# Variational Assimilation of Remote Sensing Data for Land Surface Hydrologic Applications

by

Rolf H. Reichle

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Environmental Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2000



Author .....
Department of Civil and Environmental Engineering
January 20, 2000

Certified by .....
Dennis B. McLaughlin
H.M. King Bhumibol Professor of Water Resource Management
Thesis Supervisor

Certified by .....
Dara Entekhabi
Associate Professor of Civil and Environmental Engineering
Thesis Supervisor and Committee Chair

Accepted by .....
Daniele Veneziano
Chairman, Department Committee on Graduate Students

# Variational Assimilation of Remote Sensing Data for Land Surface Hydrologic Applications

by
Rolf H. Reichle

Submitted to the Department of Civil and Environmental Engineering
on January 20, 2000, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Environmental Engineering

## Abstract

Soil moisture plays a major role in the global hydrologic cycle. Most importantly, soil moisture controls the partitioning of available energy at the land surface into latent and sensible heat fluxes. We investigate the feasibility of estimating large-scale soil moisture profiles and related land surface variables from low-frequency (L-band) passive microwave remote sensing observations using weak-constraint variational data assimilation. We extend the iterated indirect representer method, which is based on the adjoint of the hydrologic model, to suit our application. The four-dimensional (space and time) data assimilation algorithm takes into account model and measurement uncertainties and provides optimal estimates by implicitly propagating the full error covariances. Explicit expressions for the posterior error covariances are also derived. We achieve a dynamically consistent interpolation and extrapolation of the remote sensing data in space and time, or equivalently, a continuous update of the model predictions from the data. Our hydrologic model of water and energy exchange at the land surface is expressly designed for data assimilation. It captures the key physical processes while remaining computationally efficient.

The assimilation algorithm is tested with a series of experiments using synthetically generated system and measurement noise. In a realistic environment based on the Southern Great Plains 1997 (SGP97) hydrology experiment, we assess the performance of the algorithm under ideal and nonideal assimilation conditions. Specifically, we address five topics which are crucial to the design of an operational soil moisture assimilation system. (1) We show that soil moisture can be satisfactorily estimated at scales finer than the resolution of the brightness images (downscaling), provided sufficiently accurate fine-scale model inputs are available. (2) The satellite repeat cycle should be shorter than the average interstorm period. (3) The loss of optimality by using shorter assimilation intervals is offset by a substantial gain in computational efficiency. (4) Soil moisture can be satisfactorily estimated even if quantitative precipitation data are not available. (5) The assimilation algorithm is only weakly sensitive to inaccurate specification of the soil hydraulic properties. In summary, we demonstrate the feasibility of large-scale land surface data assimilation from passive microwave observations.

Thesis Supervisor: Dennis B. McLaughlin
Title: H.M. King Bhumibol Professor of Water Resource Management

Thesis Supervisor and Committee Chair: Dara Entekhabi
Title: Associate Professor of Civil and Environmental Engineering

# Acknowledgments

First of all, I would like to thank my parents, who have provided unselfishly for me throughout my studies. Without their continued support, none of this work would have been possible, and their contribution is unmatched by my own effort. Special thanks also go to my grandparents, my sister, and all my friends for their encouragement and support.

Extended gratitude also goes to my academic advisors Dennis McLaughlin and Dara Entekhabi, who offered lots of help and ideas. Starting on the project was easy for me because this thesis grew directly out of their joint research proposal. In addition to continued funding, Dennis and Dara have always given me ample freedom to pursue my research.

Moreover, I am indebted to Alan Willsky from the Department of Electrical Engineering and Computer Sciences at MIT and to John Galantowicz from Atmospheric and Environmental Research Inc. who volunteered to serve on my thesis committee and provided many helpful discussions and comments.

Financial support for my stay at MIT was mainly provided by my advisors' research funds from the NASA Land Surface Hydrology Program and by a fellowship from the Joint Center for Geoscience at MIT, which supports collaborative efforts with NASA.

A debt is also due to Professor Andrew Bennett and his coworkers from the College of Atmospheric and Oceanic Sciences at Oregon State University. Their 1999 Summer School on Inverse Methods and Data Assimilation did not only offer me a welcome summer getaway, but also provided insights into the representer method which proved crucial to finishing the thesis.

I wish to thank the Earth System Science Center at Pennsylvania State University and the Oklahoma Mesonet for their invaluable data.

I would also like to thank the members of my research group and everybody at Parsons Lab. Very special thanks go to Freddi-Jo Eisenberg, not least for bringing me cookies during the qualifying exam and for proof-reading this thesis. Chris Stubbs very often provided helpful advice about how to survive as a graduate student. Without his help, I would still be fixing computers and trying to figure out how to use Arcview and Arcinfo. Thanks are also due to Johan Valstar, whose write-ups on the representer technique made the method accessible to me. Steve Margulis often and freely shared his time and advice when I ran out of ideas. Julie Kiang helped organize the graduation party and knows I have a weakness for maps. Thanks for inspiring the perfect graduation gift. Jeremy Pal and Garbo endured my practice talk for the thesis proposal examination and helped with many a computer problem. Personal thanks also go to Sheila Frankel who successfully lobbied for providing an Ethernet connection at everybody's desk. Finally, I am grateful for having Kirsten Findell as an office mate. Her smile and good humor make our office a much better place.

Last but very certainly not least, I would like to thank Amy for bearing with me, especially during the last few months which have at times been difficult. Thanks for helping me to keep things in perspective and for being on the same side.

# Contents

# List of Figures

# List of Tables

# Part I

# Introduction and Theory

# Chapter 1

# Introduction

## 1.1 Motivation

In recent years, understanding and quantifying the global hydrologic cycle has become a priority research topic. Soil moisture, in particular, has gained a lot of attention as it constitutes a key variable in the global hydrologic cycle. Land-atmosphere processes critically depend on the state of soil moisture, as soil moisture partitions the energy fluxes available at the land surface into latent and sensible heat fluxes. In addition, soil moisture conditions are important in determining the amount of infiltration and groundwater recharge.

Improving our understanding of soil moisture and temperature conditions will help us in many ways. Global circulation models are now routinely used in weather and climate predictions, but they usually contain only inadequate representations of the physical processes at the land-atmosphere interface. A better understanding of soil moisture and temperature dynamics will therefore help us with the assessment and prediction of global change and improve our ability to produce reliable short-term weather forecasts.

Sustainable management of water resources for agricultural and urban use will be feasible if we are able to more accurately quantify soil moisture conditions and the corresponding recharge into groundwater aquifers. Predicting floods is not only a question of knowing how much precipitation will reach the ground. An accurate flood forecast also depends on a good knowledge of the prevalent soil moisture conditions. Moreover, there is a feedback of soil moisture onto precipitation [Eltahir, 1998]. But to usefully incorporate such feedback mechanism into hydrologic and meteorologic predictions, including the forecast of droughts, it is again necessary to know the prevalent state of soil moisture.

Traditionally, improving models of large-scale soil moisture dynamics has been difficult due to the lack of corresponding large-scale observations. However, the advent of remote sensing data has now made it possible to study land-atmosphere processes on large spatial scales. Ideally, the satellite data are used in conjunction with the existing land-surface models to extract the valuable information contained in both the data and the models. Such optimal merging of data and models is generally termed data assimilation. In a variational assimilation scheme, the estimates are determined by minimizing a measure of fit between the land-surface states and both prior information and new data. The measure of fit is formulated using weights that depend on the corresponding uncertainties.

The observations that are available for assimilation are not always direct measurements of the land surface variables of interest. This is especially true for satellite remote sensing data. Satellites, for instance, cannot observe soil moisture directly, and only satellite-

observed radiances may be used to infer soil moisture conditions. In this case, the model consists not only of a component for soil moisture dynamics, but also of a forward Radiative Transfer scheme relating the soil moisture fields to the remotely sensed radiances (or brightness temperatures). Alternatively, off-line soil moisture retrievals could be obtained by inverting the Radiative Transfer model. It is preferable, however, to assimilate satellite radiances directly. It is much easier to specify appropriate weights for radiances in the objective function than to estimate covariance structures for soil moisture retrievals. Moreover, the off-line inversion of the Radiative Transfer model presents an unnecessary source of error.

Unlike sparse and infrequent observations, which can only be related to particular fields at particular times and locations, estimates produced by an assimilation scheme can provide a complete description in time and in space of the entire land-surface state, including soil moisture, soil temperature, and canopy temperature. This is achieved by using a dynamic model as part of the data assimilation algorithm. The data are effectively interpolated in time and extrapolated in space by respecting the dynamical and physical constraints. From such a complete picture, land-surface processes can be examined in detail. In meteorology, the number of investigators using such estimated "data sets" is probably much greater than those using any individual data type [Errico, 1999].

It is important to emphasize the fact that data assimilation reaches beyond mere model calibration. An optimal data assimilation algorithm will consider all the useful information and the errors contained in the model and the data along with the corresponding error statistics. In addition to the consideration of measurement error, a modern assimilation scheme usually involves the assumption of imperfect models, which is reflected in parameter and model error (or process noise) terms in the model equations. Moreover, posterior error covariances can be inferred.

Whereas model calibration is typically implemented to estimate a set of parameters once and for all, data assimilation algorithms are designed to run in an operational mode, continuously estimating state variables of interest. An additional feature of modern data assimilation algorithms is the possibility to test scientific hypotheses by formulating the model together with the statistical assumptions for the errors as a null hypothesis. If the hypothesis is rejected, the data are not statistically consistent with the underlying assumptions on the model and the errors. In this case, the estimates are of little meaning, but we would have learned something about land-surface dynamics. Finally, data assimilation provides a valuable tool for assessing and validating observation systems.

Data assimilation techniques have been successfully used meteorology and oceanography. In meteorology in particular, data assimilation has led to considerable improvements in the quality of short-term weather forecasts over the past few decades. Today, six hour global forecasts of wind and temperature produced with estimates derived from data assimilation algorithms are generally as accurate in a root-mean-square sense as most individual verifying observations themselves [Errico, 1999].

Hydrologists now face the challenge to apply true data assimilation techniques to all problems where remote sensing data can provide new insights. However, this is a difficult task due to the highly nonlinear nature of land-surface processes, the size of the problem, and the lack of data and experience to determine error statistics accurately. Consequently, the implementation of data assimilation techniques always requires trade-offs between resolution, complexity, computational effort, and data availability.

This study is predominantly a feasibility study. Its main goals are (1) to develop an

optimal land surface data assimilation algorithm, and (2) to determine how useful remotely-sensed L-band ($1.4GHz$) passive microwave measurements could be for the large-scale estimation of soil moisture.

## 1.2 Environmental Data Assimilation

In this Section, we briefly review a few influential studies in the field of environmental data assimilation with a focus on large-scale applications. Our primary goal is to introduce data assimilation techniques that have been used in the environmental sciences and to establish a broader frame of reference for this thesis. We certainly do not claim to provide a full review of the topic.

This Section covers a wide range of applications, mostly drawn from meteorology and oceanography. In meteorology, in particular, massive amounts of observations have been available for decades as operational data streams. Early on it has been indispensable to develop methods that make optimal use of these data for numerical weather forecasting and model development. More recently, large-scale operational observations that are useful for oceanographers have become widely available, and advanced data assimilation systems have been successfully developed and applied.

Large-scale hydrologic data assimilation, however, is still a field very much in its infancy. This probably owes as much to the scarcity of large-scale data as to the lack of consensus about how best to model land surface processes. In Section 1.2.3 we briefly present a few studies on hydrologic parameter estimation and data assimilation. A specific survey of soil moisture data assimilation can be found in Section 1.3.

Please note that our partitioning of the discussion into meteorologic, oceanographic, and hydrologic data assimilation does not at all imply that the methods used in these fields are different or separate. In fact, almost all assimilation techniques currently used in environmental data assimilation are simplifications or variants of the weak-constraint variational technique (Section 2.1) or, equivalently, the Kalman smoother [Gelb, 1974]. For details on the equivalence and approximations of the optimal methods consult the review papers cited below. All techniques can theoretically be applied to almost any dynamic problem in the geosciences, and the most important factor in determining the choice of method is usually computational feasibility.

### 1.2.1 Data Assimilation in Meteorology

*"One of the main reasons we cannot tell what the weather will be tomorrow is that we do not know accurately enough what the weather is today. [ ... ] Data at the initial time of a numerical forecast can be supplemented, however, by observations of the atmosphere over a time interval preceding it. New observing systems [ ... ] make it absolutely necessary to find new and more satisfactory methods of assimilating meteorological observations — for the dual purpose of defining atmospheric states and of issuing forecasts from states thus defined".* This quote is taken from the preface of a volume on progress in data assimilation published in 1981 [Bengtsson et al., 1981]. Almost two decades later, the European Centre for Medium-Range Weather Forecasts (ECMWF) has implemented a fully four-dimensional data assimilation algorithm in their operational forecast system [Klinker et al., 1999].

Since the early days of numerical weather forecasting, researchers have been trying to merge data and models. Excellent descriptions, reviews, and comparisons of the various

data assimilation techniques used in meteorology have been provided by Le Dimet and Talagrand [1986], by Ghil and Malanotte-Rizzoli [1991], and by Daley [1991] on general methods, by Lorenc [1986] on variational methods, and by Todling and Cohn [1994] on sequential methods. Courtier et al. [1993] compiled a literature list on the use of adjoints, variational methods and the Kalman filter in meteorology. The list includes comments and goes back to 1955.

### Weak-constraint Variational Assimilation

Unfortunately, truly optimal operational data assimilation with a full Kalman filter, a Kalman smoother or an equivalent variational technique[1] is still not computationally feasible, and the algorithm implemented at the European Centre (ECMWF) is not yet ideal. (Why this is so will be discussed below.) However, a successful large-scale research application of a fully optimal data assimilation approach has been presented by Bennett et al. [1996]. Their study is unique in that the model is only imposed as a weak constraint. In other words, errors in the model formulation are taken into account as process noise (or model error). The optimal estimate is derived with the variational representer approach, which is presented in detail in Section 2.3.

Bennett et al. [1996] invert a global Numerical Weather Prediction (NWP) model using about 2500 scalar data from reprocessed cloud-track wind observations. However, in an operational setting the assimilated data should include all of the global quality-controlled but otherwise raw observations. The authors point out that there are about 40,000 in situ observations alone [Daley, 1991], which clearly shows the current limitations of the technique in an operational context.

### Strong-constraint Variational Assimilation

There have been numerous attempts at simplifying either the model equations (although the physics did not change, of course) or the optimal estimation equations in order to make operational data assimilation computationally feasible. The algorithm recently implemented operationally at the European Centre (ECMWF) is based on the variational scheme 4DVAR [Thépaut and Courtier, 1991; Thépaut et al., 1993; Courtier et al., 1994]. In 4DVAR, the model is assumed perfect and imposed as a strong constraint, that is model errors are neglected. Only uncertainties in the initial and boundary fields are taken into account. If model error is present, as is certainly the case, using the model as a strong constraint may result in erroneous adjustments of the estimates. In other words, the estimates of the initial and boundary conditions must compensate for any significant model errors. However, unlike the weak-constraint method proposed by Bennett et al. [1996], 4DVAR is already feasible in the operational environment of the European Centre (ECMWF).

### Simplified Kalman Filters

4DVAR is certainly an improvement over conventional Optimal Interpolation [Rabier et al., 1993], which is used in most other weather forecasting centers (see below). But neglecting model errors does constitute a serious limitation. As an alternative to 4DVAR, one could

[1] Optimality refers to the implementation of the full Kalman filter, the Kalman smoother, or a weak-constraint variational algorithm. We ignore for a moment any suboptimality resulting from nonlinearities in the physics.

think of sophisticated approximations to the Kalman filter which allow for model errors but rely on a simplified propagation equation for the forecast error covariance. Such approximations are also called low-rank approximations of the full Kalman filter. Dee [1991] suggests a Kalman filter in which a simplified version of the dynamic model is used for the forecast error covariance propagation. Alternatively, Cohn and Todling [1996] suggest a filter based on partial eigenvalue decompositions of the forecast error covariance together with adaptive tuning based on reduced resolution ideas.

#### Optimal Interpolation

To this day, almost all operational weather forecast centers still use Optimal Interpolation or variants thereof as their method of choice. Optimal Interpolation can be viewed as a simplified Kalman filter in which the propagation of the estimation error covariance is entirely neglected [Daley, 1991]. In return for the computational savings, the complicated and time-dependent error covariance fields of the atmospheric states must be accurately estimated. In practice, this is quite impossible and leads to rather suboptimal assimilation algorithms, even though the name of the method would suggest otherwise. Moreover, Optimal Interpolation is usually implemented in the spectral domain, which limits the choices of error covariance models in practical applications. Finally, the approximate solution method for the update equations in conventional Optimal Interpolation can lead to dynamic imbalances [Cohn et al., 1998].

For these reasons, NASA's Data Assimilation Office has recently developed the Physical-space Statistical Analysis System (PSAS) as an improved variant of Optimal Interpolation [Cohn et al., 1998; Chen et al., 1999]. The new method has been included into the Goddard Earth Observing System (GEOS) data assimilation package. PSAS operates in physical space rather than in the spectral domain and employs a different numerical method to solve for the updates. Since it is operating in physical space, PSAS is capable of using more advanced error covariance models than conventional Optimal Interpolation.

### 1.2.2 Data Assimilation in Oceanography

In oceanography, operational data streams have not been available in the same way as in meteorology. Therefore the focus of the investigations has been somewhat different, oriented more towards learning about ocean dynamics by using optimal methods for individual case studies whenever data are available. Ghil [1989], Bennett [1992], Ghil and Malanotte-Rizzoli [1991], Evensen [1994a], Malanotte-Rizzoli [1996], and Wunsch [1996] offer good collections, descriptions, reviews, and comparisons of the various attempts to solve inverse problems in oceanography.

#### Weak-constraint Variational Assimilation

Egbert et al. [1994] use the direct representer algorithm (Section 2.3) to estimate global tides from the TOPEX/POSEIDON altimeter data. Even though the tide model is linear, which is rarely the case for geophysical applications, the number of remote sensing data is still too big for a naive implementation of the direct representer approach. Egbert et al. [1994] therefore develop a set of steps in which they reduce the dimensionality of the problem.

Eknes and Evensen [1997] extend the representer formalism by solving a simultaneous parameter and state estimation problem with a weak-constraint formulation for an Ekman

model. Finally, Bennett et al. [1998] apply the indirect representer algorithm to assimilate data from the Tropical Atmosphere-Ocean (TAO) array into a coupled model of the tropical Pacific. They compare their weak-constraint approach with a strong-constraint algorithm and reach the important conclusion that the assumption of a perfect model must be clearly rejected in this case study.

### Simplified Kalman Filters

A low-rank approximation of the Kalman filter has been applied by Verlaan and Heemink [1997] to the tidal flow forecasting problem. Their approach is to combine a reduced rank approximation of the error covariance with a square root factorization. The use of the factorization ensures that the error covariance matrix stays positive-definite at all times, while the smaller rank reduces the computational effort.

Asif and Moura [1999] develop a computationally efficient formulation of the optimal Kalman filter which is based on the block structure that results from the discretization of the partial differential equations commonly used in the physical sciences and from the sparseness of the measurements, for example satellite scans. The authors further develop an approximate implementation of the block Kalman filter. Underlying this simplified implementation is the approximation of the inverse error covariance matrix, that is the information matrix, by a sparse block banded matrix. Such banded approximations correspond in essence to modeling the error field in the spatial estimates at each point in time as a reduced-order Markov random field. To demonstrate the concept, Asif and Moura [1999] use the optimal filter and the simplified scheme to assimilate synthetic satellite altimeter data into a linear shallow water model. The comparison shows that the suboptimal filter performs well and that the approximations of the simplified filter are reasonable.

Yet another simplification to the Kalman filter for large-scale applications has been proposed by Evensen [1994b]. In the so-called Ensemble Kalman Filter, the error covariance is propagated with a Monte Carlo method. Instead of solving the Riccati equation for the error covariance evolution, the scheme is based on propagating an ensemble of model forecasts. If a measurement becomes available, the forecast error covariance needed for the update step is estimated from this ensemble. Evensen and van Leeuwen [1996] use the Ensemble Kalman Filter to assimilate Geosat altimeter data into a two-layer quasi-geostrophic model of the Agulhas Current. The validity of the ensemble approach is obviously dependent on the size of the ensemble. It does seem daring to estimate the forecast error covariance from an ensemble of 500 model trajectories when the state vector is approximately of dimension 100, that is when the error covariance matrix contains on the order of 10,000 elements.

Very recently, Lermusiaux and Robinson [1999a] presented an assimilation scheme based on a combination of the Ensemble Kalman Filter and a reduced rank approximation. Like in the Ensemble Kalman Filter, the error covariance is propagated with a Monte Carlo approach. Before the update step, however, the covariance matrix is reduced in rank. The authors formulate an objective criterion to decide whether the addition of another member to the existing ensemble is necessary or not. The result is a suboptimal filter which tracks an evolving error subspace in space and in time. Consequently, the scheme is termed error subspace statistical estimation (ESSE).

In a companion paper, Lermusiaux and Robinson [1999b] apply their algorithm to shelf-break front simulations in the Middle Atlantic Bight. Identical twin (synthetic) experiments are conducted under the assumption of a perfect model, that is model errors are neglected.

Moreover, the synthetic observations do not contain measurement error, although a small measurement error is used in the estimation algorithm. The proposed filter compares favorably to a traditional Optimal Interpolation scheme which does not include any error covariance propagation. In case the computational demand for the proposed filter is too big for a given operational application, the authors suggest that their scheme could be used to improve the parameterizations of the Optimal Interpolation approach.

#### Multiresolution Optimal Interpolation

Fieguth et al. [1995] have developed and applied a new variant of Optimal Interpolation. The goal is to provide interpolated estimates at multiple resolutions at the time of the update or analysis step. The multiresolution algorithm is a generalization of time series state-space models for which the Kalman filter is an efficient estimator. When applying the multi-scale estimation technique, the biggest task is to build a model for the particular application that fits the covariance matrix at the finest scale. In addition to providing interpolated estimates and accompanying error variance statistics at multiple resolutions, a striking advantage of the multi-scale estimation framework is that its complexity scales linearly with the problem size. Moreover, the efficiency of the algorithm is entirely insensitive to irregularities in the sampling or spatial distribution of measurements and to heterogeneities in measurement errors or model parameters. Consequently, the approach has the potential of being an effective tool in a variety of remote sensing problems.

Fieguth et al. [1995] have applied the multiresolution estimation algorithm to the interpolation and statistical analysis of the TOPEX/POSEIDON altimeter data in the North Pacific Ocean. Another application of the multi-scale Optimal Interpolation algorithm to the mapping of temperature in the northeastern Pacific has been published by Menemenlis et al. [1997]. The authors also concern themselves with the development of a class of multi-scale statistical models appropriate for oceanographic mapping. Finally, Fieguth et al. [1998] have applied the method to map the sea level anomaly of the Mediterranean Sea based on TOPEX/POSEIDON and ERS-1 data. Unfortunately, the extension of the multiresolution framework to problems with temporal evolution presents formidable challenges. The development of temporally dynamic models is the subject of ongoing research.

### 1.2.3 Data Assimilation in Hydrology

In hydrology, inverse methods have traditionally been focusing on parameter estimation and model calibration rather than state estimation. In particular for groundwater inverse problems, measurements are scarce, and highly heterogeneous parameters such as the hydraulic conductivity are virtually unknown a priori. McLaughlin and Townley [1996] offer an excellent review of the subsurface data assimilation problem. Also, Zimmerman et al. [1998] compare seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Recently, Reid [1996] and Sun [1997] have worked on parameter estimation in groundwater contaminant transport problems. Finally, Daniel et al. [1999] have applied the multiscale estimation approach described in Section 1.2.2 to the estimation of solute travel time.

Hydrologic data assimilation as a state estimation problem has only very recently become a topic of widespread interest. In a review of hydrologic data assimilation published in 1995, McLaughlin [1995] is *"unaware of any studies which use distributed watershed models*

*to assimilate field data"*. Recent data assimilation efforts to estimate soil moisture are summarized in Section 1.3.

## 1.3 State of the Art of Soil Moisture Data Assimilation

In this Section, we attempt to assess the state of the art of data assimilation techniques used for the estimation of soil moisture. Only selected works will be discussed, with a focus on studies that apply optimal estimation techniques or at least non-trivial approximations thereof. By no means do we claim for our overview to be complete.

Two obvious classification schemes can be applied. In a first scheme, the studies can crudely be classified into a first category consisting of spatially one-dimensional physical models using synthetic data [Entekhabi et al., 1994; Milly, 1986], or small-scale field data [Mahfouf, 1991; Katul et al., 1993; Parlange et al., 1993; Galantowicz et al., 1999; Calvet et al., 1998; Callies et al., 1998; Bouyssel et al., 1999; Castelli et al., 1999], and a second category, in which large-scale field cases have been investigated [Houser et al., 1998; Bouttier et al., 1993b; Rhodin et al., 1999]. In the first category of small-scale studies, the dimensions of the state vector and the observation vector are small, and the computational effort for truly optimal estimation is easily bearable. The models in the second category of large-scale applications are horizontally distributed and of high dimensionality. Consequently, only suboptimal filters have been implemented to date.

A second classification could be based on the data types that are assimilated. With the exception of [Castelli et al., 1999], the studies of Sections 1.3.1 and 1.3.2 use either direct measurements of soil moisture or remotely sensed brightness data which are very closely related to surface soil moisture. Since such measurements of soil moisture are not yet available operationally, there have been numerous investigations on soil moisture data assimilation from low-level atmospheric parameters such as air temperature and relative humidity at $2m$ above the ground. However, these parameters are only weakly and indirectly related to surface soil moisture. The latter studies are geared towards improving numerical weather prediction and treat soil moisture rather as a tuning parameter. For this reason we describe them in the separate Section 1.3.3.

### 1.3.1 One-dimensional Optimal Estimation Approaches

If the modeled land surface system is one-dimensional and contains only a single vertical column, the dimension of the state vector is small and the application of truly optimal estimation techniques is not limited by computational resources. One such optimal technique is the Kalman filter [Gelb, 1974], which has been used by many investigators. Other investigators have applied a variational approach, which is described in detail in Section 2.1. Among the latter are Mahfouf [1991], Callies et al. [1998] and Bouyssel et al. [1999]. Since they assimilate low-level atmospheric observations to infer soil moisture, their studies are discussed in Section 1.3.3, even though optimal variational assimilation methods are used.

#### The Study by Milly [1986]

Milly [1986] presented a study to determine the optimal temporal characteristics of a remote soil moisture sensor. He uses a very simple linear soil moisture model in which the parameters are perfectly known. The forcing consists of a sequence of equally spaced Dirac delta

functions to model the precipitation input. Milly [1986] uses a full Kalman filter to evaluate the relative merits of the accuracy and the sampling frequency of the measurements.

### The Studies by Katul et al. [1993] and by Parlange et al. [1993]

Katul et al. [1993] use an Extended Kalman Filter (EKF) for the estimation of the soil moisture state in a simple bucket model. The model is obtained from the depth-integration of a one-dimensional version of Richards' equation, which results in a nonlinear state-space formulation with a scalar state. A no-flow boundary condition is imposed at the top, and the flux at the lower boundary is prescribed. The hydraulic conductivity and the soil moisture content are related through a simple exponential-type two-parameter model.

The assimilated soil moisture data are neutron-probe measurements from a small field drainage experiment carried out by the authors. In addition to the state estimation, Katul et al. [1993] also estimate the two soil hydraulic parameters, the initial estimation error variance, and the model error of the state-space formulation. These four parameters are determined through repeated runs of the Extended Kalman Filter. For every run, a set of parameters is guessed, and a goodness-of-fit objective function is evaluated. The goodness-of-fit is measured with a sum of squared differences between the predicted states and the corresponding measurements. No prior information about the parameters is used. The final set of parameters is then given by the best fit. To carry out the optimization, the authors implemented a simplex scheme.

In a similar study, Parlange et al. [1993] estimate the field scale diffusivity together with the initial estimation error variance and the model error of the state-space formulation. The starting point here is an approximate solution to the depth-integrated diffusion equation, combined with a water balance equation. Once the model equation is cast into a state-space formulation, the mechanics of the estimation algorithm are identical to the approach by Katul et al. [1993].

### The Studies by Entekhabi et al. [1994] and by Galantowicz et al. [1999]

The studies by Entekhabi et al. [1994] and by Galantowicz et al. [1999] stand out because an optimal data assimilation approach is applied to a multi-layer model of soil moisture and temperature dynamics. The authors use a Kalman filter to update the temperature and moisture profile from observations of the brightness temperature. The spatially one-dimensional model is entirely physically-based, making use of Richards' equation, the heat equation, and a model for the radiative transfer.

Entekhabi et al. [1994] show that it is possible to infer information about the temperature and the moisture at depths below the penetration depth of the microwaves. Note, however, that the focus is on the methodology. Most importantly, the data are completely synthetic and vegetation is not modeled. Only one vertical column is considered. In addition, updates from the brightness temperature and the infrared temperature data are made hourly, which is not a very realistic situation.

In a very recent study, Galantowicz et al. [1999] present an assimilation algorithm which is based on the Kalman filter and similar to the one in Entekhabi et al. [1994]. The algorithm is tested on field data, namely data from the Beltsville Agricultural Research Center (BARC), Maryland, taken during a seven-day drydown in July 1994 [Jackson et al., 1997]. Moreover, the authors test their algorithm with a four-month series of simulated operational conditions. The results indicate that the soil moisture profile can indeed be

retrieved from updates of the brightness temperature made only every three days, and that the proposed data assimilation scheme is stable.

### The Study by Calvet et al. [1998]

Calvet et al. [1998] present a comprehensive study on the feasibility of retrieving root zone soil moisture from surface soil moisture or surface soil temperature observations. They use the ISBA (Interaction between Soil, Biosphere, and Atmosphere) surface scheme of the French weather forecast system, which models soil moisture in just two layers, a very shallow surface layer and a deep reservoir.

The assimilation technique is a strong-constraint variational method. The uncertain parameter is the initial soil moisture of the deep reservoir, and the objective function to be minimized consists of the root mean square difference between the measured and the simulated values of the observed surface soil moisture content. No prior regularizing term is included in the objective. The data are from two months of field observations taken in Spring and Fall 1995 in southern France. The assimilation period is either a moving fifteen-day window or a moving five-day window during the thirty-day observation periods. In a series of assimilation experiments, observations are available to the estimation algorithm from twice daily to once every four days.

In conclusion, Calvet et al. [1998] suggest that deep soil moisture can indeed be retrieved with reasonable accuracy from surface soil moisture observations once every three days, but concede that soil moisture estimation from soil temperature measurements can at best work under dry conditions. Finally, the authors conclude that the length of the assimilation window should not be less than ten days.

### The Study by Castelli et al. [1999]

A major goal of the study by Castelli et al. [1999] is to reduce the data needs for surface flux and soil moisture estimation. Therefore, the authors only assimilate observations of ground temperature, which are readily obtained from current remote sensing platforms. The uncertain input is a time-dependent parameter which is called soil moisture index. The soil moisture index describes the limitation of evaporation due to the limited availability of soil water and is closely related to the surface heat flux.

Castelli et al. [1999] use a variational technique and include the surface energy balance as a physical constraint in the objective function. In mathematical terms, the estimation of the time-dependent soil moisture index amounts to the estimation of a state-dependent model error term. The scalar weights used in the objective function imply that this model error is not correlated in time.

Estimates of the surface heat flux and the soil moisture index are derived from the data of the First International Satellite Land Surface Climatology Project Field Experiment (FIFE). The experiments cover the summer months of 1987 and 1988, but the individual assimilation windows are limited to one day. Daily averages of the estimated surface heat flux compare well to independent latent heat flux observations. However, the authors conclude that there is a need to discriminate between soil moisture and aerodynamic contributions to the surface control over evaporation.

### Discussion

The straightforward application of optimal data assimilation algorithms to one-dimensional problems has met with fair success, and the potential for inferring soil moisture from remote sensing observations of passive microwave data has clearly been demonstrated. In light of these results, the most pressing question is how best to extend the techniques presented above to large-scale applications.

## 1.3.2 The Study by Houser et al. [1998]

The so far most comprehensive study on soil moisture data assimilation has been carried out by Houser et al. [1998]. The authors modified and extended the TOPLATS land-atmosphere transfer scheme [Famiglietti and Wood, 1994a; Famiglietti and Wood, 1994b]. TOPLATS is a spatially distributed hydrologic model to predict the diurnal dynamics of the water and energy fluxes at the land surface as well as the local vertical recharge into the underlying aquifer. Its algorithms are intentionally simpler than the ones used in operational surface-vegetation-atmosphere transfer schemes (SVATS). The basic components of TOPLATS are water balance equations for the canopy and the soil as well as an energy balance equation at the surface. The original model describes the unsaturated zone with two layers, a root zone and a transmission zone. Houser et al. [1998] added a shallow third soil layer at the top. The soil moisture in this new soil layer can possibly be inferred from remote sensing. The soil hydraulic properties are parameterized with the model of Brooks and Corey [1964] and the soil moisture dynamics are based on an approximate analytical solution of Richards' equation using infiltration and exfiltration capacities [Eagleson, 1978]. Horizontal flow exists only in the underlying saturated layer. In the unsaturated zone, lateral flow is completely neglected. The model is applied to the Walnut Gulch watershed in southeastern Arizona.

In the following, we briefly describe the data assimilation techniques that have been applied by Houser et al. [1998]. In all cases, the assimilated data are soil moisture values that have been obtained through an off-line inversion of remotely sensed microwave observations.

### Direct Insertion and Statistical Corrections

The simplest data assimilation method used by Houser et al. [1998] is Direct Insertion. Here, all observations are assumed perfect. In the update step, the model prediction is simply replaced with the measurement for all observed components of the state vector. No other assimilation is performed, nor are the observations pre-interpolated. This results in very abrupt discontinuities of the soil moisture field. Any advection of information is accomplished through the subsequent prediction steps. The propagation of the error covariance is entirely neglected. The computational savings are enormous, and the computational effort is almost the same as for a pure simulation run without using the data at all. However, the scheme is wholly suboptimal.

Houser et al. [1998] also employ a technique they call Statistical Corrections. In this approach, the mean and the variance of the observations are computed. Then the components of the predicted state vector are rescaled in order to match the statistics of the observations.

### Nudging

Next, Houser et al. [1998] test two forms of nudging. The idea behind nudging is to add an artificial forcing term to the model equation such as to drive the state continuously towards the observations. This artificial forcing is not obtained from covariance propagation and the filter is thus suboptimal. Two nudging techniques are used. In "Nudging Towards a Gridded Analysis", the observations are pre-interpolated to the model grid. This means that the scheme can only be applied *within* a region of observations. After the pre-interpolation, "observations" are available for every state. In the second nudging technique, termed "Nudging to Individual Observations," no interpolation is carried out. In both nudging techniques, the artificial forcing is entirely empirical. Houser et al. [1998] implemented all possible combinations of the two nudging techniques together with the method of Statistical Corrections. The authors were most satisfied with "Nudging to Individual Observations" both inside and outside the region of observations.

### Optimal Interpolation

Optimal Interpolation (or Statistical Interpolation) is a special case of the Kalman Filter [Daley, 1991]. In Optimal Interpolation, the error covariance propagation equations are omitted. Houser et al. [1998] approximate the predicted error covariance (or background error covariance) for the computation of the gain in the following way. First, the difference between observed and TOPLATS-simulated values is computed. Second, the covariance of this difference is calculated. Third, an analytical covariance model is fitted to the data. Houser et al. [1998] use a model for a climatological background derived by Thiebaux [1976]. Although Optimal Interpolation is always less computationally demanding than a full Kalman Filter, the computational effort for the application of Houser et al. [1998] was still by far too large. The authors therefore reduced the number of measurements in an ad hoc fashion. They followed two approaches. In the first approach, a random subset of observation is chosen for each grid point, and together with the observation closest to the given grid point, only these measurements are assimilated. All other measurements are discarded. In the second approach, "super-observations" are obtained by averaging the available measurements over a coarser spatial grid and thus reducing their number.

### Discussion

Except for Optimal Interpolation, all data assimilation techniques used by Houser et al. [1998] are empirical. By empirical we mean that the estimation equations are neither derived from the optimization of an objective criterion, nor are they consistent simplifications such that the approximation error could be quantified in some way. An example for the former would be a full Kalman Filter, an example for the latter would be an eigenvalue decomposition with only the largest eigenvalues retained in the estimation algorithm. Optimal Interpolation in its full form could be called semi-empirical. It is an optimal interpolator in space at each isolated time step, provided the correct background error covariance is known. The suboptimality with respect to the Kalman Filter is due to the fact that the error covariance is not propagated. Therefore Optimal Interpolation is only optimal if the Kalman Filter happens to operate in steady-state and if the background error covariance of the Optimal Interpolation algorithm is equal to the steady-state Kalman Filter error covariance prediction. However, in the way Optimal Interpolation is implemented by Houser

et al. [1998], namely through the ad hoc reduction of the number of measurements, the estimation algorithm is certainly empirical.

The degree of optimality in the data assimilation approaches implemented by Houser et al. [1998] is not clear and strongly depends on the particular application. In the case of nudging, the degree of optimality also depends on the choice of the many parameters. In the Optimal Interpolation scheme as implemented by Houser et al. [1998], the ad hoc reduction of the number of measurements defies a strict assessment of the approximation.

### 1.3.3 Soil Moisture Estimation from Atmospheric Observations

#### The Study by Mahfouf [1991]

Mahfouf [1991] introduced a technique to estimate soil moisture from the assimilation of low-level atmospheric parameters such as relative humidity and air temperature. The main purpose of the investigation is to come up with a better initialization of soil moisture in atmospheric models and consequently with better short- and medium-range weather forecasts.

The basic idea is that errors in the predicted meteorologic quantities may be related to errors in soil moisture. This is hypothesized to be particularly true for low-level air temperature and humidity, which are linked to surface and deep soil moisture by the sensible and latent heat fluxes. In other words, the assimilated data are not measurements of soil moisture, but rather observations of low-level atmospheric parameters, namely screen level temperature and relative humidity.

Two assimilation algorithms are developed. The first approach is a variational algorithm based on the minimization of an objective function. The objective function consists of the weighted sum of squared differences between the observed and the estimated low-level air temperature and relative humidity. No regularizing prior term for the uncertain initial soil moisture is included in the performance index, and the model is imposed as a strong constraint. The objective function is minimized with a standard Gauss-Newton method. This optimal scheme takes the modeled nonlinear relationship between the low-level parameters and soil moisture fully into account and is therefore computationally expensive.

The second approach is a statistical algorithm based on linear regression. This sequential scheme is suboptimal but computationally efficient and compatible with current operational assimilation. To get soil moisture estimates, the errors in the low-level atmospheric parameters are linearly related to the soil moisture errors in two layers, a very shallow top layer and a deeper reservoir. The linear relationship is described with a set of so-called nudging coefficients, which are in turn determined by an Optimal Interpolation analysis [Daley, 1991]. This implies that the nudging coefficients depend on the observation and forecast error statistics of the low-level atmospheric parameters as well as on the forecast error statistics of soil moisture. Mahfouf [1991] infers the necessary forecast error covariances with a Monte Carlo technique.

Mahfouf [1991] successfully assimilates field data into a one-dimensional version of a mesoscale numerical weather prediction model for three 48-hour periods. The results indicate that under certain atmospheric conditions it is indeed possible to estimate soil moisture from low-level atmospheric variables. The author also states that the variational scheme is preferable to the sequential scheme, but that the latter appears good enough to be given serious consideration for implementation in current operational numerical weather prediction systems.

### The Study by Bouttier et al. [1993a]

Bouttier et al. [1993a] further investigate the sequential soil moisture estimation technique introduced by Mahfouf [1991]. As an alternative to the sequential approach of Mahfouf [1991], Bouttier et al. [1993a] provide an approximate analytic formulation of the nudging coefficients. This avoids the computationally expensive Monte Carlo simulations needed to come up with reasonable forecast error covariances. In the analytic formulation, the nudging coefficients are given as explicit expressions of the surface characteristics, most importantly of the vegetation parameters.

Moreover, Bouttier et al. [1993a] assess the sensitivity of the estimates to the vegetation, the soil texture, and the wind conditions. Vegetation is identified as the most crucial parameter. In a companion paper, Bouttier et al. [1993b] implement their simplified analytical filter in a mesoscale model. The study area covers $400 \times 400\, km^2$ in the southwest of France. Disturbing the initial soil moisture conditions from a reference simulation, Bouttier et al. [1993b] show that their nudging technique is able to restore soil moisture to the reference value within 48 hours.

### The Study by Hu et al. [1999]

Hu et al. [1999] apply the sequential nudging technique of Mahfouf [1991] to 16 sites selected to sample a range of climates and land covers across the globe. Their goal is to derive a single set of nudging coefficients, which is then applied to a test site. The authors report computational instability when assimilating air temperature and relative humidity directly. Apparently, the strong correlation between air temperature and relative humidity makes the coefficient matrix of the equation for the nudging coefficients close to singular. The problem is overcome with a Principal Component Analysis.

Finally, Hu et al. [1999] conclude that numerical weather prediction can be improved by the nudging technique, but that nudging is unable to determine accurately the soil moisture within soil layers that are accessible to the atmosphere. Moreover, the improvement in the weather forecast holds only if the meteorologic model simulates precipitation poorly. If, on the other hand, precipitation is simulated well but surface radiation is modeled poorly, the nudging technique could erroneously adjust soil moisture.

### The Studies by Callies et al. [1998] and by Bouyssel et al. [1999]

Following up on the variational approach of Mahfouf [1991], Callies et al. [1998] and Bouyssel et al. [1999] further investigate the feasibility of off-line soil moisture estimation for operational weather forecasting. The authors use one-dimensional versions of the soil and atmospheric boundary layer models of the German and the French weather services, respectively. Apart from minor differences, both soil models consist of a two-layer force-restore approximation for the soil temperature and of a two-layer soil moisture model. Callies et al. [1998] and Bouyssel et al. [1999] assimilate near-surface atmospheric measurements of air temperature and relative humidity to estimate the initial soil moisture conditions for a one-day and a two-day assimilation window, respectively. Whereas Bouyssel et al. [1999] choose ideal anticyclonic conditions with a clear sky and low advection, Callies et al. [1998] specifically choose a day reflecting non-perfect conditions.

While Callies et al. [1998] use the original strong constraint variational approach of Mahfouf [1991], Bouyssel et al. [1999] add a regularizing prior misfit term for the uncertain

initial condition to the objective function. Neither study takes model errors into account. Note that the variational approach of Callies et al. [1998] is approximate because some of the fluxes in the atmospheric model are kept constant while the initial soil moisture is changed.

Both studies closely investigate the shape of the cost function and find that the initial soil moisture in the two layers cannot be estimated unambiguously. Lacking a regularizing term in the objective function, Callies et al. [1998] resort to fixing the soil moisture of the lower layer. Bouyssel et al. [1999] relate the ambiguity to the relatively short assimilation window and expect the problem to be solved by using longer assimilation periods.

In conclusion, the authors confirm the usefulness of the assimilation of soil moisture for short-term weather forecasts even under non-perfect conditions. Callies et al. [1998] state that *"a significant part of the information carried by the data cannot be explained by the need for higher energy input at the surface (stronger radiation) but must be attributed to an incorrectly modeled Bowen ratio probably resulting from a bad soil moisture specification"*. However, their retrieved soil moisture seems to be too low for the season, which reveals the estimated soil moisture as a tuning parameter for improving numerical weather prediction rather than a physical quantity.

### The Studies by Rhodin et al. [1999]

In a recent study, Rhodin et al. [1999] apply the technique of Callies et al. [1998] to a regional weather forecast model. For the assimilation of soil moisture, all horizontal correlations are neglected and the three-dimensional problem is treated as a collection of completely entirely independent single-column assimilation problems. This offers huge computational savings, but large-scale structures in the errors of the soil moisture fields, arising for example from geologic processes, have to be neglected.

### Discussion

Assimilating low-level atmospheric observations for the estimation of soil moisture offers great opportunities to improve short- and medium-range weather forecasts. Most importantly, the data are readily available within operational data assimilation system used for numerical weather prediction. However, the soil moisture values estimated in this way lack physical meaning. Indeed, Callies et al. [1998] deduce that *"soil moisture retrieval by the present method should be considered as a parametric approach"*.

Moreover, by dividing the domain into completely independent columns for the sake of soil moisture assimilation, the approach followed by Rhodin et al. [1999] does not allow for any explicit horizontal correlation of the initial condition of soil moisture. In their approach, soil moisture is correlated horizontally only through the spatial correlation of the low-level atmospheric parameters. This is clearly undesirable from a hydrologist's point of view and makes soil moisture even more of a tuning parameter.

