and 11.12).

The microbolometer focal plane array

THE CHOICE OF MICROBOLOMETER TECHNOLOGY FOR THE MUST INSTRUMENT

Thermal detection is a technology that has been around for several decades. The amazing evolution of microelectronics, micro-machining, and thin films

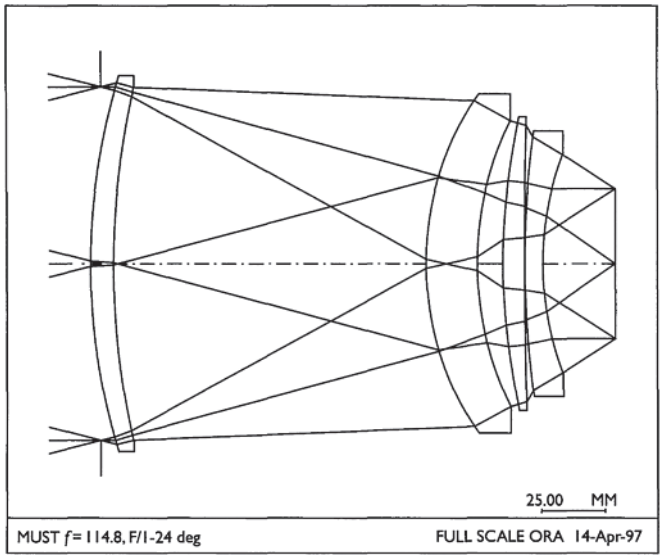


Figure 11.12 MUST module optical concept.

technologies has allowed recently two-dimensional arrays of thermal detectors to be built, at low cost, with better sensitivity and response speed for use in staring cameras dedicated to commercial and military applications, such as the 320 × 240 pixel thermal vision cameras. NEDT lower than 100 mK in the 8–12 μm range have been reported by several suppliers of detectors.

Three main types of thermal detectors are available, according to the physical parameter that senses the temperature change: the thermopile (electromotive voltage generated by Seebeck effect at junction as in thermocouples), the pyroelectric (polarization change by variation of the dielectric constant), and the bolometer (carrier density and mobility change that results in a resistance change). A trade-off analysis of these different technologies has been carried out and the microbolometer is today considered as the best candidate for the MUST instrument (performance, electronic chain complexity, thermal constraints, space environment adaptation, technology durability, potential evolution, and manufacturers availability.

THE MUST FOCAL PLANE ARCHITECTURE

The pixel size of 35 μm results from a compromise between the need to have a small pixel for the optics and the noise performances and technological constraint. In order to be compatible with standard CMOS, three elementary

butted modules of 512 pixels (18mm) are considered for the MUST focal plane. The two spectral channel separations needed for the temperature retrieval is performed by an in-field separation with two linear arrays on a common substrate, each with a dedicated bandpass filter in front of it (Figure 11.13).

The performances of microbolometers are closely related to the quality of the thermal insulation of their detectors and heat exchanges by convection have to be minimised. The internal volume of the detector package, sealed by a ZnSe window, needs to be evacuated for on-ground operation and then opened just before the launch. As thermal detectors behave as temperature sensors, their absolute response is therefore sensitive to the thermal variations of their environment. As usually done, the MUST focal plane will use thermo-electrical heat pumps to regulate its temperature. In addition, masked structures (blind pixels) will be used to compensate the residual temperature variations.

Overview of the MUST payload performance and main characteristics (Table 11.6)

Table 11.6 MUST instrument performances

Ground sampling	250 m at nadir, 460 m at swath edge
Swath width	1,420 km
Spectral channels	10.3–11.3 μm, 11.3 μm–12.3 μm
MTF	24% along track, 40% across track
Temperature resolution	0.24 K (lower band), 0.43 K (upper band)
Absolute temperature accuracy	1 K
Mass	60 kg
Power consumption	100W
Data transmission rate	3.5 Mbps

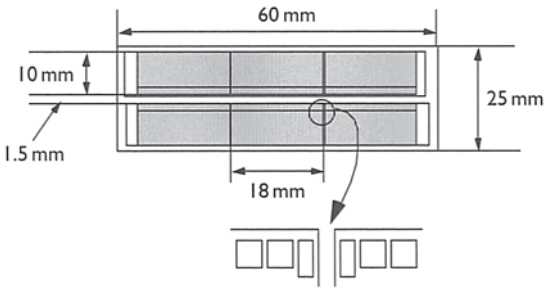


Figure 11.13 The MUST focal plane architecture.