Finally note that the indirect estimation of soil moisture from low-level atmospheric parameters is unsuitable for cloudy conditions or situations with predominantly large-scale advection. In these situations, air temperature and relative humidity are not related to local soil moisture.

## 1.4 Challenges in Hydrologic Data Assimilation

More research is necessary in order to set up an operational soil moisture data assimilation package. In particular, investigations should be undertaken along the following lines.

- The data assimilation approach should be truly optimal and four-dimensional, that is the large-scale correlation of land surface states should be considered explicitly. Introducing such structure only indirectly through correlations in low-level atmospheric parameters as done by Rhodin et al. [1999] is clearly inadequate from a hydrologist's point of view. Trying to estimate forecast state error covariances for Optimal Interpolation approaches is similarly unsatisfactory. In other words, the data assimilation algorithm should be optimal and provide for some form of large-scale error covariance propagation.

- In order to deal with the complexity of real world applications, a land-surface model suitable for hydrologic data assimilation has to be developed. Such a model must capture the key physical processes, but at the same time be highly computationally efficient.

- It is desirable to resolve the soil moisture profile in the field studies to a greater extent. A finer discretization in the vertical allows for a much better description of the nonlinear behavior than two-layer or three-layer models can provide.

- Off-line inversion of the remotely-sensed radiances into land surface states such as soil moisture can lead to inconsistencies in the model physics. It is therefore preferable to assimilate the remote sensing data directly into the hydrologic model.

- The temperature profile of the soil strongly affects the remotely sensed brightness temperature. When brightness temperatures are assimilated, it is necessary to model soil temperature along with soil moisture. The dynamics of the temperature profile can easily be described with the heat equation or approximations thereof, and observations of the surface temperature are readily available. Therefore, the land surface model should include soil temperature, and the estimation algorithm should provide for the assimilation of soil surface temperature measurements.

- Another problem that needs to be addressed is the mismatch between the scale of the hydrologic model and the scale of the observations. In particular, ground-based observations of soil moisture are generally point measurements, whereas remotely sensed observations are satellite footprints with a resolution of typically tens of kilometers. The same disparity in scales is true for precipitation measurements, one of the most important inputs for a soil moisture model. Moreover, inputs to hydrologic models are often available at scales finer than the scales of satellite remote sensing data. This creates a need for optimal downscaling methodologies. Hence, a consistent multiscale framework needs to be developed in order to accommodate measurements at different scales.

- Last but not least it is desirable to include more detailed models of the vegetation, as vegetation is probably the most important factor for the calculation of the latent and sensible heat fluxes at the land surface.

We address all of the above topics in our research. However, due to the highly nonlinear structure of the physical processes at the land surface, and given the high dimensionality of real world applications, a compromise will have to be made between a desirable physical foundation of the model and crude simplifications in order to achieve computational feasibility.

# Chapter 2

# Data Assimilation

The goal of this Chapter is to develop a comprehensive set of estimation equations for assimilating remote sensing data into a model of land surface dynamics. The hydrologic model itself is described in detail in Chapters 3 and 4.

In Section 2.1 we first define the state and measurement equations in a general form. From this general formulation, we then derive the nonlinear estimation equations (Euler-Lagrange equations) using a variational technique. Sections 2.2 and 2.3 provide an overview of the representer algorithm which is used to solve for the estimates.

The focus of Section 2.4 is to derive the posterior covariances of the state and the measurements within the representer approach. In Section 2.5 we briefly discuss the nature of the representer approach as a data space search engine and the opportunities for data compression and for the a posteriori assessment of the observing system.

## 2.1 General Formulation of the Estimation Problem

### 2.1.1 State and Measurement Equations

We formulate the state in vector form, assuming that the model equations have been discretized in space but not in time. In other words, the components of the state vector correspond to state variables at discrete spatial nodes but depend continuously on time. There are two parts to the state vector: the vector $X(t)$ of length $N_X$ which obeys a set of implicit algebraic equations, and the vector $Y(t)$ of length $N_Y$ which obeys a set of ordinary differential equations. In a land surface model, the storage terms of some canopy states, for example the canopy temperature, are typically neglected. Therefore such canopy states obey diagnostic (algebraic) equations. Soil moisture and temperature, on the other hand, have significant memory through storage of water and energy in the soil. Such prognostic variables are subject to ordinary differential equations. For details on the exact definition of $X$ and $Y$ within the land surface model see Section 4.2. Together, $X$ and $Y$ obey the *state equation*

$$\begin{aligned} 0 &= \phi(X, Y; \alpha) + D_\nu P_\nu \nu \\ \frac{\partial Y}{\partial t} &= \varphi(X, Y; \alpha) + D_\omega(Y) P_\omega \omega \end{aligned} \tag{2.1}$$

The operators $\phi(X, Y; \alpha)$ and $\varphi(X, Y; \alpha)$ depend nonlinearly on the state and on the uncertain parameter vector $\alpha$ of length $N_\alpha$. In the soil moisture application, these parameters

could for example be the saturated hydraulic conductivities. The state equation also includes the process noise terms $\nu(t)$ and $\omega(t)$, which are time-dependent vectors of length $N_\nu$ and $N_\omega$, respectively.

The diagonal matrices $D_\nu$ and $D_\omega$ of size $N_X$ and $N_Y$, respectively, account for scaling between the states and the process noise. To facilitate the estimation, it is important that all variables be scaled. On the other hand, we would like to directly relate the process noise to a physical flux. As a result, it is necessary to account for scaling between the states and the process noise. We also include a formal dependence $D_\omega(Y)$ of the scaling matrix on the state, which we will need for the soil moisture application. From the spatial discretization, we usually get a matrix multiplying the time derivative. We include this factor in the operator $\varphi$ and the matrix $D_\omega$.

Finally, the process noise does not necessarily affect all components of the state vector. The $N_X \times N_\nu$ matrix $P_\nu$ and the $N_Y \times N_\omega$ matrix $P_\omega$ serve the purpose of projecting the process noise onto just those components of the state vector that we consider subject to model errors. In this sense, the choice of $P_\nu$ and $P_\omega$ partially reflect our assumptions on the model's shortcomings. Both $P_\nu$ and $P_\omega$ contain ones and zeros only.

The state equation (2.1) is subject to the *initial condition*

$$Y|_{t=0} = Y_0(\beta) \tag{2.1a}$$

which is parameterized by the uncertain vector $\beta$ of length $N_\beta$. Section 4.3 explains why we need such a nonlinear parameterization. Without loss of generality, we have set the initial time to zero. The final time of the estimation interval is denoted with $t_f$.

Measurements, and in particular remote sensing data, are not necessarily direct observations of the state. We therefore introduce a nonlinear *measurement equation.*

$$Z = M[X, Y] + v \tag{2.2}$$

All $N_Z$ individual measurements are collected into the data vector $Z$ and are corrupted by the measurement noise $v$. The measurement operator $M[\cdot]$ is a vector-valued, nonlinear functional operating on vector-valued functions (e.g. the state vector). For later use, we rewrite the measurement operator according to

$$M[X, Y] = \int_0^{t_f} [\delta] f(X, Y) dt \tag{2.3}$$

where

$$[\delta] \equiv \operatorname{diag}\{\delta(t - t_1), \delta(t - t_2), \ldots, \delta(t - t_{N_Z})\} \tag{2.3a}$$

Both $M[X, Y]$ and $f(X, Y)$ are vectors of length $N_Z$. Whereas $M[X, Y]$ is a (vector-valued) functional, $f(X, Y) \equiv f(X(t), Y(t))$ is a (vector-valued) function of the state vector evaluated at time $t$. For all practical purposes, both $M[\cdot]$ and $f(\cdot)$ can be thought of as the measurement operator. The scalar Dirac delta function for time $t_k$ is denoted with $\delta(t - t_k)$, and $\operatorname{diag}\{\cdot\}$ stands for a diagonal matrix with the argument of $\operatorname{diag}\{\cdot\}$ on the diagonal and zeros elsewhere. The formulation of (2.3) implies that the measurement $Z_k$ has been taken at time $t_k$. Note that the $t_k$ are not necessarily all different. Two measurements can be taken at the same time, for example as different pixels of the same remote sensing image.

But note that (2.3) constrains the set of possible measurements to point measurements in time, i.e. we cannot express measurements which integrate the state over time. For the soil moisture application, we will not need such measurements.

For the statistics of the process noise $\nu$ and $\omega$, the measurement noise $v$, and the parameters $\alpha$ and $\beta$ we assume

$$\begin{aligned} \overline{\nu(t)} &\equiv 0 & \overline{\nu(t_1)\nu(t_2)^T} &= C_\nu(t_1,t_2) \\ \overline{\omega(t)} &\equiv 0 & \overline{\omega(t_1)\omega(t_2)^T} &= C_\omega(t_1,t_2) \\ \overline{v} &= 0 & \overline{vv^T} &= C_v \\ \overline{(\alpha-\overline{\alpha})} &= 0 & \overline{(\alpha-\overline{\alpha})(\alpha-\overline{\alpha})^T} &= C_\alpha \\ \overline{(\beta-\overline{\beta})} &= 0 & \overline{(\beta-\overline{\beta})(\beta-\overline{\beta})^T} &= C_\beta \end{aligned} \tag{2.4}$$

where the overbar is the expectation operator. We generally denote the covariance of a random variable $\xi$ with $C_\xi$. The superscript $T$ denotes the matrix transpose. The prior values $\overline{\alpha}$ and $\overline{\beta}$ are our best guesses for the parameters prior to using the data $Z$. Moreover, we assume that all cross-covariances between $\nu$, $\omega$, $v$, $\alpha$, and $\beta$ vanish.

$$\begin{aligned} \overline{\omega(t)\nu(t')^T} &\equiv 0 & \overline{v\nu(t)^T} &\equiv 0 & \overline{\alpha\nu(t)^T} &\equiv 0 & \overline{\beta\nu(t)^T} &\equiv 0 \\ \overline{v\omega(t)^T} &\equiv 0 & \overline{\alpha\omega(t)^T} &\equiv 0 & \overline{\beta\omega(t)^T} &\equiv 0 \\ \overline{v\alpha^T} &= 0 & \overline{v\beta^T} &= 0 & \overline{(\alpha-\overline{\alpha})(\beta-\overline{\beta})^T} &= 0 \end{aligned} \tag{2.4a}$$

It is important to keep in mind that we only consider second-order statistics and implicitly assume distributions to be Gaussian or at least close to Gaussian. The computation of the reduced objective (Section 2.3.6) and of the posterior covariances (Section 2.4) will allow for tests of this crucial assumption.

Finally note that the measurement error covariance $C_v$ can usually be made block-diagonal for remote sensing applications. In the soil moisture application, the snapshots of brightness temperature at different observation times are usually uncorrelated, although the measurement errors in each image are typically spatially correlated. Arranging the measurements within $Z$ appropriately allows us to express the covariance $C_v$ as a block-diagonal matrix, with the blocks containing the spatial correlation matrix of the measurement error.

### 2.1.2 Objective Function

Without loss of generality, we assume that there are no measurements exactly at the initial or the final time, that is we assume $t_k \in (0, t_f)$ and use $[0, t_f]$ as the time window for the inversion. The performance index for the estimation problem is the objective function

$$\begin{aligned} \tilde{J} = &(Z - M[X,Y])^T C_v^{-1} (Z - M[X,Y]) \\ &+ (\alpha-\overline{\alpha})^T C_\alpha^{-1} (\alpha-\overline{\alpha}) + (\beta-\overline{\beta})^T C_\beta^{-1} (\beta-\overline{\beta}) \\ &+ \int_0^{t_f}\int_0^{t_f} \nu(t')^T C_\nu^{-1}(t',t'')\nu(t'')dt'dt'' + \int_0^{t_f}\int_0^{t_f} \omega(t')^T C_\omega^{-1}(t',t'')\omega(t'')dt'dt'' \end{aligned} \tag{2.5}$$

which will be minimized with respect to $\nu(t)$, $\omega(t)$, $\alpha$, and $\beta$ subject to the state equation (2.1) as a constraint. The first term accounts for the misfit between the data and the

model predictions of the measurements. The weights are given by the inverse covariance matrices of the measurement errors. The second and third terms penalize the deviation of the parameter vectors from their prior values $\overline{\alpha}$ and $\overline{\beta}$. The weights are given by the inverse covariance matrices $C_\alpha^{-1}$ and $C_\beta^{-1}$ of the respective parameters.

Finally, the last two terms accounts for the model error. Note that the weight here is an inverse covariance function, which is defined by the operator identity

$$\int_0^{t_f} C_\xi(t,t')C_\xi^{-1}(t',t'')dt' = \delta(t-t'')I_{N_\xi} \tag{2.6}$$

where $I_{N_\xi}$ is the $N_\xi \times N_\xi$ identity matrix and $\delta(\cdot)$ is again the Dirac delta function.

### 2.1.3 Euler-Lagrange Equations

In order to derive the Euler-Lagrange equations, we adjoin the state equation (2.1) and the initial condition (2.1$a$) as constraints to the objective function $\tilde{J}$ (2.5). This step introduces the adjoint variables $\mu(t)$, a vector of length $N_X$, as well as $\lambda(t)$ and $\lambda_0$, both vectors of length $N_Y$.

$$\begin{aligned} J = \; & \tilde{J} \\ & - 2\int_0^{t_f} \mu^T \left(\phi(X,Y;\alpha) + D_\nu P_\nu \nu\right) dt \\ & + 2\int_0^{t_f} \lambda^T \left(\frac{\partial Y}{\partial t} - \varphi(X,Y;\alpha) - D_\omega(Y)P_\omega\omega\right) dt + 2\lambda_0^T \left(Y|_{t=0} - Y_0(\beta)\right) \end{aligned} \tag{2.7}$$

The minimization of the objective function is a straightforward application of the calculus of variations [Courant and Hilbert, 1953; Lanczos, 1966]. Details of the derivation are outlined in Appendix A.1. We eventually get a set of equations for the estimates $\hat{X}$, $\hat{Y}$, $\hat{\nu}$, $\hat{\omega}$, $\hat{\alpha}$, and $\hat{\beta}$ of the state, the process noise, and the parameters, respectively. We call this set the *Euler-Lagrange equations.*

Forward Equation

$$\begin{aligned} 0 &= \phi(\hat{X},\hat{Y};\hat{\alpha}) + D_\nu P_\nu \hat{\nu} \\ \frac{\partial \hat{Y}}{\partial t} &= \varphi(\hat{X},\hat{Y};\hat{\alpha}) + D_\omega(\hat{Y})P_\omega\hat{\omega} \end{aligned} \tag{2.8}$$

$$\hat{Y}|_{t=0} = Y_0(\hat{\beta}) \tag{2.8a}$$

Backward Equation

$$
\begin{aligned}
0 &= \widehat{\frac{\partial \phi}{\partial X}}^T \mu + \widehat{\frac{\partial \varphi}{\partial X}}^T \lambda + \widehat{\frac{\partial f}{\partial X}}^T [\delta] C_v^{-1} \left( Z - M[\hat{X}, \hat{Y}] \right) \\
-\frac{\partial \lambda}{\partial t} &= \widehat{\frac{\partial \phi}{\partial Y}}^T \mu + \widehat{\frac{\partial \varphi}{\partial Y}}^T \lambda + \frac{\partial [D_\omega(\hat{Y}) P_\omega \hat{\omega}]}{\partial Y}^T \lambda \\
&\quad + \widehat{\frac{\partial f}{\partial Y}}^T [\delta] C_v^{-1} \left( Z - M[\hat{X}, \hat{Y}] \right)
\end{aligned}
\tag{2.9}
$$

$$
\boxed{\lambda|_{t=t_f} = 0} \tag{2.9a}
$$

Parameter Update

$$
\begin{aligned}
\hat{\alpha} &= \overline{\alpha} + C_\alpha \int_0^{t_f} \left( \widehat{\frac{\partial \phi}{\partial \alpha}}^T \mu + \widehat{\frac{\partial \varphi}{\partial \alpha}}^T \lambda \right) dt \\
\hat{\beta} &= \overline{\beta} + C_\beta \widehat{\frac{\partial Y_0}{\partial \beta}}^T \lambda|_{t=0}
\end{aligned}
\tag{2.10}
$$

Noise Update

$$
\begin{aligned}
\hat{\nu} &= \int_0^{t_f} C_\nu(t,t') P_\nu^T D_\nu^T \mu(t') dt' \\
\hat{\omega} &= \int_0^{t_f} C_\omega(t,t') P_\omega^T D_\omega(\hat{Y}(t'))^T \lambda(t') dt'
\end{aligned}
\tag{2.11}
$$

Note that all partial derivatives of $\phi$, $\varphi$, $f$, and $Y_0$ are evaluated at the estimates $\hat{X}$, $\hat{Y}$, $\hat{\alpha}$, and $\hat{\beta}$, for which we use the shortcut notation

$$
\widehat{\frac{\partial \phi}{\partial X}} \equiv \left. \frac{\partial \phi}{\partial X} \right|_{\hat{X},\hat{Y};\hat{\alpha}} \qquad \widehat{\frac{\partial \varphi}{\partial \alpha}} \equiv \left. \frac{\partial \varphi}{\partial \alpha} \right|_{\hat{X},\hat{Y};\hat{\alpha}} \qquad \widehat{\frac{\partial f}{\partial Y}} \equiv \left. \frac{\partial f}{\partial Y} \right|_{\hat{X},\hat{Y}} \quad \text{etc.} \tag{2.12}
$$

Equation (2.8) resembles the state equation and is solved forward in time subject to the initial condition (2.8*a*). This so-called *forward equation* uses the estimates of the parameters and is forced with the estimate of the process noise. The *parameter update* (2.10) and the *process noise update* (2.11) are in turn determined by the solution of the adjoint equation (2.9), which is subject to the terminal condition (2.9*a*). Since the adjoint equation is solved backward in time, we call it the *backward equation.* Note that the backward equation is linear in the adjoint variables $\mu$ and $\lambda$. At measurement times $t_k$, the backward equation is forced with the misfit between the data and the estimates of the observations. But to compute the latter, we need the state estimates from the forward equation (2.8).

Obviously, the Euler-Lagrange equations present a highly coupled, nonlinear set of equations. Notice that even for linear state and measurement equations the Euler-Lagrange equations are coupled through the data misfit term in the backward equation. This coupling leaves us with a two-point boundary value problem. Sections 2.2 and 2.3 provide details on how the Euler-Lagrange equations are solved.

## 2.2 Solving the Nonlinear Euler-Lagrange Equations: Iterated Representers

A simple solution strategy for the Euler-Lagrange equations is to iterate simultaneously on all the nonlinear terms as well as on the data misfit coupling term in the backward equation. Breaking the coupling in this way leaves us with two initial value problems at each iteration, which are easily integrated. However, this approach failed to converge for the soil moisture problem. Similar difficulties for oceanographic applications have been reported by Bennett [1992].

Alternatively, we can solve the nonlinear estimation problem as a sequence of linear estimation problems [Bennett, 1992]. Rather than linearizing all the coupling terms, we can linearize the Euler-Lagrange equations with a standard Taylor series expansion around a given trajectory and at the same time keep the coupling through the data misfit term in the backward equation. We are then left with having to solve a linear two-point boundary value problem in every iteration. Each of these two-point boundary value problems corresponds to solving the estimation problem for a set of linearized state and measurement equations. Equivalently, if we consistently linearize the state and measurement equations with a standard Taylor series expansion around the given trajectory, and if we then derive the estimation equations for these linearized state and measurement equations, we find that they are exactly the linearized Euler-Lagrange equations. Below we outline this approach in detail.

An elegant technique for solving a linear Euler-Lagrange system is the representer approach [Bennett, 1992; Bennett, 1999]. In this approach the Euler-Lagrange equations are decoupled by introducing a series expansion solution of the estimate around so-called representer fields. The representers are the prior cross-covariances of the measurement predictions with the state, which implies that there are as many representer fields as there are measurements. Fortunately, we do not have to compute every representer field in order to get the best estimate. It is sufficient to compute a sequence of suitable linear combinations of the representer fields. This approach is called the indirect representer method, and we describe the technique in detail in Section 2.3. We use the indirect representer method in combination with the iteration on the nonlinearity to obtain the estimates $\hat{X}$, $\hat{Y}$, $\hat{\alpha}$, and $\hat{\beta}$. After the estimates have been computed, we can use the representer method to compute some posterior covariances (Section 2.4). A summary of the algorithm can be found in Figure 2.1 and in Section 2.3.5.

### 2.2.1 Tangent-linear Model

In meteorology and oceanography the state equation linearized around a trajectory is called the tangent-linear model. Our approach for solving the nonlinear Euler-Lagrange equations (2.8)–(2.11) is based on successive tangent-linearizations of the state and measurement equations. In this Section, we outline the linearization procedure.

Let $X^\eta$, $Y^\eta$, $\omega^\eta$, $\alpha^\eta$, and $\beta^\eta$ denote the best estimates of the previous iteration level $\eta$. At the current iteration level $\eta + 1$, we linearize the state and measurement equations around the trajectories $X^\eta(t)$, $Y^\eta(t)$, and $\omega^\eta(t)$ as well as around the previous parameter

estimates $\alpha^\eta$ and $\beta^\eta$. Simplifying the linearization in $\omega^\eta$ according to

$$\begin{aligned} D_\omega(Y)P_\omega\omega &\approx D_\omega(Y^\eta)P_\omega\omega^\eta + \left.\frac{\partial[D_\omega(Y)P_\omega\omega]}{\partial Y}\right|_\eta (\omega-\omega^\eta) + \left.\frac{\partial[D_\omega(Y)P_\omega\omega]}{\partial Y}\right|_\eta (Y-Y^\eta) \\ &\equiv D_\omega(Y^\eta)P_\omega\omega + \left.\frac{\partial[D_\omega(Y)P_\omega\omega]}{\partial Y}\right|_\eta (Y-Y^\eta) \end{aligned}$$

and neglecting terms of second and higher order, we obtain the *tangent-linear state equation*

$$\begin{aligned} 0 =& \phi(X^\eta, Y^\eta; \alpha^\eta) + \left.\frac{\partial\phi}{\partial X}\right|_\eta (X-X^\eta) + \left.\frac{\partial\phi}{\partial Y}\right|_\eta (Y-Y^\eta) + \left.\frac{\partial\phi}{\partial\alpha}\right|_\eta (\alpha-\alpha^\eta) \\ &+ D_\nu P_\nu \nu \\ \frac{\partial Y}{\partial t} =& \varphi(X^\eta, Y^\eta; \alpha^\eta) + \left.\frac{\partial\varphi}{\partial X}\right|_\eta (X-X^\eta) + \left.\frac{\partial\varphi}{\partial Y}\right|_\eta (Y-Y^\eta) + \left.\frac{\partial\varphi}{\partial\alpha}\right|_\eta (\alpha-\alpha^\eta) \\ &+ D_\omega(Y^\eta)P_\omega\omega + \left.\frac{\partial[D_\omega(Y)P_\omega\omega]}{\partial Y}\right|_\eta (Y-Y^\eta) \end{aligned} \tag{2.13}$$

$$Y|_{t=0} = Y_0(\beta^\eta) + \left.\frac{\partial Y_0}{\partial\beta}\right|_\eta (\beta-\beta^\eta) \tag{2.13a}$$

The *tangent-linear measurement equation* reads

$$\begin{aligned} Z &= M[X^\eta, Y^\eta] + L[X-X^\eta, Y-Y^\eta] + v \\ L_k[X,Y] &\equiv \int_0^{t_f} \left\{ \left.\frac{\partial f_k}{\partial X}\right|_\eta X(t) + \left.\frac{\partial f_k}{\partial Y}\right|_\eta Y(t) \right\} \delta(t-t_k)dt \\ &\equiv \left.\frac{\partial f_k}{\partial X}\right|_{\eta,t_k} X(t_k) + \left.\frac{\partial f_k}{\partial Y}\right|_{\eta,t_k} Y(t_k) \end{aligned} \tag{2.14}$$

We used (2.3) to get this explicit expression for the linear (vector-valued) functional $L[\cdot]$, which can be interpreted as the slope of the measurement operator. For the partial derivatives we use the obvious short-cut notation

$$\left.\frac{\partial\phi}{\partial X}\right|_\eta \equiv \left.\frac{\partial\phi}{\partial X}\right|_{X^\eta,Y^\eta;\alpha^\eta} \qquad \left.\frac{\partial\varphi}{\partial\alpha}\right|_\eta \equiv \left.\frac{\partial\varphi}{\partial\alpha}\right|_{X^\eta,Y^\eta;\alpha^\eta} \qquad \left.\frac{\partial f}{\partial Y}\right|_\eta \equiv \left.\frac{\partial f}{\partial Y}\right|_{X^\eta,Y^\eta} \qquad \text{etc.}$$

Note that the linear operator $L[\cdot]$ of the tangent-linear measurement equation changes with each iteration. In particular, $L[\cdot]$ of the current iteration level $\eta+1$ is obtained by evaluating the partial derivative of $f(\cdot)$ at the best estimate of the previous iteration level. Consequently, $L[\cdot]$ should carry a superscript $\eta+1$. However, we opt to drop this superscript because it is easy to infer from the context.

### 2.2.2 Linearized Euler-Lagrange Equations

We obtain a "linearized" version of the objective function by introducing the linearizations of (2.13) and (2.14) into (2.7). Following the general procedure outlined in Appendix A.1 for the derivation of the nonlinear Euler-Lagrange equations, we can easily derive the linearized

Euler-Lagrange equations from the "linearized" objective function.

$$
\begin{aligned}
0 &= \left.\frac{\partial \phi}{\partial X}\right|_\eta^T \mu^{\eta+1} + \left.\frac{\partial \varphi}{\partial X}\right|_\eta^T \lambda^{\eta+1} \\
&\quad + \left.\frac{\partial f}{\partial X}\right|_\eta^T [\delta] C_v^{-1} \left(Z - M[X^\eta, Y^\eta] - L[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta]\right) \\
-\frac{\partial \lambda^{\eta+1}}{\partial t} &= \left.\frac{\partial \phi}{\partial Y}\right|_\eta^T \mu^{\eta+1} + \left.\frac{\partial \varphi}{\partial Y}\right|_\eta^T \lambda^{\eta+1} + \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta^T \lambda^{\eta+1} \\
&\quad + \left.\frac{\partial f}{\partial Y}\right|_\eta^T [\delta] C_v^{-1} \left(Z - M[X^\eta, Y^\eta] - L[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta]\right)
\end{aligned}
\tag{2.15}
$$

$$
\lambda|_{t=t_f}^{\eta+1} = 0 \tag{2.15a}
$$

$$
\begin{aligned}
\alpha^{\eta+1} &= \overline{\alpha} + C_\alpha \int_0^{t_f} \left( \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta^T \mu^{\eta+1} + \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta^T \lambda^{\eta+1} \right) dt \\
\beta^{\eta+1} &= \overline{\beta} + C_\beta \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta^T \lambda|_{t=0}^{\eta+1}
\end{aligned}
\tag{2.16}
$$

$$
\begin{aligned}
\nu^{\eta+1} &= \int_0^{t_f} C_\nu(t,t') P_\nu^T D_\nu^T \mu^{\eta+1}(t') dt' \\
\omega^{\eta+1} &= \int_0^{t_f} C_\omega(t,t') P_\omega^T [D_\omega(Y^\eta(t'))]^T \lambda^{\eta+1}(t') dt'
\end{aligned}
\tag{2.17}
$$

$$
\begin{aligned}
0 =& \phi(X^\eta, Y^\eta; \alpha^\eta) + \left.\frac{\partial \phi}{\partial X}\right|_\eta \left(X^{\eta+1} - X^\eta\right) + \left.\frac{\partial \phi}{\partial Y}\right|_\eta \left(Y^{\eta+1} - Y^\eta\right) + \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta \left(\alpha^{\eta+1} - \alpha^\eta\right) \\
&+ D_\nu P_\nu \nu^{\eta+1} \\
\frac{\partial Y^{\eta+1}}{\partial t} =& \varphi(X^\eta, Y^\eta; \alpha^\eta) + \left.\frac{\partial \varphi}{\partial X}\right|_\eta \left(X^{\eta+1} - X^\eta\right) + \left.\frac{\partial \varphi}{\partial Y}\right|_\eta \left(Y^{\eta+1} - Y^\eta\right) + \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta \left(\alpha^{\eta+1} - \alpha^\eta\right) \\
&+ D_\omega(Y^\eta) P_\omega \omega^{\eta+1} + \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta \left(Y^{\eta+1} - Y^\eta\right)
\end{aligned}
\tag{2.18}
$$

$$
Y|_{t=0}^{\eta+1} = Y_0(\beta^\eta) + \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta \left(\beta^{\eta+1} - \beta^\eta\right) \tag{2.18a}
$$

Having completed iteration level $\eta$, we need to solve these linearized Euler-Lagrange equations for the new estimates $\mu^{\eta+1}$, $\lambda^{\eta+1}$, $\alpha^{\eta+1}$, $\beta^{\eta+1}$, $\nu^{\eta+1}$, $\omega^{\eta+1}$, $X^{\eta+1}$ and $Y^{\eta+1}$. The iteration can be initialized by setting the initial best estimates $X^\eta$ and $Y^\eta$ for $\eta = 0$ equal to the prior trajectories $\overline{X}$ and $\overline{Y}$ of the nonlinear model, which are the solution of the

nonlinear Euler-Lagrange equations when no data are available. In particular, $\overline{X}$ and $\overline{Y}$ are defined through

$$0 = \phi(\overline{X}, \overline{Y}; \overline{\alpha}) \qquad \frac{\partial \overline{Y}}{\partial t} = \varphi(\overline{X}, \overline{Y}; \overline{\alpha}) \qquad \overline{Y}|_{t=0} = Y_0(\overline{\beta}) \tag{2.19}$$

Note that $\overline{\mu} \equiv \overline{\lambda} \equiv 0$.

As mentioned in the introduction to this Section, the linearized Euler-Lagrange equations (2.15)–(2.18) can also be obtained by linearizing the nonlinear estimation equations (2.8)–(2.11). However, care must be taken that the linearization be carried out consistently, because there is an infinite number of ways in which the Euler-Lagrange equations can be linearized. Re-deriving the linearized Euler-Lagrange equations from the tangent-linear model ensures that the meaning of the linearized Euler-Lagrange equations as the iterated estimation equations of the original problem is preserved.

Although there is no proof or guarantee that the sequence defined above converges, experience has shown that it converges in practice for forward models that are reasonably close to linear. For a quasigeostrophic model, Bennett [1992] provides a theorem on convergence in a doubly-periodic domain. The sequence is then bounded, and so must have points of accumulation or cluster points, but not necessarily unique limits.

Solving the nonlinear problem as a series of linear estimation problems is clearly motivated by the fact that we are certain to obtain a solution to the linear estimation problem at each iterate. Before implementing the full estimation technique, it is therefore a good idea to check whether the tangent-linear model converges to the solution of the nonlinear forward model, which is equivalent to solving the estimation problem with no data. Without fulfilling this prerequisite there is little hope that the sequence will converge when data are assimilated.

## 2.3 Solving the Linear Euler-Lagrange Equations: Indirect Representers

The focus of this Section is to illustrate the representer solution of the *linearized* Euler-Lagrange equations (2.15)–(2.18). The representer approach is a very elegant way to decouple and solve the linear Euler-Lagrange equations. The approach is based on series (or representer) expansions for the estimates of the state and its adjoint variable. The series expansions linearly superimpose the measurement updates from each individual observation, implying that the representer solution is inherently the solution to a *linear* estimation problem.

In essence, the representer solution reduces the size of the space in which the objective is minimized from infinity to a finite dimension equal to the number of observations. This promises better convergence behavior than a direct minimization of the objective function with a gradient search. Fortunately, the *indirect representer method* allows us to solve for the estimates without explicitly computing all of the individual representer fields, that is we do not necessarily have to solve the basic equation (forward or backward) $2N_Z$ times (per nonlinear iteration). The indirect representer method therefore provides us with a fast way to get the estimate [Bennett et al., 1996]. If we opt for computing the individual representer fields after the nonlinear iteration has converged and the estimate has been obtained, the representer approach also yields a lot of information about the posterior error covariance (Section 2.4).

We conclude this Section with a short discussion of the hypothesis test associated with the value of the reduced objective function (Section 2.3.6) and a summary of the iterated indirect representer algorithm (Section 2.3.5).

### 2.3.1 Representer Expansion

The representer approach is essentially a series expansion solution of the estimation problem, where the number of unknown coefficients exactly matches the number of available measurements. For the new estimates of the current iteration level $\eta + 1$, we define

$$X^{\eta+1}(t) = \overline{X}^{\eta+1}(t) + \sum_{k=1}^{N_Z} b_k \Xi^k(t) \qquad Y^{\eta+1}(t) = \overline{Y}^{\eta+1}(t) + \sum_{k=1}^{N_Z} b_k \Upsilon^k(t) \tag{2.20}$$

$$\mu^{\eta+1}(t) = \overline{\mu}^{\eta+1}(t) + \sum_{k=1}^{N_Z} b_k \Omega^k(t) \qquad \lambda^{\eta+1}(t) = \overline{\lambda}^{\eta+1}(t) + \sum_{k=1}^{N_Z} b_k \Lambda^k(t) \tag{2.21}$$

The $4N_Z$ representer functions $\Xi^k(t)$, $\Upsilon^k(t)$, $\Omega^k(t)$ and $\Lambda^k(t)$ for the state and its adjoint are time-dependent vectors of length $N_X$ and $N_Y$, respectively. We write $\Upsilon_i^k$ in order to refer to the $i$-th component of the $k$-th representer function $\Upsilon^k$. The $N_Z$ scalar representer coefficients $b_k$ are constant in time. Note that the coefficients are the same for both the state and the adjoint representers.

The expansion is carried out around the prior fields $\overline{X}^{\eta+1}$, $\overline{Y}^{\eta+1}$, $\overline{\mu}^{\eta+1}$, and $\overline{\lambda}^{\eta+1}$, which are the solutions of the linearized Euler-Lagrange equations (2.15)-(2.18) when no data are available. In particular, $\overline{X}^{\eta+1}$ and $\overline{Y}^{\eta+1}$ are defined through

$$\begin{aligned} 0 =& \phi(X^\eta, Y^\eta; \alpha^\eta) + \left.\frac{\partial \phi}{\partial X}\right|_\eta \left(\overline{X}^{\eta+1} - X^\eta\right) + \left.\frac{\partial \phi}{\partial Y}\right|_\eta \left(\overline{Y}^{\eta+1} - Y^\eta\right) + \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta (\overline{\alpha} - \alpha^\eta) \\ \frac{\partial \overline{Y}^{\eta+1}}{\partial t} =& \varphi(X^\eta, Y^\eta; \alpha^\eta) + \left.\frac{\partial \varphi}{\partial X}\right|_\eta \left(\overline{X}^{\eta+1} - X^\eta\right) + \left.\frac{\partial \varphi}{\partial Y}\right|_\eta \left(\overline{Y}^{\eta+1} - Y^\eta\right) + \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta (\overline{\alpha} - \alpha^\eta) \\ &+ \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta \left(\overline{Y}^{\eta+1} - Y^\eta\right) \end{aligned} \tag{2.22}$$

$$\overline{Y}|_{t=0}^{\eta+1} = Y_0(\beta^\eta) + \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta (\overline{\beta} - \beta^\eta) \tag{2.22a}$$

and $\overline{\mu}^{\eta+1} \equiv \overline{\lambda}^{\eta+1} \equiv 0$. Recall that $X^\eta$ and $Y^\eta$ are the best estimates from the previous iteration and serve as the trajectories around which we linearize.

It is easy to see that the prior fields for iteration level $\eta + 1$ as defined above are the tangent-linear approximations of the prior trajectory (2.19) of the nonlinear problem. Obviously, one cannot simultaneously linearize around the previous estimate and the nonlinear prior trajectory (2.19). In order to achieve a dynamically consistent estimate, we must linearize around the previous best estimate, and the accuracy of the prior must be compromised. Therefore the nonlinear estimate cannot be strictly optimal. However, if the model is reasonably linear, the error so introduced is bearable.

### On the Notation

Please note again that the superscript $k$ that is used for the representer fields stands for the number of the measurement that corresponds to the representer function in question. This superscript is an integral part of the name of the representer function. Individual vector components of a given representer field are denoted with subscripts, usually $i$ or $j$. For instance, $\Xi_i^k$ denotes the $i$-th component of the representer field that corresponds to the $k$-th measurement. In contrast, the representer coefficients carry a subscript $k$. This subscript represents the number of the measurement to which the representer coefficient corresponds. It also serves as the index of the vector component when the representer coefficients $b_k$ are collected into the vector $b$, in analogy to assembling the individual measurements in the data vector $Z$.

Moreover, it is important to note that the representer fields and the representer coefficients like the estimates and the prior trajectories of the states change with every iteration. Consequently, $b_k$, $\Xi^k$, $\Upsilon^k$, $\Omega^k$ and $\Lambda^k$ should have an additional superscript $\eta+1$. Again, we opt to drop this superscript because it is easy to infer from the context. Unlike for the states, we never have to use representer fields or coefficients from different iteration levels in the same equation. Finally note that the priors of the original uncertain parameters $\overline{\alpha}$, $\overline{\beta}$, and $\overline{\nu}(t) \equiv \overline{\omega}(t) \equiv 0$ never change!

## 2.3.2 Representer Equations

By inserting the adjoint representer expansion (2.21) into the backward equation (2.15), and choosing

$$b = C_v^{-1}\left(Z - M[X^\eta, Y^\eta] - L[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta]\right) \tag{2.23}$$

we can derive an equation for the adjoint representers.

Adjoint Representer Equations

$$\begin{aligned} 0 &= \left.\frac{\partial \phi}{\partial X}\right|_\eta^T \Omega^k + \left.\frac{\partial \varphi}{\partial X}\right|_\eta^T \Lambda^k + \delta(t-t_k)\left.\frac{\partial f_k}{\partial X}\right|_\eta^T \\ -\frac{\partial \Lambda^k}{\partial t} &= \left.\frac{\partial \phi}{\partial Y}\right|_\eta^T \Omega^k + \left.\frac{\partial \varphi}{\partial Y}\right|_\eta^T \Lambda^k + \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta^T \Lambda^k + \delta(t-t_k)\left.\frac{\partial f_k}{\partial Y}\right|_\eta^T \end{aligned} \tag{2.24}$$

$$\boxed{\Lambda^k|_{t=t_f} = 0} \tag{2.24a}$$

Recall that the $k$-th measurement is taken at time $t_k$. If we compare the adjoint representer equation (2.24) with the backward equation (2.15), we can see that the two equations are very similar. Whereas the backward equation is forced with the posterior data misfit, the adjoint representer equation is forced with a single unit impulse at one measurement time and location. We can therefore solve for the adjoint representer fields without having to know the estimate in advance.

Next, we derive an equation for the state representers by inserting the parameter update (2.16), the process noise update (2.17), and the representer expansions (2.20) and (2.21) into the forward equation (2.18) and its initial condition (2.18$a$). Using the definition of

the prior fields (2.22), we find

State Representer Equations

$$
\begin{aligned}
0 &= \left.\frac{\partial \phi}{\partial X}\right|_{\eta} \Xi^k + \left.\frac{\partial \phi}{\partial Y}\right|_{\eta} \Upsilon^k + \int_0^{t_f} D_\nu P_\nu C_\nu(t,t') P_\nu^T D_\nu^T \Omega^k(t') dt' \\
&\qquad + \left.\frac{\partial \phi}{\partial \alpha}\right|_{\eta} C_\alpha \int_0^{t_f} \left( \left.\frac{\partial \phi}{\partial \alpha}\right|_{\eta}^T \Omega^k(t') + \left.\frac{\partial \varphi}{\partial \alpha}\right|_{\eta}^T \Lambda^k(t') \right) dt' \\
\frac{\partial \Upsilon^k}{\partial t} &= \left.\frac{\partial \varphi}{\partial X}\right|_{\eta} \Xi^k + \left.\frac{\partial \varphi}{\partial Y}\right|_{\eta} \Upsilon^k + \int_0^{t_f} D_\omega(Y^\eta(t)) P_\omega C_\omega(t,t') P_\omega^T D_\omega(Y^\eta(t'))^T \Lambda^k(t') dt' \\
&\qquad + \left.\frac{\partial \varphi}{\partial \alpha}\right|_{\eta} C_\alpha \int_0^{t_f} \left( \left.\frac{\partial \phi}{\partial \alpha}\right|_{\eta}^T \Omega^k(t') + \left.\frac{\partial \varphi}{\partial \alpha}\right|_{\eta}^T \Lambda^k(t') \right) dt' \\
&\qquad + \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_{\eta} \Upsilon^k
\end{aligned}
\tag{2.25}
$$

$$
\boxed{\Upsilon^k|_{t=0} = \left.\frac{\partial Y_0}{\partial \beta}\right|_{\eta} C_\beta \left.\frac{\partial Y_0}{\partial \beta}\right|_{\eta}^T \Lambda^k|_{t=0}} \tag{2.25a}
$$

As will be shown in Sections 2.4.1 and A.2.1, the state representer function for the $k$-th measurement is equal to the prior cross-covariance of the model prediction for the $k$-th measurement and the states at all times and locations. Most importantly, the representer fields encapsulate only prior information and do not depend on the data. Finally note that we can solve for the state representers as soon as we have integrated the adjoint representer fields.

### 2.3.3 Representer Coefficients and Representer Matrix

In order to derive the representer equations, we defined the representer coefficients as

$$
b = C_v^{-1} \left( Z - M[X^\eta, Y^\eta] - L[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta] \right) \tag{2.23}
$$

Using the representer expansion (2.20) for the estimate and the linearity of $L$, we can easily rewrite (2.23) as

$$
b = U^{-1} \left( Z - M[X^\eta, Y^\eta] - L[\overline{X}^{\eta+1} - X^\eta, \overline{Y}^{\eta+1} - Y^\eta] \right) \tag{2.26}
$$

where we define

$$
U \equiv C_v + R \qquad \text{and} \qquad [R]_{kl} \equiv L_k[\Xi^l, \Upsilon^l] \tag{2.27}
$$

Both $U$ and the *representer matrix* $R$ are symmetric $N_Z \times N_Z$ matrices. We can calculate the entries of the representer matrix (and hence $U$) if we compute all $N_Z$ individual representer fields $\Xi^k$ and $\Upsilon^k$, and subsequently apply the operator $L$ of (2.14).

Obviously, knowledge of the representer matrix suffices to compute the estimate. If we know the representer matrix, we can solve (2.26) for the representer coefficients. Once we have the representer coefficients, we can use the definition (2.23) and integrate the backward equation (2.15). After the backward integration, we can easily calculate the parameter update (2.16) and the noise update (2.17), and finally solve the forward equation (2.18) for the state estimates. The coupling between the forward and the backward equations of the original two-point boundary value problem (2.15)–(2.18) has now been broken by the representers. The price we pay for breaking the coupling is of course the increased computational burden.

The representer matrix encapsulates all the relevant prior information. In contrast to (2.23), Equation (2.26) expresses the representer coefficients entirely in prior terms, except for the explicit appearance of the data vector $Z$. For a linear estimation problem we could even precompute the representer matrix before we know the data, and then solve for the representer coefficients when the data become available. Solving for the optimal estimates via explicit computation of the representer matrix is called the direct representer method.

### 2.3.4 Indirect Representers

In particular for data-rich remote sensing applications, the computational burden for the calculation of the representer matrix is very heavy, because all $N_Z$ individual representer fields must be integrated. But recall that we really only need the representer coefficients to compute the estimated fields. Once we have the representer coefficients, we can integrate the backward equation (2.15) and subsequently the forward equation (2.18). Fortunately, we do not need to know the representer matrix explicitly in order to get the representer coefficients. The conjugate gradient technique for solving linear equations like (2.26) does the trick.

A conjugate gradient solver [Press et al., 1992] successively approximates the solution of a linear equation. In our case, this means constructing successive approximations of the representer coefficient vector $b$. Let us rewrite (2.26) as $Ub = \text{rhs}$, where rhs is obviously the known prior data misfit. Given an approximation $\zeta$ of the representer coefficient vector $b$, the conjugate gradient technique only uses the product $U\zeta$ to refine the approximation. If we can supply the conjugate gradient algorithm with the product $U\zeta$ for any given vector $\zeta$, we will eventually find the representer coefficients. We never need to supply the matrix $U$ itself to the conjugate gradient solver.

It turns out that we can compute the product $U\zeta$ without explicit knowledge of the matrix $U$ (or equivalently $R$) [Bennett et al., 1996]. All we need to do is solve the representer equations for a linear combination of inhomogeneities. For a given vector $\zeta$, we solve the following equations for the fields $\Xi$ and $\Upsilon$.

$$
\begin{aligned}
0 &= \left.\frac{\partial \phi}{\partial X}\right|_{\eta}^{T} \Omega + \left.\frac{\partial \varphi}{\partial X}\right|_{\eta}^{T} \Lambda + \sum_{k=1}^{N_z} \delta(t-t_k) \left.\frac{\partial f_k}{\partial X}\right|_{\eta}^{T} \zeta_k \\
-\frac{\partial \Lambda}{\partial t} &= \left.\frac{\partial \phi}{\partial Y}\right|_{\eta}^{T} \Omega + \left.\frac{\partial \varphi}{\partial Y}\right|_{\eta}^{T} \Lambda + \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_{\eta}^{T} \Lambda + \sum_{k=1}^{N_z} \delta(t-t_k) \left.\frac{\partial f_k}{\partial Y}\right|_{\eta}^{T} \zeta_k \\
\Lambda|_{t=t_f} &= 0
\end{aligned}
$$

$$0 = \left.\frac{\partial\phi}{\partial X}\right|_\eta \Xi + \left.\frac{\partial\phi}{\partial Y}\right|_\eta \Upsilon + \int_0^{t_f} D_\nu P_\nu C_\nu(t,t') P_\nu^T D_\nu^T \Omega(t') dt'$$
$$+ \left.\frac{\partial\phi}{\partial\alpha}\right|_\eta C_\alpha \int_0^{t_f} \left( \left.\frac{\partial\phi}{\partial\alpha}\right|_\eta^T \Omega(t') + \left.\frac{\partial\varphi}{\partial\alpha}\right|_\eta^T \Lambda(t') \right) dt'$$

$$\frac{\partial\Upsilon}{\partial t} = \left.\frac{\partial\varphi}{\partial X}\right|_\eta \Xi + \left.\frac{\partial\varphi}{\partial Y}\right|_\eta \Upsilon + \int_0^{t_f} D_\omega(Y^\eta(t)) P_\omega C_\omega(t,t') P_\omega^T D_\omega(Y^\eta(t'))^T \Lambda(t') dt'$$
$$+ \left.\frac{\partial\varphi}{\partial\alpha}\right|_\eta C_\alpha \int_0^{t_f} \left( \left.\frac{\partial\phi}{\partial\alpha}\right|_\eta^T \Omega(t') + \left.\frac{\partial\varphi}{\partial\alpha}\right|_\eta^T \Lambda(t') \right) dt' + \left.\frac{\partial[D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta \Upsilon$$
$$\Upsilon|_{t=0} = \left.\frac{\partial Y_0}{\partial\beta}\right|_\eta C_\beta \left.\frac{\partial Y_0}{\partial\beta}\right|_\eta^T \Lambda|_{t=0}$$

Having computed the fields $\Xi$ and $\Upsilon$, we have by way of linearity

$$R\zeta = L[\Xi, \Upsilon]$$

It is now simple to compute $R\zeta + C_v\zeta \equiv U\zeta$, which is all we need to supply to the conjugate gradient solver. Note that in the indirect representer approach we are effectively computing linear combinations of the representer fields, namely $\Xi = \sum_k \zeta_k \Xi^k$ and $\Upsilon = \sum_k \zeta_k \Upsilon^k$.

Numerical issues aside, the conjugate gradient solver for our linear system is guaranteed to converge after at most $N_Z$ iterations, that is we need to compute at most $N_Z$ linear combinations of representer fields. This means that we can only do better compared with the direct approach of assembling the representer matrix, for which we need exactly $N_Z$ (individual) representer fields. The actual number of iterations for the conjugate gradient solver depends of course on the problem at hand. For the land surface application, the savings are substantial. Detailed results on the efficiency of the indirect representer method are presented in Section 8.2.1.

Finally, preconditioning with an approximation of $U$ (and hence the representer matrix $R$) can further improve the convergence speed [Press et al., 1992]. Bennett [1999] discusses several options for a preconditioner. If only one data type is assimilated, preconditioning does not seem necessary at all.

### 2.3.5 Summary of the Iterated Indirect Representer Algorithm

Figure 2.1 shows a flowchart summarizing the iterated indirect representer algorithm. For simplicity, we omit the diagnostic state vector $X$ in the Figure and in the discussion. We start by computing the prior state trajectory $\overline{Y}$ from the nonlinear state equation with the model errors and the uncertain parameters set to their prior values (2.19). Next, we initialize the previous best estimate of the first iteration as $Y^0 \equiv \overline{Y}$ and enter the loop of iterations on the nonlinearity in the state equation. The iteration on the nonlinearity is the outer loop of the algorithm.

During each iteration, the state and measurement equations are linearized around the trajectory of the previous best estimate $Y^\eta$ (2.13)–(2.14). First, we compute the prior trajectory $\overline{Y}^{\eta+1}$ of the current iteration (2.22). From this prior trajectory, we derive the

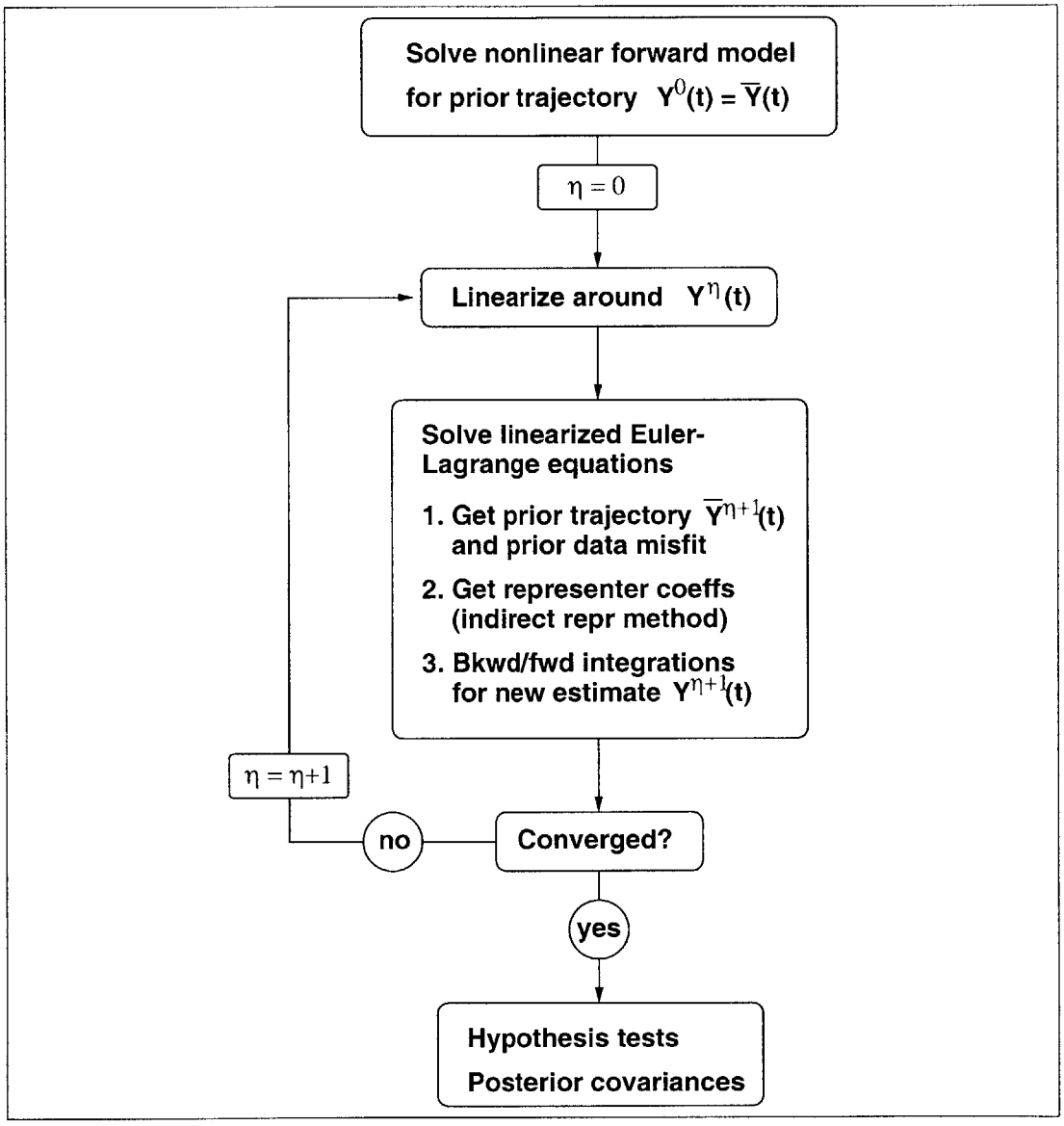


Figure 2.1: Flowchart for the iterated indirect representer method. A summary of the algorithm is presented in Section 2.3.5. The diagnostic state $X$ has been omitted for clarity.

corresponding prior data misfit. Next, we calculate the representer coefficients of the current iteration with the indirect representer method (Section 2.3.4). Note that this constitutes an inner loop in which we successively refine our best estimate of the representer coefficients with a conjugate gradient algorithm. During each of the inner iterations, we essentially compute one linear combination of representer fields. After having obtained the representer coefficients, we finally compute the new estimate $Y^{\eta+1}$ using the definition of the representer coefficients (2.23) in the linearized Euler-Lagrange equations (2.15)–(2.18), which are then decoupled.

The convergence check is carried out by comparing the previous best estimate $Y^{\eta}$ to the current best estimate $Y^{\eta+1}$. In particular, we check whether $||Y^{\eta} - Y^{\eta+1}|| < \epsilon_Y$ for a given threshold $\epsilon_Y$. Additional convergence checks are done for the uncertain parameters, for instance $||\beta^{\eta} - \beta^{\eta+1}|| < \epsilon_\beta$. After convergence has been achieved, we may compute posterior covariances (Section 2.4) and carry out hypothesis tests on the reduced objective function (Section 2.3.6) or on the posterior data residuals (Section 2.4.1). Note that the hypothesis test on the reduced objective function does not require any posterior covariance calculations and can be done at negligible cost.

### 2.3.6 Reduced Objective Function

The objective function at the minimum is given by

$$\begin{aligned}\hat{J}^{\eta+1} =& (Z - M[X^{\eta}, Y^{\eta}] - L[\overline{X}^{\eta+1} - X^{\eta}, \overline{Y}^{\eta+1} - Y^{\eta}])^T U^{-1} \cdot \\ & \cdot (Z - M[X^{\eta}, Y^{\eta}] - L[\overline{X}^{\eta+1} - X^{\eta}, \overline{Y}^{\eta+1} - Y^{\eta}]) \\ =& (Z - M[X^{\eta}, Y^{\eta}] - L[\overline{X}^{\eta+1} - X^{\eta}, \overline{Y}^{\eta+1} - Y^{\eta}])^T b\end{aligned} \tag{2.28}$$

and can be evaluated at negligible cost once the representer coefficients are known [Bennett, 1992]. If our assumptions about the model physics and the prior statistics are true, and if we invoke the central limit theorem in case we do not claim that the priors are exactly normally distributed, then the reduced objective is a chi-squared random variable with $N_Z$ degrees of freedom

$$\hat{J}^{\eta+1} \sim \chi^2_{N_Z} \qquad \overline{\chi^2_{N_Z}} = N_Z \qquad \mathrm{var}(\chi^2_{N_Z}) = 2N_Z \tag{2.29}$$

We can therefore formulate our assumptions about the model physics (2.1), the measurement process (2.2), and the prior statistics (2.4) as a null hypothesis and apply the chi-square test to see whether these assumptions are statistically consistent with the data. Should our null hypothesis fail the test, we are forced to discard the estimated fields. In this case, however, we would have learned something about land surface dynamics.

Note that the hypothesis test, much like the estimates, relies on the successive tangent-linearizations. In the case of nonlinear model dynamics, the test can only be valid approximately.

## 2.4 Posterior Covariances

After the nonlinear iteration has converged, we can opt for calculating the posterior covariances of the state and the observations. These may be obtained by computing the individual representer functions $\Xi^k$ and $\Upsilon^k$ (rather than the linear combinations $\Xi$ and $\Upsilon$

of the indirect representer approach) [Bennett, 1992]. This Section provides the expressions for the posterior covariances as functions of the representers.

Since the problem at hand is nonlinear, the linearized posterior covariances derived in this Section can at best be approximations of the true posterior covariances. For example, the prior fields $\overline{X}^{\eta+1}$ and $\overline{Y}^{\eta+1}$ depend on the linearization around the previous estimate, which in turn depends on the data for all but the first iteration. In the derivation of the posterior covariances, however, we treat the previous estimate as a fixed input around which we linearize, although strictly speaking the previous estimate is itself a random field.

It is convenient to define short-cut notation for the various deviations that occur. In general, we use $Y$ for the true state and $Y^{\eta+1}$ for the estimate. We denote with $\overline{Y}^{\eta+1}$ the prior state as defined in (2.22). For the deviations we define

$$X(t) - \overline{X}^{\eta+1}(t) \equiv X'(t) \qquad Y(t) - \overline{Y}^{\eta+1}(t) \equiv Y'(t) \tag{2.30}$$

$$X(t) - X^{\eta+1}(t) \equiv \tilde{X}(t) \qquad Y(t) - Y^{\eta+1}(t) \equiv \tilde{Y}(t) \tag{2.31}$$

and from the representer expansion we have

$$X^{\eta+1}(t) - \overline{X}^{\eta+1}(t) = \sum_{k=1}^{N_Z} b_k \Xi^k(t) \qquad Y^{\eta+1}(t) - \overline{Y}^{\eta+1}(t) = \sum_{k=1}^{N_Z} b_k \Upsilon^k(t) \tag{2.20}$$

In other words, $Y'$ is the (linearized) prior error and $\tilde{Y}$ is the (posterior) estimation error of the state.

Similarly, the (linearized) true measurement predictions are $M[X^\eta, Y^\eta] + L[X - X^\eta, Y - Y^\eta]$, the prior measurement predictions are $M[X^\eta, Y^\eta] + L[\overline{X}^{\eta+1} - X^\eta, \overline{Y}^{\eta+1} - Y^\eta]$, the estimates for the measured quantities are $M[X^\eta, Y^\eta] + L[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta]$, and the actual measurements are $Z$. For the deviations we define

$$v \equiv Z - (M[X^\eta, Y^\eta] + L[X - X^\eta, Y - Y^\eta]) \tag{2.14}$$

$$\hat{v} \equiv Z - \left(M[X^\eta, Y^\eta] + L[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta]\right) \tag{2.32}$$

$$\begin{aligned}\tilde{v} &\equiv (M[X^\eta, Y^\eta] + L[X - X^\eta, Y - Y^\eta]) - \left(M[X^\eta, Y^\eta] + L[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta]\right) \\ &\equiv L[X - X^{\eta+1}, Y - Y^{\eta+1}]\end{aligned} \tag{2.33}$$

Note that $\hat{v} = v + \tilde{v}$. In other words, $v$ is the true measurement error, $\tilde{v}$ is the (posterior) estimation error of the measurement prediction, and $\hat{v}$ is the estimate of the measurement error, or equivalently, the posterior data residual.

Like the representer fields, the deviations defined above are different at each iteration. Again, we opt to drop the superscript $\eta + 1$, as it is clear from the context.

### 2.4.1 Posterior Covariances

It is a very important fact, not only in the derivation of the posterior error covariances, that the state representers are equal to the linearized (prior) cross-covariances of the measurement predictions and the states.

$$\overline{L_k[X', Y']X'(t)} = \Xi^k(t) \qquad \overline{L_k[X', Y']Y'(t)} = \Upsilon^k(t) \tag{2.34}$$

A proof is given in Appendix A.2.1. Using (2.34) it is easy to derive the posterior covariance of the state vector.

$$\boxed{\begin{aligned}
[C_{\tilde{X}\tilde{X}}(t_1,t_2)]_{ij} &\equiv \overline{\tilde{X}_i(t_1)\tilde{X}_j(t_2)} = [C_{X'X'}(t_1,t_2)]_{ij} - \sum_k \sum_l \Xi_i^k(t_1)[U^{-1}]_{kl}\Xi_j^l(t_2) \\
[C_{\tilde{X}\tilde{Y}}(t_1,t_2)]_{ij} &\equiv \overline{\tilde{X}_i(t_1)\tilde{Y}_j(t_2)} = [C_{X'Y'}(t_1,t_2)]_{ij} - \sum_k \sum_l \Xi_i^k(t_1)[U^{-1}]_{kl}\Upsilon_j^l(t_2) \\
[C_{\tilde{Y}\tilde{X}}(t_1,t_2)]_{ij} &\equiv \overline{\tilde{Y}_i(t_1)\tilde{X}_j(t_2)} = [C_{Y'X'}(t_1,t_2)]_{ij} - \sum_k \sum_l \Upsilon_i^k(t_1)[U^{-1}]_{kl}\Xi_j^l(t_2) \\
[C_{\tilde{Y}\tilde{Y}}(t_1,t_2)]_{ij} &\equiv \overline{\tilde{Y}_i(t_1)\tilde{Y}_j(t_2)} = [C_{Y'Y'}(t_1,t_2)]_{ij} - \sum_k \sum_l \Upsilon_i^k(t_1)[U^{-1}]_{kl}\Upsilon_j^l(t_2)
\end{aligned}} \tag{2.35}$$

The prior error covariance is denoted with $C_{X'X'}(t_1,t_2)$ etc. Details of the derivation are presented in Appendix A.2.2. The estimation error covariance for the state indicates how close to the truth the estimator thinks it is. Its diagonal provides us with error bars around the state estimate.

The posterior covariance for the estimates of the measured variable at observation times is

$$\boxed{C_{\tilde{v}} \equiv \overline{\tilde{v}\tilde{v}^T} = R - RU^{-1}R} \tag{2.36}$$

Details of the derivation can be found in Appendix A.2.2. Again, $C_{\tilde{v}}$ indicates how close to the truth the estimator thinks it is. We can use the diagonal of $C_{\tilde{v}}$ to draw error bars around the estimates of the measurements.

Finally, we find for the covariance of the posterior data residuals

$$\boxed{C_{\hat{v}} \equiv \overline{\hat{v}\hat{v}^T} = C_{\tilde{v}} + C_v - C_vU^{-1}R - (C_vU^{-1}R)^T} \tag{2.37}$$

For details of the derivation see Appendix A.2.2. We can use $C_{\hat{v}}$ to normalize the data residuals after the estimation and to check the algorithm's underlying statistical assumptions. This test is particularly powerful when field data are assimilated, because it provides a way to check whether the estimator is operating in accordance with its statistical assumptions. It must be stressed, however, that this test relies on the linearizations that have been applied. Since the model is nonlinear, both $\tilde{v}$ and $\hat{v}$ already have bias (e.g. $\overline{M[X,Y]} \neq M[\overline{X},\overline{Y}]$).

Even if we are only interested in the covariance of the posterior data residuals (2.37), we need a good approximation of the matrix $U^{-1} = (R + C_v)^{-1}$ (2.27). This task is very computationally demanding, because we need to calculate individual representer functions. Luckily, we only have to do this once. After the iteration on the nonlinearity of the forward model has converged, we can switch from the indirect representer technique to computing individual representers as needed. Depending on the problem at hand, we will likely be able to get a reasonable approximation of $U^{-1}$ by calculating only a subset of all representer functions. Note that data compression (Section 2.5) could be particularly helpful to reduce the number of individual representers and therefore the dimension of $U$.

### 2.4.2 Prior Covariance

In order to get error bars around the state estimate or to reinitialize the algorithm for subsequent assimilation windows, we need to compute at least the posterior variance of the

state, which is given by the diagonal elements of the estimation error covariance (2.35). Although the storage of the $N_Z$ representer fields $\Upsilon^k(t)$ poses a serious problem, a difficult part in computing the estimation error variance is the *prior* variance of the state, namely $[C_{X'X'}(t,t)]_{ii} \equiv \sigma^2_{X'_i}(t)$ and $[C_{Y'Y'}(t,t)]_{ii} \equiv \sigma^2_{Y'_i}(t)$.

For the initialization of a subsequent assimilation interval, we are only interested in the posterior error covariance of the state at the final time of the previous assimilation window. This also means that we only need the prior error covariance at the final time. If we only need the prior state covariance at certain times and for certain states, we can again use the representer technique.

Introducing a pseudo-observation for a direct measurement of the state $Y_j(t_f)$ at the final time allows us to compute the corresponding pseudo state representer function $\Upsilon^{*j}$. Now recall that the state representer function is the covariance of the state with the measurement prediction, which in the case of our pseudo-observation is the state at the final time (Sections 2.4.1 and A.2.1). Therefore, the pseudo representer function $\Upsilon^{*j}(t_f)$ at the final time includes the prior state cross-covariance at the final time.

$$[C_{Y'Y'}(t_f,t_f)]_{ij} = \Upsilon^{*j}_i(t_f) \tag{2.38}$$

It is important to note that the computation of the pseudo state representer function $\Upsilon^{*j}$ does *not* require the specification of a corresponding measurement error covariance or a pseudo measurement value. In the representer technique, all the prior knowledge is encapsulated in the representer functions and the representer matrix, which are completely separate from the measurement error covariance. In fact, by comparing (2.9) and (2.24) we see that the representer functions are computed using unit impulses instead of the data misfit terms in the backward equation.

## 2.5 Data Compression and Observing System Assessment

Satellite images often contain a lot of redundant information. When such images are transmitted, image compression is routinely used to minimize the number of bits being transferred. Similarly, treating each pixel of the image as a separate data point in the assimilation algorithm may be very inefficient.

Whether or not data compression techniques help reduce the computational effort required by the data assimilation scheme depends on the assimilation method. If the objective function (2.5) is minimized directly with a gradient search technique, the search happens effectively in the space of the uncertain inputs. If model error is included, this space is more or less equivalent to the state space and hence very big (Section 8.3). Moreover, compressing the data does not lead to significant computational savings because the state space remains unchanged when the data are compressed.

In the representer technique, on the other hand, the search consists mainly of solving (2.26) for the representer coefficients, where the dimension of the space involved is equal to the number of observations. We can say that the search is carried out in the data space and that the unobservable modes have been discarded a priori [Bennett, 1992].

Since in land-surface data assimilation there are generally a lot fewer observations than states, the representer method is an elegant way to reduce the size of the estimation problem by searching only the data space. In addition, the representer method offers the intriguing possibility to save even more computational effort by compressing the data a priori [Bennett

et al., 1993]. If each satellite image is compressed linearly, for example with a singular value decomposition or a two-dimensional Fourier Transform, the assimilated data vector $Z$ can simply be defined as a collection of linear combinations of the original observations.

The degree of meaningful compression and the corresponding savings depend of course on the problem at hand. Note that data compression does not generally mean that observations are discarded. In fact, we typically use all the original observations, but instead of looking at the individual measurements which are supplied by the sensor, we focus on the dominant linear combinations.

It is also important to emphasize that the indirect representer method and a priori data compression may achieve related savings. Data compression might therefore not be of much additional help if the estimates are computed with the indirect representer method (Section 2.3.4). In contrast, only data compression may make the calculation of the posterior error covariances feasible, because for the posterior covariances the individual representer functions are needed (Section 2.4.1), and their number is equal to the length of the (compressed) data vector $Z$. Results related to this question are presented in Section 8.2.1.

Within the (direct) representer technique, we can also assess the observing system a posteriori by carrying out an eigenvalue decomposition of the matrix $U \equiv (C_v + R)$ (2.27). Note that the inverse is unstable if the matrix $U$ is poorly conditioned. If this is the case, the measurement functional could be rotated and the modes corresponding to very small eigenvalues could be discarded to stabilize the assimilation algorithm. This technique amounts to compressing the data a posteriori. For details see [Bennett, 1992].

# Chapter 3

# A Land Surface Model for Data Assimilation

Selecting a hydrologic model for the land surface data assimilation problem is not an easy task. The model must capture the key physical processes adequately, but at the same time it must be efficient enough to make large-scale optimal estimation computationally feasible. In addition, the variational data assimilation approach also requires a differentiable model. From the variety of models described in the literature, we develop a land surface scheme especially designed for data assimilation purposes. Our land surface scheme is a simple model for moisture and heat transport in the unsaturated soil zone and at the land-atmosphere boundary, together with a Radiative Transfer model relating the soil moisture and temperature to the remotely sensed brightness temperature.

A key assumption is to neglect lateral flow in the unsaturated zone, which is reasonable for terrain with moderate relief and on the spatial scales under consideration. The model domain thus breaks down into a collection of one-dimensional vertical cells or pixels (Figure 4.1). In this Chapter we describe the one-dimensional vertical components of the land surface model and the Radiative Transfer scheme. In Chapter 4 we then describe in detail how these one-dimensional model components can be used in a fully four-dimensional (space and time) land surface data assimilation algorithm.

First, in Section 3.1, we present the one-dimensional model for the moisture and heat dynamics. We use Richards' equation for the moisture transport and the force-restore approximation for the soil temperature. A conceptually similar model of the soil processes has been presented by Ács et al. [1991]. The vegetation model is similar to the Simplified Biosphere Model (SSiB) developed by Xue et al. [1991]. In Section 3.2 we describe the Radiative Transfer (RT) model. The RT model consists of a soil part, including (1) the Dobson mixing model for the wet soil dielectric constant, (2) the Fresnel equations for the soil microwave emissivity, and (3) the gradient RT model or the grey body approximation for the effective soil temperature, and of a vegetation part.

A comprehensive list of all symbols used in the hydrologic model can be found in Appendix B.1. Note the notational convention to label most of the empirical constants in the various parameterizations with $\kappa$ for scalars and with $\beta$ for distributed parameters. The constants are superscripted with the variable which is being parameterized and subscripted with a number in case more than one empirical constant is needed.

## 3.1 Moisture and Heat Transport Model

In this Section we describe a simple one-dimensional model for moisture and heat transport in the unsaturated soil zone and at the land-atmosphere boundary. The moisture dynamics are modeled with Richards' equation, whereas the temperature submodel relies on the force-restore approximation of the heat equation. The dynamics of the moisture and the temperature are coupled via the heat capacity, which depends on the moisture content, and via the evapotranspiration rate/latent heat flux at the land surface. The transport of soil water vapor is entirely neglected. The downward flux of water out of the bottom layer is described by gravitational drainage only. The vegetation submodel is designed after the Simplified Biosphere Model (SSiB) Xue et al. [1991].

### 3.1.1 Soil Moisture Submodel

Vertical unsaturated flow is described with a modified version of Richards' equation neglecting soil vapor movement and flow due to thermal gradients.

$$\frac{\partial \theta_g}{\partial t} = \frac{\partial}{\partial z} K_u \frac{\partial}{\partial z}(\psi_g + z) - S_g \tag{3.1}$$

The volumetric moisture content is denoted with $\theta_g$ $[m^3/m^3]$, the matric head with $\psi_g$ $[m]$, and the unsaturated hydraulic conductivity with $K_u$ $[m/s]$. The sink term $S_g$ $[1/s]$ accounts for root water uptake through transpiration.

At the surface, the boundary condition is the net flux resulting from precipitation falling through the canopy $P_l$ $[m/s]$ and from evaporation from the ground surface $E_g$ $[kg/m^2/s]$

$$q_l = -P_l + E_g/\rho_w \tag{3.2}$$

where $\rho_w$ $[kg/m^3]$ is the density of liquid water. Note that $P_l$ is defined as a positive quantity. The gravity drainage from the bottom layer is given by

$$q_b = -K_u(z_{\text{bottom}}) \tag{3.3}$$

The soil hydraulic properties are parameterized with the model by Clapp and Hornberger [1978]

$$\psi_g = \psi^{\text{CH}} W_g^{-B^{\text{CH}}} \qquad W_g = \left(\frac{\psi_g}{\psi^{\text{CH}}}\right)^{-1/B^{\text{CH}}} \tag{3.4}$$

$$K_u = K_s W_g^{2B^{\text{CH}}+3} \qquad K_u = K_s \left(\frac{\psi_g}{\psi^{\text{CH}}}\right)^{-(2B^{\text{CH}}+3)/B^{\text{CH}}} \tag{3.4a}$$

where the soil wetness or saturation $W_g$ $[-]$ is related to the volumetric soil moisture content $\theta_g$ and the porosity $\theta_s$ through $W_g = \theta_g/\theta_s$. The hydraulic conductivity at saturation is denoted with $K_s$. The "matric head at saturation" $\psi^{\text{CH}}$ and the parameter $B^{\text{CH}}$ are empirical constants.

Recent results on robust solvers for Richards' equation can be found in [Miller et al., 1998]. We discretize Richards' equation following the mass conservative scheme developed by Celia et al. [1990]. For details on the discretized equations see [Simunek et al., 1997]. Here we only introduce the notation for the vertical discretization. The locations of the $N_z$

vertical nodes are collected into the vector $z$ $[m]$, and the thickness of the layers between the nodes is denoted with $\Delta$ $[m]$.

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{N_z} \end{bmatrix} \qquad \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_{N_z-1} \end{bmatrix} = \begin{bmatrix} z_2 - z_1 \\ z_3 - z_2 \\ \vdots \\ z_{N_z} - z_{N_z-1} \end{bmatrix} \tag{3.5}$$

The vertical coordinate $z$ $[m]$ increases upward and the locations of the bottom and the top nodes are $z_1$ and $z_{N_z}$, respectively. Note that the vector $\Delta$ has $N_z - 1$ elements.

From the node locations we derive the locations $\bar{z}$ $[m]$ of the mid-points in between the nodes and the thickness $\bar{\Delta}$ $[m]$ of the layers between the mid-points, which is also the thickness of the layers "around" the finite difference nodes.

$$\begin{bmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \vdots \\ \bar{z}_{N_z} \\ \bar{z}_{N_z+1} \end{bmatrix} = \begin{bmatrix} z_1 \\ \frac{z_1+z_2}{2} \\ \vdots \\ \frac{z_{N_z-1}+z_{N_z}}{2} \\ z_{N_z} \end{bmatrix} \qquad \begin{bmatrix} \bar{\Delta}_1 \\ \bar{\Delta}_2 \\ \vdots \\ \bar{\Delta}_{N_z-1} \\ \bar{\Delta}_{N_z} \end{bmatrix} = \begin{bmatrix} \bar{z}_2 - \bar{z}_1 \\ \bar{z}_3 - \bar{z}_2 \\ \vdots \\ \bar{z}_{N_z} - \bar{z}_{N_z-1} \\ \bar{z}_{N_z+1} - \bar{z}_{N_z} \end{bmatrix} = \begin{bmatrix} \frac{\Delta_1}{2} \\ \frac{\Delta_1+\Delta_2}{2} \\ \vdots \\ \frac{\Delta_{N_z-2}+\Delta_{N_z-1}}{2} \\ \frac{\Delta_{N_z-1}}{2} \end{bmatrix} \tag{3.6}$$

It is convenient to define the bottom and the top point as the first and the last mid-points. The vector $\bar{z}$ is thus of length $N_z + 1$, and the vector $\bar{\Delta}$ is of length $N_z$.

### 3.1.2 Soil Temperature Submodel

The temperature submodel is based on the force-restore method [Bhumralkar, 1975; Deardroff, 1978; Lin, 1980; Dickinson, 1988; Hu and Islam, 1995]. The force-restore approximation relies on the analytical solution of the heat equation under periodic forcing, which is used to parameterize the almost periodic daily ground heat flux. In this way, a very simple and efficient but reasonably accurate description of the temperature dynamics can be achieved. Only two layers are considered, a surface layer of thickness $\delta_g$ $[m]$ at temperature $T_g$ $[K]$, and a deeper layer, which serves as a heat reservoir. The deeper layer is at temperature $T_d$ $[K]$, the depth-average temperature of the soil. The prognostic equation for the temperature $T_g$ in the surface layer is

$$\frac{\partial T_g}{\partial t} = \Gamma_g \left[ \Gamma'_g G_g / C_g - (T_g - T_d) \right] \tag{3.7}$$

Here $G_g$ $[W/m^2]$ is the ground heat flux evaluated at the surface, and the coefficients $C_g$ $[J/m^3/K]$, $\Gamma_g$ $[1/s]$, and $\Gamma'_g$ $[s/m]$ are discussed below.

At the ground surface, the energy balance (in $[W/m^2]$) consists of the net shortwave radiation $R_{gs}^{\text{net}}$, the net longwave radiation $R_{gl}^{\text{net}}$, the latent heat flux $LE_g$, and the sensible heat flux $H_g$. The latent heat of vaporization $L$ $[J/kg]$ is used to convert the mass flux $E$ into a latent heat flux.

$$G_g = R_{gs}^{\text{net}} + R_{gl}^{\text{net}} - LE_g - H_g \tag{3.8}$$

The coefficients $\Gamma_g$ $[1/s]$ and $\Gamma'_g$ $[s/m]$ are given by

$$\Gamma_g = \frac{\omega_d}{\alpha_g} \qquad \Gamma'_g = \sqrt{\frac{2}{K_T \omega_d}} = \frac{2}{d_g \omega_d} \tag{3.9}$$

where the coefficient $\alpha_g$ is discussed below, and the damping depth $d_g$ $[m]$ of the daily temperature forcing is

$$d_g = \sqrt{\frac{2\lambda_g}{\omega_d C_g}} = \sqrt{\frac{2K_T}{\omega_d}} \tag{3.10}$$

The angular frequency $\omega_d = 2\pi/86400/s$ enters from the consideration of the analytical formulation with diurnally periodic forcing. The volumetric heat capacity is denoted with $C_g$ $[J/m^3/K]$ and the thermal conductivity with $\lambda_g$ $[W/m/K]$. Both depend on the soil moisture content and thus vary with time (Section 3.1.7). Lastly, we define thermal diffusivity $K_T$ $[m^2/s]$ as

$$K_T = \frac{\lambda_g}{C_g}. \tag{3.11}$$

As discussed in Section 3.1.7 we use a constant thermal diffusivity, implying that both $\Gamma_g$ and $\Gamma'_g$ are constant.

Following [Hu and Islam, 1995], the different versions of the force-restore method can be described by different expressions for the dimensionless coefficient $\alpha_g$. We will use the version proposed by Hu and Islam [1995], which minimizes the approximation error of the soil temperature at depth $\delta_g$. The computationally efficient polynomial approximation to the optimal $\alpha_g$ is

$$\alpha_g = 1 + \kappa_1^{\alpha_g}(\delta_g/d_g) + \kappa_2^{\alpha_g}(\delta_g/d_g)^2 + \kappa_3^{\alpha_g}(\delta_g/d_g)^3 + \kappa_4^{\alpha_g}(\delta_g/d_g)^4 \tag{3.12}$$

$$\kappa_1^{\alpha_g} = 0.943 \qquad \kappa_2^{\alpha_g} = 0.223 \qquad \kappa_3^{\alpha_g} = 1.68 \cdot 10^{-2} \qquad \kappa_4^{\alpha_g} = -5.27 \cdot 10^{-3}$$

which is valid for $0 \leq \delta_g/d_g \leq 5$. The closed form for greater values of $\delta_g/d_g$ can be found in [Hu and Islam, 1995]. If we keep in mind that $\delta_g$ is the modeled depth of the surface layer in which the temperature varies diurnally, and that $d_g$ is the damping depth of the forcing, we see that $\delta_g/d_g$ should not be far from unity to be consistent. Therefore, the polynomial approximation (3.12) will be appropriate for all practical purposes.

It should be noted that the force-restore versions of Bhumralkar [1975] and of Lin [1980], using $\alpha_g = 1 + 2\delta_g/d_g$ and $\alpha_g = 1 + \delta_g/d_g$, respectively, can be considered special cases of (3.12). They yield similar results for large (Bhumralkar) and small (Lin) values of $\delta_g/d_g$, respectively. For details see [Hu and Islam, 1995].

### 3.1.3 Vegetation Submodel

The vegetation is modeled with a standard resistance network for the latent and sensible heat fluxes (Figure 3.1).

#### Canopy Heat Submodel

We assume zero heat capacity for the canopy. The canopy energy balance reads

$$R_{cs}^{\text{net}} + R_{cl}^{\text{net}} - LE_{ct} - LE_{ce} - H_c = 0 \tag{3.13}$$

consisting of the net shortwave radiation $R_{cs}^{\text{net}}$, the net longwave radiation $R_{cl}^{\text{net}}$, the latent heat fluxes $LE_{ct}$ and $LE_{ce}$ from the dry and wet portions of the canopy (canopy transpiration and canopy evaporation), and the sensible heat flux $H_c$. All fluxes are measured in $[W/m^2]$. This energy balance is solved for the canopy temperature $T_c$ $[K]$.

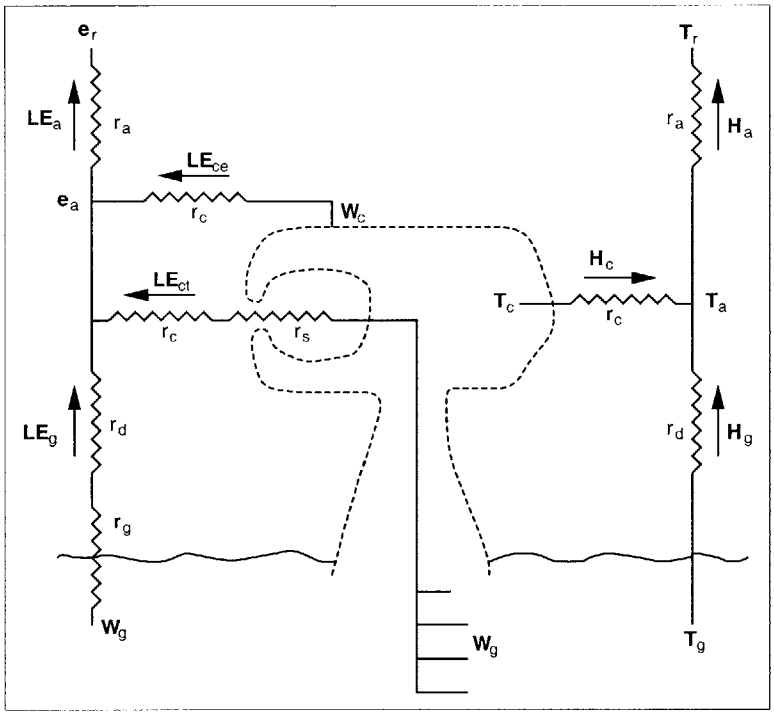


Figure 3.1: Resistance network (after [Xue et al., 1991])

### Canopy Water Submodel

If vegetation is present, the canopy intercepts the precipitation. The balance equation for the water $W_c$ $[m]$ stored in the canopy (per unit area) reads

$$\frac{\partial W_c}{\partial t} = P_i - E_{ce}/\rho_w - D_c \tag{3.14}$$

The intercepted flux is denoted with $P_i$ $[m/s]$. The evaporation from the wet canopy is $E_{ce}$ $[kg/m^2/s]$, and $D_c$ $[m/s]$ accounts for dripping off the canopy. We use $D_c = W_c/t_c^{\text{drip}}$, with a typical dripping time $t_c^{\text{drip}}$ $[s]$ on the order of $0.5d$ [Thompson and Pollard, 1995].

A simple parameterization for the intercepted precipitation and the evaporation from the wet canopy is

$$\begin{aligned} P_i &= \min\left(f_c P_r, (W_c^{\max} - W_c)/\Delta t\right) \\ E_{ce} &= \min\left(E_{ce}^{\text{pot}}, \rho_w W_c/\Delta t\right) \end{aligned} \tag{3.15}$$

We denote with $f_c$ $[-]$ the fraction of the land surface which is shaded by the canopy. $P_r$ $[m/s]$ is the precipitation at screen (or reference) height $z_r$ $[m]$ as measured by the meteorologic station. (Recall that precipitation is defined as a positive quantity.) $E_{ce}^{\text{pot}}$ is the potential evaporation from the wet canopy assuming unlimited water supply. $W_c^{\max}$ $[m]$ is the canopy storage capacity, and $\Delta t$ $[s]$ is the numerical time step. Equation (3.15) approximately constrains the water stored in the canopy to be nonnegative and less than the storage capacity, $W_c \lesssim W_c^{\max}$. We write the canopy storage capacity as

$$W_c^{\max} = f_c \text{LSAI} \kappa^{W_c} \tag{3.16}$$

where LSAI [−] is the sum of the one-sided leaf area index LAI [−] and the stem area index SAI [−]. The parameter $\kappa^{W_c}$ $[m]$ is on the order of $\kappa^{W_c} \approx 10^{-4} \ldots 10^{-3} m$ [Abramopoulos et al., 1988]. Finally, we have for the throughfall

$$P_t = P_r + D_c - P_i \tag{3.17}$$

### Canopy Air Submodel

We also assume zero heat capacity for the canopy air. The energy balance reads

$$H_a = H_c + H_g \tag{3.18}$$

where $H_a$ $[W/m^2]$ is the overall sensible heat flux to the atmosphere. This energy balance is solved for the temperature of the canopy air $T_a$ $[K]$. The mass balance within the canopy air space is

$$LE_a = LE_g + LE_{ct} + LE_{ce} \tag{3.19}$$

where $LE_a$ $[W/m^2]$ is the overall evapotranspiration to the atmosphere. This mass balance is solved for the vapor pressure of the canopy air $e_a$ $[mb]$.

## 3.1.4 Radiation Balance

All radiation fluxes are in $[W/m^2]$. The downwelling shortwave and longwave radiation at reference height above the canopy are denoted with $R_{rs}$ and $R_{rl}$, respectively. For the net fluxes at the ground surface and at the canopy we have [Deardroff, 1978]

$$R_{gs}^{\text{net}} = (1 - f_c)(1 - a_g) R_{rs} \tag{3.20}$$

$$R_{gl}^{\text{net}} = (1 - f_c)\epsilon_g (R_{rl} - \sigma T_g^4) + f_c \frac{\epsilon_c \epsilon_g}{\epsilon_c + \epsilon_g - \epsilon_c \epsilon_g} \sigma (T_c^4 - T_g^4) \tag{3.21}$$

$$R_{cs}^{\text{net}} = f_c (1 - a_c) R_{rs} \tag{3.22}$$

$$R_{cl}^{\text{net}} = f_c \left[ \epsilon_c R_{rl} + \frac{\epsilon_c}{\epsilon_c + \epsilon_g - \epsilon_c \epsilon_g} \epsilon_g \sigma T_g^4 - \frac{\epsilon_c + 2\epsilon_g - \epsilon_c \epsilon_g}{\epsilon_c + \epsilon_g - \epsilon_c \epsilon_g} \epsilon_c \sigma T_c^4 \right] \tag{3.23}$$

where $a_g$, $a_c$, $\epsilon_g$, and $\epsilon_c$ [−] denote the albedos (shortwave) and emissivities (longwave) of the ground and the canopy, respectively, and $\sigma$ $[W/m^2/K^4]$ is the Stefan-Boltzmann constant.

The parameterization for the ground surface albedo follows [Idso et al., 1975].

$$a_g = \kappa_1^{a_g} - \kappa_2^{a_g} W_g(z_{\text{top}}) \qquad \kappa_1^{a_g} = 0.25 \qquad \kappa_2^{a_g} = 0.125 \tag{3.24}$$

The canopy albedo is on the order of $a_c \approx 0.16 \ldots 0.2$ for grassland and crops.

We compute the downwelling longwave radiation from the air temperature

$$R_{rl} = \epsilon_r \sigma T_r^4 \tag{3.25}$$

and the expressions for the air and the ground surface emissivities are

$$\epsilon_r = \kappa_1^{\epsilon_r} + \kappa_2^{\epsilon_r} e_r \qquad \kappa_1^{\epsilon_r} = 0.74 \qquad \kappa_2^{\epsilon_r} = 0.0049 \tag{3.26}$$

$$\epsilon_g = \kappa_1^{\epsilon_g} + \kappa_2^{\epsilon_g} \theta_g \qquad \kappa_1^{\epsilon_g} = 0.9 \qquad \kappa_2^{\epsilon_g} = 0.18 \tag{3.27}$$

$T_r$ $[K]$ is the atmospheric temperature at screen height, and $e_r$ $[mb]$ is the water vapor pressure at screen height. The formulation of the soil surface emissivity is taken from [Chung and Horton, 1987]. The expression (3.26) for the atmospheric emissivity is known as the *Idso*-formula. For the volumetric soil moisture affecting the long wave radiation we use the top node moisture content $\theta_g = \theta_g(z_{\text{top}})$. The canopy emissivity is on the order of $\epsilon_c \approx 0.95 \ldots 1$.