Table 11.7 MUST economic assessment by region

	Minimal costs per year	*Minimal costs x 1.5 (per year)*	*Sales per year*
European market	6 M€	9 M€	2.4–4.5 M€
European + North American market	7.1 M€	10.6 M€	6.6–11 M€
European + North American market + Middle East market	8.2 M€	12.3 M€	9.1–14 M€

11.3.6 Mission economic assessment

The present paper does not detail the different steps of the study leading to the mission economic statement. The mission economic assessment results from the price policy, assessment of addressable market and sales, and estimated costs of the system. The mission is assumed to be viable if yearly sales exceed yearly costs by at least 50% (commercial margins and return on investment). The resulting business case is sketched in Table 11.7.

The profitability of the MUST System is clearly dependent on the size of the areas in the world covered by the mission. Considering only a European coverage, and here a European market, the MUST System cannot be profitable. The System reaches to a slight profitability when the North America and Near East areas are added. This situation will become more and more applicable when the covered areas and relative markets increase. The only condition for profitability of an additional LD station is that the cost for this addition is lower than the additional sales provided by this station. This is always true as it has been demonstrated for higher potentiality markets (North America) and lower potentiality markets (Near East).

11.4 Conclusions

The MUST project has been conducted with a large emphasis on the consideration of the user requirements and on the assessment of the benefits to be retrieved from MUST information products. In this respect, the work carried out with the user community is highly beneficial to both parties. It is beneficial to the mission and system designers to understand the user expectations and consequently design the system. It is beneficial to the users to learn about the system capacities, have a frame of reference for their reflection on potential applications, and accordingly imagine new applications for it. The generation of simulated information products also proves to be

very helpful for supporting the discussions on requirements and benefits and to raise new applications for the information products. The design of the MUST instrument is just sized to the user needs. This has been obtained thanks to the close co-operation with the user community. The resulting instrument is consequently light, uses well-adapted microbolometer technology, and could be produced at a low cost.

Economically, the MUST mission appears to be potentially profitable, once the areas covered by the satellite are wide enough. This could be the case when Europe, North America, and Near East are covered. This will be reinforced by additional coverage (South America, Asia, Australia, and Africa).

This conclusion may also be influenced by considering some markets that could be important, but which have not been considered in this study:

- Sea surface temperature-based products, mainly in coastal areas and inland waters because this is an application for which a better resolution than AVHRR is attractive. The concerned applications are pollution monitoring and fisheries support.
- Urban area temperature monitoring applications that are believed to be wider than the ones considered by the user groups: monitoring for the purpose of planning transportation networks, monitoring for operational "on line" control of the traffic induced pollution, etc.
- Scientific applications.

The MUST mission may then be recommended as a passenger payload on a host spacecraft. Another possibility is, of course, the payload to be part of an overall mission dealing with Land resources. This could be the case of the envisaged ESA "Coastal Zone Earth Watch" mission.

References

Brasa, A., F.M.Santa Ollala, and V.Caselles (1996) Maximum and actual evapotranspiration for barley (*Hordeum vulgare* L.) through NOAA satellite images in Castilla-La Mancha, Spain. *J. Agric. Eng. Res.* 63:283–94.

Huete, A.R. (1988) A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 27:47–57.

Jackson, R.D., R.J.Reginato, and S.B.Idso (1977) Wheat canopy temperature: a practical tool for evaluating water requirements. *Water Resour. Res.* 13 (3): 651–6.

Jackson, R.D., S.B.Idso, R.J.Reginato, and P.J.Pinter Jr (1981) Canopy temperature as a crop water stress indicator. *Water Resour. Res.* 17:1133–8.

Laguette, S., A.Vidal, and P.Vossen (1995) Using satellite data into productivity model. *Proc. IGARSS'95,* Florence, Italy, 10–14 July pp. 1465–7.

Laguette, S., A.Vidal, and P.Vossen (1997) Télédétection et estimation des rendements en blé en Europe. *Ingénieries-EAT* 12: pp. 19–33.

Monteith, J.L. (1972) Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 19:747–66.

Moran, M.S., T.R.Clarke, Y.Inoue, and A.Vidal (1994) Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sens. Environ.* 49 (3): 246–263.

Perrier, A. (1975) Etude physique de l'évapotranspiration dans les conditions naturelles. I. Evaporation et bilan d'énergie des surfaces naturelles. *Ann. Agron.* 26 (1): 1–18. II. Expressions et paramètres donnant 1'évapotranspiration réelle d'une surface "mince". *Ann. Agron.* 26 (2): 105–23. III. Evapotranspiration réelle et potentielle des couverts végétaux. *Ann. Agron.* 26 (3): 229–43.

Seguin, B. and B.Itier (1983) Using midday temperature to estimate daily evaporation from satellite thermal IR data. *Int. J. Remote Sens.* 4 (2): 371–83.

Valor E. and V.Caselles (1996) Mapping land surface emissivity from NDVI. Application to European, African and South-American areas. *Remote Sens. Environ.* 57:167–84.