With both the emissivities for the ground and the canopy being close to unity, we can expand the above equations into series in $(1 - \epsilon_g)$ and $(1 - \epsilon_c)$. Using

$$\epsilon_c + \epsilon_g - \epsilon_c\epsilon_g \equiv 1 - (1 - \epsilon_c)(1 - \epsilon_g)$$

and neglecting terms of second order or higher we get

$$\begin{aligned} R_{gl}^{\text{net}} &= (1 - f_c)\epsilon_g(R_{rl} - \sigma T_g^4) + f_c(\epsilon_c + \epsilon_g - 1)\sigma(T_c^4 - T_g^4) \qquad (3.21a) \\ &= (1 - f_c)\epsilon_g R_{rl} - [\epsilon_g + f_c(\epsilon_c - 1)]\sigma T_g^4 + f_c(\epsilon_c + \epsilon_g - 1)\sigma T_c^4 \\ R_{cl}^{\text{net}} &= f_c\left[\epsilon_c\epsilon_r\sigma T_r^4 + (\epsilon_c + \epsilon_g - 1)\sigma T_g^4 - (2\epsilon_c + \epsilon_g - 1)\sigma T_c^4\right] \qquad (3.23a) \end{aligned}$$

These expressions are used in the model.

### 3.1.5 Sensible and Latent Heat Fluxes

All sensible and latent heat fluxes are determined with a resistance formulation.[1]

$$H_a = \rho_a c_a \frac{T_a - T_r}{r_a} \qquad (3.28)$$

$$H_g = (1 - f_c)\rho_a c_a \frac{T_g - T_a}{r_d} \qquad (3.29)$$

$$H_c = 2\text{LSAI}\, f_c \rho_a c_a \frac{T_c - T_a}{r_c} \qquad (3.30)$$

$$LE_a = \frac{\rho_a c_a}{\gamma}\frac{(e_a - e_r)}{r_a} \qquad (3.31)$$

$$LE_g = (1 - f_c)\frac{\rho_a c_a}{\gamma}\frac{(e_s(T_g) - e_a)}{r_g + r_d} \qquad (3.32)$$

$$LE_{ce}^{\text{pot}} = f_c f_{ce}\text{LSAI}\frac{\rho_a c_a}{\gamma}\frac{(e_s(T_c) - e_a)}{r_c} \qquad (3.33)$$

$$LE_{ct}^{\text{pot}} = f_c(1 - f_{ce})\text{LAI}\frac{\rho_a c_a}{\gamma}\frac{(e_s(T_c) - e_a)}{r_c + r_s} \qquad (3.34)$$

$$LE_{ct} = \chi_g(W_g)LE_{ct}^{\text{pot}} \qquad (3.35)$$

All latent and sensible heat fluxes are expressed in $[W/m^2]$. The air density is denoted with $\rho_a$ $[kg/m^3]$, the specific heat of air at constant pressure with $c_a$ $[J/kg/K]$, and the psychometric constant with $\gamma$ $[mb/K]$. The temperature of the canopy is $T_c$ $[K]$. The vapor pressure and the temperature within the canopy air space are $e_a$ $[mb]$ and $T_a$ $[K]$,

[1] Note that we can rewrite the evapotranspiration rate as $E_a = \rho_a(q_a - q_r)/r_a$ by using $\gamma = c_a p_r/L\epsilon$ and $q_r = \epsilon e_r/p_r$, where $q_r$ is the specific humidity, $p_r$ the air pressure, and $\epsilon$ the ratio of the gas constants, $\epsilon = R_{\text{dry air}}/R_{\text{vapor}} = 0.622$.

respectively. Furthermore, $e_r$ $[mb]$ and $T_r$ $[K]$ are the vapor pressure and the air temperature at screen or reference height as given by the meteorologic measurements, and $e_s(T)$ $[mb]$ is the saturation vapor pressure at temperature $T$, calculated by the empirical formula

$$e_s(T) = 6.11 \exp\left\{\frac{17.4(T - T_0)}{T - 34.16}\right\} \tag{3.36}$$

where $T_0 = 273.15K$. Moreover, the fraction of the canopy which is wet and from which water is directly evaporated is denoted with $f_{ce}$ $[-]$ and modeled after [Abramopoulos et al., 1988]

$$f_{ce} = (W_c/W_c^{\max})^{\kappa^{fce}} \tag{3.37}$$

The resistances $r_x$ and the stress function $\chi_g(W_g)$, which parameterizes water-limited transpiration, are discussed in Section 3.1.6.

### 3.1.6 Resistances

The resistances can be partitioned into the aerodynamic resistances $r_a, r_d, r_c$ and the surface resistances $r_g, r_s$. All resistances are measured in $[s/m]$. Water-limited (stressed) transpiration is parameterized with the stress function $\chi_g(W_g)$ $[-]$.

#### Aerodynamic Resistances

In our scheme we have three aerodynamic resistances: the atmospheric resistance between the canopy air space and the reference level in the atmosphere $r_a$, the resistance between the leaves and the canopy air space $r_c$, and the resistance between the soil surface and the canopy air space $r_d$ (Figure 3.1).

The atmospheric resistance for *neutral* conditions (in terms of buoyancy) can be obtained by elimination of the friction velocity $u_*$ $[m/s]$ from the expression (3.39). The friction velocity is a parameter in the logarithmic wind profile (3.38).

$$u_r = \frac{u_*}{K} \ln\left[\frac{z_r - d_c}{z_0}\right] \tag{3.38}$$

$$r_a = \frac{\ln\left[\frac{z_r - d_c}{z_0}\right]}{K u_*} = \frac{\ln^2\left[\frac{z_r - d_c}{z_0}\right]}{K^2 u_r} \tag{3.39}$$

Here, $u_r$ $[m/s]$ is the wind velocity at the reference height $z_r$ $[m]$, $K = 0.4$ is the von Karman constant, and $z_0$ $[m]$ is the roughness length. The zero displacement height $d_c$ $[m]$ accounts for the geometric effect of vegetation stands. If vegetation of height $h_c$ $[m]$ is present, we use

$$d_c = \kappa^{d_c} h_c \qquad \text{and} \qquad z_0 = \kappa^{z_0} h_c \tag{3.40}$$

with $\kappa^{d_c} = 0.63$ and $\kappa^{z_0} = 0.13$ [Abramopoulos et al., 1988]. Although strictly speaking the zero displacement height vanishes for bare soils, we can still use (3.40) by setting $h_c = z_0/\kappa^{z_0}$ with a roughness length on the order of $z_0 \approx 2.5mm$, which is typical for bare soils. Note that in this case $z_r \gg d_c$ and the error introduced by using (3.40) for bare soil is negligible.

For nonneutral conditions, stability corrections can be taken into account. These corrections are usually parameterized with the gradient Richardson number or the Monin-Obukhov length. Detailed expressions can be found in [Ács et al., 1991]. We will confine ourselves to the assumption of neutral conditions.

For the other two aerodynamic resistances, we use parameterizations from [Sellers and Dorman, 1987].

$$r_c = \frac{\beta^{r_c}}{\sqrt{u_c}} \tag{3.41}$$

$$r_d = \frac{\beta^{r_d}}{u_c} \tag{3.42}$$

The wind speed $u_c$ $[m/s]$ at the canopy top can be obtained from the wind profile (3.38) as

$$u_c = u_r \ln\left[\frac{h_c - d_c}{z_0}\right] \ln^{-1}\left[\frac{z_r - d_c}{z_0}\right] \tag{3.43}$$

Note that a measurement of zero wind speed would lead to numerical problems because the resistances diverge. Such problems are easily overcome by preprocessing and setting all measurements of zero wind speed to a minimum of $0.01m/s$. This is well below the instrument resolution and the error inflicted is negligible.

### Surface Resistances

The surface resistance of bare soil $r_g$ $[s/m]$ parameterizes the effect of soil-water limited evaporation. If the upper soil layer is not saturated, the water evaporates at some depth in the soil and must reach the surface through diffusion. We work with the following formulation,

$$r_g = \kappa^{r_g} \exp\left(\beta^{r_g}(W^{r_g} - W_g)\right) \tag{3.44}$$

which depends on three parameters. At saturations $W_g \gtrsim W^{r_g}$ the surface resistance $r_g \lesssim \kappa^{r_g} \equiv 10s/m$ becomes negligible when compared to $r_d$. The third parameter is the maximum surface resistance $r_g^{\max}$ at zero saturation, which determines $\beta^{r_g} = 1/W^{r_g} \ln(r_g^{\max}/\kappa^{r_g})$. Typical values are $W^{r_g} \approx 0.25 \ldots 0.6$ and $r_g^{\max} \approx 3000 \ldots 7000s/m$, depending on the texture of the soil in question. Finer soils have higher surface resistances. Figure 3.2 shows the bare soil resistance as a function of the volumetric soil moisture content. Also shown are two formulations by Kondo et al. [1990] and by van de Griend and Owe [1994], which we will not discuss here.

The stomatal resistance $r_s$ $[s/m]$ describes the closure of the plants' stomate due to environmental impacts [Lhomme et al., 1998]. The most important factors determining the transpiration under unstressed conditions (no water limitation) are shortwave (or photosynthetically active) radiation, temperature, and vapor deficit, although some authors believe that the dependence on vapor deficit is an artificial effect. We only retain the dependence of $r_s$ on the shortwave radiation, which essentially shuts off transpiration at night. We use a formulation equivalent to the one in [Dorman and Sellers, 1989]

$$r_s = r_s^{\min} \chi_c(R_{rs}) \tag{3.45}$$

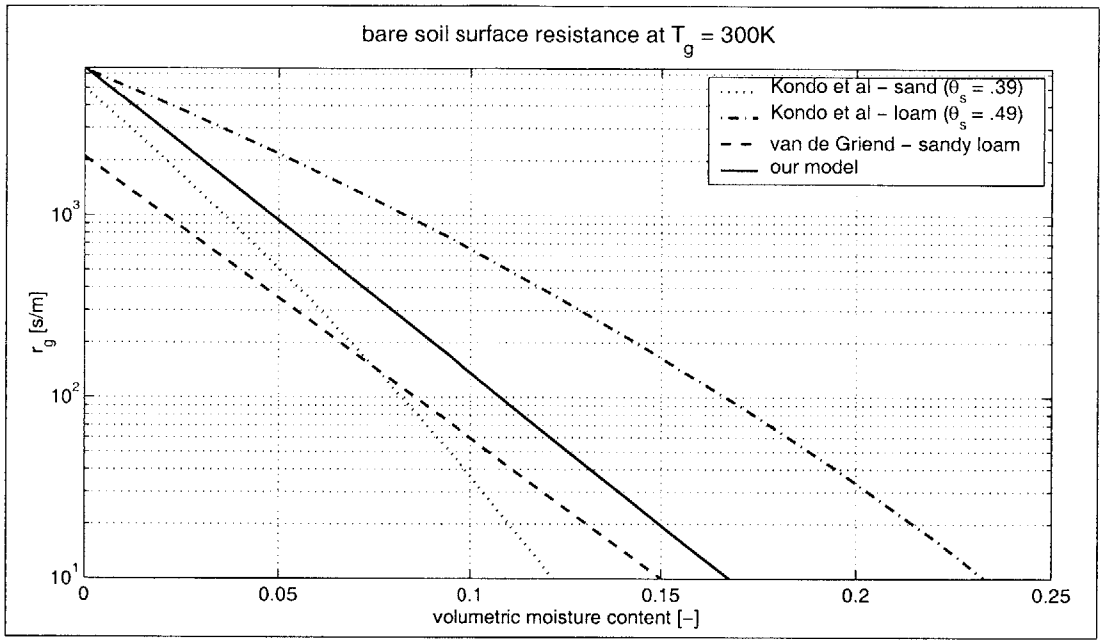


Figure 3.2: Bare soil resistance from (3.44) as a function of the volumetric soil moisture content using $W^{r_g} = 0.6$ and $r_g^{\max} = 6500 s/m$. For comparison, we also show the parameterizations by Kondo et al. [1990] and by van de Griend and Owe [1994].

where

$$\chi_c(R_{rs}) = \begin{cases} \kappa_1^{\chi_c} & R_{rs} < 0 \\ \frac{R_{rs}/\kappa_2^{\chi_c}+1}{R_{rs}/\kappa_2^{\chi_c}+1/\kappa_1^{\chi_c}} & R_{rs} \geqq 0 \end{cases} \tag{3.46}$$

Typical values are $r_s^{\min} \approx 30 \ldots 50 s/m$, $\kappa_1^{\chi_c} \approx 50 \ldots 100$, and $\kappa_2^{\chi_c} \approx 100 \ldots 200 W/m^2$.

The factor $\chi_g(W_g)$ in (3.35) parameterizes the conditions of water-limited transpiration. The total stress is given by the sum over the individual stress terms at the $i = 1 \ldots N_z$ nodes of the vertically discretized soil moisture model.

$$\chi_g(W_g) = \sum_{i=1}^{N_z} \chi_{gi}(W_{gi}) \qquad \chi_{gi}(W_{gi}) = \frac{f_{Ri}}{1 + \psi_{gi}/\psi^{\text{wilt}}} = \frac{f_{Ri}}{1 + \frac{\psi_i^{\text{CH}}}{\psi^{\text{wilt}}} \frac{1}{W_{gi}^{B^{\text{CH}}}}} \tag{3.47}$$

We used (3.4) to obtain the last equality. We denote with $\psi^{\text{wilt}}$ $[m]$ the matric head at which the plants start wilting. A typical number is the equivalent of $-15 bar$ [Dickinson et al., 1993]

$$\psi^{\text{wilt}} = -15 bar/\rho_w/g = -153 m \tag{3.48}$$

The acceleration due to gravity is $g = 9.81 m^2/s$. The root distribution factor $f_{Ri} > 0$ measures the root density attributed to node $i$. We use a simple exponential model for the continuous root density $\rho_R$ $[1/m]$ with maximum density at the top.

$$\rho_R(z) = \left(\frac{1}{d_R}\right) \frac{\exp(\frac{z}{d_R})}{\exp(\frac{z_{N_z}}{d_R}) - \exp(\frac{z_1}{d_R})} \tag{3.49}$$

where $d_R > 0$ $[m]$ is a typical rooting depth. For the root distribution factor $f_{Ri}$ we obtain

$$f_{Ri} = \int_{\bar{z}_i}^{\bar{z}_{i+1}} \rho_R(z)dz = \frac{\exp(\frac{\bar{z}_{i+1}}{d_R}) - \exp(\frac{\bar{z}_i}{d_R})}{\exp(\frac{z_{N_z}}{d_R}) - \exp(\frac{z_1}{d_R})} \tag{3.50}$$

Note that $\rho_R(z)$ is normalized on the interval $[z_1, z_{N_z}]$ and that $\sum_{i=1}^{N_z} f_{Ri} = 1$.

Finally, we can give an expression for the sink term as it appears in the discretized form of Richards' equation (3.1).

$$\bar{\Delta}_i S_{gi} = \chi_{gi}(W_{gi}) E_{ct}^{\text{pot}} / \rho_w \tag{3.51}$$

### 3.1.7 Thermal Properties of the Soil

The volumetric heat capacity of the surface layer depends on the soil moisture content according to [Simunek et al., 1997]

$$C_g = \rho_{gb} c_g + \theta_g \rho_w c_w \tag{3.52}$$

where $c_g$ $[J/kg/K]$ is the specific heat of the dry soil, $\rho_w$ $[kg/m^3]$ is the density of water, and $c_w$ $[J/kg/K]$ is the specific heat of water. The bulk density of the dry soil $\rho_{gb}$ $[kg/m^3]$ is given by

$$\rho_{gb} = (1 - \theta_s)\rho_g \tag{3.53}$$

where $\rho_g$ $[kg/m^3]$ is the density of the soil particles. A generally accepted number for all soil types is $\rho_g = 2.65 \cdot 10^3 kg/m^3$. For the volumetric moisture content and the porosity which enter the parameterization we choose values averaged over the top layer of thickness $\delta_g$ of the force-restore approximation.

Following [Chung and Horton, 1987], the thermal conductivity $\lambda_g$ $[W/m/K]$ can be parameterized as

$$\lambda_g = \beta_1^{\lambda_g} + \beta_2^{\lambda_g} \theta_g + \beta_3^{\lambda_g} \sqrt{\theta_g} \tag{3.54}$$

The $\beta_i^{\lambda_g}$ depend on the soil type (Table B.5). Figure 3.3 shows the volumetric heat capacity, the thermal conductivity, the thermal diffusivity, and the damping depth as functions of the saturation for three soil types [Chung and Horton, 1987]. We adopt the approximation of a constant heat diffusivity $K_T$ which is justifiable for not too dry conditions, $W_g \gtrsim 0.1$. This is important because for constant $K_T$ the force-restore parameters $\alpha_g$, $d_g$, $\Gamma_g$, and $\Gamma'_g$ are also constant and need not be computed at each time step.

### 3.1.8 Thickness and Number of Soil Layers

In the article by Ács et al. [1991], the thickness of the uppermost soil layer for the moisture submodel is $10cm$, whereas the thickness of the top layer for the temperature submodel is $\delta_g = 2cm$. This disparity is somewhat inconsistent with the model formulation, because the thermal properties of the soil depend on the soil water content. A top moisture layer of $10cm$ thickness can only represent an average moisture value, and it cannot resolve the moisture content of the top $2cm$, which is used in the temperature submodel.

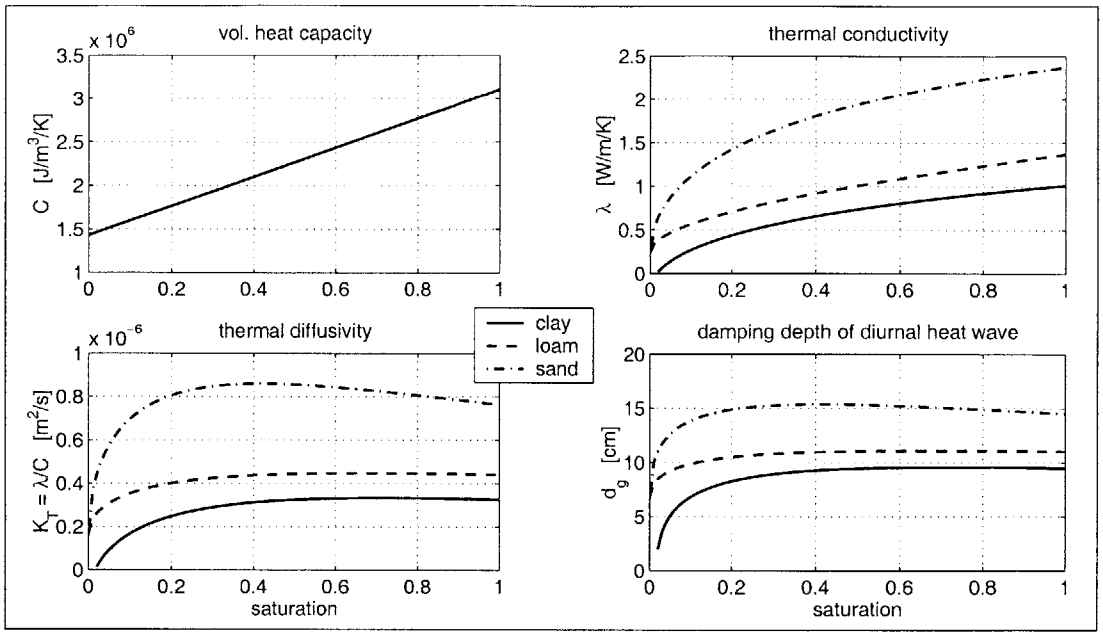


Figure 3.3: Soil thermal properties: volumetric heat capacity (upper left), thermal conductivity (upper right), thermal diffusivity (lower left), and damping depth (lower right) as functions of the saturation $W_g$. The parameters are $\rho_g = 2.65 \cdot 10^3 kg/m^3$, $\theta_s = 0.4$, $c_g = 900 J/kg/K$.

To be consistent *within* the force-restore approximation, the layer for the temperature $T_g$ should also correspond as much as possible to the layer in which the temperature actually fluctuates. A measure for the thickness $\delta_g$ of this layer is the damping depth $d_g$ of the diurnal forcing. Recall from Figure 3.3 that the damping depth $d_g$ varies between $7cm$ and $15cm$ for various soil types and over a wide range of moisture conditions. Similarly, Hu and Islam [1995] cite damping depths $d_g$ that range from $5cm$ (clay) to $20cm$ (rock).

Now note that appreciable errors due to the neglect of higher harmonics can occur in the force-restore method if the upper soil thickness $\delta_g$ is less than the damping depth of the diurnal forcing [Hu and Islam, 1995]. This suggests the use of a thicker top soil layer for the temperature submodel. On the other hand, the ground surface temperature $T_g$ from the force-restore method is used in the computation of the boundary fluxes and will also be used for the Radiative Transfer model (Section 3.2). We are therefore interested in the temperature of a shallow layer. Moreover, the diurnal heat wave is much stronger than the higher harmonics, so the errors due to the neglect of higher harmonics should be bearable.

The best overall compromise for moisture and heat transport seems to be a single value of $5cm$ for both the top moisture layer and the $\delta_g$-layer of the force-restore approximation. For the lower layers in the moisture submodel, we use a few layers which reflect the measurement levels of existing time-domain reflectometry (TDR) profile measurement devices. Generally, we work with six layers ranging from 0–$5cm$, 5–$15cm$, 15–$30cm$, 30–$45cm$, 45–$60cm$, and 60–$90cm$. This allows for the assimilation of data from field experiments such as the Southern Great Plains 1997 (SGP97) experiment (Section 5.3.1).

## 3.2 Radiative Transfer Model

In this Section we describe a model for the Radiative Transfer (RT). It relates the soil moisture and temperature to the remotely sensed radiobrightness temperature $T_B$, which is a measure of the microwave energy emitted by the soil. In Section 3.2.1 we first describe a model for microwave emission from bare soil, and we then give a complete description in the presence of a vegetation cover in Section 3.2.2.

### 3.2.1 Microwave Emission from Bare Soil

For microwave emission from bare soil, various coherent and incoherent Radiative Transfer models can be found in the literature [Njoku and Kong, 1977; Wilheit, 1978; Schmugge and Choudhury, 1981; Choudhury et al., 1982; Raju et al., 1995]. Most recently, Galantowicz et al. [1999] discussed and compared several options such as conventional RT, a gradient RT approximation, a grey body approximation, and a coherent model for a stratified medium.

In a non-coherent model, the observed radiobrightness temperature $T_B$ is factored into the microwave emissivity $\varepsilon_g$ and the effective soil temperature $T_g^{\text{eff}}$.

$$T_B^{\text{bare}} = \varepsilon_g T_g^{\text{eff}} \tag{3.55}$$

The model then consists of three parts:

1. The wet soil dielectric constant is computed from the dielectric constants of dry soil and water depending on the volumetric soil moisture content (Dobson mixing formula).
2. The microwave emissivity $\varepsilon_g$ is obtained from the wet soil dielectric constant (Fresnel equations).
3. The effective temperature $T_g^{\text{eff}}$ is determined from the soil temperature profile and possibly the dielectric constant.

In the following subsections, we will discuss each of these steps separately.

#### Wet Soil Dielectric Constant

The dielectric constant of the wet soil $k_g$ is obtained from the dielectric constants of the dry soil and the soil water, which are evaluated at the microwave frequency of interest. For L-band observations, for instance, the frequency is $\nu_r = 1.4GHz$. Ulaby et al. [1986] present among other models a semi-empirical mixing model which makes use of soil texture information in order to account for the bound water contribution in an aggregate way. We call their model the Dobson mixing model because it is an improved version of the semi-empirical mixing model presented by Dobson et al. [1985]. In particular, the model for the wet soil dielectric constant reads

$$k_g = \left\{ 1 + (1 - \overline{\theta_s})(k_{gd}^{\kappa_1^{k_g}} - 1) + \overline{\theta_g}^{\beta^{k_g}} (k_w^{\kappa_1^{k_g}} - 1) \right\}^{(1/\kappa_1^{k_g})} \tag{3.56}$$

with

$$\beta^{k_g} = \kappa_2^{k_g} - \kappa_3^{k_g} f_S - \kappa_4^{k_g} f_C \tag{3.57}$$

where $f_S$ [−] and $f_C$ [−] are the sand and clay fractions of the soil in question. The $\kappa_i^{k_g}$ are empirical constants. Their values can be found in Table B.7. The porosity $\overline{\theta_s}$ and the volumetric moisture content $\overline{\theta_g}$ are appropriate averages over depth (Section 3.2.1). For L-band observations, Galantowicz et al. [1999] used averaging depths of $1.5cm$ and $2cm$ for gradient RT and grey body RT, respectively (Section 3.2.1). In our coarse discretization (Section 3.1.8) we best use an average moisture content and an average porosity over the top layer.

The dielectric constant of the dry soil $k_{gd}$ depends only very weakly on the density of the soil particles. A good approximation for all soil types is [Dobson et al., 1985]

$$k_{gd} = 4.67 \tag{3.58}$$

The dielectric constant of water $k_w$ [−] at the microwave frequency $\nu_r$ [$Hz$] and temperature $T$ [$K$] is determined with a Debye model [Ulaby et al., 1986].

$$k_w = k_{w\infty} + \frac{(k_{w0} - k_{w\infty})}{1 + (2\pi\nu_r\tau_w)^2}(1 + i \cdot 2\pi\nu_r\tau_w) \tag{3.59}$$

where $k_{w\infty} = 4.9$ is the high frequency limit of $k_w$. The static limit $k_{w0}$ [−] as well as the relaxation time of water $\tau_w$ [$s$] are given by Taylor series expansions.

$$k_{w0} = \kappa_1^{k_{w0}} + \kappa_2^{k_{w0}}(T - T_0) + \kappa_3^{k_{w0}}(T - T_0)^2 + \kappa_4^{k_{w0}}(T - T_0)^3 \tag{3.60}$$

$$\tau_w = \frac{1}{2\pi}\left\{\kappa_1^{\tau_w} + \kappa_2^{\tau_w}(T - T_0) + \kappa_3^{\tau_w}(T - T_0)^2 + \kappa_4^{\tau_w}(T - T_0)^3\right\} \tag{3.61}$$

The coefficients $\kappa_i^{\tau_w}$ and $\kappa_i^{k_{w0}}$ are listed in Table B.7. The reference temperature is $T_0 = 273.15K$. We choose for the temperature $T = T_g$ from the force-restore approximation (Section 3.2.1).

Figure 3.4 shows the dependence of the wet soil dielectric constant on soil moisture for the five soil types discussed by Dobson et al. [1985] together with the wet soil dielectric constant derived from two dielectric mixing models presented by Birchak et al. [1974], which we will not discuss here. Figure 3.5 shows the dielectric constant of water at frequency $\nu_r = 1.4GHz$ as a function of temperature. Note that the temperature effect is of the same order as the texture effect (Figure 3.4).

### Microwave Emissivity

The microwave emissivity for a smooth soil surface is readily obtained from the Fresnel equations. The microwave emissivities of the soil for horizontal polarization $\varepsilon_{gh}^{\text{smooth}}$ and for vertical polarization $\varepsilon_{gv}^{\text{smooth}}$ are for a look-angle $\phi_r$ from nadir (in air)

$$\varepsilon_{gh}^{\text{smooth}} = 1 - \left|\frac{k_g\cos\phi_r - \sqrt{k_g - \sin^2\phi_r}}{k_g\cos\phi_r + \sqrt{k_g - \sin^2\phi_r}}\right|^2 \tag{3.62}$$

$$\varepsilon_{gv}^{\text{smooth}} = 1 - \left|\frac{\cos\phi_r - \sqrt{k_g - \sin^2\phi_r}}{\cos\phi_r + \sqrt{k_g - \sin^2\phi_r}}\right|^2 \tag{3.63}$$

$$\varepsilon_{gn}^{\text{smooth}} = 1 - \left|\frac{\sqrt{k_g} - 1}{\sqrt{k_g} + 1}\right|^2 \tag{3.64}$$

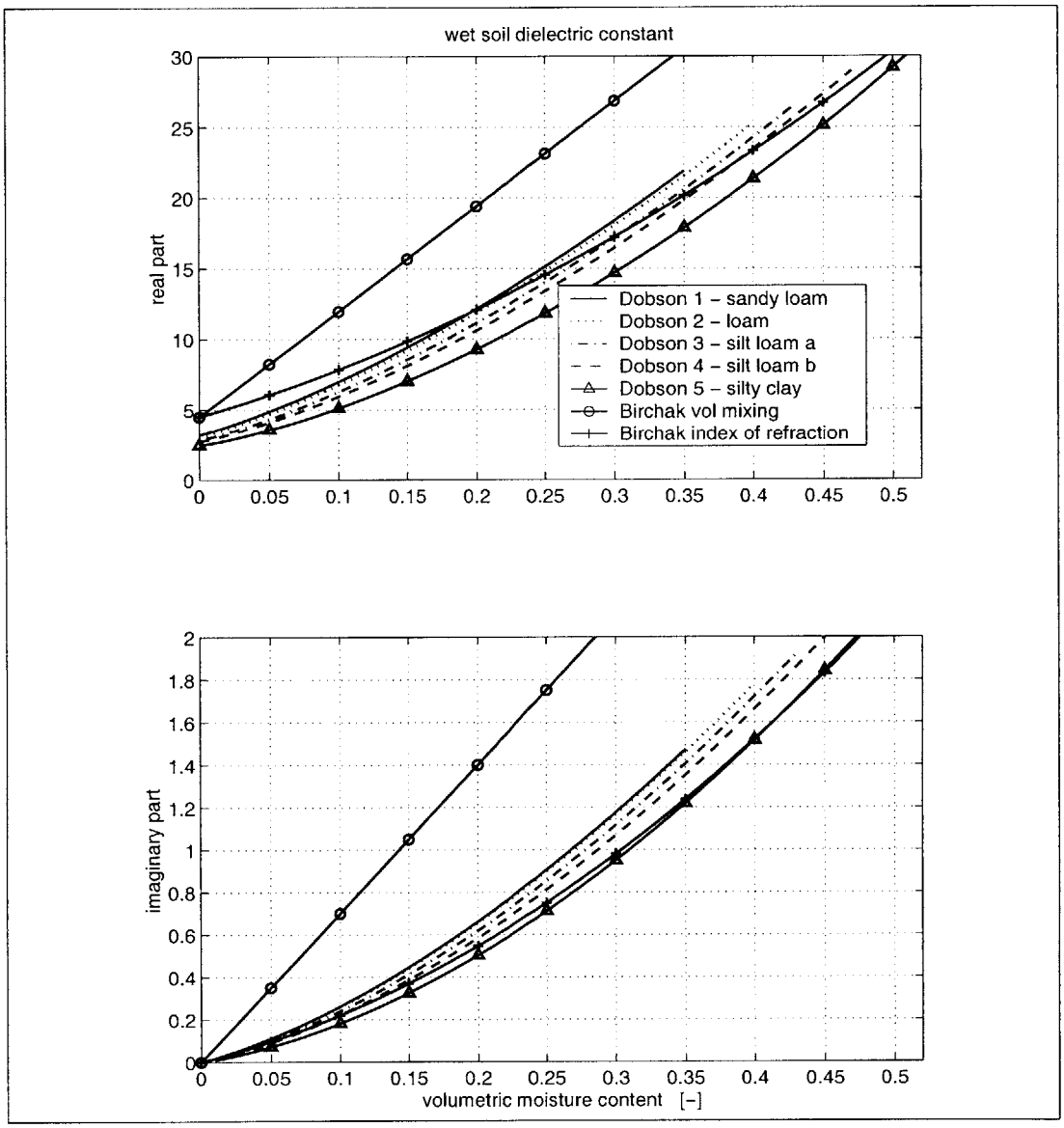


Figure 3.4: Wet soil dielectric constant for the five soil types discussed by Dobson et al. [1985] together with the wet soil dielectric constant derived from two dielectric mixing models presented by Birchak et al. [1974] at $T_g = 300K$ and $1.4GHz$ as a function of the volumetric moisture content.

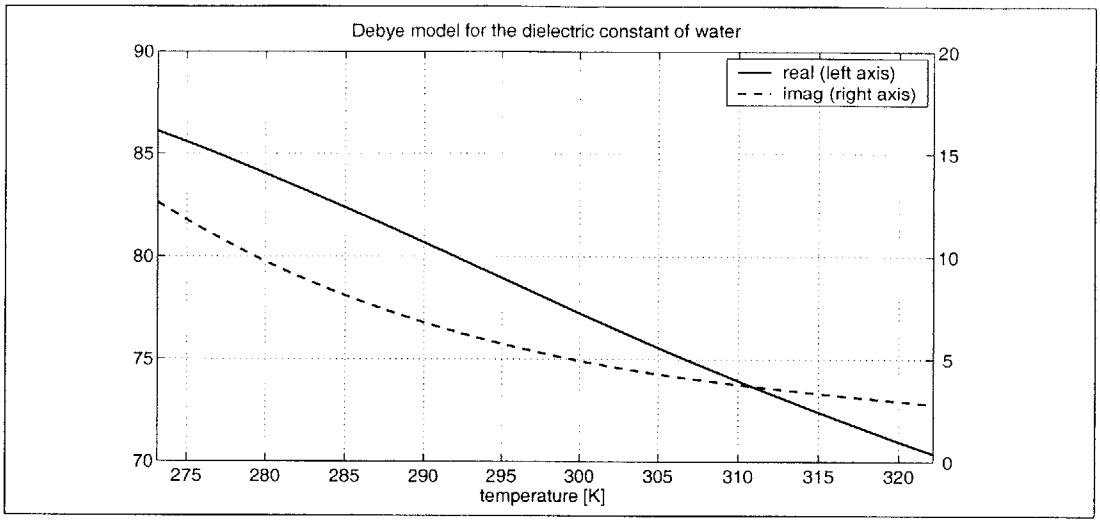


Figure 3.5: Dielectric constant of water $k_w$ at frequency $\nu_r = 1.4GHz$ as a function of temperature $T$ $[K]$.

The last equation shows the microwave emissivity for incidence from nadir ($\phi_r = 0$). For microwave frequencies and look-angles $\phi_r \approx 10°$ the nadir incidence formula is a very good approximation for both the horizontal and vertical polarization formulae. Note also that strictly speaking the emissivities are zero if the look-angle $\phi_r$ is larger than the critical angle at which total reflection occurs. We do not consider this case and assume that $\phi_r$ is small enough.

Figure 3.6 shows the microwave emissivities for the five soil types discussed by Dobson et al. [1985] together with the emissivities derived from two dielectric mixing models presented by Birchak et al. [1974], which we will not discuss here.

**Roughness Effects** The previous discussion assumed the soil surface to be smooth, but this is rarely the case in nature. Roughness effects can be taken into account with a simple one-parameter model developed by Choudhury et al. [1979]. The overall effect of roughness is to increase the emissivity, that is to decrease the reflectivity. In the model, the reflectivity is modified by an exponential term including the roughness parameter $\beta^{\varepsilon_g}$ $[-]$

$$\varepsilon_{gp} = 1 - (1 - \varepsilon_{gp}^{\text{smooth}}) \exp(-\beta^{\varepsilon_g} \cos^2 \phi_r) \tag{3.65}$$

Using (3.62), the emissivity for incidence from nadir ($\phi_r = 0$) is

$$\varepsilon_{gn} = 1 - \left| \frac{\sqrt{k_g} - 1}{\sqrt{k_g} + 1} \right|^2 \exp(-\beta^{\varepsilon_g}) \tag{3.66}$$

### Effective Temperature

The effective temperature generally depends on the soil temperature and moisture profiles, where the moisture profile enters primarily through the dielectric constant. For our objectives, there are primarily two choices of models: the gradient RT model and the grey body RT approximation, which are both non-coherent models.

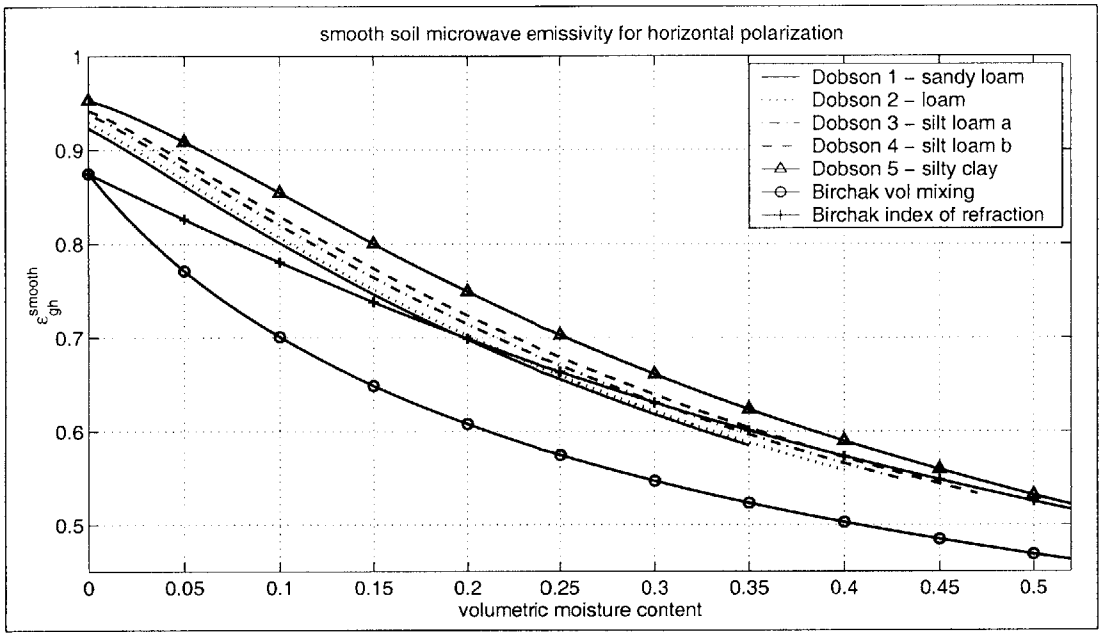


Figure 3.6: Microwave emissivities $\varepsilon_{gh}^{\text{smooth}}$ for horizontal polarization ($\phi_r = 10°$) for the five different soils in [Dobson et al., 1985] and for two models presented by Birchak et al. [1974] at $T_g = 300K$.

**Gradient RT Model** In the gradient RT model, the effective temperature is given by

$$T_g^{\text{eff}} = T(z_{\text{grad}}) - \frac{(1 - c_{\text{grad}})\cos\phi_g}{\alpha_e} \left.\frac{dT}{dz}\right|_{z_{\text{grad}}} \tag{3.67}$$

where

$$\alpha_e = \frac{4\pi\nu_r}{c_{\text{light}}} \text{Im}\sqrt{k_g}$$

is the attenuation coefficient and

$$z_{\text{grad}} = -\frac{c_{\text{grad}}\cos\phi_g}{\alpha_e}$$

is an effective depth for the gradient RT model. The in-soil propagation angle $\phi_g$ is obtained from Snellius' law.

$$\phi_g = \arcsin(\sin\phi_r/\sqrt{k_g}) \tag{3.68}$$

The speed of light is $c_{\text{light}} = 3 \cdot 10^8 m/s$, and finally, $c_{\text{grad}}$ is a gradient RT parameter. Galantowicz et al. [1999] use $c_{\text{grad}} = 1.03$. Figure 3.7 shows the gradient RT effective depth $z_{\text{grad}}$ for the same soils and models as Figure 3.6.

**Grey Body RT Model** In a grey body approximation, one simply uses

$$T_g^{\text{eff}} = T(z_{\text{grey}}) \tag{3.69}$$

The parameter $z_{\text{grey}}$ determines the depth at which the temperature is evaluated. For L-band observations, Galantowicz et al. [1999] use a depth of $z_{\text{grey}} = 1.5cm$.

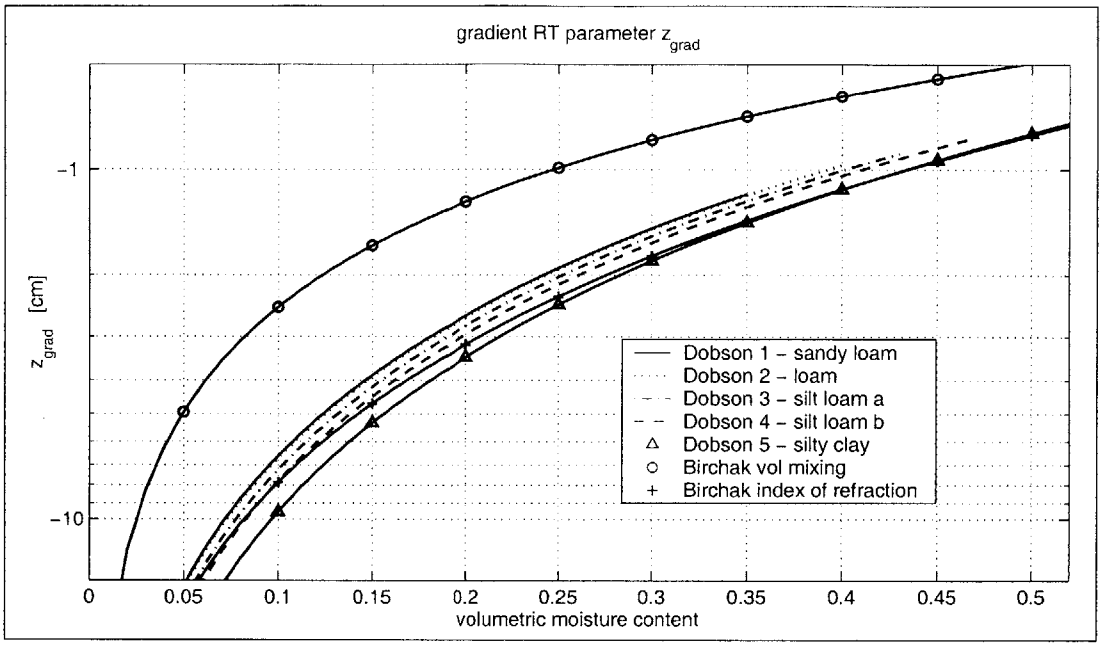


Figure 3.7: Gradient RT effective depth $z_{\text{grad}}$ for the five different soils in [Dobson et al., 1985] and for two models presented by Birchak et al. [1974].

**Choice of a Model** The appropriate choice of the RT model depends on the damping depth $d_g$ and the resulting discretization $\delta_g$ chosen in the force-restore approximation of the heat equation (Section 3.1.2). Recall that the damping depth $d_g$ describes the penetration depth of the diurnal heat wave, whereas $z_{\text{grey}}$ and $z_{\text{grad}}$ relate to the penetration depths of the microwaves.

According to Galantowicz et al. [1999], the gradient RT model is superior to the grey body approximation over the full diurnal cycle if the parameters are calibrated with daytime data. However, the force-restore model provides an approximate temperature gradient only at a depth $d_g \approx \delta_g$. Now recall from Figure 3.7 that for all soils and most moisture conditions, $z_{\text{grad}} \lesssim 5cm$. Compared to the typical damping depth and discretization $d_g \gtrsim \delta_g \approx 5cm$ (Section 3.1.2), we see that the use of the gradient RT model might be inappropriate. From the force-restore method, we cannot get the temperature gradient we need in the gradient RT approximation. In summary, we therefore use the grey body RT approximation with $T_g^{\text{eff}} = T_g$.

### 3.2.2 Microwave Emission in the Presence of a Canopy

If vegetation is present, the radiation observed by the instrument is given by

$$\begin{aligned} T_B =& (1 - f_c)\varepsilon_g T_g^{\text{eff}} \\ &+ f_c \left[\varepsilon_g T_g^{\text{eff}} \alpha_c + (1 - \alpha_c) T_c + (1 - \alpha_c) T_c (1 - \varepsilon_g) \alpha_c\right] \end{aligned} \tag{3.70}$$

The first term (proportional to $(1 - f_c)$) is the direct microwave emission from the bare or unshaded portion of the soil. The other three terms (proportional to $f_c$) are

- emission from the soil attenuated by the canopy,

| Vegetation type | $\delta_c \cos\phi_r$ [−] | $W_v$ $[kg/m^2]$ | $\beta^{\delta_c}$ $[m^2/kg]$ |
|---|---|---|---|
| Corn | - | - | 0.115 |
| Corn | 0.452 | 4.0 | 0.113 |
| Corn | 0.163 | 1.2 | 0.133 |
| Corn | 0.611 | 6.0 | 0.102 |
| Soybean | - | - | 0.086 |
| Soybean | 0.087 | 1.0 | 0.087 |
| Sorghum | 0.613 | 5.4 | 0.105 |
| Winter Rye | 0.080 | 0.7 | 0.114 |
| Short Grass | 0.093 | 0.3 | 0.300 |
| Tall Grass | 0.288 | 0.4 | 0.720 |
| Tall Grass | - | 0.5 | 0.150 |
| Rangeland, Pasture | - | - | 0.098 |

Table 3.1: L-band (wavelength $21cm$) parameters for the canopy transmissivity [Jackson and Schmugge, 1991].

- direct (upward) emission from the canopy (neglecting canopy reflection, that is emission equals absorption equals one-minus-transmission),
- ground surface reflection of emission from the canopy and attenuated by the canopy.

The model follows [Ulaby et al., 1986] (p. 1552), where we neglected the scattering albedo [Jackson and Schmugge, 1991]. We can rewrite (3.70) as

$$\begin{aligned} T_B &= [1 - f_c(1-\alpha_c)]\,\varepsilon_g T_g^{\text{eff}} + f_c T_c\,[(1-\alpha_c)(1+(1-\varepsilon_g)\alpha_c)] \\ &= \left\{[1 - f_c(1-\alpha_c)]\,T_g^{\text{eff}} - [f_c\alpha_c(1-\alpha_c)]\,T_c\right\}\varepsilon_g + f_c(1-\alpha_c^2)T_c \end{aligned} \tag{3.70a}$$

The canopy microwave attenuation $\alpha_c$ [−] depends on the canopy optical thickness $\delta_c$ [−].

$$\alpha_c = \exp(-\delta_c) \tag{3.71}$$

The optical thickness is parameterized as

$$\delta_c = \beta^{\delta_c} W_v / \cos\phi_r \tag{3.72}$$

where $W_v$ $[kg/m^2]$ is the vegetation water content, and $\phi_r$ is the look-angle from nadir. For the L-band ($\nu_r = 1.4GHz$), some of the parameters given by Jackson and Schmugge [1991] can be found in Table 3.1.

# Chapter 4

# Implementing Land Surface Data Assimilation

In this Chapter we describe in detail how the one-dimensional land-surface model of Chapter 3 is used in a fully four-dimensional (space and time) data assimilation algorithm. We also match the general estimation formulation of Chapter 2 with the variables of the hydrologic model.

A key assumption in the hydrologic model is to neglect lateral flow in the unsaturated zone, which is reasonable for terrain with moderate relief and on the spatial scales under consideration. The model domain thus breaks down into a collection of one-dimensional vertical cells or pixels (Figure 4.1). The horizontal resolution of the land surface model is determined by the availability of the micro-meteorologic data and of the land surface parameters, such as land cover and soil texture maps, and by computational resources. Since we aim to estimate the land surface states at this resolution, we refer to this scale as the *estimation pixels.*

From a simulation perspective, the estimation pixels are uncoupled but for larger-scale patterns of the inputs such as the meteorologic forcings as well as the land cover and soil texture classes. Lateral unsaturated moisture and heat fluxes between different pixels are neglected. When the model is incorporated into the data assimilation algorithm, however, the pixels are also coupled through the statistics of the uncertain inputs. The model error, representing for example errors in surface forcings, and the initial condition parameters in different pixels are assumed to be random fields which are correlated over time and space. The assimilation algorithm is therefore fully four-dimensional (space and time).

The horizontal resolution of the brightness images is solely determined by the sensor. We call this scale the *observation pixels.* In the foreseeable future, the brightness images will only be available at a resolution much lower than the resolution of the land surface model, that is each observation pixel generally contains several estimation pixels. We therefore need to develop the assimilation algorithm together with a general downscaling capability.

This Chapter is organized as follows. We first describe the model domain and the spatial discretization (Section 4.1). Next, we define the state vector for the land surface application (Section 4.2). The initial condition parameterization is explained in Section 4.3, the state equation is defined in Section 4.4, and the model error terms are identified in Section 4.5. Section 4.7 contains the definition of the assimilated data and the measurement operator. We also present the general downscaling scheme, which is based on an appropriate definition of the measurement operator. Next, Section 4.8 reviews the covariance models

of the uncertain inputs that we use in the assimilation. After a brief discussion of how the adjoint and the tangent-linear models are derived (Section 4.9), we conclude this Chapter with a description of the temporal discretization of the model equations (Section 4.10).

## 4.1 Model Domain and Spatial Discretization

The model domain for the estimation of soil moisture and temperature consists of $N_{ep} \equiv N_x \cdot N_y$ vertical columns or *estimation pixels* (Figure 4.1). Typical scales for the size of the estimation pixels are $1 \ldots 10km$. Each column has $N_z$ vertical nodes for the saturation. The vertical resolution is typically six layers ($N_z = 7$). Starting at the surface, the thickness increases downwards ($5cm$, $10cm$, $15cm$, $15cm$, $15cm$, and $30cm$). For the soil surface temperature there is only one node per column, likewise for the canopy temperature, the interception storage, the canopy air temperature, and the canopy vapor pressure.

In the unsaturated zone, horizontal moisture and heat fluxes over the scale of an estimation pixel are insignificant compared to vertical fluxes. We therefore neglect horizontal unsaturated fluxes, which makes the model very computationally efficient. Even though the columns are uncoupled as far as lateral moisture and energy exchange is concerned, they are in fact connected through horizontal correlations of the micro-meteorologic inputs and the soil and vegetation parameters. Meteorologic and geologic processes generally vary over a scale larger than the scale of an estimation pixel for soil moisture. Such correlations provide the large scale patterns observed by remote sensing instruments. Within the data assimilation algorithm, the estimation pixels are also coupled through the covariance structure of the uncertain inputs (Section 4.8).

## 4.2 State Vector

The state variables for all pixels are collected into the state vectors $X$ and $Y$ (4.1). The canopy temperature $T_c$, the vapor pressure $e_a$ and the temperature $T_a$ in the canopy air space, which are subject to algebraic equations, are concatenated into the state vector $X$. Likewise, the saturation $W_g$, the soil temperature $T_g$, and the interception water content $W_c$, which obey ordinary differential equations, are collected in the vector $Y$. The superscript in parentheses indicates the number of the estimation pixel.

$$X = \begin{bmatrix} T_c^{(1)}/T_0 \\ e_a^{(1)}/e_{20} \\ T_a^{(1)}/T_0 \\ \cdots\cdots \\ T_c^{(2)}/T_0 \\ e_a^{(2)}/e_{20} \\ T_a^{(2)}/T_0 \\ \cdots\cdots \\ \vdots \end{bmatrix} \qquad Y = \begin{bmatrix} \begin{pmatrix} W_{g1}^{(1)} \\ \vdots \\ W_{gN_z}^{(1)} \end{pmatrix} \\ T_g^{(1)}/T_0 \\ W_c^{(1)}/(W_c^{\max,(1)} + \gamma_s) \\ \cdots\cdots\cdots\cdots\cdots \\ \vdots \end{bmatrix} \tag{4.1}$$

All temperatures are scaled with the reference temperature $T_0 = 273.15K$. The vapor pressure $e_a$ in the canopy air space is scaled with the water vapor pressure $e_{20} = 23.4mb$ at $20°C$. The interception storage is dimensionless, but needs to be scaled to be of order

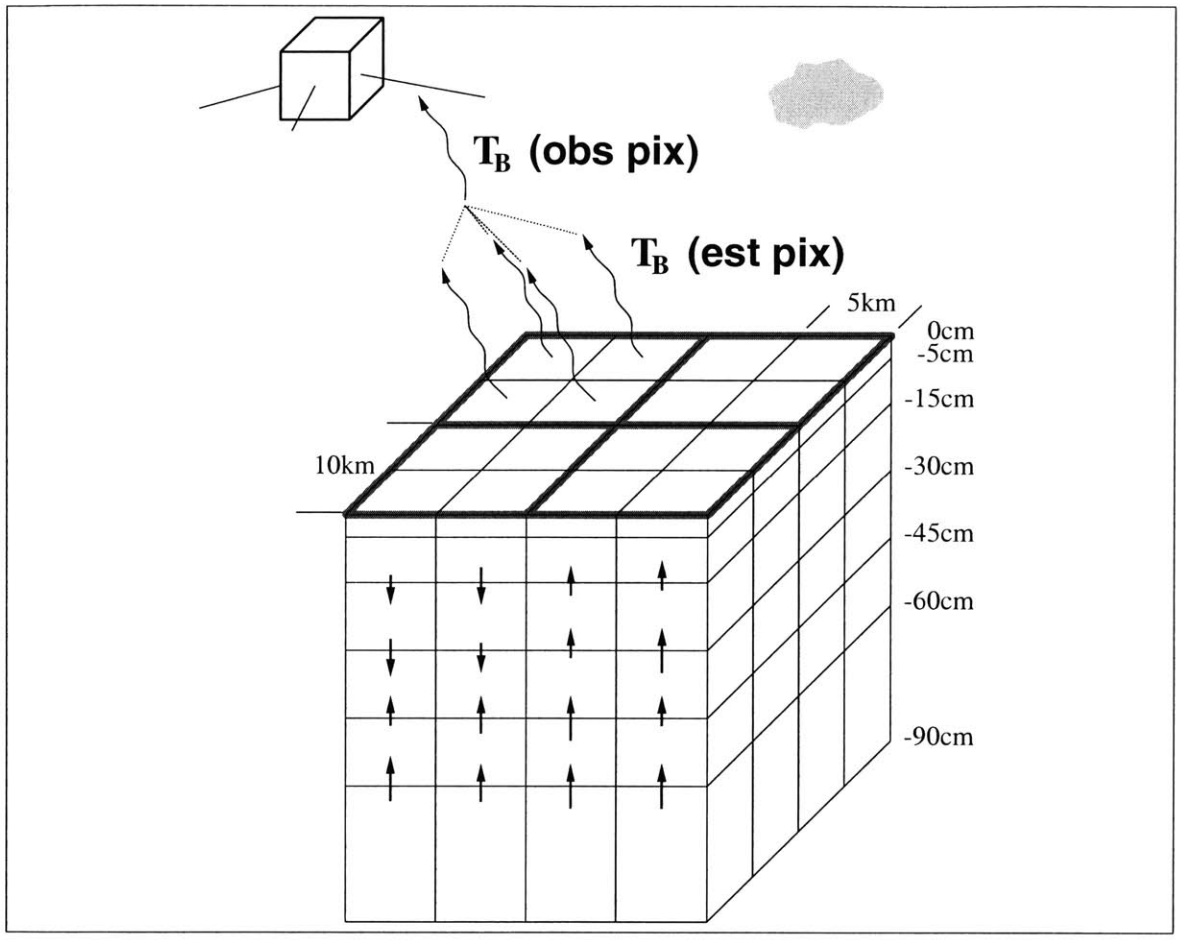


Figure 4.1: Schematic of the model grid with typical length scales. Note the huge ratio of horizontal to vertical scales. In the unsaturated zone, vertical fluxes dominate. Lateral unsaturated fluxes of moisture and energy are neglected (Section 4.1). Soil moisture estimates are obtained on the $5km$-scale of the estimation pixels (fine outline). Remotely sensed brightness measurements are available on the $10km$-scale of the observation pixels (thick outline). The downscaling is achieved by including in the measurement operator the computation of the average of the brightness temperature over an observation pixel. For L-band brightness temperatures, the arithmetic average is appropriate [Drusch et al., 1999]. This average is then subtracted from the data to provide the data misfit. By respecting the information contained in the fine-scale model inputs, the algorithm implicitly distributes the innovation from the measurement over all estimation pixels within a given observation pixel (Section 4.7).

one. This is done with $(W_c^{\max} + \gamma_s)$, where $W_c^{\max}$ depends on the pixel (3.16) and $\gamma_s$ is a constant scaling factor. The saturation $W_g$ is already dimensionless and on the order of one. It does not require further scaling.

The time-dependent vector $Y(t)$ is of length $N_Y = N_x \cdot N_y \cdot (N_z + 2)$. The length of $X(t)$ is $N_X = N_x \cdot N_y \cdot 3$. Altogether, there are $N_x \cdot N_y \cdot (N_z + 5)$ states at every time step. For a two-week period with $15min$ time steps, we need $N_t \approx 1300$ time steps. Using $N_x = 16$, $N_y = 32$, and $N_z = 7$, the total number of states is on the order of $10^7$.

## 4.3 Initial Condition

In this Section, we match the land-surface variables to the initial condition parameterization (2.1$a$).

### 4.3.1 Initial Saturation

The parameterization for the initial saturation $W_g(t = 0)$ maps the entire profile of each pixel onto a single uncertain parameter $\gamma_g$, the transformed total initial water storage, as follows. First, we decompose the initial saturation profile of each pixel into the total initial water storage in the column $W_g^{\text{store}}$ and the shape of the initial profile $W_g^{\text{shape}}$, which is a vector of length $N_z$.

$$W_{gi}(t = 0) = W_g^{\text{store}} \cdot W_{gi}^{\text{shape}} \tag{4.2}$$

The total initial water storage is defined as

$$W_g^{\text{store}} = \sum_{i=1}^{N_z} \bar{\Delta}_i \theta_{si} W_{gi}(t = 0) \tag{4.3}$$

For a definition of the thickness around the nodes $\bar{\Delta}$ see (3.6). The shape of the initial saturation can be derived from a hydrostatic profile or from the estimate at the final time of a previous data assimilation interval.

Since the initial storage is a bounded variable, the probability distribution of its error cannot conceivably be Gaussian. We therefore introduce a transform which maps the bounded variable onto the entire real axis. This transform is also an elegant way of telling the estimator that the storage is bounded, which greatly improves the estimator's convergence behavior. We map the initial storage $W_g^{\text{store}}$ with the transform

$$\gamma_g(W_g^{\text{store}}) = \tan\left[\pi\left(\frac{W_g^{\text{store}} - W_{g,\min}^{\text{store}}}{W_{g,\max}^{\text{store}} - W_{g,\min}^{\text{store}}} - 0.5\right)\right] \tag{4.4}$$

onto a transformed initial storage $\gamma_g$ varying over the entire real axis. The parameters $W_{g,\min}^{\text{store}}$ and $W_{g,\max}^{\text{store}}$ can be derived from (4.3). In practice, however, it is advisable to treat $W_{g,\min}^{\text{store}}$ and $W_{g,\max}^{\text{store}}$ as empirical parameters and to constrain the initial condition parameter $W_g^{\text{store}}$ well within its physically possible limits. Starting a simulation at very low or very high initial saturations can entail serious numerical problems in the forward model.

### 4.3.2 Initial Soil Temperature and Interception Storage

The initial condition for the scaled soil temperature $T_g/T_0$ can be used directly. There is no need for a nonlinear transform.

Similar to the saturation, the initial interception storage must be transformed onto the entire real axis. We use the transform

$$\gamma_c(W_c) = \tan\left[\frac{\pi}{\gamma_s'}\left(\frac{W_c(t=0)}{W_c^{\max}+\gamma_s} - 0.5\right)\right] \tag{4.5}$$

Both $\gamma_s$ and $\gamma_s'$ are constant scaling factors.

### 4.3.3 Initial Condition Parameterization

The complete initial condition parameterization reads

$$\beta = \begin{bmatrix} \gamma_g^{(1)} \\ T_g^{(1)}/T_0 \\ \gamma_c^{(1)} \\ \cdots\cdots \\ \gamma_g^{(2)} \\ T_g^{(2)}/T_0 \\ \gamma_c^{(2)} \\ \cdots\cdots \\ \vdots \end{bmatrix} \qquad Y_0(\beta) = \begin{bmatrix} \begin{pmatrix} W_{g1}^{(1)}(\beta_1) \\ \vdots \\ W_{gN_z}^{(1)}(\beta_1) \end{pmatrix} \\ T_g^{(1)}(\beta_2)/T_0 \\ W_c^{(1)}(\beta_3)/W_c^{\max,(1)} + \gamma_s) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \vdots \end{bmatrix} \tag{4.6}$$

The superscript in parentheses again indicates the number of the estimation pixel.

## 4.4 State Equation

Like the state vector, the state equation is a collection of equations from each pixel. The exact notation follows easily from the definition of the state vector (4.1).

The state equation for $X$ is given by a collection of mass and energy balance equations from each pixel. In particular, the energy balance for the canopy (3.13) yields $T_c$, the water mass balance equation for the canopy air space (3.19) yields $e_a$, and the energy balance for the canopy air space (3.18) yields $T_a$. For a one-pixel problem, we define

$$\begin{aligned} \phi_1 &= (R_{cs}^{\text{net}} + R_{cl}^{\text{net}} - LE_{ct} - LE_{ce} - H_c)/f_c \\ \phi_2 &= LE_a - LE_g - LE_{ct} - LE_{ce} \\ \phi_3 &= H_a - H_c - H_g \end{aligned} \tag{4.7}$$

The notation for multi-pixel problems follows immediately from (4.1). Note that the canopy energy balance in $\phi_1$ is divided by the fractional vegetation cover. This is important to get proper scaling of $\phi$ for varying vegetation densities. The mass and energy balance equations are implicit and are solved for the states with a Newton-Raphson method [Press et al., 1992].

The state equation for $Y$ is the collection of the discretized version of Richards' equation (3.1), the force-restore equation (3.7), and the interception equation (3.15) for each

estimation pixel. For a one-pixel problem, we define

$$\begin{aligned}
\varphi_i &= \left( \sum_{j=1}^{N_z} A_{ij}\psi_j + b_j \right) / \bar{\Delta}_i / \theta_{si} \qquad \text{for} \quad 1 \le i \le N_z \\
\varphi_{N_z+1} &= \Gamma_g \left[ \Gamma'_g G_g / C_g - (T_g - T_d) \right] / T_0 \\
\varphi_{N_z+2} &= (P_i - E_{cc}/\rho_w - D_c)/(W_c^{\max} + \gamma_s)
\end{aligned} \tag{4.8}$$

Again, the notation for multi-pixel problems follows immediately from (4.1). The matrix $A$ and the vector $b$ in the definition of $\varphi_i$, $(1 \le i \le N_z)$ follow from the discretization of Richards' equation according to the scheme by Celia et al. [1990] (see also [Simunek et al., 1997]).

$$A = \begin{pmatrix}
d_1 & e_1 & 0 & \dots & 0 \\
e_1 & d_2 & e_2 & \ddots & \vdots \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & e_{N_z-2} & d_{N_z-1} & e_{N_z-1} \\
0 & \dots & 0 & e_{N_z-1} & d_{N_z}
\end{pmatrix} \tag{4.9}$$

$$d = -\begin{bmatrix} 0 \\ e_1 \\ \vdots \\ e_{N_z-1} \end{bmatrix} - \begin{bmatrix} e_1 \\ \vdots \\ e_{N_z-1} \\ 0 \end{bmatrix} \tag{4.10}$$

$$e_i = (K_{ui} + K_{u,i+1})/(2\Delta_i) \qquad \text{for} \quad 1 \le i \le N_z - 1 \tag{4.11}$$

$$\begin{aligned}
b_1 &= -\bar{\Delta}_1 S_{g1} + (K_{u1} + K_{u2})/2 + q_b \\
b_i &= -\bar{\Delta}_i S_{gi} + (K_{u,i+1} - K_{u,i-1})/2 \qquad \text{for} \quad 2 \le i \le N_z - 1 \\
b_{N_z} &= -\bar{\Delta}_{N_z} S_{gN_z} - (K_{u,N_z-1} + K_{uN_z})/2 - q_t
\end{aligned} \tag{4.12}$$

## 4.5 Model Error

The model error or process noise accounts for errors in the model formulation, such as simplistic parameterizations or unresolved processes. In addition, it also accounts for errors in the micro-meteorologic forcings or incorrect parameter values.

In the general formulation of the estimation problem of Chapter 2, the model error is written as an additive term. Likewise, the land-surface fluxes enter the hydrologic model additively. It is therefore convenient to formulate the model error as uncertainties in the *physical* fluxes such as the latent and sensible heat fluxes. This has the added benefit that we can get a reasonable estimate of the magnitude of the model error. However, the formulation of the model error as uncertainties in the land-surface fluxes does not necessarily mean that the fluxes are the only source of uncertainty.

Formulating the model error as uncertainties in the physical fluxes determines the scaling factors $D_\nu$ and $D_\omega$ in the state equation (2.1). For a single pixel, we have $D_\nu = I_{N_X}$, where $I_{N_X}$ is the $N_X \times N_X$ identity matrix, and

$$D_\omega(Y) = \mathrm{diag}\left\{-1/\bar{\Delta}_1/\theta_{s1} \ldots -1/\bar{\Delta}_{N_z}/\theta_{sN_z} \quad \Gamma_g\Gamma_g'/C_g(Y)/T_0 \quad 1/(W_c^{\max}+\gamma_s)\right\} \quad (4.13)$$

where $\mathrm{diag}\{\cdot\}$ stands for a diagonal matrix with the argument of $\mathrm{diag}\{\cdot\}$ on the diagonal and zeros elsewhere.

The projection matrices $P_\nu$ and $P_\omega$ in the state equation (2.1) are made up of ones and zeros. They determine which fluxes are deemed uncertain. We usually restrict the model error to affect only the fluxes at the land-surface boundary. For bare soil, this means that only the flux into and out of the top node for soil moisture is uncertain. If there is vegetation, the uncertain top flux boundary condition for the saturation more directly affects all nodes that are connected to the surface by plants' roots. Owing to the rather shallow depth of the domain, all nodes are typically connected to the surface through the vegetation.

## 4.6 Uncertain Parameters

Land-surface hydrologic models are heavily parameterized, and few, if any, of the parameters are known accurately. The soil hydraulic parameters, for example, or the many parameters of the Radiative Transfer model are prime candidates for parameter estimation. However, estimating such parameters is no simple task. The parameter estimation problem is likely to be very ill-posed, and the benefit of estimating soil hydraulic parameters in an operational context is doubtful anyway.

At this point, we do not implement the parameter estimation in $\alpha$. Our assumption is that the parameters of the hydrologic model are already well calibrated when the model is used in the data assimilation algorithm. Furthermore, we account for model deficiencies with the model error terms. But note that we do treat the parameterized initial condition as uncertain (Section 4.3).

## 4.7 Assimilated Data, Measurement Equation, and Downscaling

In this study, the only data assimilated into the hydrologic model are brightness temperatures. We define the data vector as

$$Z = \begin{bmatrix} T_{B1}^{\mathrm{obs}}(t^1)/T_0 \\ \vdots \\ T_{BN_{op}}^{\mathrm{obs}}(t^1)/T_0 \\ \cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots \\ T_{B1}^{\mathrm{obs}}(t^{N_t^{\mathrm{obs}}})/T_0 \\ \vdots \\ T_{BN_{op}}^{\mathrm{obs}}(t^{N_t^{\mathrm{obs}}})/T_0 \end{bmatrix} \quad (4.14)$$

The first $N_{op}$ entries correspond to the image taken at the observation time $t^1$, the next $N_{op}$ entries correspond to the image taken at the observation time $t^2$, etc. With a total of $N_t^{\text{obs}}$ images and $N_{op}$ observation pixels within each image the number of measurements is $N_Z = N_{op} \cdot N_t^{\text{obs}}$. Note that the observed brightness temperatures are scaled with the reference temperature $T_0$. Within the general framework of Chapter 2, it is straightforward to assimilate other types of data, for example remotely sensed surface temperatures.

Soil moisture data assimilation can be formulated as a general downscaling problem. Satellite observations of L-band brightness temperatures will likely be available at a scale much larger than the scale of the micro-meteorologic, soil texture, and land cover data. Over the continental United States, the latter data are typically available on scales of $1km$ or less, whereas the projected resolution for a space-born passive L-band sensor is on the order of $50km$ in the near future and possibly $10km$ in ten years.

Clearly, it is desirable to get soil moisture estimates at a scale finer than the resolution of the brightness images. In our formulation, the fine-scale micro-meteorologic, soil texture, and land cover data constitute valuable information incorporated into the hydrologic model. We can use such fine-scale information to effectively downscale the brightness images. By defining the measurement operator appropriately, this downscaling is easily formulated within the assimilation algorithm. In essence, we run the hydrologic model on the finer scale and define the measurement operator such that it maps the fine-scale model predictions of brightness to the coarse scale of the remotely sensed brightness data.

As an example, Figure 4.1 shows the hydrologic model defined on a scale of $5km$ (estimation pixels). By using the Radiative Transfer model (3.70) described in Section 3.2, we can derive model predictions of the brightness temperature on the $5km$-scale. In contrast, brightness data are assumed to be available on a scale of $10km$ (observation pixel). In order to compute the data misfit term $Z - M[X, Y]$, the data assimilation algorithm needs model predictions of brightness temperature on the $10km$-scale of the measurements. This is easily achieved by averaging the fine-scale brightness predictions over the observation pixels. For L-band observations, the arithmetic average is appropriate [Drusch et al., 1999; Liou et al., 1998]. In mathematical terms, we this can be summarized as

$$M_k[X, Y] \equiv \frac{1}{N} \sum_{i_k} T_{Bi_k}(t_k)/T_0 \tag{4.15}$$

The $T_{Bi_k}(t_k)$ are the predicted brightness temperatures at time $t_k$ of the $N$ estimation pixels $i_k$ within the observation pixel corresponding to measurement $k$. We use the Radiative Transfer model (3.70) on the scale of the estimation pixels to obtain the $T_{Bi_k}$. Like the observations, the predicted brightness temperatures are scaled with the reference temperature $T_0$.

In other words, with the measurement $Z_k$ we only supply to the algorithm the average brightness temperature over the estimation pixels within the corresponding observation pixel. But we also supply micro-meteorologic, soil texture, land cover, and other model inputs on the finer scale of the estimation pixels, that is we tell the algorithm how the estimation pixels behave relative to each other. Respecting such information through the model physics, the algorithm is then able to distribute the coarse-scale brightness update onto the finer scale estimation pixels accordingly. Note again that the downscaling procedure is carried out implicitly within the algorithm. In summary, the downscaling scheme follows naturally from an appropriate definition of the (forward) measurement operator.

## 4.8 Error Covariances

In principle, the formulation of the variational data assimilation algorithm allows the use of any covariance for the model and parameter errors. In practice, however, one is limited by the computational effort required to store the covariance matrices and to solve the update equations (2.16) and (2.17). Depending on the choice of covariance, the calculation of the convolution integrals for the update is easily the most computationally demanding part of the assimilation algorithm.

The computational burden of solving the update equations is greatly reduced by assuming stationarity for the error covariances. In this case, we can speed up the numerical evaluation of the convolution integral significantly by using the Fast Fourier Transform [Press et al., 1992]. Another technique to efficiently solve the convolution integral relies on solving a corresponding differential equation [Bennett et al., 1997].

It is important to note that stationarity for the model error does *not* imply that the fluxes themselves obey stationarity. For example, the moisture flux into the soil is the sum of precipitation and evaporation. Precipitation in particular is certainly not stationary in time or in space. But the stationarity assumption is only made for the *error* in the fluxes and all other contributions to the model error. The fluxes themselves are not assumed to be stationary.

We choose exponential correlation functions to model all covariances of the uncertain inputs. The initial condition parameter covariance is

$$C_\xi^{(i)(j)} = \sigma_\xi^2 \exp\left(-\frac{\sqrt{(x^{(i)}-x^{(j)})^2+(y^{(i)}-y^{(j)})^2}}{l_\xi^{xy}}\right) \tag{4.16}$$

where $\xi$ represents the transformed initial soil moisture storage $\gamma_g$ (4.4), the (scaled) initial soil temperature, or the transformed initial interception storage $\gamma_c$ (4.5). The spatial coordinates of estimation pixel $(i)$ are denoted with $x^{(i)}$ and $y^{(i)}$, and $l_\xi^{xy}$ is the isotropic horizontal correlation length. All cross-correlations between physically different components of the uncertain parameters are assumed to vanish. For example, the error in the initial storage is always uncorrelated with the error in the initial soil temperature, regardless of the location.

For the model error, we assume that the covariance is separable in time and space. Again, physically different components of the model error are assumed to be uncorrelated. The same is assumed for saturation components belonging to different layers. For one component, the model error covariance can be written as

$$C_\xi^{(i)(j)}(t,t') = \sigma_\xi^2 \exp\left(-\frac{|t-t'|}{l_\xi^t}\right) \exp\left(-\frac{\sqrt{(x^{(i)}-x^{(j)})^2+(y^{(i)}-y^{(j)})^2}}{l_\xi^{xy}}\right) \tag{4.17}$$

where $\xi$ represents for example the component of $\nu$ corresponding to the model error in the canopy energy balance, or the component of $\omega$ corresponding to the error in the moisture flux into and out of the top node for the saturation.

For $\omega$, we also implemented a non-stationary model error covariance which is effectively white in time.

$$C_\xi^{(i)(j)}(t,t') = \sigma_\xi^2(t)\delta(t-t') \exp\left(-\frac{\sqrt{(x^{(i)}-x^{(j)})^2+(y^{(i)}-y^{(j)})^2}}{l_\xi^{xy}}\right) \tag{4.18}$$

The whiteness in time turns the convolution integral in the update equations into a simple multiplication. The time-dependent variance $\sigma_\xi^2(t)$ is meant to be zero most of the time, and rather big at select times, for example at times of heavy precipitation. In this way we can account for much larger errors during precipitation events without introducing too much overall uncertainty. The approach is physically motivated by the fact that quantitative precipitation measurements over large horizontal scales are usually quite uncertain (Section 7.2). Alternatively, one can choose the assimilation window in such a way that the initial condition is re-estimated after each major precipitation event (Section 7.1).

## 4.9 Tangent-linear and Adjoint Equations

Both the adjoint (backward) equation (2.15) and the tangent-linear (forward) model (2.18) use derivatives of the forward operators with respect to the states. It is straightforward, albeit laborious, to derive these adjoint operators from the definition of the states (4.1), the initial condition parameterization (4.6), the forward operators (4.7) and (4.8), the scaling factor $D_\omega(Y)$ (4.13), and the measurement operator (4.15). We consumed a lot of paper and pencils and accomplished this task manually. Alternatively, automatic adjoint compilers are available [Giering and Kaminski, 1998]. Since such tools have been developed only recently, they are not yet easy to use.

## 4.10 Temporal Discretization

So far, time has been treated as a continuum. We must, however, also discretize the time coordinate in order to implement the algorithm on a digital computer.

In the estimation algorithm, the nonlinearity of the forward model is resolved by iterating on the tangent-linear model (Section 2.2). This iteration is initialized with the prior model trajectory. Therefore we need a good initial prior model trajectory to keep the computational effort for the tangent-linear iteration at bay. For the initial prior fields, we implemented a nonlinear solver which uses a Picard iteration to handle the nonlinearity and a variable time step to deal with the intermittent nature of the forcings. The variable time step scheme follows closely the one implemented in HYDRUS, a sophisticated model for flow and transport in the unsaturated zone [Simunek et al., 1997]. The length of the time step is governed by the number of iterations needed in the Picard iteration for the most recent matric head. Moreover, the time step is reduced to seconds if there is precipitation.

In contrast, the tangent-linear model and the backward equation are implemented with a fixed time step. This time step is determined by the frequency of the micro-meteorologic inputs. Since the equations are linear and all the coefficient matrices are constant, we can solve the system directly using a time-implicit scheme. We choose the time-implicit scheme for its stability at longer time steps.

# Part II

# Assessing the Performance of the Assimilation Algorithm

# Chapter 5

# Synthetic Experiments

In the following Chapters, we present a series of synthetic (or twin) experiments to test our land surface data assimilation algorithm. Synthetic experiments are assimilation runs with synthetically generated parameter, model, and measurement errors, allowing us to compare the estimated and the prior fields to the (synthetic) true fields. Such experiments are ideally suited to evaluate the performance of the algorithm as all of the uncertain inputs are known. Furthermore, synthetic experiments allow us to assess the potential impact of the proposed L-band passive microwave satellite. All experiments are designed to mimic the conditions during the 1997 Southern Great Plains (SGP97) experiment in central Oklahoma (Section 5.3.1) which provides a realistic setup.

In this Chapter, we first explain briefly how synthetic experiments are conducted and explain in more detail why we choose synthetic experiments. Next, we describe the experiment area and the hydrology of the experiment period (Section 5.2). Finally, we describe in detail the data that we use as inputs to the hydrologic model (Section 5.3). The synthetic experiments themselves are described in Chapters 6 and 7.

## 5.1 Synthetic Experiments and Performance Assessment

### 5.1.1 Design of Synthetic Experiments

For a synthetic experiment, one realization of the uncertain inputs is generated with a random number generator and suitable mathematical methods to obtain the desired correlation structures. The model is then integrated with this realization of the uncertain inputs, and the output is the corresponding realization of the (synthetic) true state variables. Next, true values of the measured variables are obtained at specified times and locations by applying the measurement equation to the (synthetic) true states. Finally, synthetic measurements are generated by adding synthetically generated measurement error to the true fields of the measured variables.

In the assimilation procedure, the estimation algorithm is denied knowledge of the given (true) realization of uncertain inputs. Instead, it is only supplied with the statistics of the uncertain inputs, that is the prior mean and covariances. The algorithm is also supplied with the noisy synthetic observations, which contain some information about the particular realization of the true fields. From the prior statistics and the noisy synthetic observations, the assimilation algorithm produces a best estimate of the true fields.

If the assimilation algorithm is only supplied with the prior statistics of the uncertain

inputs but not with the synthetic observations, it will default to the so-called prior solution. This prior solution or prior state trajectory is derived from a (forward) model integration with all parameters and model errors set to their prior mean values. In particular, the prior states do not contain any information about the given realization of the true solution. When data are assimilated, the prior solution is the starting point for the algorithm to derive the best estimates of the true states (Section 2.2).

### 5.1.2 Why Synthetic Experiments?

Synthetic experiments are irreplaceable tools for assessing the performance of the assimilation algorithm. By design, the true solution and the statistics of the uncertain parameters are perfectly known. If we use the same hydrologic model and the same model inputs for the generation of the (synthetic) true fields and for the estimation, the assimilation algorithm is operating under ideal conditions. Non-linear issues aside, the estimate must have certain features that can be tested with the hypothesis tests described in Sections 2.3.6 and 2.4.1. Consequently, coding errors can easily be detected and the effects of nonlinearities can be assessed. Moreover, observing system characteristics can be evaluated and optimized. In Chapter 6, we describe a series of assimilation experiments under such ideal conditions.

When field data are assimilated, the hydrologic model will only be a crude approximation of the "model" that nature is using, and the error statistics we specify will likely be poor approximations of the true error characteristics. By specifying different parameters or statistics in the estimation process than have been used for the generation of the synthetic uncertain inputs, such nonideal situations can be investigated. With synthetic experiments, we can therefore investigate the sensitivity of the estimation algorithm to wrongly specified error statistics.

Similarly, inputs such as the soil hydraulic parameters will only be poorly known. By using different soil hydraulic parameters for the generation of the (synthetic) true fields and for the estimation process, we can emulate the realistic condition of assimilating data into a model that contains structural errors and bias. To compensate for the discrepancy, we will likely have to increase the model error variance in the assimilation algorithm. If we conduct synthetic experiments under such nonideal conditions, we can investigate the influence of the quality of the hydrologic model on the assimilation. Synthetic experiments allow for the evaluation of any number of such scenarios.

A final word of caution is in order. We believe that synthetic experiments are useful and irreplaceable tools for testing the assimilation algorithm and for investigating the sensitivity of the estimation process to various factors. Synthetic experiments cannot, however, replace the ultimate test of the algorithm with field observations. Such a field application must eventually be carried out.

## 5.2 Experiment Area and Period

The area of the synthetic experiments is located within the SGP97 experiment area. We choose the domain to coincide loosely with the swath of the airborne ESTAR brightness temperature measurements (Section 5.3.1). Figure 5.1 shows the horizontal grid of 16 by 32 estimation pixels together with the county lines and the meteorologic stations of the Oklahoma Mesonet. At a resolution of $5km \times 5km$, the experiment area covers a total of $12,800km^2$ in an $80km \times 160km$ rectangle. The estimation pixels are numbered from 1 to

512 starting in the southwestern corner and going north until the boundary of the domain is reached. The numbering continues at the bottom of the next column to the east of the first column and so on until the northeastern corner of the domain is reached.

The vertical resolution of the saturation is as shown in Figure 4.1 (Section 4.1). Each column has $N_z = 7$ vertical nodes for the saturation $W_g$ at $0cm$, $-5cm$, $-15cm$, $-30cm$, $-45cm$, $-60cm$, and $-90cm$. There is only one node per column for each of the other five states (canopy temperature $T_c$, vapor pressure $e_a$, temperature in the canopy air space $T_a$, soil temperature $T_g$, and interception water content $W_c$). Per pixel, we have therefore 12 states, and the state vector has a dimension of $512 \cdot 12 = 6144$ at every time step. (Note that the total number of scalar data in the reference experiments is also 6144 by coincidence.) The total number of 15 minute time steps in the synthetic experiment is 1280 (see below), and the total number of states is therefore $6144 \cdot 1280 \approx 10^7$.

All experiments extend over a two-week period from June 18, 1997 (day 169) to July 2, 1997 (day 183). Figure 5.2 shows time series of the area average micro-meteorologic inputs as derived from the Oklahoma Mesonet data (Section 5.3.2). The top panel shows the area average precipitation. Initially, there is a four day drydown across the entire domain. On day of year 174, significant rain is falling across the entire area. In contrast, the two major precipitation events of days 177 and 179 are concentrated in the northern half of the domain. The total rain over the two-week period measured at the stations in the northern half was between $3cm$ and $8cm$, whereas only $0.5cm$ to $2cm$ were observed at the southern stations. The area average cumulative precipitation over the two-week period is $2.8cm$.

The other panels of Figure 5.2 show the area average air temperature, wind speed, incoming shortwave radiation, and relative humidity. These fields are fairly homogeneous across the experiment area. Note that the area average quantities are only shown for illustration. In the synthetic experiments we use spatially distributed data. For details see Section 5.3.2.

## 5.3 Inputs to the Hydrologic Model

In this Section, we briefly describe the sources and the character of the various inputs to the hydrologic model that we use in the synthetic experiments.

### 5.3.1 The Southern Great Plains 1997 (SGP97) Hydrology Experiment

The Southern Great Plains 1997 (SGP97) Hydrology Experiment took place in the sub-humid environment in Oklahoma over the one-month period of June 18 – July 17, 1997 [Jackson, 1997; Jackson et al., 1999]. Its main objectives are

1. to examine the estimation of surface soil moisture and temperature using remote sensing at a hierarchy of scales,

2. to examine the feasibility of estimating vertical profiles of soil moisture and temperature by combining in situ data, remote sensing measurements at the surface, and modeling techniques,

3. and to evaluate the influence of soil moisture on the local surface energy budget and the influence of mesoscale variability in the surface energy budget on the development of the convective boundary layer.

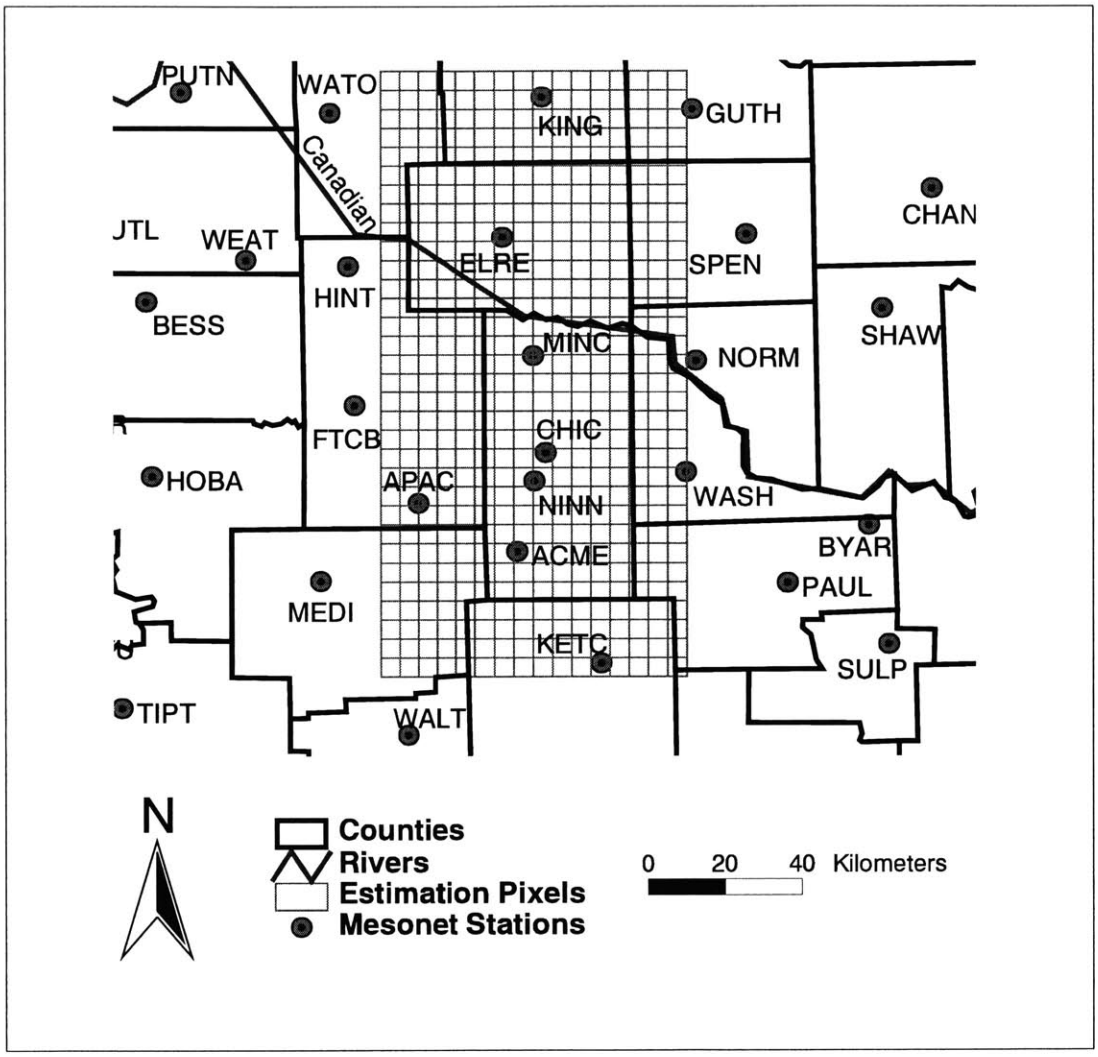


Figure 5.1: Area for the synthetic experiment. The horizontal grid of 16 by 32 pixels ($80km \times 160km$) is shown together with the locations of the Oklahoma Mesonet stations. The horizontal resolution is $5km \times 5km$. Oklahoma City is just to the north of Mesonet station NORM, on the eastern boundary of the domain.

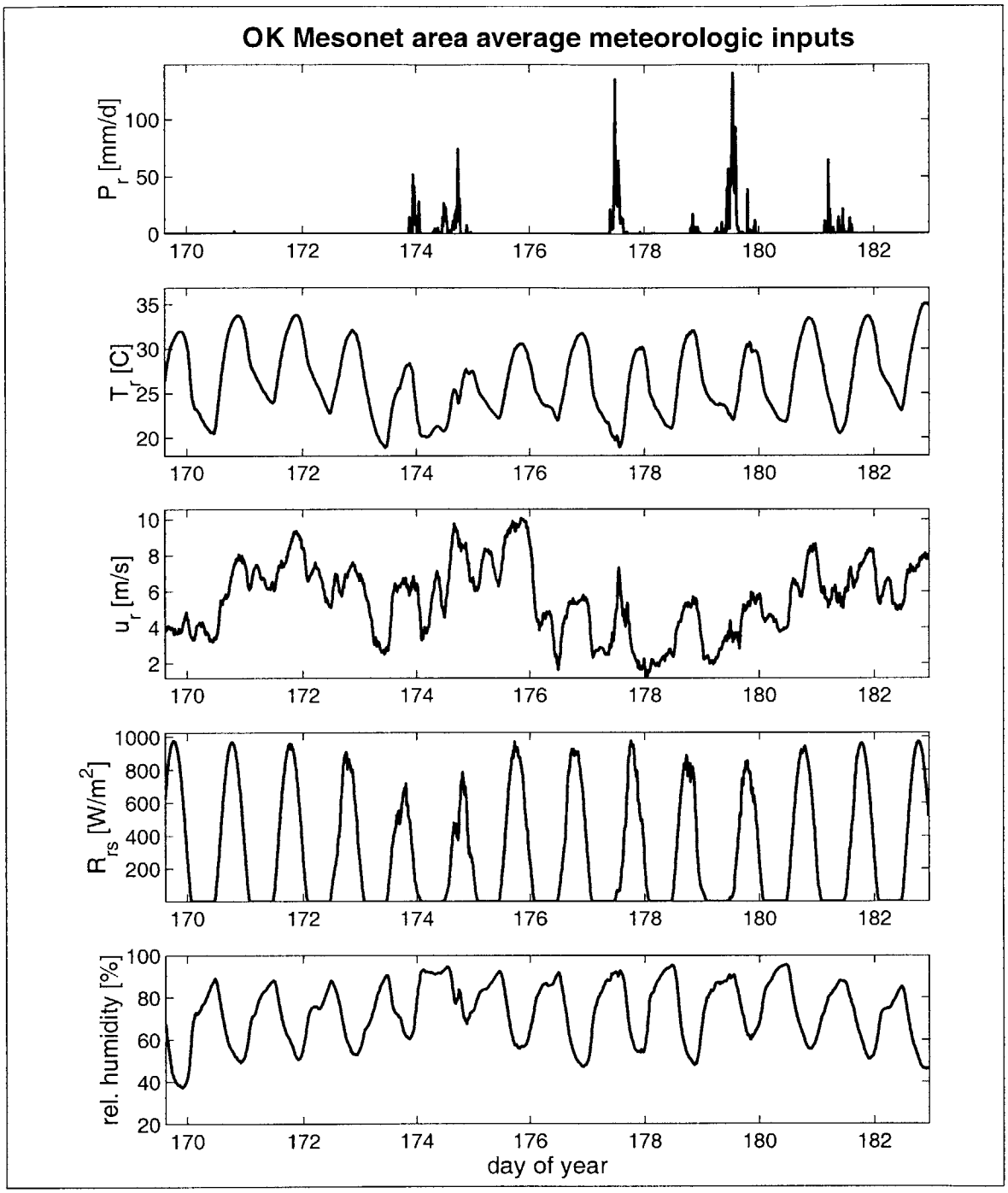


Figure 5.2: Area average micro-meteorologic inputs as derived from the Oklahoma Mesonet data. The panels show (from top to bottom) the area average precipitation, air temperature, wind speed, incoming shortwave radiation, and relative humidity. The two major precipitation events are concentrated in the northern half of the domain. The area average quantities are shown for illustration only. In the experiments, we use spatially distributed data.