Vidal, A. and C.Devaux-Ros (1995) Evaluating forest fire hazard with a Landsat TM derived water stress index. *Agric. For. Meteorol.* 77:207–24.

Vidal, A. and A.Perrier, (1988) Analysis of a simplified relation used to estimate daily evapotranspiration from satellite thermal IR data. *Int.J. Remote Sens.* 10 (8): 1327–37.

Vidal, A., F.Pinglo, H.Durand, C.Devaux-Ros, and A.Maillet (1994) Evaluation of a temporal fire risk index in mediterranean forests from NOAA thermal IR. *Remote Sens. Environ.* 49 (3): 296–303.

Vidal, A., C.Devaux-Ros, and M.S.Moran (1997) Atmospheric correction of Landsat TM thermal band using surface energy balance. *Remote Sens. Rev.* 15: 23–33.

Epilogue

The *raison d'etre* for this book is predicated upon our belief that TIR remote sensing data are of high utility for enhancing research in land surface processes. Moreover, we see where TIR data have not been exploited to their full advantage for land surface processes research by the science community, principally because of some common misconceptions regarding their availability, the methods and techniques used for information extraction and analysis, and their interpretability for input into models. As noted in the Preface, the development of this book stems from a workshop held in La Londe, France, in 1993, where a group of scientists congregated to develop a more cohesive framework for illustrating the merits of TIR data for identifying, characterizing, and quantifying surface thermal energy fluxes as key drivers to land surface processes. It became obvious as a result of this workshop that TIR data had in many respects an "image problem" due to these data being perceived as "recalcitrant" to work with by the broader Earth science research community. It is our hope that this volume considerably mitigates this misapprehension and helps TIR remote sensing to improve its "image" as an extremely useful tool for use in analyzing, quantifying, and modeling a host of land-surface energy flux-related characteristics.

As can be seen from the content of the various chapters of this book, however, TIR data certainly cannot be viewed directly as a panacea for resolving some of the more challenging questions related to the surface energy budget within the whole purview of land surface processes research. Without question, the utility and applicability of TIR data must be couched as a tool in the overall compendium of research techniques and analysis methods that can be used to address surface energy flux questions as part of land processes investigations; that is, the overall value of TIR data increases multifold when these data are integrated with other types of remote sensing data (e.g. multispectral or hyperspectral data) and with data derived from *in situ* instruments. TIR data, therefore, help to synergistically enhance land surface processes research by providing information that adds knowledge, and scientific value, to the whole exploratory experience.

Above all, we hope this treatise exists either directly or indirectly as a "how to" book for using and analyzing TIR data for land surface processes research. Throughout the chapters encompassed within the book, there are theoretical presentations, techniques, algorithms, caveats, and in-depth discussions that illustrate the "how to's" and "how not's" of collecting, preand post-processing, interpreting, analyzing, quantifying, and modeling of TIR remote sensing data, as well as elucidating the issues and concerns associated with TIR sensor calibration. Hence, this book may be viewed in many ways as an "instruction manual" on how to apply TIR remote sensing data in land surface processes research. We invite readers to shed any uncertainty they may have in dealing with TIR data, and take this book in hand as a guide for utilizing TIR data in their own research. We anticipate that in doing so, readers will allay their trepidation of working with TIR data and find that TIR remote sensing really does provide information that is both new and exciting, as well as inherently useful, to their research initiatives.

Lastly, one of the real concerns in putting this book together is that we see where TIR remote sensing instruments are somewhat of a "threatened species" in regard to their existence on future satellite platforms. For example, as of this writing, the present design of the follow-on to the NASA Landsat ETM+ (Landsat 7) platform does not include a TIR sensor. The elimination of a thermal spectral range will present a vacuum in the history of TIR data from the Landsat Thematic Mapper series of satellites that dates back to the mid-1980s. Moreover, there are no current plans for the design or launch of a high spatial resolution (i.e. ≤20m) satellite-mounted TIR sensor. Although moderate to coarse spatial resolution (i.e. 60m-lkm or more) TIR data are presently available (e.g. from the NASA Terra system), there is a critical need to obtain TIR data at spatial scales for quantifying surface energy fluxes of discrete surfaces (i.e. surfaces that can be identified as individual components of the overall heterogeneous landscape, such as pavements in cities, forest tracts, or cropland characteristics). Without this kind of fine resolution TIR data, we believe that land surface processes research will not be able to fully capture the surface energy fluxes that ultimately define or drive thermal energy dynamics across the landscape. We trust then, that this book will whet the appetite of those individuals who have not employed TIR data actively in their research and, therefore, precipitate the wider usage of TIR data. We also hope this book will cause those who have exploited these data to their research advantage to come together with us in voicing their opinion on the strong need for the continued collection of these data, as well as for the development of improved and enhanced satellite sensors and airborne that will make TIR data a universal and essential aspect of land surface processes research in the future.

Index