To achieve these objectives, a host of ground-based and remotely sensed data were gathered. This includes measurements of ground-based soil moisture and soil properties, airborne boundary layer measurements, atmospheric soundings, vegetation parameters, airborne remote sensing measurements, ground-based remote sensing observations, satellite remote sensing measurements, surface flux data, and surface hydrometeorologic measurements. Moreover, micro-meteorologic data were provided by the Oklahoma Mesonet (Section 5.3.2), and soil texture classes were compiled by the Earth System Science Center (ESSC) at Pennsylvania State University (Section 5.3.3).

In the synthetic experiments of Chapters 6 and 7, we use the Oklahoma Mesonet data for the micro-meteorologic inputs, the ESSC database for the soil texture classification, and the land cover map to obtain the vegetation class for each pixel. In addition, for the Radiative Transfer model we use the maps of the vegetation water content $W_v$, the surface roughness parameter $\beta^{\varepsilon_g}$, and the vegetation parameter $\beta^{\delta_c}$.

The land cover data are derived from Landsat Thematic Mapper (TM) images and come in $30m$ resolution. The parameters of the Radiative Transfer model are available at a resolution of $800m$. According to their type, the data were aggregated to the resolution of the estimation grid by assigning the class of the block majority to the estimation pixel or by averaging over the estimation pixels. Figure 5.3 shows a map of the land cover classes aggregated to the $5km$ resolution. The projection is Universal Transverse Mercator (UTM), Zone 14, which is used throughout this thesis. The predominant land cover class is pasture. In the northern half, there are also large areas of wheat cultivation. The remaining land cover classes include forage, shrub, urban, and water. For the synthetic experiment, we neglect the latter two classes and use the pasture properties for the corresponding pixels as well as for the pixels for which there are no data available. This change does not affect the results of the synthetic experiments. When field data are assimilated, however, the water bodies and urban areas need to be treated separately.

#### Remotely Sensed Brightness Temperature from ESTAR

During the SGP97 field campaign, the Electronically Scanned and Thinned Array Radiometer (ESTAR) was flown daily on the NASA P3 aircraft. ESTAR successfully recorded sixteen images of L-band brightness temperatures ($1.4GHz$) on a swath of roughly $50km \times 200km$ at a resolution of $800m$. Together with the ground-based soil moisture [Famiglietti et al., 1999] and flux measurements, these invaluable data will allow a first test of the soil moisture assimilation algorithm on field data.

### 5.3.2 Micro-meteorologic Inputs: The Oklahoma Mesonet

The micro-meteorologic inputs to our land-surface model consist of six data types: precipitation $P_r$, incoming shortwave radiation $R_{rs}$, air temperature $T_r$, vapor pressure $e_r$, wind speed $u_r$, and depth average soil temperature $T_d$. All of the above are directly available or easily derived from the Oklahoma Mesonet data [Brock et al., 1995].

The Oklahoma Mesonet is a statewide network of 115 automated weather observing stations. It is unique in its dense spatial coverage, with an average distance between sites of about $31km$ [Basara et al., 1999]. Among other parameters, each site records at five minute intervals relative humidity at $1.5m$ height above the ground, air temperature at $1.5m$, average wind speed at $10m$, rainfall, barometric pressure, and solar radiation. In

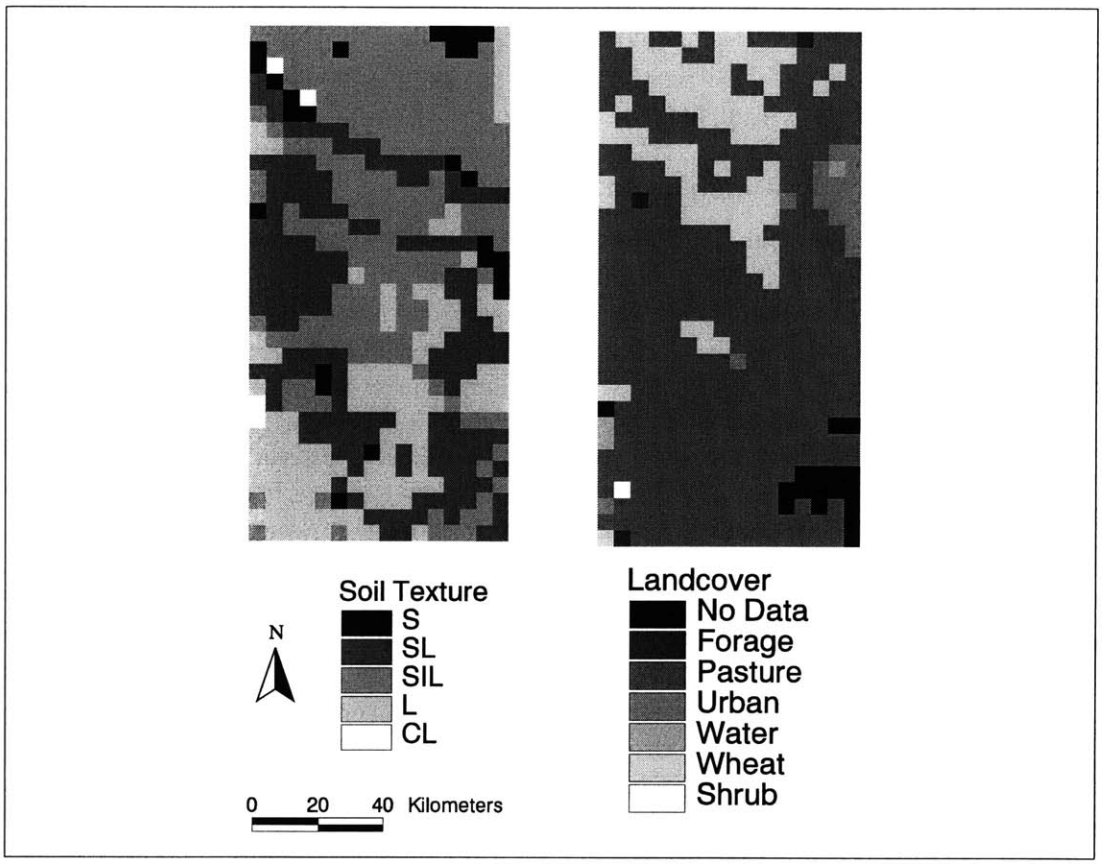


Figure 5.3: Surface soil texture and land cover classes for the synthetic experiment. The surface soil texture map has been derived from the ESSC data base (Section 5.3.3), and the land cover map has been obtained from SGP97 data (Section 5.3.1). The dominant soil texture classes are sand loam (SL), silt loam (SIL), and loam (L). There are also some areas where sand (S) and clay loam (CL) can be found. The dominant land cover class is pasture. In the northern half of the domain, there are also large areas of wheat. The urban, water, and no-data pixels are treated as pasture in the synthetic experiment. The horizontal resolution is $5km \times 5km$.

addition, bare and vegetated soil temperature at $10cm$ depth are measured at 15 minute intervals.

There are a number of supplemental paramcters available. Of most interest to us are observations of the bare and vegetated soil temperatures at $5cm$ and at $30cm$, which are taken at nearly half of the stations. Recently, many stations have been upgraded with heat dissipation sensors and time-domain reflectometry (TDR) probes to record soil moisture profiles [Basara et al., 1999]. Unfortunately, at the time of this writing there were still a few unresolved issues in the calibration of the moisture probes. In the future, such operational ground-based soil moisture data have great potential for assimilation into a land-surface model and for the validation of such algorithms.

#### Preprocessing the Micro-meteorologic Inputs

Before using the Mesonet data as inputs to our land-surface model, we apply a few preprocessing steps. The very few missing data points are filled in with neighboring observations or earlier observations at the same site. The observed relative humidity is converted into the vapor pressure using the measured air temperature. The deep soil temperature of the force-restore equation (3.7) is obtained from a monthly average of the air temperature.

The five minute interval for the micro-meteorologic inputs is attractive from a modeling point of view, as it allows to resolve the dynamics very finely. On the other hand, the model runs become very computationally demanding. We therefore average the five minute data to 15 minute inputs. Consequently, the basic time step of the tangent-linear and the adjoint models is 15 minutes. Note, however, that the initial prior state trajectory is obtained by solving the nonlinear land surface model with a variable time step (Section 4.10). During and after a rain event, the time step of the nonlinear model is typically reduced to a few seconds.

Even though the network of the Oklahoma Mesonet is very dense, the resolution is still too coarse for direct use in the hydrologic model. We interpolate the data with inverse square distance weights to the grid of estimation pixels. The weights for the seventeen stations of the synthetic experiment are shown in Figure 5.4.

It is important to note that we did not place emphasis on the optimal preprocessing of the micro-meteorologic data. Our focus is on the assimilation algorithm rather than on the calibration of the hydrologic model. Most importantly, the quality of the meteorologic inputs will be a lot poorer in future operational applications of land-surface hydrologic data assimilation. By including model error, the data assimilation algorithm is expressly designed to account for such deficiencies (Section 4.5).

### 5.3.3 Soil Properties

The database of the Earth System Science Center (ESSC) at Pennsylvania State University contains a variety of geographically referenced data sets. For the convenience of SGP97 researchers, the ESSC derived soil properties and land cover data from the State Soil Geographic (STATSGO) data set compiled by the Natural Resources Conservation Service (NRCS) of the U.S. Department of Agriculture. In our synthetic experiments, we use the dominant soil texture data, the sand and clay fraction data, and the bulk density data, all of which are available at $1km$ resolution for each of 11 standard soil layers.

The data sets come in the Universal Transverse Mercator (UTM), Zone 14 projection, which we use throughout this study. The aggregation of the data to the estimation grid

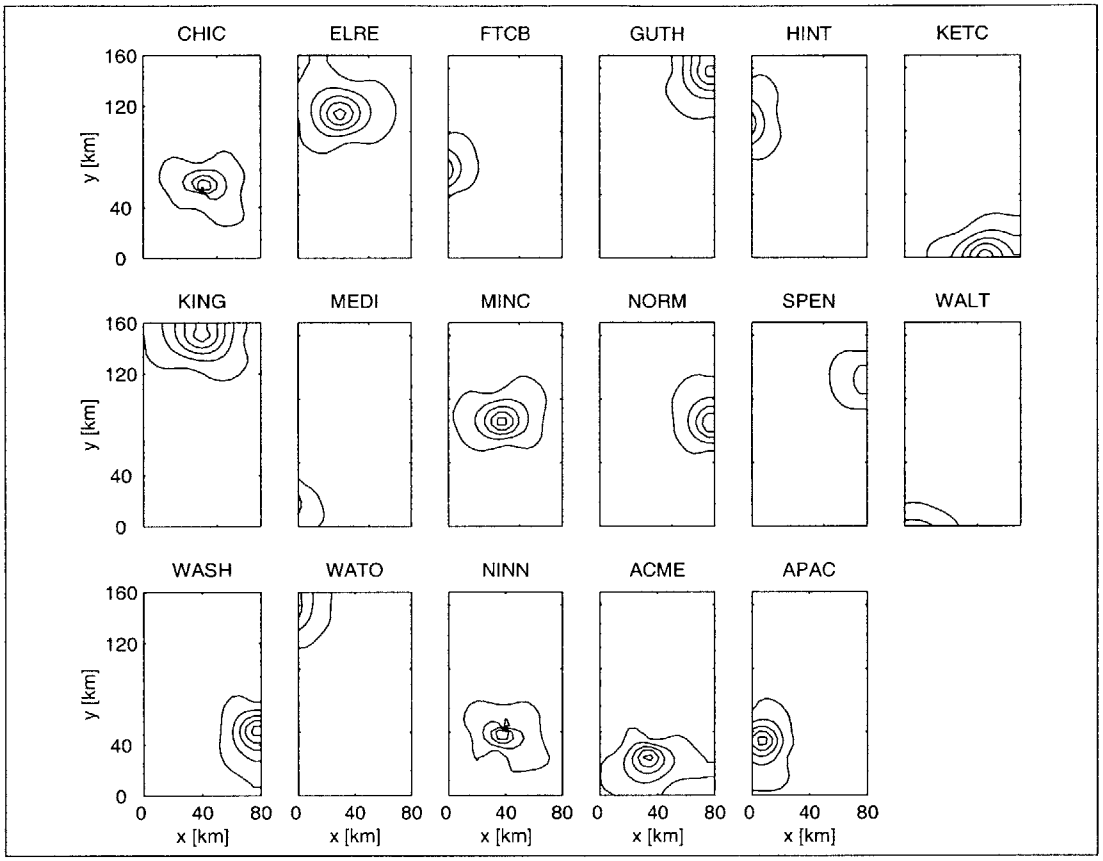


Figure 5.4: Inverse square distance weights for the interpolation of the micrometeorologic inputs. The 0.1, 0.3, 0.5, 0.7, and 0.9 contour lines are shown. The locations of the seventeen Oklahoma Mesonet stations used in the synthetic experiments of Chapters 6 and 7 are also shown in Figure 5.1. The average distance between stations is only $31km$.

was performed according to the data type by assigning the class of the block majority or by averaging. Figure 5.3 shows a map of the surface soil texture classes at $5km$ resolution. The dominant textures are sand loam (SL), silt loam (SIL), and loam (L). There are also some spots with sand (S) and clay loam (CL). Note the streaky pattern of both the soil texture and the land cover classes in the northwestern corner of the domain. This pattern can be recognized in the soil moisture fields during drydowns (Figure 6.3).

### 5.3.4 Other Data Sources

Not all of the inputs necessary to run the hydrologic model are provided by the SGP97 data set. Each soil texture and land cover class still needs to be translated into the corresponding numerical parameters. Vegetation parameters for the Little Washita catchment, which is located within the SGP97 experiment area, are tabulated in [Kustas and Jackson, 1999]. Inputs to land-surface schemes within General Circulation Models can be found in [Sellers, Los, Tucker, Justice, Dazlich, Collatz and Randall, 1996]. Similar data for the roughness length $z_0$, the fractional vegetation cover $f_c$, the vegetation albedo $a_c$, the minimum stomatal resistance $r_s^{\min}$, and the leaf and stem area indices LAI and SAI can be found in [Dickinson et al., 1993]. Typical values for the roughness length $z_0$ are also tabulated in [Dorman and Sellers, 1989]. Yang et al. [1998] give typical rooting depths. The soil hydraulic parameters are based on the data published by Clapp and Hornberger [1978]. Finally, the parameters $\beta^{r_c}$ and $\beta^{r_d}$ of the resistance network are given in [Sellers and Dorman, 1987].

The choice of the above parameters amounts to calibrating the hydrologic model, a task which we performed only in a rudimentary fashion. Our focus is on the performance of the assimilation algorithm and not on model calibration. For the synthetic experiments presented here the calibration matters little. However, for the future assimilation of field data such as the ESTAR brightness observations, the model must be reasonably well calibrated.

# Chapter 6

# Assimilation Under Ideal Conditions

In this Chapter, we present the results from a series of synthetic experiments under ideal assimilation conditions. By ideal conditions we mean that the same hydrologic model, the same model inputs, and the same error statistics are used for the generation of the (synthetic) true fields and for the assimilation. Experiments under more realistic nonideal conditions are presented in Chapter 7.

Even though the estimates derived in ideal experiments are unrealistically good, such experiments are valuable. First, assimilation under ideal conditions allows us to test whether the estimation algorithm is working properly. If conditions were not ideal, any coding error can be equally well blamed on the nonideal nature of the assimilation, and we would be left with an inconclusive result. Second, it is much easier to portray and to understand the basic features of the assimilation algorithm when conditions are ideal. Finally, we can gain valuable information on the sensitivity of the estimation procedure to various inputs.

In Sections 6.1 and 6.2 we present the estimation results of two reference experiments. The setup of the reference experiments is without downscaling, that is the observation pixels coincide with the estimation pixels, and synthetic brightness observations are available once daily. The focus of Reference Experiment I is on the estimation of the initial condition, whereas Reference Experiment II is designed to examine the estimation of the model error. In another series of experiments, we address two topics which are crucial to the design of an operational soil moisture data assimilation system. First, we investigate the quality of the estimates as the horizontal resolution of the brightness data decreases (Section 6.3). Finally, we address the impact of the satellite repeat cycle on soil moisture estimation (Section 6.4). The computational requirements for the assimilation are discussed in Chapter 8.

## 6.1 Reference Experiment I

### 6.1.1 Experiment Design

For a synthetic experiment, we generate synthetic fields for the uncertain inputs. This allows us to specify fields of land surface variables which serve as the "true" states. These true states are derived with a single simulation or forward integration of the land surface model. Starting from a synthetically generated initial condition field, the land surface model is driven with the measured micro-meteorologic inputs and synthetically generated process

noise. The resulting fields are then used as the true fields to which we can compare the estimates and the prior trajectories. See Section 5.1 for a more detailed description of synthetic experiments.

We use normal distribution functions when generating the (synthetic) true fields of the uncertain inputs, namely the uncertain initial condition parameter and the model error. As described in Section 4.5 we use exponential correlation functions for the temporal and spatial covariances. In this first experiment, Reference Experiment I, we choose a relatively large uncertainty for the initial condition of the saturation. The initial condition fields were generated around a spatially uniform prior top node saturation of 0.5. Alternatively, we could have spun up the model by integrating it for an extended period of time in order to derive possibly more realistic and heterogeneous prior initial conditions. But since we are conducting a synthetic experiment, the difference in the prior mean values matters little and the most important feature is the variance of the uncertain inputs.

The variance of the transformed saturation parameter $\gamma_g$ has been selected such that the initial condition distribution for the top node saturation is almost uniform between a minimum allowed saturation of 0.3 and a maximum allowed saturation of 0.75. For the given realization, the sample mean is 0.55 and the sample standard deviation is 0.1. A snapshot of the true initial top node saturation just two hours after the initial time is shown in Figure 6.3. The correlation length for the initial condition parameters is $50km$. The initial saturation profile is set to be hydrostatic, that is the gravity force is exactly balanced by a pressure gradient, that is a gradient in soil saturation.

Finally, we have fixed the initial condition for the soil temperature at the air temperature at the initial time, which is at 10:00h local time (CDT) in the morning. Trials have shown that varying or estimating the initial soil temperature does not alter the results. This is because the memory of the soil temperature is only on the order of a few hours. An incorrectly specified initial soil temperature therefore makes very little difference over the two-week assimilation window.

In contrast to the high uncertainty in the initial saturation, we choose a comparably small model error for this first experiment. We account for model errors in the top flux boundary conditions for the saturation and the soil temperature by choosing appropriate ones and zeros in the matrix $P_\omega$. We also account for model error in the canopy energy balance (3.13). The other energy balance equations are treated as perfect models and the matrix $P_\nu$ is determined accordingly. The standard deviation for the top moisture flux condition is the equivalent of $10W/m^2$ in latent heat flux. Likewise, the standard deviation of the forcing terms in the force-restore equation and in the canopy energy balance equation is set to $10W/m^2$. Note that the uncertain top flux boundary condition for the saturation affects all nodes that are connected to the surface by plants' roots. Owing to the rather shallow depth of the domain, all nodes are typically connected to the surface through the vegetation.

For the horizontal correlation of the model error we choose a small value of $1.5km$. This implies that only model errors in directly neighboring pixels are effectively correlated. In this way we can keep the computational effort for the evaluation of the convolution integral in the model error update small (Section 8.1.2). The correlation time for all model errors is 10 hours.

Figure 6.1 shows the area average true states for Reference Experiment I. The top panel shows the area average soil saturation at four out of the seven nodes. There is a general trend towards drier conditions across the domain over the two weeks of the experiment.

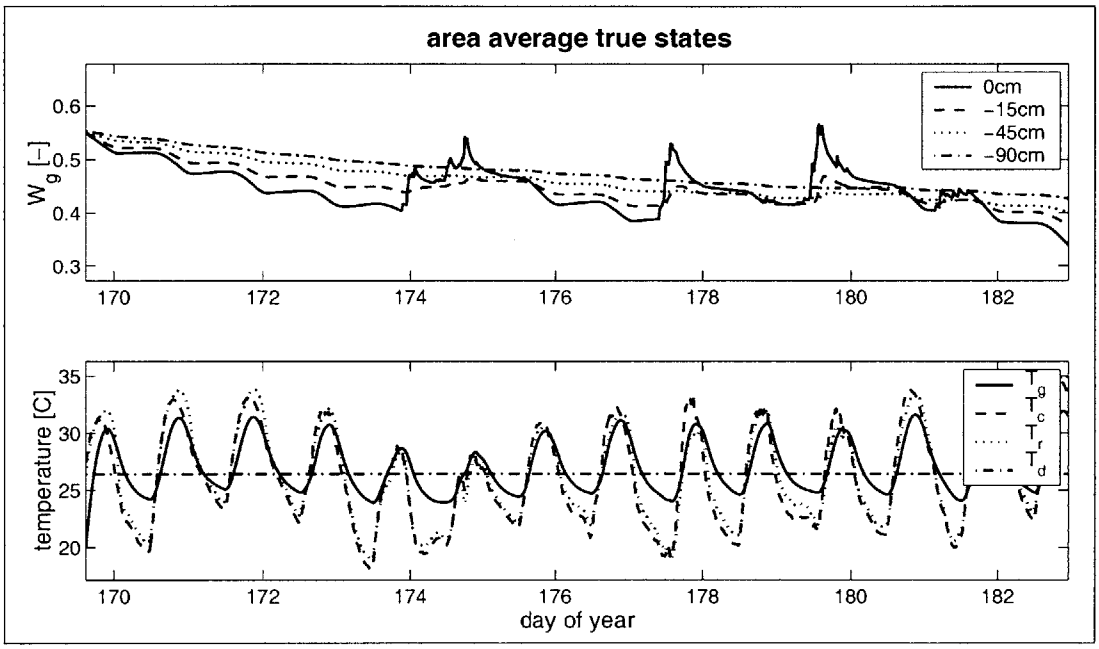


Figure 6.1: Area average true states for Reference Experiment I. The top panel shows the area average saturation at four out of seven layers. The precipitation events shown in Figure 5.2 are clearly discernible. The lower panel shows the area average soil temperature $T_g$, the canopy temperature $T_c$, and the deep soil temperature $T_d$ together with the observed area average air temperature $T_r$. The deep soil temperature is derived by temporally averaging the measured air temperature as part of the micro-meteorologic inputs (Section 5.3.2). See Figure 5.2 for the area average micro-meteorologic inputs.

This drying trend is related to the fact that the southern half of the domain experienced only very little rainfall (Section 5.2). The lower panel shows the area average true soil and canopy temperatures. For comparison, the observed area average air temperature and the deep soil temperature are also shown. The deep soil temperature is specified as the time average of the observed air temperature (Section 5.3.2). As expected, the amplitude of the canopy temperature is larger than the amplitude of the soil temperature. There is also a short time lag between the canopy temperature and the soil temperature. The temperature of the canopy air space (not shown) tracks the canopy temperature very closely.

For the synthetic experiment, we neglect the interception water storage by setting $\kappa^{W_c}$ and therefore the capacity of the interception water reservoir $W_c^{\max}$ to zero. The interception process is highly intermittent and interception happens on a time scale much shorter than the typical scale for the other state variables. Moreover, interception is almost by definition a threshold process and therefore not differentiable. For these reasons, the interception water is very difficult to estimate stably. Since the interception reservoir for pasture and crops is relatively small, neglecting the interception introduces only a small error. However, we feel that more research in this direction is necessary.

From the true states, synthetic L-band ($1.4GHz$) brightness data are derived on the scale of $5km$, which is also the scale of the estimation pixels. In this case the observation pixels coincide with the estimation pixels, and we refer to this scenario as a setup without downscaling. For twelve days, brightness images of the entire domain are generated once daily at 9:45h local time (CDT) in the morning. To each scalar true brightness temperature we add a spatially and temporally uncorrelated measurement error with a standard deviation of $5K$. Note that the absence of spatial correlation is not a constraint imposed by the algorithm. Uncorrelated measurement errors are attractive for a first application because they make coding easier and more transparent. In the future, spatially correlated measurement errors can be introduced rather straightforwardly. With 12 images and 512 brightness pixels per image, the total number of scalar data in the experiment is $12 \cdot 512 = 6144$.

Figure 6.2 summarizes the temporal setup of the reference experiments and other experiments which are discussed later. For the reference experiments, we use the entire two-week period as the assimilation window and measurements from all twelve observation times. Note that for some of the pixels some observations are taken during rain events. While it is raining, the water film covering the vegetation and the soil defies taking accurate measurements of the passive L-band microwave radiation. In an operational setting, a quality control routine must screen such faulty data. In this context, the rainfall interception model becomes important. For the synthetic experiments presented here, observations taken during rain events pose no particular problem.

### 6.1.2 Estimation of the True Fields

Figure 6.3 shows the true, the prior, and the estimated top node saturation for Reference Experiment I at six different times during the two-week period. In addition to the persistence of the initial condition, the evolution of the true top node saturation is governed by the precipitation inputs and the soil texture and land cover classes. This can best be observed in the prior fields because they start from a spatially uniform saturation. During the initial drydown (day 172.66) and during the second drydown (day 176.66), the soil texture and land cover properties lead to clearly visible streaky patterns which are most prominent in the northwestern corner of the experiment area. See Figure 5.3 for maps of the texture

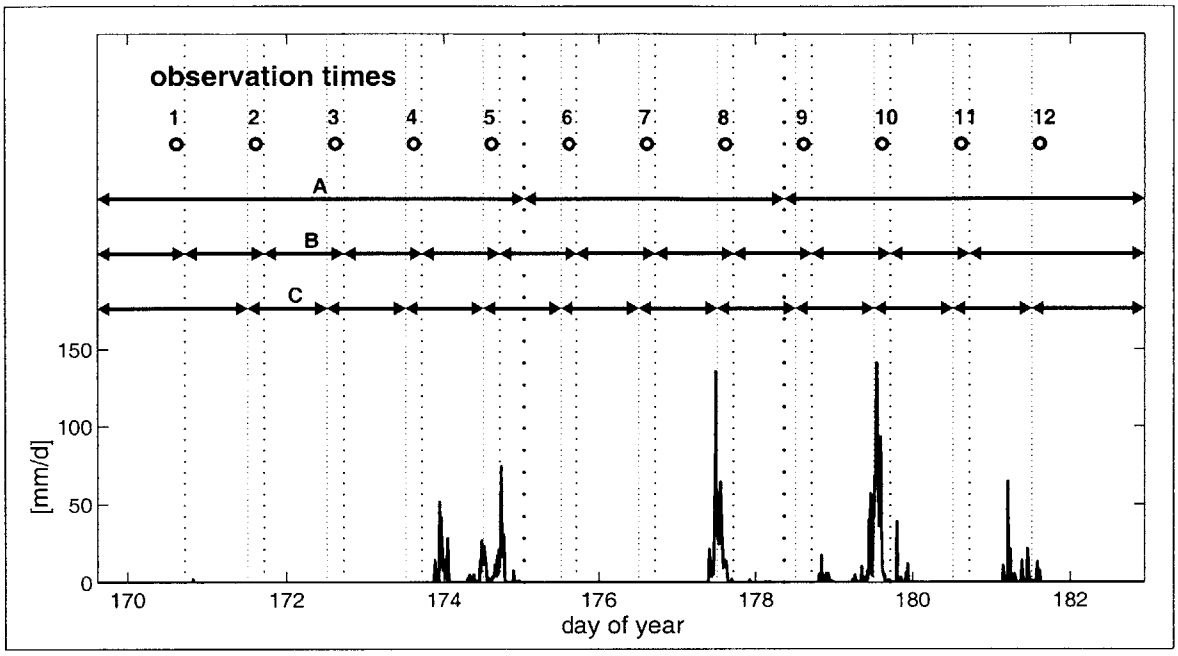


Figure 6.2: This Figure summarizes the temporal setup of the synthetic experiments. For reference, the area average precipitation time series is also shown. The experiment period covers two weeks from June 18, 1997 (day 169) to July 2, 1997 (day 183). There are up to twelve observation times at 9:45h local time (CDT) in the morning. In the reference experiments (Sections 6.1 and 6.2), all twelve observation times are used and there is just one assimilation (inversion) window which covers the entire two-week period. The same is true for the downscaling experiments (Section 6.3). The experiments on the brightness observation repeat cycle (Section 6.4) also use just one assimilation window. For the 3-day repeat cycle experiment, only the measurements at observation times 1, 4, 7, and 10 are assimilated. Likewise, only the measurements at observation times 1 and 7 are assimilated for the 6-day repeat cycle. Finally, the vertical dotted lines indicate the inversion time windows that are used in the three experiments with multiple assimilation windows (Section 7.1). In experiment A, we use three assimilation windows as indicated by the coarsely dotted vertical lines and the horizontal arrows. In experiments B and C, we use twelve assimilation windows for which the observation times are at the end and at the beginning of the window, respectively.

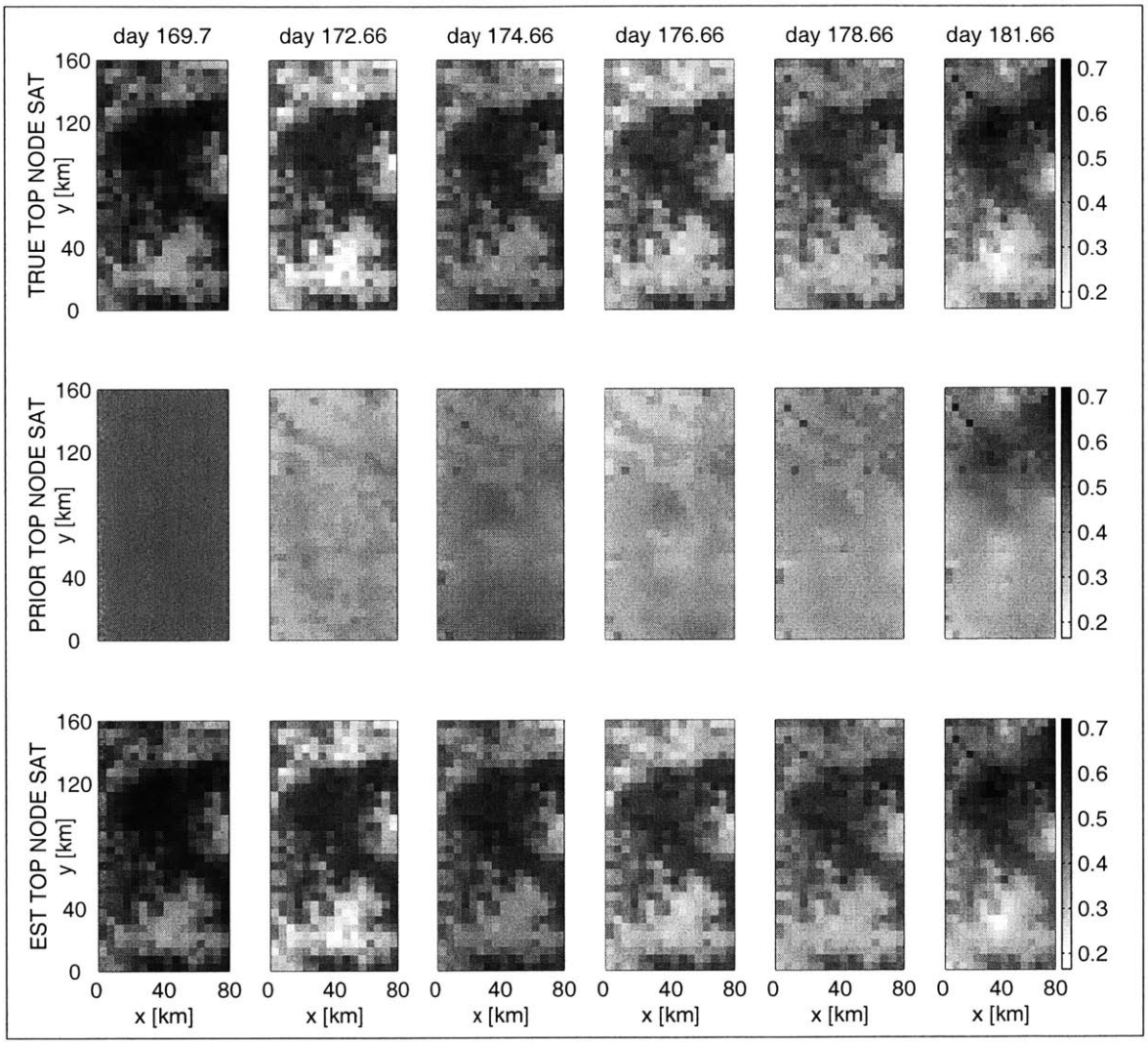


Figure 6.3: Top node saturation for Reference Experiment I. The first row shows the true top node saturation at six different times during the assimilation interval. In the second and third rows the prior and the estimate of the top node saturation, respectively, are depicted for the same six times.

and land cover classes. Owing to the very heterogeneous distribution of rain (Section 5.2), the northern half of the domain is obviously much wetter than the southern half at the end of the two-week period.

The value of the observations for retrieving the uncertain initial condition is obvious. Without assimilating the brightness data, our best guess for the top node saturation is the prior solution shown in the middle row of Figure 6.3, which is quite far from the truth. Once we assimilate the brightness data, we are able to accurately estimate the initial condition. With the initial condition being the dominant source of uncertainty in this experiment, we also get good estimates of the entire saturation time series.

Figure 6.4 shows time series of the area average errors in the top node saturation and in the soil temperature. The errors shown are root-mean-square errors (rmse) of the prior and the estimated fields with respect to the true fields. In the legend we also indicate the time average of the area average rmse's. Note that the soil moisture errors are in terms of the relative saturation, which varies between zero and one. In order to derive the errors in terms of volumetric moisture percent, the numbers have to be scaled with the porosity. For reference, the area average porosity is 0.46.

It is obvious that on average the estimated saturation fields are much closer to the truth than the prior fields. This again confirms that there is enough information in the brightness observations to infer the true initial condition for the saturation. Another interesting feature of Figure 6.4 is how the error of the estimate varies with time. In particular, the error decreases each time it rains. In other words, precipitation events tend to wipe out the memory of the system including errors.

Note that the errors in the soil temperature estimates are unrealistically small, because the experiment was not designed to test the soil temperature estimate thoroughly. In brief, the soil temperature initial condition was assumed perfectly known, and there was only a small uncertainty in the forcing of the soil temperature equation. These constraints lead to an overly optimistic estimate of the soil temperature if compared to field conditions.

It is important to point out why the area average prior saturation error in Figure 6.4 decreases with time. This is due to the specific setup of Reference Experiment I. In this particular setup, the initial condition becomes unimportant after some time because of the nonlinearity of the system. If, for instance, we start from too wet an initial saturation, then the evapotranspiration will also be higher than in the true case, therefore reducing the difference between the prior and the true fields with time. Conversely, if we start too dry, then evapotranspiration will be suppressed, and the prior will again edge towards the truth. Since in this experiment we use the same model and the same micro-meteorologic forcings for the generation of the (synthetic) true field and for the estimates, and since we only have a small model error, we know that the prior, the estimate, and the true fields must finally converge.

If this setup were to hold true in nature, we could indeed reasonably well estimate the saturation at the final time without a complex assimilation algorithm, in fact we could even do without the brightness data. In reality, however, we will of course never know the model physics and the forcings well enough to rely on this naive idea. Nature will always be much more complicated than our models, and we will need to estimate the initial condition over and over again. The important point for the interpretation of Reference Experiment I is therefore to look at how well we are doing in estimating the initial condition. Clearly, we can much improve the prior trajectories and we can estimate the initial condition to satisfying accuracy if we assimilate the brightness observations.

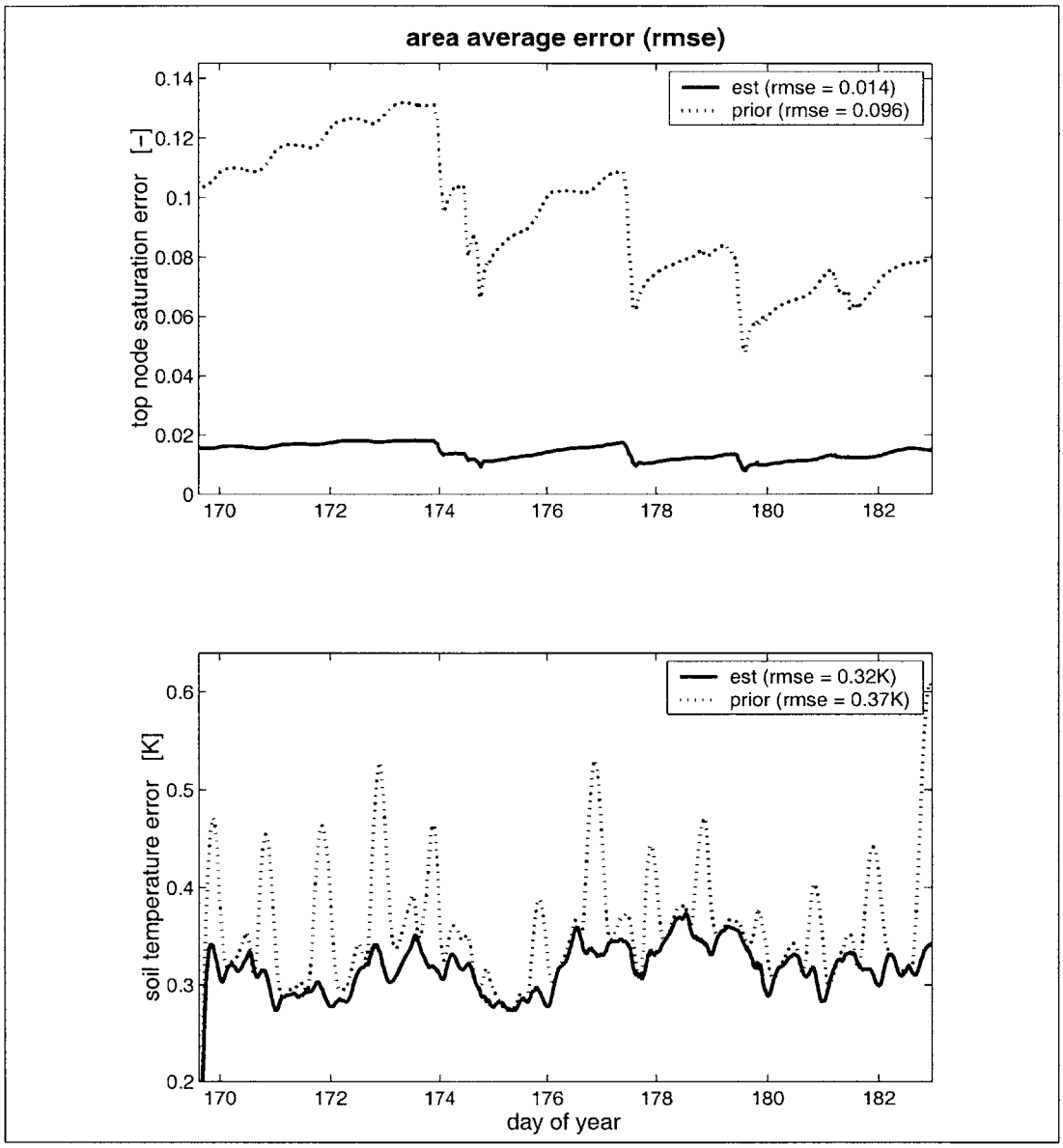


Figure 6.4: Area average errors for Reference Experiment I. The root-mean-square errors (rmse) of the prior and the estimated top node saturation and soil temperature with respect to the (synthetic) true fields are shown. In the legend we also indicate the temporal average of the area average rmse. Note that the soil moisture errors are in terms of saturation. To derive the errors in terms of volumetric moisture percent, the numbers have to be scaled with the porosity. The area average porosity is 0.46. Obviously, the assimilation greatly improves the errors in the top node saturation over the prior errors. Note that this experiment is not designed to test the soil temperature estimate. Also note that the decrease in the prior rmse is an artefact of the setup of Reference Experiment I (see text).

### 6.1.3 Profile Estimation and Validity of the Land Surface Model

For all subsurface nodes, we get the same excellent estimates as for the surface node (see Figure 6.14 for Reference Experiment II in Section 6.2.1). However, the fact that the profile estimates are this good does not in itself mean that the profile information is necessarily contained in the brightness observations. Indeed, recall that the shape of the initial condition profile is fixed as hydrostatic in both the generation of the true solution and in the assimilation. Moreover, in our ideal setup we use the same hydrologic model for the generation and for the estimation of the true fields. Finally, by using only a small model error in the top flux boundary condition, we assume an almost perfect model. For these reasons we naturally get excellent results for the profile estimates.

Whether subsurface information can in fact be retrieved from measurements related to the states in the top few centimeters is mostly determined by the accuracy and the physical realism of the land surface model. Without assimilating measurements that are directly related to the subsurface states, the only way information can be propagated to the deeper soil is via the hydrologic model. However, the time scales for the evolution of the profile are longer than the intervals that could reasonably be covered by a single assimilation window. Consequently, experiments to show the benefit of brightness assimilation for the estimation of the saturation profile depend on the successful development of an operational framework for soil moisture data assimilation.

Experiments to test the validity of the profile estimates could be verified with independent observations of the subsurface states. But the meaning of the large-scale (surface and subsurface) saturation is itself not obvious, considering that each pixel covers an area the size of tens of square kilometers. Certainly one cannot expect to go out to the field and verify the estimates with point measurements of the profile saturation. The large-scale saturation profile may rather be understood as an aggregate measure of how much water is stored at depth across the pixel. Such a vertically distributed and nonlinear reservoir may be used to determine how the land surface interacts with the atmosphere. In other words, the reservoir may tell us how much large-scale evapotranspiration and how much large-scale infiltration can be sustained by the land surface.

But suppose that we assimilate brightness data for a long time in an operational fashion. After enough time has elapsed, the posterior data residuals will show whether the subsurface dynamics of the hydrologic model and the prior statistics are consistent with the data. If this is the case, we have proof that the land surface model captures the multilayer subsurface reservoir and its feedback onto the surface states with reasonable accuracy. We can then interpret the saturation profiles in the manner suggested. If, on the other hand, the posterior data residuals remain biased or have structure even after months of assimilation, the model turns out not to be adequate in its description of land surface dynamics. In this case, we would have learned something about land surface dynamics. The assimilation then prompts us to go back and improve the hydrologic model accordingly.

### 6.1.4 Reduced Objective Function

Variational data assimilation is based on minimizing an objective function. Consequently, it is important to closely examine the value of the objective. Figure 6.5 shows the prior and the reduced objective as a function of the iteration on the tangent-linearization. In the first iteration, the objective function decreases significantly from its prior value of 31,749 to 6705. Unlike in a gradient search, the objective function need not decrease monotonically

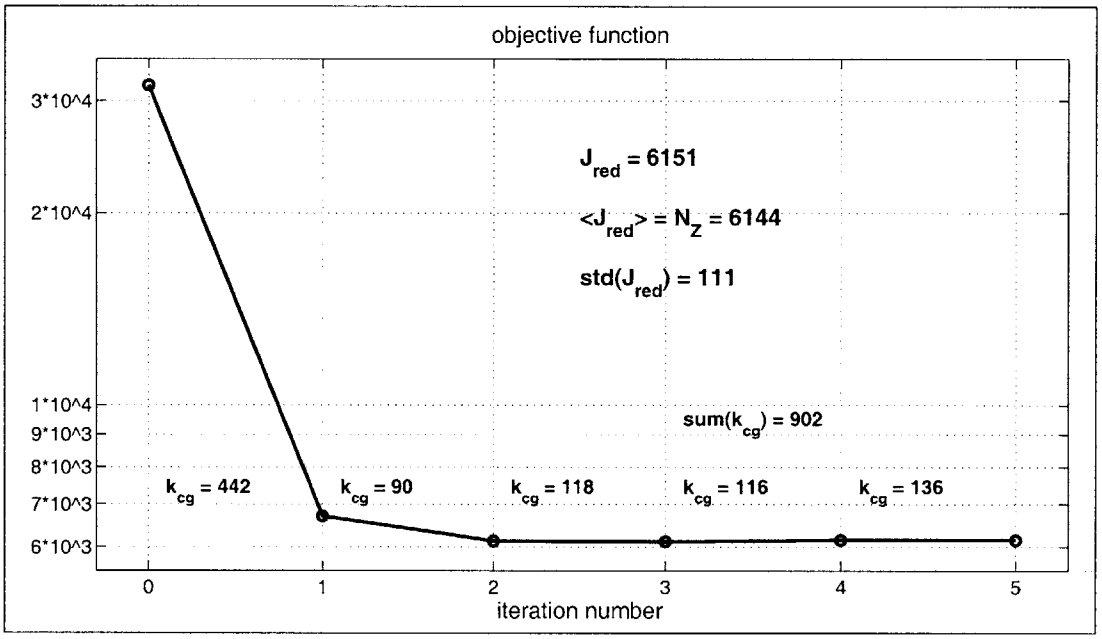


Figure 6.5: Objective function versus iteration number for Reference Experiment I. The reduced objective function after convergence is 6151. The number of data points is 6144, which is also the expected value of the reduced objective function. The standard deviation of the reduced objective function is 111. The values of $k_{cg}$ indicate the number of linear combinations of representer functions that needed to be evaluated during the conjugate gradient iteration of the indirect representer approach (Chapter 8).

with the number of iterations. In fact, upon closer inspection we see that the reduced objective equals 6132 after the second iteration, 6120 after the third iteration, 6150 after the fourth iteration, and finally 6151 after convergence. During the first iteration, the linearized trajectory is adjusted to a first-cut estimate using all observations. Had the problem been linear from the outset, we would have been done after this first iteration. During the remaining iterations, the assimilation algorithm mainly ensures that the estimate obeys the nonlinear state equation by making small corrections to the estimated trajectory. In other words, during the early iterations the estimates are not yet dynamically consistent, and the value of the objective function may decrease or grow while dynamic consistency is achieved during the iteration process.

Note that the final reduced objective function $\hat{J} = 6151$ compares well with the expected value of $\overline{J} = 6144$, which is equal to the number of data points (Section 2.3.6). The standard deviation of the objective function in this case is $\sigma_J = 111$. Since the errors were generated synthetically with perfectly known covariances, and since the true fields were generated with the same model that is used in the inversion, the reduced objective must by design indicate consistency of the prior assumptions with the (synthetic) data. This is an important validation step in the practical implementation of the algorithm.

### 6.1.5 Posterior Data Residuals

After the estimates have been derived, it is advisable to also take a close look at the posterior data residuals. The data residuals are the difference between the estimates of the measured quantities and the observations (Equation (2.32), Section 2.4). For an optimal estimate, the residuals must not show spatial or temporal patterns, that is they ought to be white in space and time. Moreover, their distribution should be close to normal with a mean of zero. If all of the above holds true, we can assume with reasonable confidence that the estimation process was optimal.

First, we examine the mean values (Figure 6.6). For all of the twelve individual residual brightness images that have been assimilated, we find a mean whose 95% confidence interval includes zero. But the mean for all residuals with a 95% confidence interval is $0.14 \pm 0.12K$, which does not include zero. Indeed, Figure 6.6 suggests that there is a slight positive bias. The fact that the sample mean of all residuals is not compatible at 5% significance level with the hypothesis of a zero mean could just be a spurious statistical fluctuation of the given realization. But the effect is more likely to have its origin in the nonlinearities of the hydrologic model. We will return to this point below when we test for normality. Note that the residuals of Reference Experiment II do not show this bias (Section 6.2).

Before further investigating the posterior data residuals, it is best to standardize them. Ideally, this is done with the posterior error covariance of the data residuals (2.37). However, computational limitations prohibit the exact calculation of this quantity. We therefore normalize with the sample standard deviation of the data residuals. The sample standard deviation of the residuals of the individual images varies between $4.4K$ and $5.1K$. The sample standard deviation for all residuals is $4.8K$. Recall that the measurement error standard deviation is $5K$.

Figure 6.7 shows the posterior data residuals for all twelve brightness images that have been assimilated in the experiment. The residuals of each image are standardized with the sample mean and standard deviation of the corresponding observation time. The residuals show no obvious spatial structure, which indicates that the estimation algorithm works

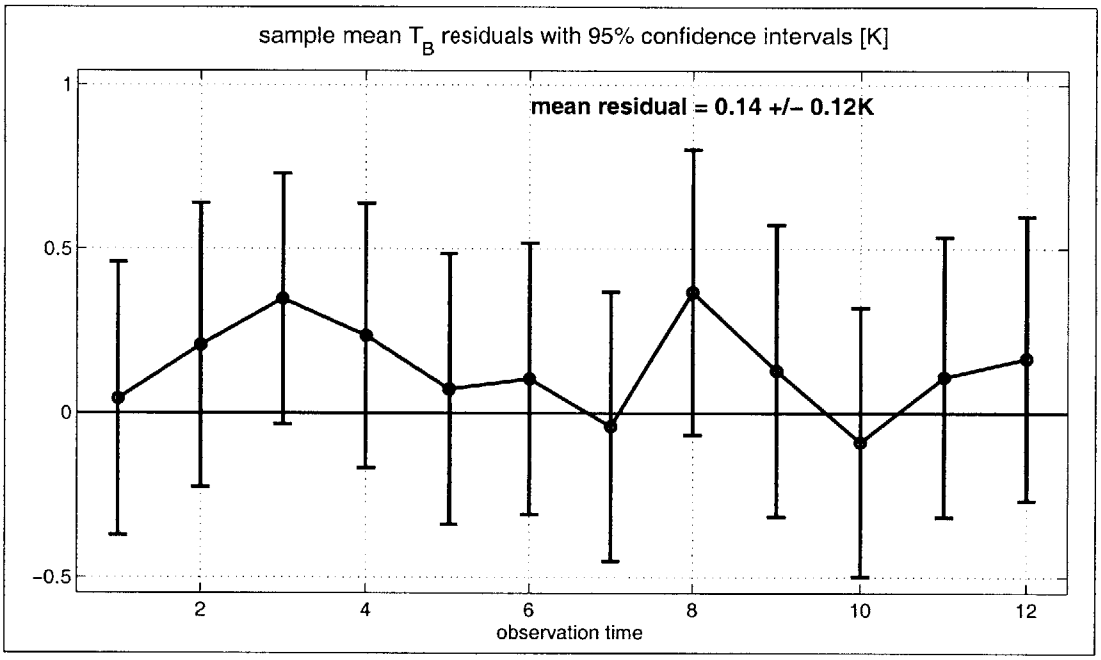


Figure 6.6: Sample mean values of the posterior data residuals for Reference Experiment I. The mean residuals for the twelve images that have been assimilated are shown together with the 95% confidence intervals. Even though the confidence interval for each of the images includes zero, the confidence interval for the sample mean of all images does not. This is likely due to the nonlinearities in the hydrologic model.

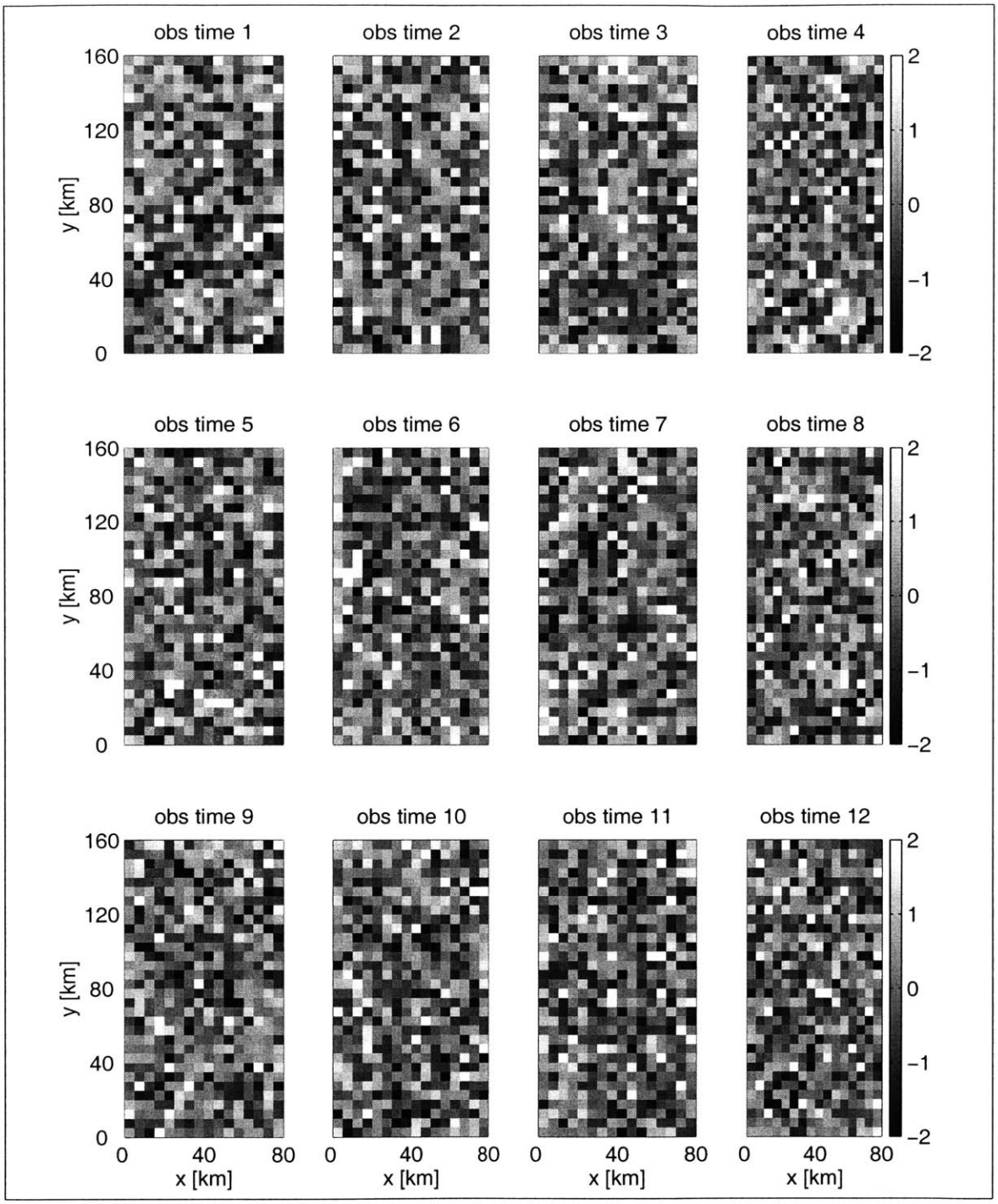


Figure 6.7: Standardized posterior data residuals for Reference Experiment I for the twelve brightness images that have been assimilated. The residuals of each image are standardized with the sample mean and standard deviation of the corresponding observation time. The residuals show no obvious spatial structure, which indicates that the estimation algorithm works optimally.

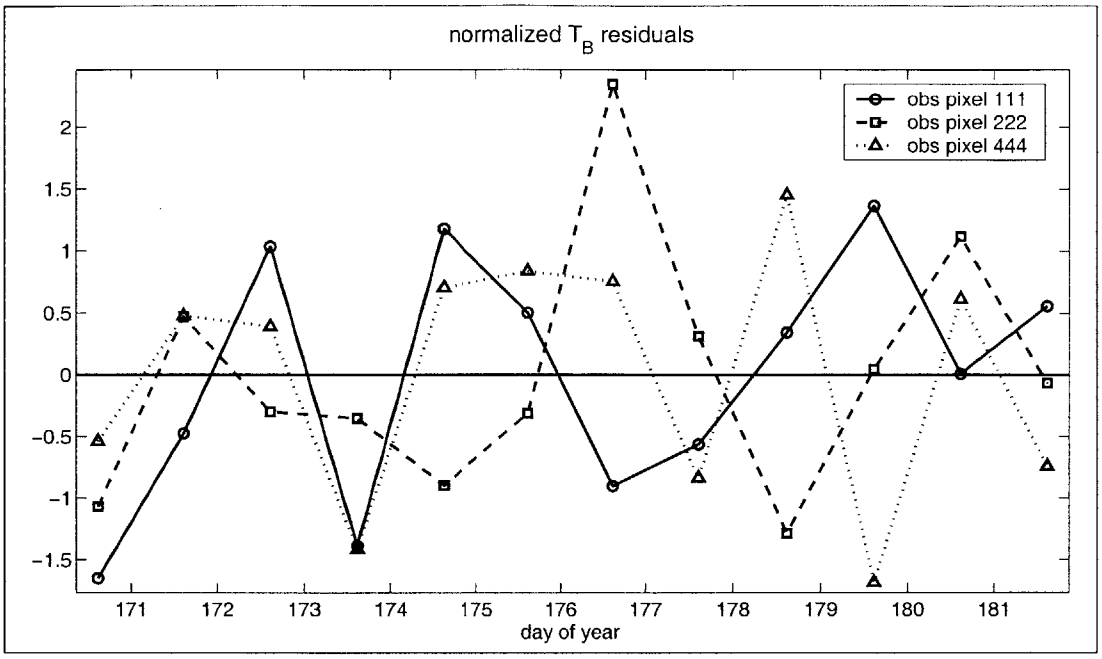


Figure 6.8: Standardized posterior data residuals for Reference Experiment I at three observation pixels from different areas within the domain. The residuals have no obvious temporal correlation, which indicates that the estimation algorithm works optimally.

optimally. Likewise, Figure 6.8 shows the time series of the data residuals for three pixels from different regions of the domain. The posterior data residuals also show no obvious temporal correlation. A test for whiteness using the autocorrelation function [Jenkins and Watts, 1968] confirms this result, but it must be noted that the short length of the time series defies an accurate statistical assessment.

Finally, Figure 6.9 shows the sample cumulative distribution functions (cdf) of the standardized posterior data residuals for two brightness images. For comparison, we also show the theoretical cumulative distribution function of the standard normal distribution. We can test the posterior residuals for normality with the Kolmogorov-Smirnov test [Benjamin and Cornell, 1970]. In brief, the Kolmogorov-Smirnov test compares the maximum difference between the sample cdf and the assumed theoretical cdf.

For two out of the twelve observation times, we must reject the hypothesis of a normal distribution of the data residuals at a 5% significance level. The upper panel of Figure 6.9 shows the sample cdf at the second observation time, for which we must reject the normality hypothesis. In contrast, the lower panel shows the sample cdf of the residuals at the tenth observation time. Here, the hypothesis of a normal distribution is not inconsistent with the data. Since the model and the statistics used for the generation of the (synthetic) true fields and for the estimation are identical in this experiment, and since the assimilation algorithm used is optimal in the linear case, we must attribute the deviation from normality to the nonlinear effects. Given the strong nonlinearities in the land surface model and in the measurement operator, it is not surprising that the posterior data residuals are not perfectly normally distributed.

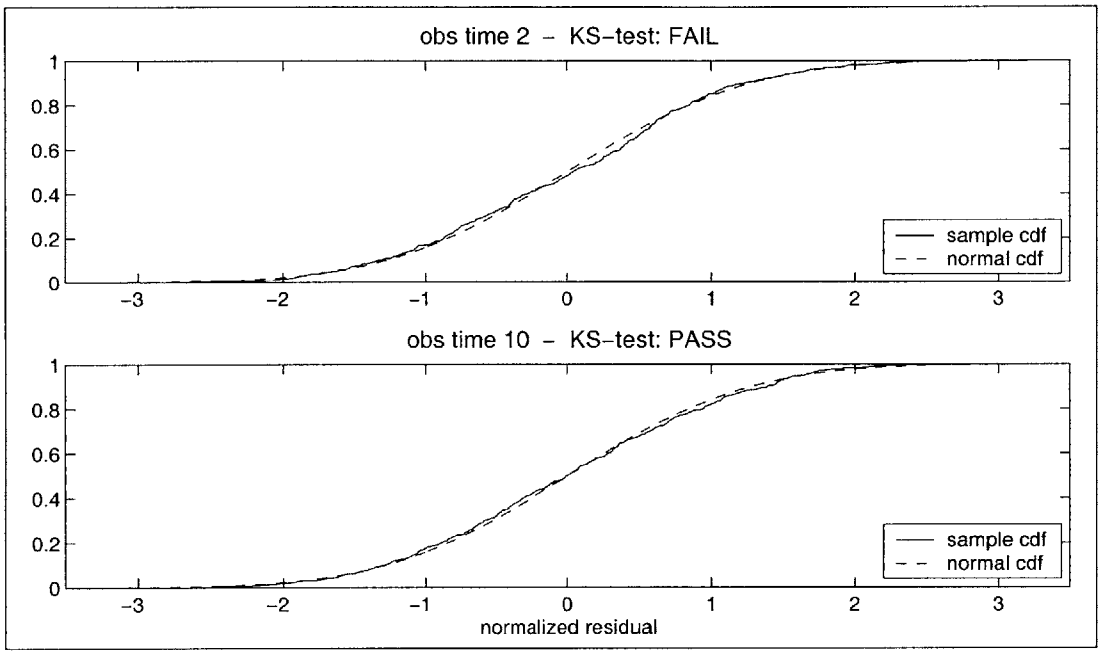


Figure 6.9: Sample cumulative distribution function (cdf) of the standardized posterior data residuals for Reference Experiment I for two of the twelve brightness images that have been assimilated (solid line). Also shown is the theoretical cumulative distribution function of the standard normal distribution (dashed line). For both images, the sample cdf is close to normal. However, the residuals at the second observation time do not pass the Kolmogorov-Smirnov test for normality at a 5% significance level. The data residuals for one other observation time out of twelve also fail this test. We attribute the deviation from normality to the nonlinearities in the hydrologic model and the measurement process.

### 6.1.6 Adjoint Variables

Figure 6.10 illustrates the temporal behavior of the adjoint variables. Time series of the adjoint saturation and the adjoint soil temperature are shown for a representative pixel. For comparison, we also plot the time series of the estimated brightness temperature together with the (synthetic) observations. Note that in this experiment the observation pixels and the estimation pixels coincide.

The forcing of the adjoint equation only consists of impulses at the observation times (2.15). These impulses are proportional to the misfits between the observed and the estimated brightness temperature. They are also proportional to the sensitivity derivative of the brightness temperature with respect to the corresponding state variable. Starting from a final condition of zero at the end of the assimilation interval (day 183), the adjoint equation is integrated backward in time. The adjoint therefore equals zero until, going backwards in time, we hit the last measurement time (day 181.6). At the last measurement time, the misfit between the observed and the estimated brightness is negative for this particular pixel. Since the soil saturation is inversely correlated with the brightness temperature, that is $\partial T_B / \partial W_g \leq 0$, the negative data misfit leads to a positive impulse for the adjoint saturation. The opposite is true for the adjoint soil temperature. Further integration backward in time causes the adjoint states to decay and grow exponentially according to the model physics while they are repeatedly forced with data misfit impulses at observation times.

In Figure 6.10 we can also observe how the brightness misfit forcing affects the saturation profile. Since the brightness temperature directly depends on the saturation in the top two nodes at $0cm$ and $-5cm$ via the top layer microwave emissivity and via the heat capacity, the misfit forcing impulse leads to instantaneous effects in the corresponding adjoint saturation components. In contrast, the lower components of the saturation do not directly influence the brightness temperature. Therefore, the adjoint saturation of the lower nodes only experience a delayed effect after the misfit forcing has been propagated downward through the (adjoint) model physics.

Finally, Figure 6.10 also illustrates the much longer memory of soil moisture compared to soil temperature. After each measurement time, the adjoint soil temperature decays (backwards) to zero within twelve hours, whereas the adjoint saturation does not reach zero before the misfit impulse at the next earlier measurement time comes in.

## 6.2 Reference Experiment II

To complement Reference Experiment I of Section 6.1, we now present another synthetic experiment using different error statistics for the uncertain parameters and a different realization, that is a different seed for the random number generator. The focus of Reference Experiment II is on the model error. Unless otherwise stated, the inputs for this experiment are the same as for Reference Experiment I. The initial condition is still uncertain, but with somewhat reduced variance compared to Reference Experiment I. For this experiment, the initial top node saturation varies between 0.3 and 0.65 (Figure 6.11). The sample mean is 0.52 and the standard deviation is 0.07. The correlation length for the initial saturation is still $50km$, and the initial saturation profiles are again hydrostatic.

The structure of the model error is the same as in Reference Experiment I, but the standard deviations are drastically increased. The standard deviation for the top moisture flux condition is now the equivalent of $50W/m^2$ in latent heat flux. Likewise, the standard

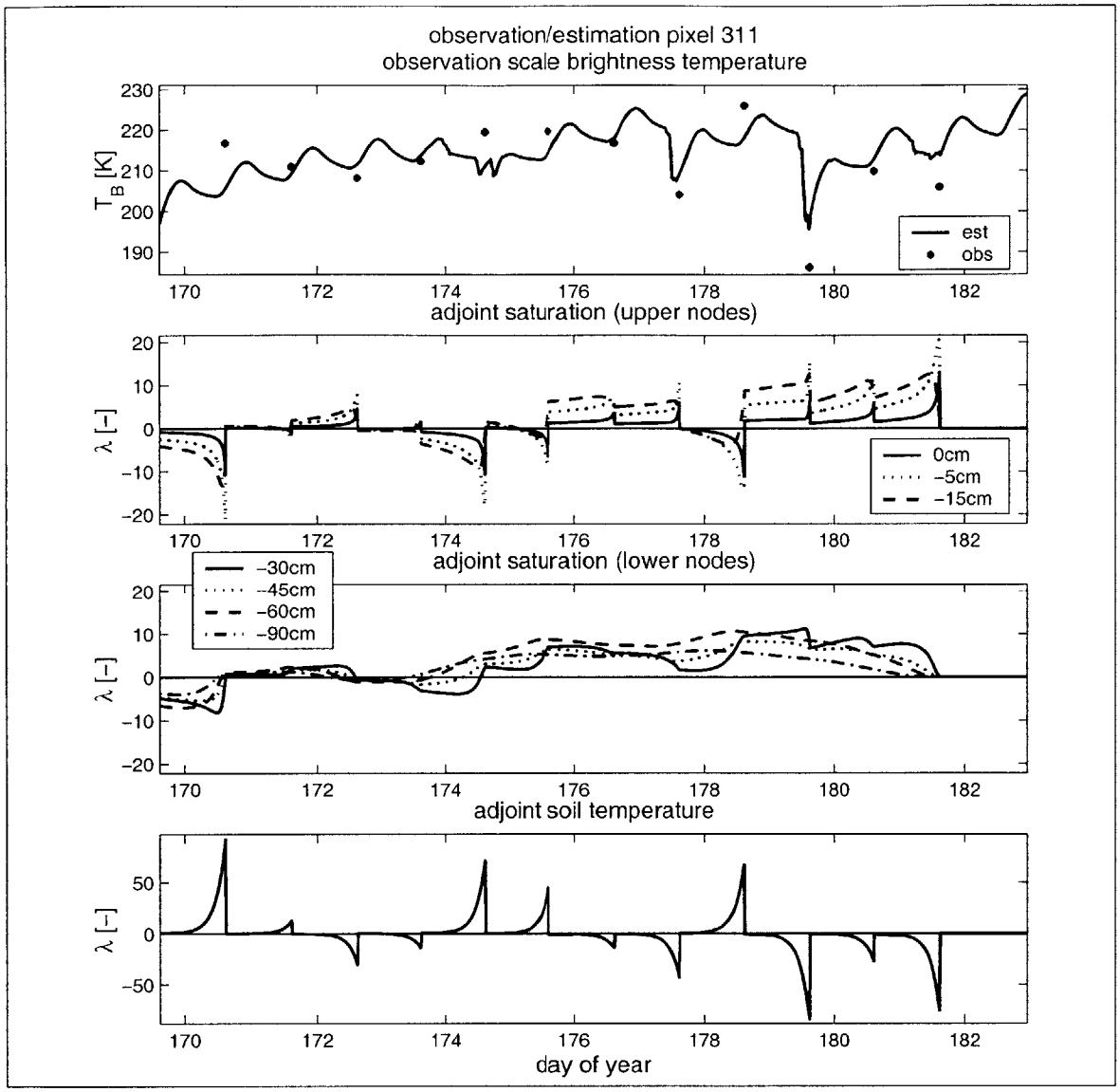


Figure 6.10: Adjoint variables for Reference Experiment I. The top panel shows the estimated brightness temperature for a representative observation pixel along with the (synthetic) observations. Note that for this experiment, observation and estimation pixels coincide. The second and the third panels show the adjoint saturation for the upper three nodes and the lower four nodes, respectively. The last panel shows the adjoint soil temperature. The adjoint equation is solved backward in time starting from zero at the final time (day 183). It is forced at observation times with the misfit between the observations and the estimates of the measured brightness temperatures. If the observed brightness temperature is higher than the estimated brightness, the saturation adjoint variable will be forced towards negative values, because the brightness temperature increases with decreasing saturation. The opposite is true for the adjoint soil temperature. Note how the upper two nodes of the adjoint saturation at $0cm$ and $-5cm$ are directly and instantaneously influenced by the data misfit forcing, while the rest of the nodes experience the forcing with a delay when the shock is propagated through the model physics.

deviation of the forcing terms in the force-restore equation and in the canopy energy balance equation is set to $50W/m^2$. Moreover, the horizontal correlation of the model error has been increased to $6km$. This implies that the model errors are correlated over a few neighboring pixels. This correlation increases the share of the process noise update to approximately 25% of the total CPU time (Section 8.1.2).

From the new realizations of the uncertain parameters we again derive (synthetic) true fields and (synthetic) brightness data. The $5km$ resolution of the estimation and the observation pixels is unchanged, and we again use twelve brightness images (Figure 6.2 and Section 6.1).

### 6.2.1 Estimation of the True Fields

Figure 6.11 shows the true, the prior, and the estimated top node saturation for Reference Experiment II at six different times during the two-week period. As in Reference Experiment I, the soil moisture estimates are greatly improved through the assimilation of the brightness data. In Figure 6.12 we plot the area average errors in the top node saturation with respect to the true fields. The time and area average error is now 2.9% in saturation, compared to 1.4% in Reference Experiment I. This loss of quality in the estimate can be attributed to the much stronger model errors.

The strong model error is also responsible for the rapid increase of the error in the estimates at the end of the assimilation window. The last brightness image is assimilated on day 181.6 (Figure 6.2). After this time, we essentially forecast the top node saturation, which is of course very difficult to do with such strong model errors. Note that for this experiment the area average prior error does not decrease over the two-week period. Since we continuously add strong model error, the prior fields will not converge artificially to the truth. Reference Experiment II therefore describes a much more realistic scenario.

The lower panel of Figure 6.12 shows the area average root-mean-square errors of the prior and the estimated soil temperature. The stronger model error in the force-restore equation leads to a higher area average prior soil temperature error of $1.6K$ compared to $0.37K$ in Reference Experiment I. Moreover, the estimate is only a slight improvement over the prior guess, because L-band brightness temperature observations once a day are too infrequent compared to the short memory of soil temperature.

Figure 6.13 examines the estimate of the model errors. For a representative pixel, we plot the estimate, the prior, and the true model error time series. For comparison, we also show the 24 hour moving average of the true model error. Recall that the prior model error is zero. The top panel of Figure 6.13 shows the model error estimate for the moisture flux upper boundary condition. While the true model error correlation time is 10 hours, we only assimilate brightness updates once per day. Therefore the estimate is necessarily much smoother than the true time series. But if we compare the estimate to the 24 hour moving average of the true model error, we see that the estimate is in fact quite good.

As shown in the second panel of Figure 6.13, we get a poorer estimate for the model error in the soil energy balance. For the model error in the canopy energy balance, shown in the third panel, we do even worse. Obviously, the daily L-band brightness images do not contain much information on the canopy states. Although we have only plotted the results for a single pixel, we reach the same conclusion when we look at all pixels. Finally note that all model error estimates vanish after the last observation time plus one correlation time (day 181.6 plus 10 hours). The observations do not contain any information on times

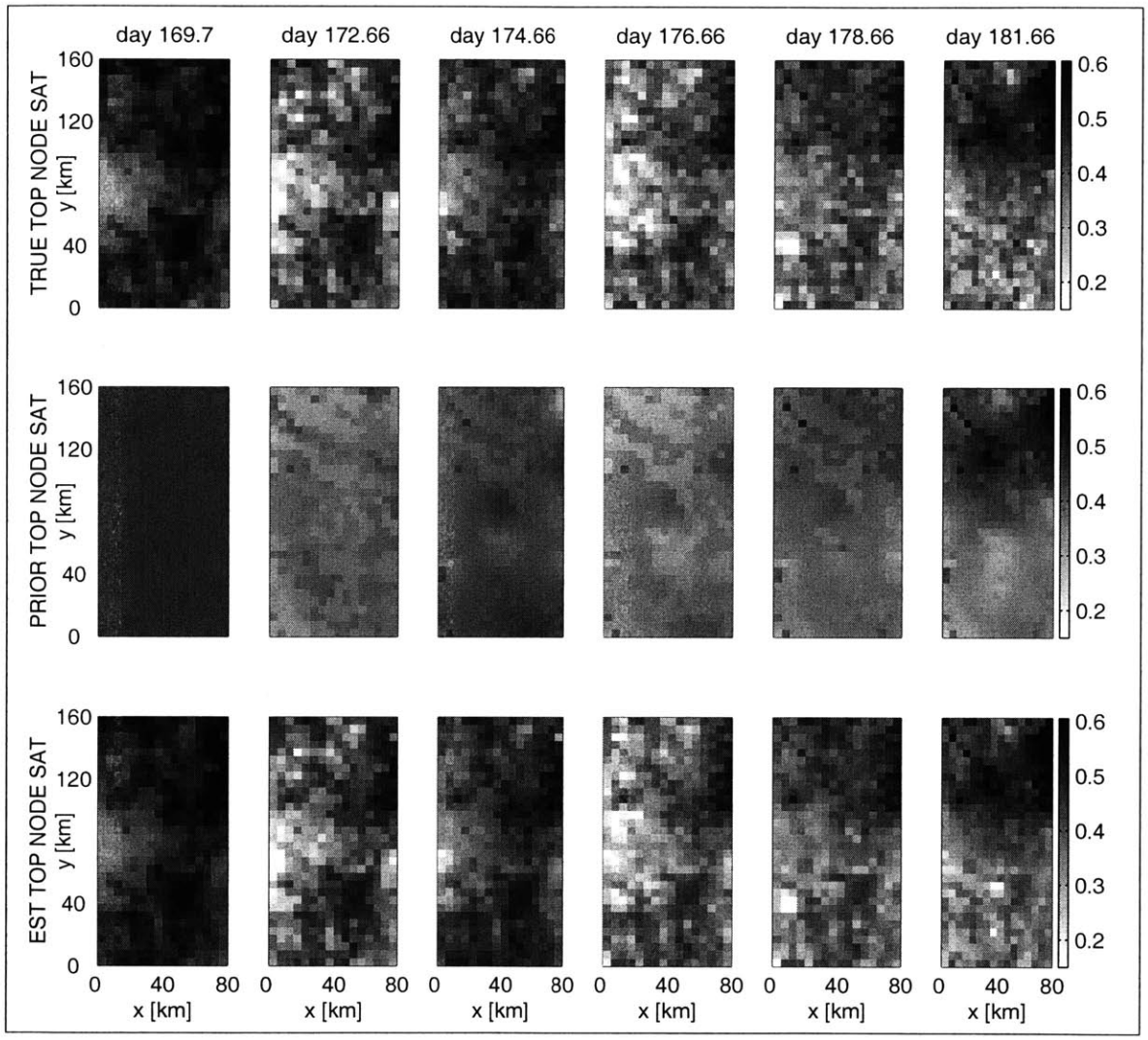


Figure 6.11: Top node saturation for Reference Experiment II. The first row shows the true top node saturation at six different times during the assimilation interval. In the second and third rows the prior and the estimate of the top node saturation, respectively, are depicted for the same six times.

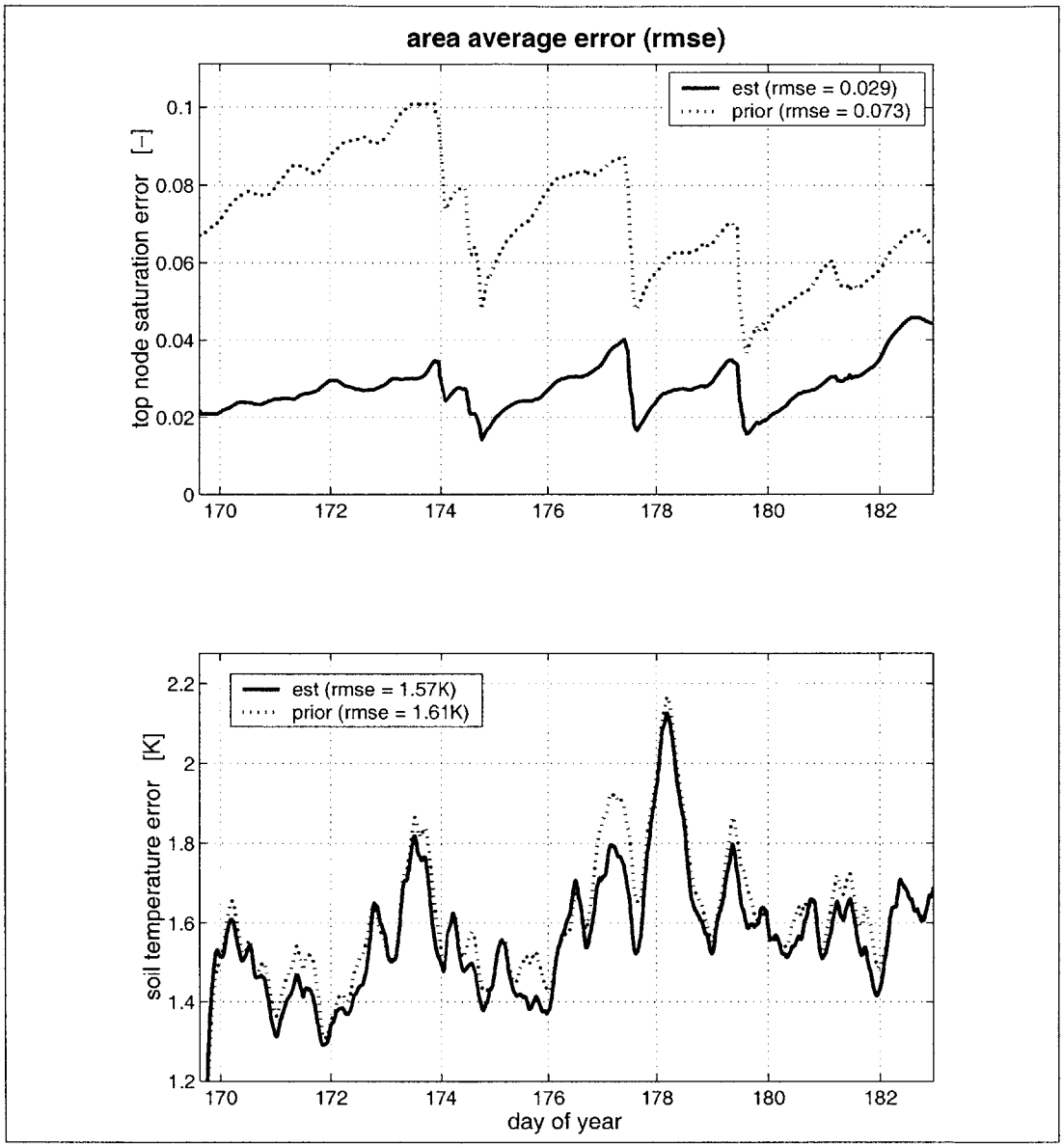


Figure 6.12: Area average errors for Reference Experiment II. The root-mean-square errors (rmse) of the prior and the estimated top node saturation and soil temperature with respect to the (synthetic) true fields are shown. In the legend we also indicate the temporal average of the area average rmse. Note that the soil moisture errors are in terms of saturation. Obviously, the assimilation greatly improves the errors in the top node saturation over the prior errors. On the other hand, daily brightness observations are not enough to improve the prior guess of the soil temperature by much.

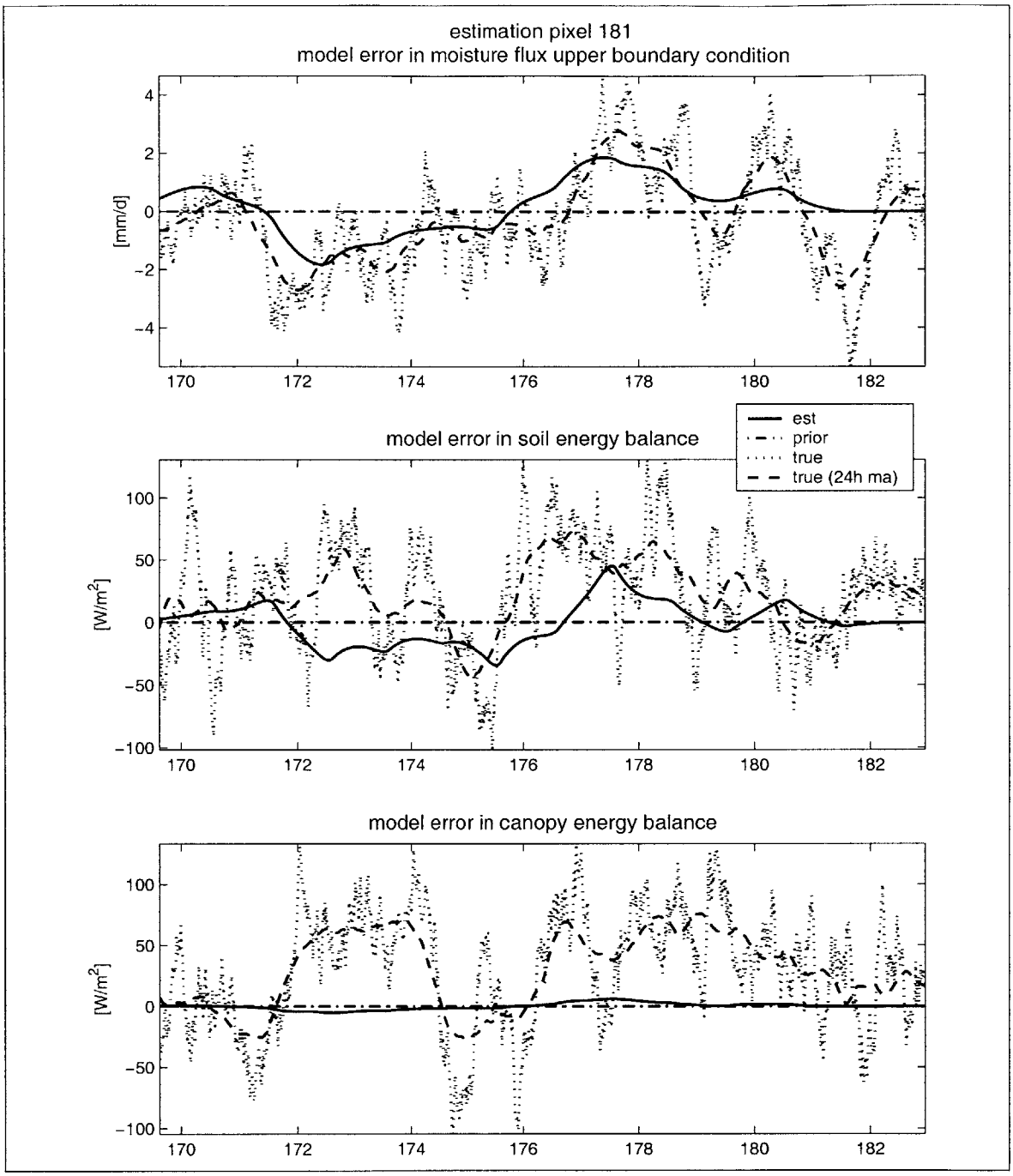


Figure 6.13: Model errors for Reference Experiment II. The estimated, the prior, and the true model errors for a representative pixel are shown. For comparison, we also plot the 24 hour moving average of the true time series. The upper panel shows the model error in the upper moisture flux boundary condition. Considering the difference in time scales between the true model error variability and the frequency of the brightness data, we get a very good estimate. In contrast, we have somewhat less skill in estimating the model error in the soil energy balance (second panel). We do even worse on the model error in the canopy energy balance (third panel).

that are further into the future.

For the subsurface nodes, we again get the same excellent estimates as for the surface node. To illustrate this point, Figure 6.14 shows the estimated, the prior, and the true profile saturation for a representative pixel. However, the limitations discussed in Section 6.1.3 still apply. In summary, we get an overly optimistic profile estimate because we use a fixed shape for the initial profile and because the assimilation window is too short for significant divergence to occur between the shape of the prior and the true profiles.

### 6.2.2 Reduced Objective Function and Posterior Data Residuals

Figure 6.15 shows the reduced objective function versus the iteration number. The converged value of the reduced objective is 5985 and lies within 1.5 standard deviations from the expected value of 6144. From the reduced objective function we have therefore no indication that the assimilation was not optimal.

Finally, a close look at the posterior data residuals yields qualitatively the same results as for Reference Experiment I (Figures 6.7, 6.8, and 6.9). The raw mean for all residuals with a 95% confidence interval is $0.04 \pm 0.11K$. For all but one of the residual brightness images we find a mean whose 95% confidence interval includes zero. The slight bias observed in the Reference Experiment I is absent here. The residuals are uncorrelated in space and in time, but the residuals of the individual images are not exactly normally distributed. We can again attribute this deviation to the nonlinear nature of the problem.

## 6.3 Downscaling Experiments

We now investigate the downscaling capability of the assimilation algorithm. As outlined in Section 4.7, we can effectively increase the resolution of the brightness images by making use of the fact that the inputs to the hydrologic model are available at a finer scale. The additional information is implicitly deduced from our knowledge of soil and land cover parameters as well as from the meteorologic forcings.

In the downscaling experiments, we use the same setup as in Reference Experiment I. In particular, we continue to estimate the land surface states at $5km$ resolution. In contrast to the reference setup, we now generate brightness observations at resolutions of $10km$ and $20km$, respectively. In the former scenario, each observation pixel contains four estimation pixels. Consequently, we call this setup the (1:4) downscaling scenario. In the case of brightness observations at $20km$ resolution, there are 16 estimation pixels within each observation pixel, and we call this setup the (1:16) downscaling scenario.

### 6.3.1 Estimation of the True Fields

In Figure 6.16, we compare the estimates of the top node saturation for both downscaling scenarios. For comparison, the true and the prior top node saturation are shown in the first two columns. All columns depict the saturation fields for the same three times during the experiment. For each downscaling estimate, we also show the outline of the corresponding observation pixels.

In both downscaling scenarios, the algorithm can adequately estimate the large-scale spatial distribution of the saturation. More importantly, structures at scales well below the scale of the observations can be resolved satisfactorily (Figure 6.16). This means that

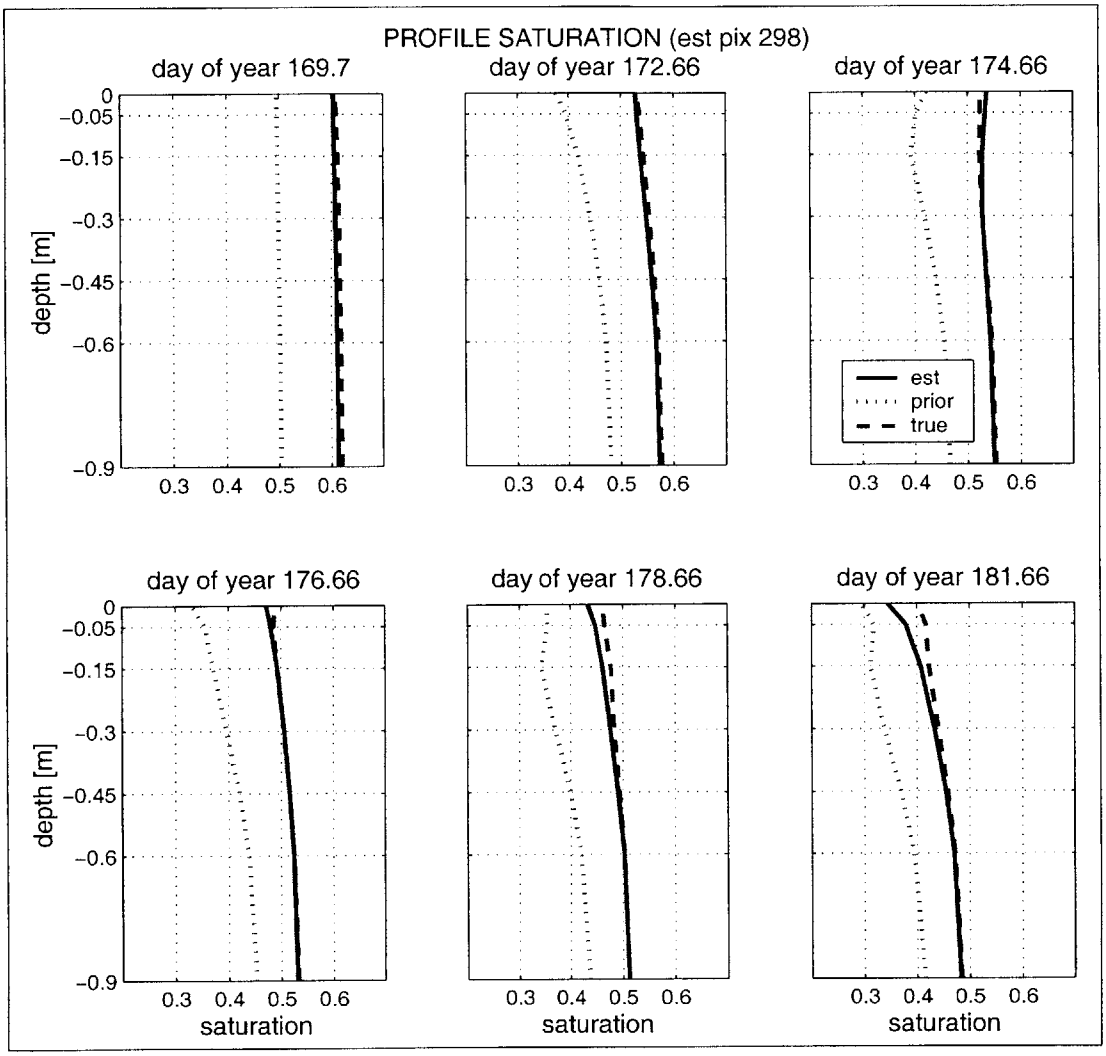


Figure 6.14: Profile saturation for Reference Experiment II. The true, the prior, and the estimated profile saturation are shown at six different times during the assimilation interval. Note that the profile estimates are overly optimistic because of the setup of the synthetic experiment.

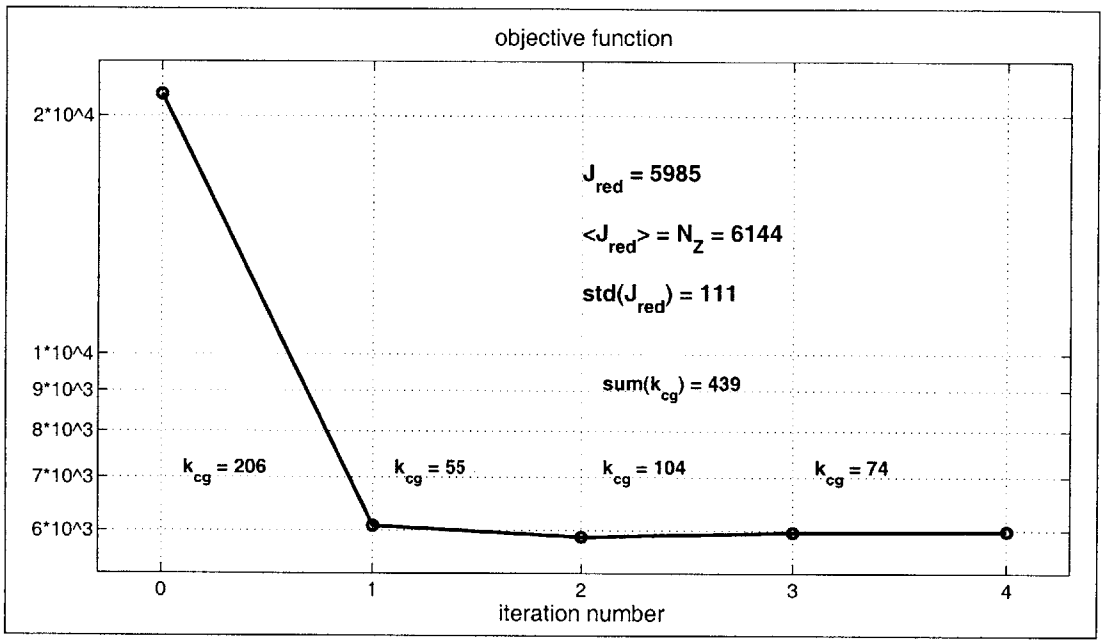


Figure 6.15: Objective function versus iteration number for Reference Experiment II. The reduced objective function after convergence is 5985. The number of data points is 6144, which is also the expected value of the reduced objective function. The standard deviation of the reduced objective function is 111. The values of $k_{cg}$ indicate the number of linear combinations of representer functions that needed to be evaluated during the conjugate gradient iteration of the indirect representer approach (Chapter 8).

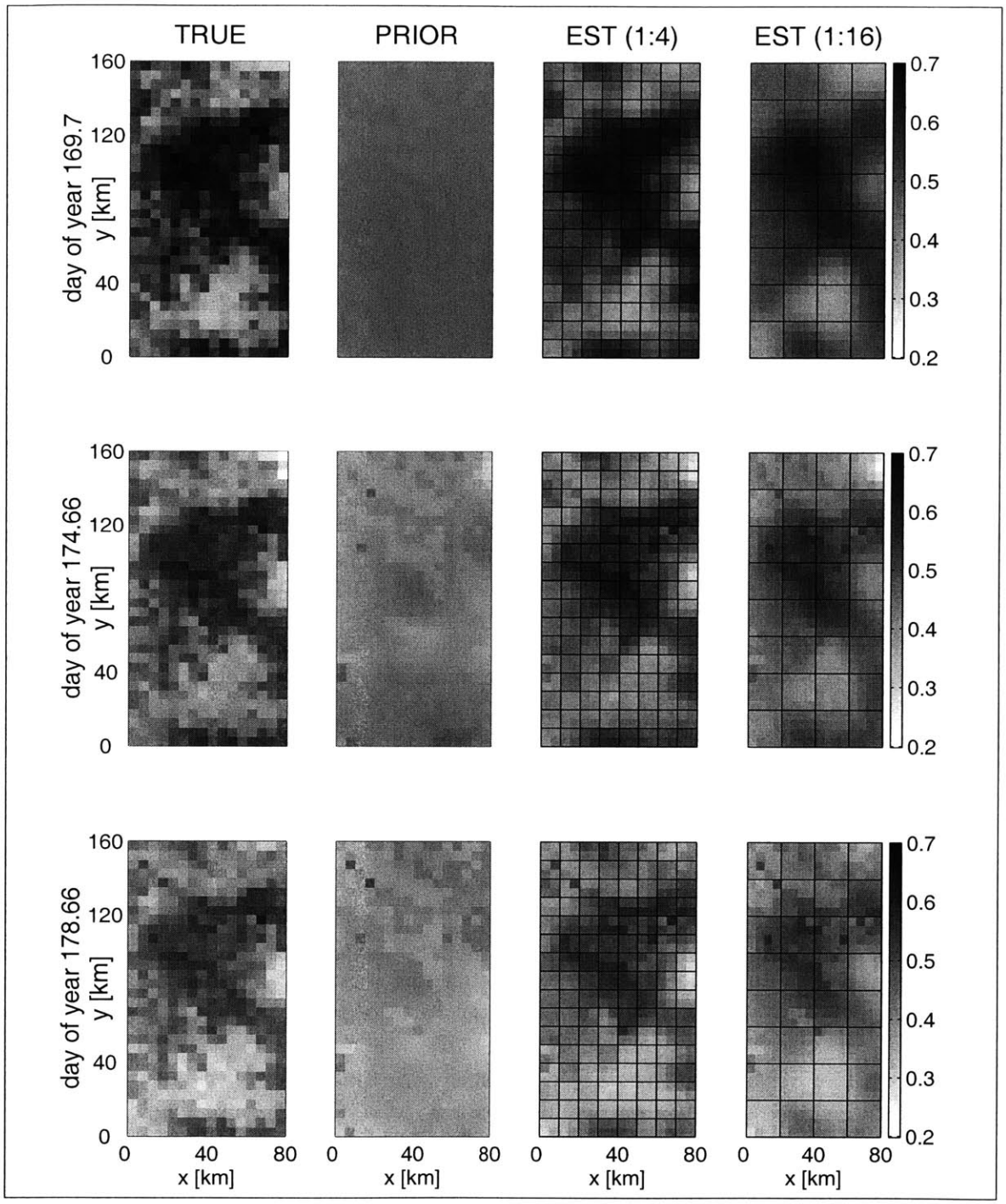


Figure 6.16: Top node saturation for the downscaling experiments. The first column shows the true top node saturation at three different times during the assimilation interval. In the second column, the prior top node saturation is depicted for the same three times. In columns three and four, the estimated top node saturation is shown for downscaling ratios of (1:4) and (1:16), resp. In the (1:4) scenario, each observation pixel covers four estimation pixels. The observation pixels are drawn with solid grid lines. The assimilation algorithm is clearly capable of capturing structures that are finer than the scale of the observations.

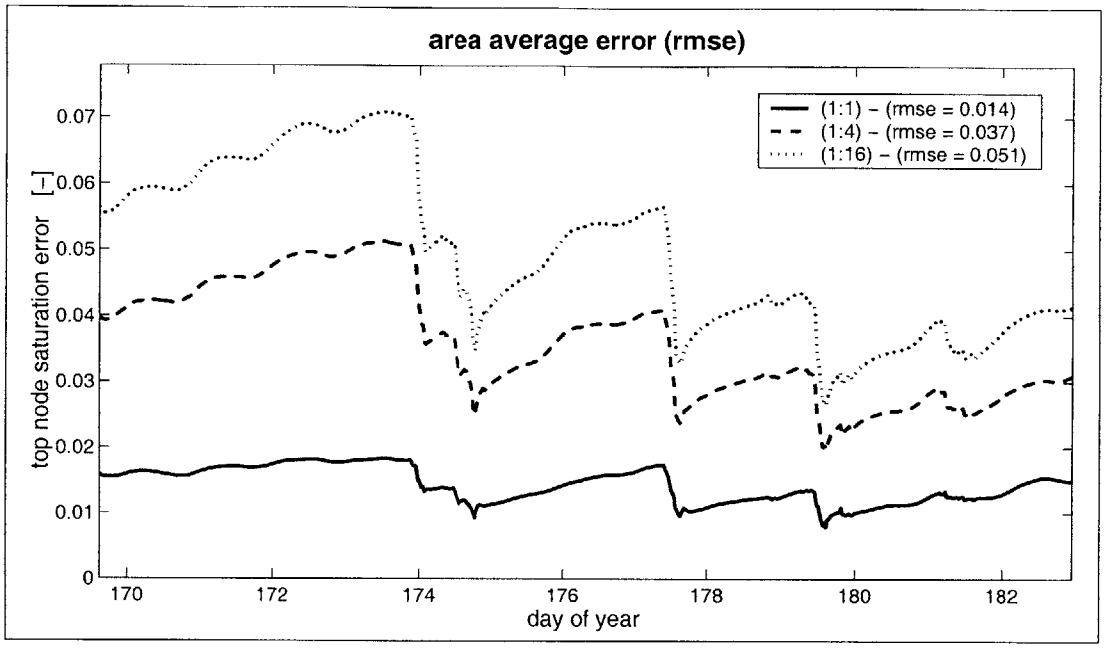


Figure 6.17: Area average errors for the downscaling experiments. The root-mean-square errors (rmse) in the estimated top node saturation with respect to the (synthetic) true fields are shown for Reference Experiment I (1:1) and two downscaling scenarios. The prior error is shown in Figure 6.4. In the legend we also indicate the temporal mean of the area average rmse. Note that the soil moisture errors are in terms of saturation (see Figure 6.4). As expected, the errors increase with decreasing resolution of the brightness data, but even for the coarser brightness images (1:16), the errors are still acceptable. The decrease with time in the errors of the two downscaling scenarios can be explained in the same way as the decrease in the prior error of Figure 6.4 (Section 6.1).

brightness images with resolutions of a few tens of kilometers are definitely useful even if the scale of interest is on the order of a few kilometers, provided we have fine-scale information on the micro-meteorologic forcings and on the model parameters.

Figure 6.17 offers a complementary view by showing time series of the area average top node saturation errors for all downscaling scenarios. The (1:1) downscaling scenario corresponds to Reference Experiment I of Section 6.1. Spatio-temporal averages of the errors are given in the legend. Note again that the errors are in terms of the saturation (see Figure 6.4). As expected, the area average error in the top node saturation increases with decreasing resolution of the brightness data. Even for the coarser (1:16) downscaling scenario, the root-mean-square error is always below 0.07, which translates into approximately 3% volumetric soil moisture. Of course the experiments so far have been assimilation exercises under ideal conditions, and the numbers are subject to change when field data are assimilated.

Finally, Figure 6.18 shows the reduced objective function for both downscaling scenarios. In both cases, the reduced objective function lies within about 1.5 standard deviations from the expected value. This confirms that the algorithm is working properly under the ideal

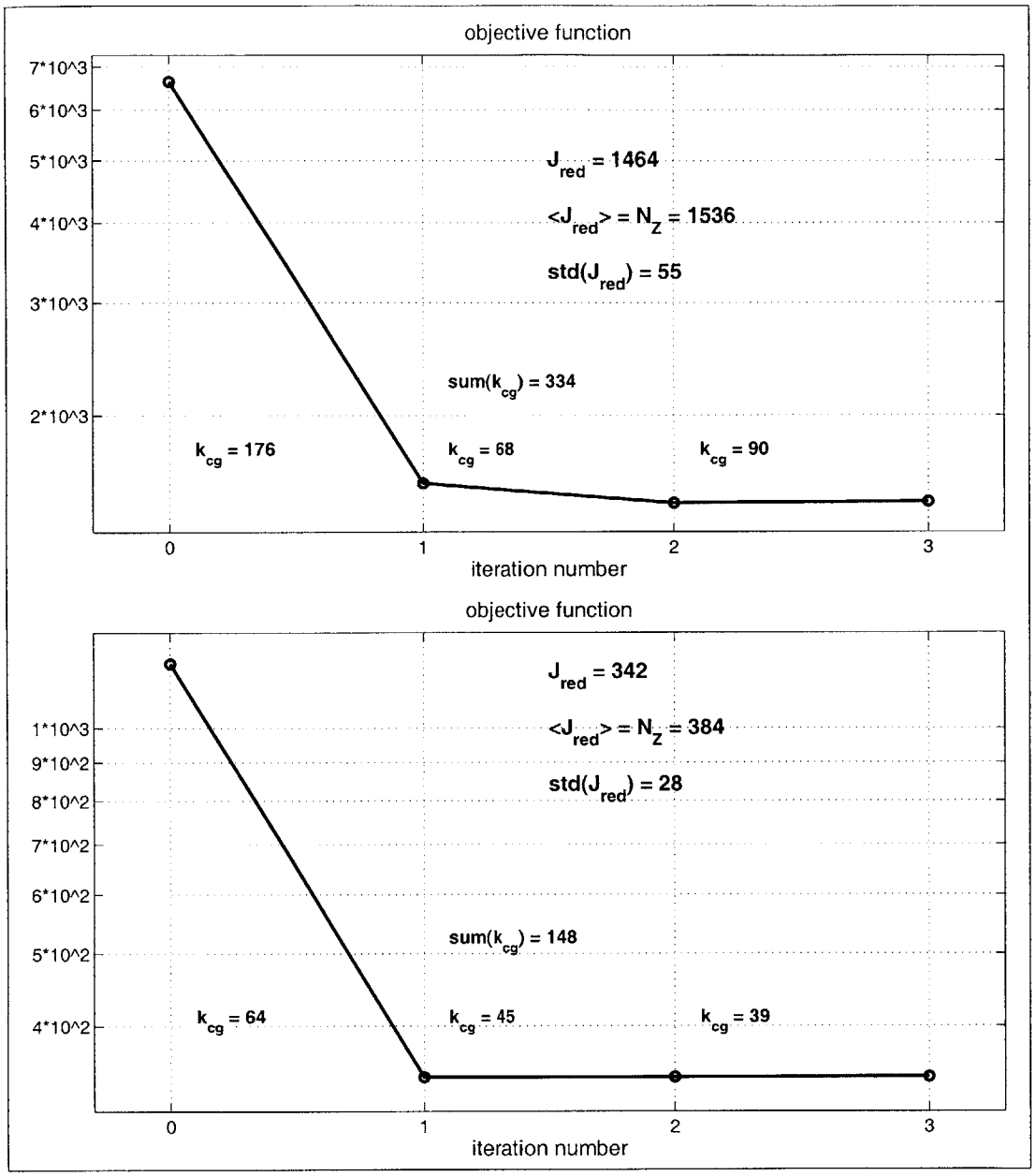


Figure 6.18: Objective function versus iteration number for the downscaling experiments. The upper and lower panels show the objective for the (1:4) and the (1:16) downscaling experiments, respectively. The reduced objective function after convergence is 1464 and 342, respectively. The number of data points is 1536 (384), which is also the expected value of the reduced objective function. The standard deviation of the reduced objective function is 55 (28). The values of $k_{cg}$ indicate the number of linear combinations of representer functions that needed to be evaluated during the conjugate gradient iteration of the indirect representer approach (Chapter 8).

conditions of these experiments.

### 6.3.2 Determining the Horizontal Resolution

Ideally, we would like to determine the horizontal resolution of the model and hence of the estimation pixels only according to the availability of the model inputs such as the micro-meteorologic forcings or the soil and land cover parameters. In practice, however, the model resolution will also depend on the resolution of the brightness images, as well as on the computational feasibility. If we only get brightness data at $50km$ resolution, then it probably does not make much sense to work with a model resolution below $10km$. Clearly, a brightness average over $50km$ will have very little additional information when assimilated into a model of $1km$ resolution, and the assimilation procedure would effectively return the prior fields as the best estimate. Moreover, the computational effort increases with increasing resolution of the model (Section 8.2.2). Even if the model parameters are available at a very fine resolution, we may not be able to afford this fine a model resolution when we carry out the assimilation.

## 6.4 Satellite Repeat Cycle Experiments

### 6.4.1 Motivation and Experiment Design

Another question of practical relevance is the sensitivity of the estimates to the satellite revisit frequency. Since L-band microwave sensors require rather large antennae, and since the antenna size to achieve a given horizontal resolution increases with increasing orbit height, only polar orbiting satellites can be equipped with such sensors. This implies the disadvantage of a rather long revisit interval, which is typically on the order of three to four days for one polar orbiting satellite. By installing sensors on a few satellites, the repeat cycle may be somewhat reduced. However, with current technology and funding, a repeat cycle of one day, that is one image per day, appears to be the best one can hope for.

The most important factor determining soil moisture is certainly precipitation. But provided we have good observations of precipitation, most of the uncertainty in the soil moisture stems from poor knowledge of the initial condition and of the land surface fluxes. Now consider that soil moisture has a relatively long memory compared to soil temperature and other land surface variables. We can therefore hope to estimate the initial condition with rather infrequent observations, and therefore remove some of the uncertainty in the soil moisture estimates. The question is then how much the uncertainty in the land surface fluxes affects the soil moisture estimates. To examine this problem, we have conducted two experiments in which fewer brightness observations were assimilated.

In the first experiment with a 3-day brightness repeat cycle, only four (synthetic) brightness images at observation times 1, 4, 7, and 10 were assimilated. In another experiment, the assimilated data were further reduced to a 6-day repeat cycle, and only the measurements at observation times 1 and 7 were assimilated. See Figure 6.2 for details on the observation times. Otherwise the setup of the experiments is identical to the setup of Reference Experiment II. In particular, the repeat cycle sensitivity experiments were carried out with no downscaling and just one assimilation window covering the entire two-week period.

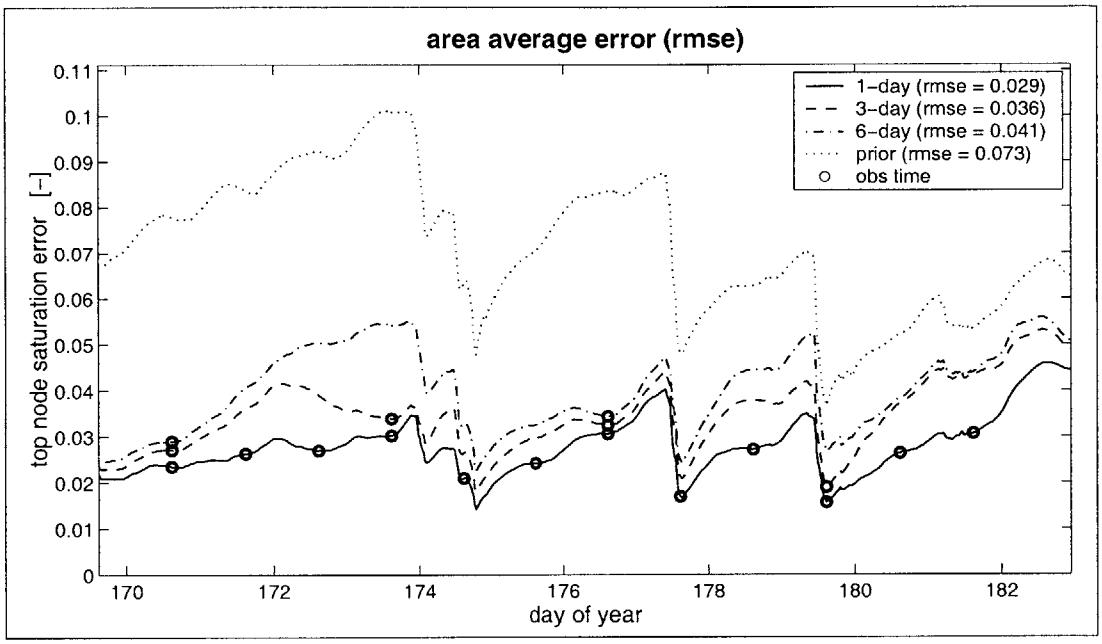


Figure 6.19: Area average errors for the repeat cycle experiments. The root-mean-square errors (rmse) of the estimated top node saturation with respect to the (synthetic) true fields are shown for Reference Experiment II (1-day repeat cycle) and two experiments with brightness repeat cycles of three and six days. The circles show the times at which observations are assimilated for each of the three experiments. In the legend we also indicate the temporal mean of the area average rmse. Note that the soil moisture errors are in terms of saturation (see Figure 6.4). As expected, the errors increase with decreasing availability of the brightness data.

### 6.4.2 Estimation of the True Fields

Figure 6.19 shows the area average top node saturation errors for the different brightness repeat cycles. Note again that the errors are in terms of saturation (see Figure 6.4). In the legend, temporal averages of the area average errors are also given. As expected, the time and area average error increases as we assimilate fewer brightness images.

At observation times 1 and 7, the errors for all experiments are nearly identical, because at these times brightness data are assimilated in all three experiments. Moreover, at observation times 4 and 10, the errors of the 3-day repeat cycle experiment and of Reference Experiment II are very close, because at these times brightness observations are assimilated in both experiments. After such shared observation times, however, the error of the 6-day and the 3-day repeat cycle experiments increases markedly. With observations available only every three or six days, we cannot expect to accurately estimate the model errors, which vary on the scale of their correlation time of 10 hours.

Since in all experiments we assimilate brightness data at the first observation time, we can reasonably well detect the initial condition for all repeat cycles. This explains why the estimates of all experiments are still much better than the prior trajectory. Even if we assimilate brightness data only every six days, we can reduce the prior error significantly, provided of course we have good precipitation data.

If we have only poor measurements of precipitation, the uncertainty in soil moisture will be dominated by the lack of precipitation information. Since the time scale of variability of soil moisture is governed by the interstorm periods, the minimum repeat cycle is then determined by the frequency of rainstorms. In the absence of accurate precipitation data, we must assimilate brightness observation at least once per interstorm period in order to get a satisfactory estimate of soil moisture. In Section 7.2, we examine whether it is feasible to estimate soil moisture even under the extreme condition of not having any quantitative precipitation data.

Finally, Figure 6.20 shows the reduced objective function for the repeat cycle experiments. In both cases, the value of the reduced objective function suggests that the algorithm works optimally.

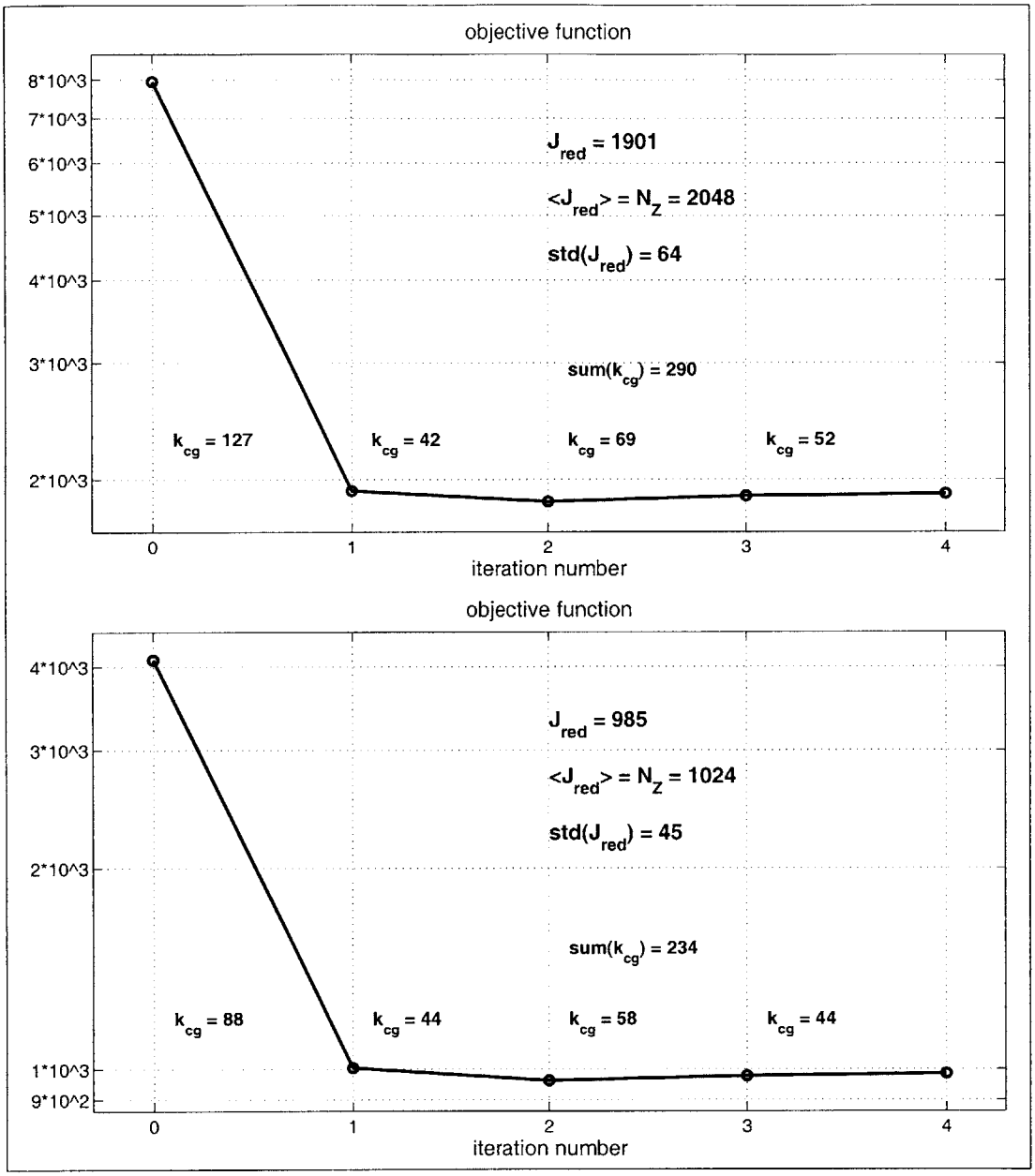


Figure 6.20: Objective function versus iteration number for the repeat cycle experiments. The upper and lower panels show the objective function for the 3-day and the 6-day brightness repeat cycle experiments, respectively. The reduced objective function after convergence is 1901 and 985, respectively. The number of data points is 2048 (1024), which is also the expected value of the reduced objective function. The standard deviation of the reduced objective function is 64 (45). The values of $k_{cg}$ indicate the number of linear combinations of representer functions that needed to be evaluated during the conjugate gradient iteration of the indirect representer approach (Chapter 8).

# Chapter 7

# Assimilation Under Nonideal Conditions

In this Chapter, we present the results from a series of synthetic experiments under nonideal assimilation conditions. By nonideal conditions we generally mean that we use different inputs or parameters for the generation of the (synthetic) true fields and for the estimation algorithm. In contrast to the ideal experiments of Chapter 6, nonideal conditions allow us to subject the assimilation algorithm to harder and more realistic tests.

It is important to note that under nonideal conditions, the assimilation algorithm is not operating optimally in the strict sense of the word. Even in the ideal cases of Chapter 6, the nonlinearities in the hydrologic model and in the measurement equation lead to distortions in the normality of the error distribution functions, thus limiting our chances to conduct a fully optimal assimilation. In the nonideal scenarios, however, we explicitly specify wrong error models, which is what we do involuntarily in any field application. The resulting estimation procedure must therefore be suboptimal.

This argument appears to be contradicting our initial goal to develop a truly optimal assimilation algorithm. If the algorithm will always be suboptimal, why not use a simple and fast scheme like Optimal Interpolation right away? But recall that Optimal Interpolation is not optimal even in the ideal linear case unless we manage to come up with the exact (time-dependent) state error covariances at each update, which is quite impossible in practice. In contrast, the representer algorithm only requires the covariances of the uncertain parameters, which are much easier to specify. By using the representer algorithm, which is truly optimal in the ideal linear case, we will arguably get estimates that are much closer to the truth.

In the following experiments, we address three topics which are of major importance for the development of an operational soil moisture data assimilation system. First, we investigate the quality of the estimates when multiple assimilation windows of variable length are used (Section 7.1). This also raises the question of how we can reinitialize the variational algorithm in subsequent assimilation windows. Second, we assess the performance of the estimation when the observed precipitation data are withheld in the assimilation (Section 7.2). Finally, we investigate the influence of the soil hydraulic parameters (Section 7.3). The computational requirements of the assimilation experiments are discussed in Chapter 8.

## 7.1 Multiple Assimilation Windows

Maybe the biggest advantage of the variational method over sequential techniques is the implicit and thus cheap propagation of the error covariances. But in an operational setup, this is also a disadvantage. If we want to use the variational method operationally, we have to choose appropriate assimilation windows and, most importantly, we have to reinitialize these windows repeatedly. In particular, we need to calculate the posterior state error covariance at the final time of the assimilation window such that we can use it as the prior initial condition covariance for the subsequent window. Unless we can exactly compute the posterior state error covariance, we are likely to end up with a suboptimal algorithm. Unfortunately, the cost of providing accurate posterior state error covariances is prohibitive, even if we are only interested in the posterior covariances at the final time.

In practice, however, this does not necessarily mean that we are going to do very poorly even if we cannot reinitialize the assimilation windows optimally. Since the exact error covariances are implicitly propagated in the variational method, a suboptimally specified initial error covariance may evolve into the optimal error covariance after some time. This means that we should specify the assimilation window such that the observation time is near the end of the window. The wrongly specified initial error covariance will then be implicitly propagated for some time, and the error introduced at the initial time will have less impact.

This is also how we would want to proceed from a practical point of view. As soon as a brightness image becomes available, we would like to improve our estimate of the current soil moisture conditions in order to issue a forecast. We would then specify an assimilation window that starts some time in the past and ends at the current time. Of course we are not limited to having just one observation time within the assimilation window. Ideally, we would like to assimilate many past observations, although in practice we are certainly limited by the computational burden that this entails.

Since soil moisture variability is foremost governed by rain, it appears to make sense to take the precipitation history into account when choosing the assimilation windows. As strong rain events tend to wipe out the soil moisture variability at least near the surface, a natural choice is to use assimilation windows that coincide roughly with the interstorm periods. This should also make it easier to estimate the initial saturation.

### 7.1.1 Experiment Design

We now present the results of three experiments which differ in the choice of the assimilation windows covering the two-week experiment period. The temporal setup is illustrated in Figure 6.2. In experiment A, we use three assimilation windows that roughly coincide with the interstorm periods. In the other two experiments, we cut the two-week period into twelve short assimilation windows, each of which contains one observation time. For experiment B, we choose the assimilation windows such that the observation time is always at the end of the window. In experiment C, the assimilation windows are chosen such that the observation time is at the beginning of the window, with the exception of the first window. As discussed above, experiment B is the setup we prefer on theoretical and practical grounds.

All three experiments are based on Reference Experiment II, that is all inputs are the same unless mentioned otherwise. The reinitialization of the assimilation windows is done in

an intentionally simple and ad hoc fashion. At the beginning of each assimilation window, we use the state estimate at the final time of the previous assimilation window to derive the prior mean of the initial condition parameters as well as the shape of the initial saturation profile. For the prior covariance of the initial error we use a scaled version of the initial condition covariance of Reference Experiment II. In particular, the correlation length is always $50km$, regardless of the evolution of the system. The scaling of the covariance was found necessary to ensure convergence. Using the initial condition uncertainty of Reference Experiment II for the now shorter assimilation windows appears to lead to poor conditioning of the representer matrix, and consequently convergence could not always be achieved.

We scale the initial condition variance with a factor depending on the length of the assimilation window, or equivalently, the number of observation times within the window. Moreover, the scaling factor also depends on the number of the assimilation window. Earlier windows have a relatively higher initial condition variance. In particular, in the first experiment the three windows contain (in order) 5, 3, and 4 out of 12 observation times, and we scale the initial condition covariance of the three windows with 0.63, 0.19, and 0.17, respectively. In the other two experiments, each window contains exactly one observation time, and we scale the initial covariance of Reference Experiment II with $0.5/n_w$, where $n_w = 1 \ldots 12$ indicates the number of the assimilation window. For example, the initial condition covariance of the second assimilation window is 0.25 times the initial condition covariance of Reference Experiment II.

### 7.1.2 Estimation of the True Fields

Figure 7.1 shows the area average root-mean-square errors (rmse) of the estimated top node saturation for the three experiments with three and twelve assimilation windows, together with the rmse of Reference Experiment II. For experiment A, the time and area average top node saturation error is 3%, and for experiments B and C the errors are 3.2% and 3.8%, respectively. These numbers compare to an error of 2.9% in Reference Experiment II. As expected, the area average error increases as the number of assimilation windows increases. With each additional assimilation window we introduce more approximations by naively reinitializing the initial saturation covariance.

Certainly the most interesting result, however, is the difference in the errors of experiments B and C. This difference illustrates the impact of the implicit error covariance propagation in the variational scheme. In experiment B, the choice of the assimilation windows lets the initial error covariance evolve dynamically for almost 24 hours before the observation time. In experiment C, by contrast, the poorly specified initial error covariance is propagated for only a couple of hours before the observation time. The update therefore relies on a crude approximation of the error covariance, which results in poorer estimates. Note that experiment C corresponds closely to an Optimal Interpolation scheme, in which the error covariances are not propagated. In summary, the estimates for Reference Experiment II and for experiments A and B are very similar. This suggests that the suboptimality introduced by the naive reinitialization is not severe, provided the assimilation windows are chosen such that the initial error covariance can evolve for at least one day.

### 7.1.3 Assessing the Optimality of the Estimates

We can further examine the degree of suboptimality by looking at the value of the reduced objective function. Figure 7.2 shows the reduced objective function after convergence for

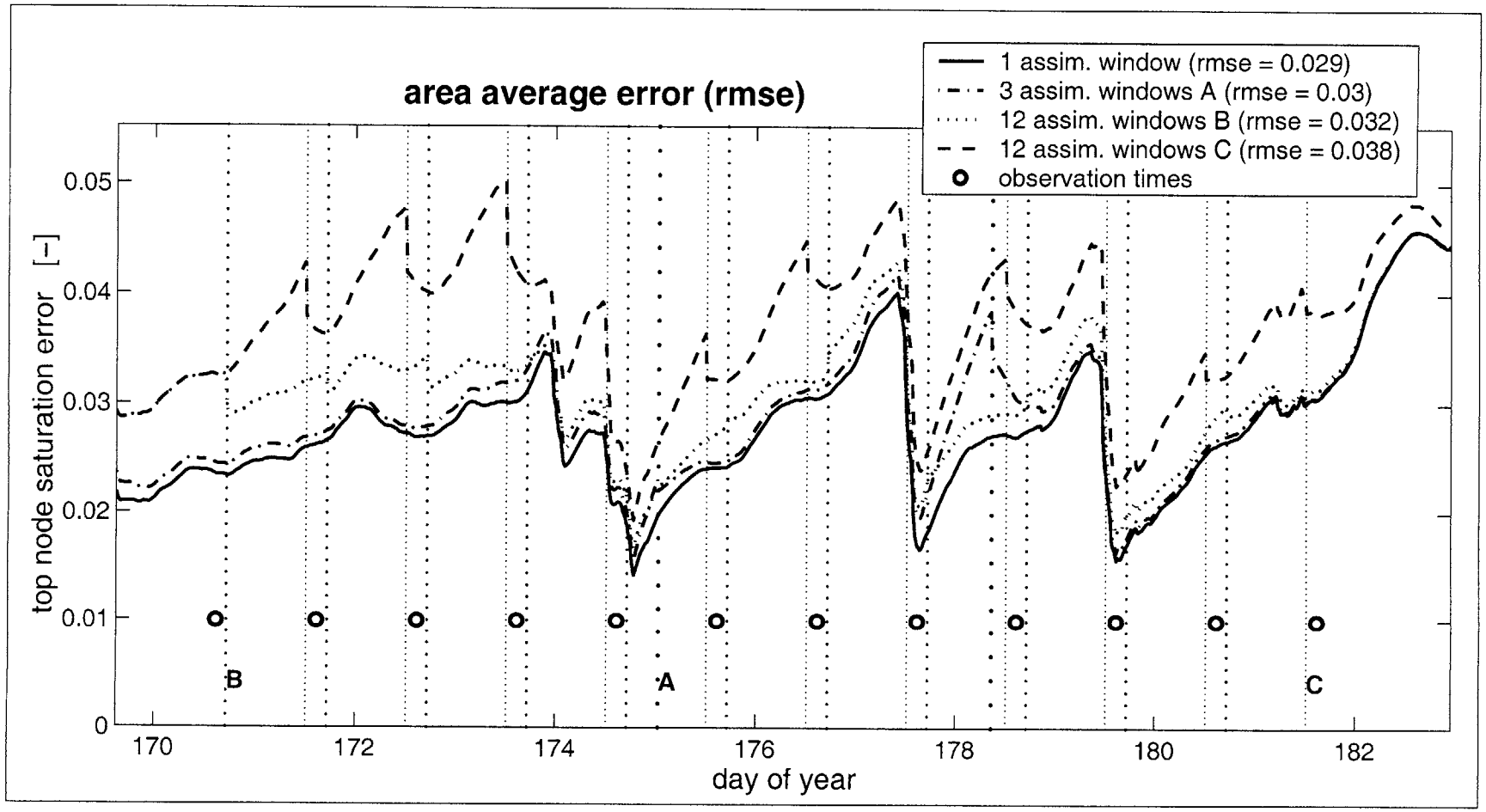


Figure 7.1: Area average errors for multiple assimilation windows. The errors in the estimated top node saturation are shown for Reference Experiment II (Figure 6.12) and the three experiments with shorter assimilation windows. The errors are in the root-mean-square sense with respect to the (synthetic) true solution. The prior root-mean-square error (rmse) is shown in Figure 6.12. In the legend we also indicate the temporal average of the area average rmse. Note that the soil moisture errors are in terms of saturation. The coarsely dotted vertical lines A delimit the three assimilation windows of the first experiment. Likewise, the more finely dotted vertical lines B and C delimit the twelve assimilation windows of the other two experiments. The observation times are indicated with circles. For experiment B the observation times are always at the end of the window, whereas for experiment C the observation times are at the beginning of the window.

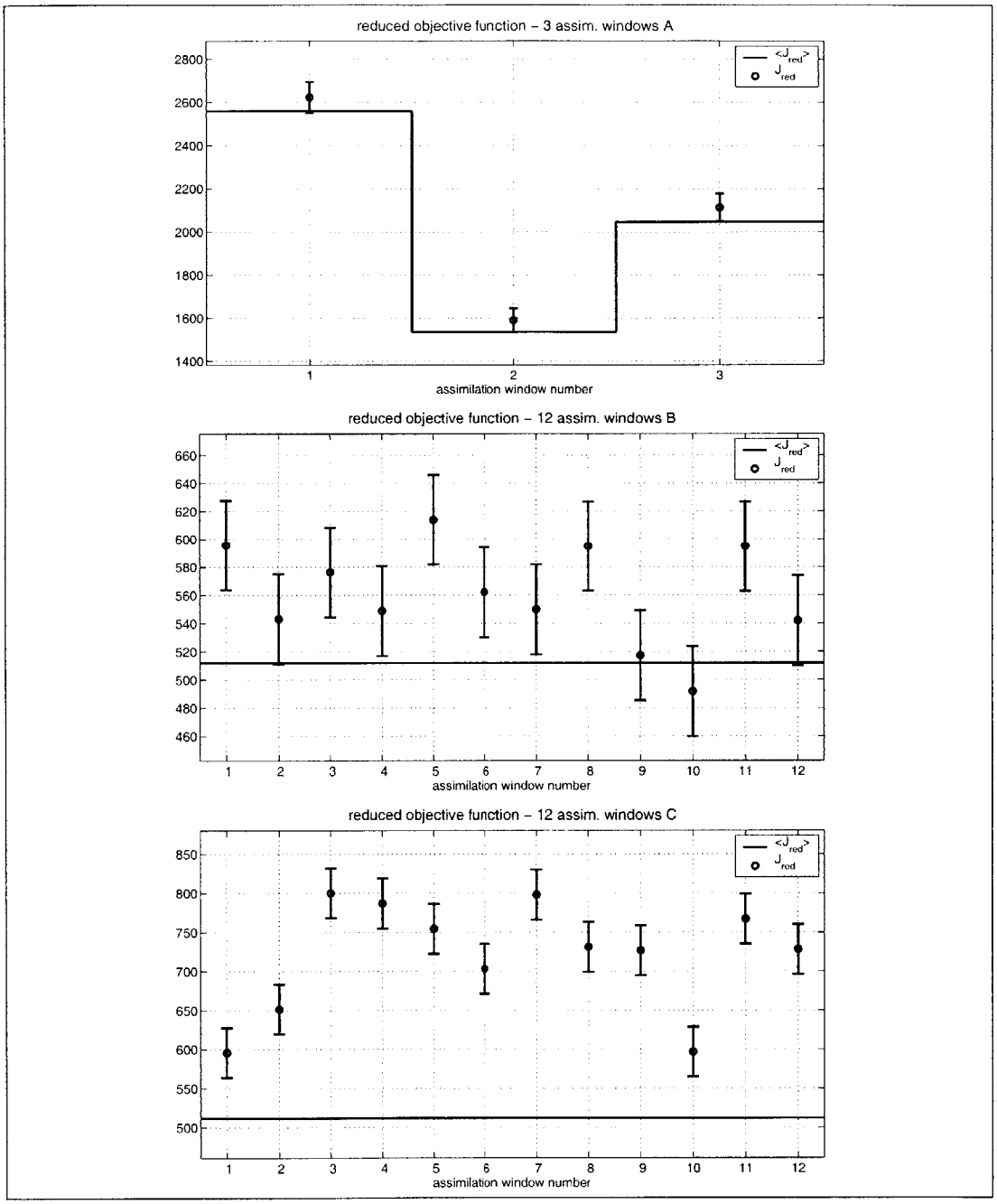


Figure 7.2: Objective function for multiple assimilation windows. The reduced objective function after convergence is plotted versus the number of the assimilation window for experiments A (top), B (middle), and C (bottom). The expected values, which equal the number of data assimilated in the window, are also shown. The error bars around the reduced objective are plus/minus one standard deviation, which is equal to the square-root of twice the number of data assimilated. For experiment A, the reduced objective function does not necessarily indicate suboptimal assimilation. Experiments B and C, however, are clearly suboptimal.

each assimilation window. We also show the expected value of the reduced objective function, which equals the number of scalar data that have been assimilated in each window. The standard deviation of the reduced objective function is indicated with error bars.

For experiment A, the reduced objective function of the three assimilation windows lies within one standard deviation from the respective expected values. Even though all three values are above the expected value, the reduced objective function does not necessarily indicate that the assimilation was suboptimal. For experiment B, eight out of twelve values of the reduced objective function are more than one standard deviation above the expected value, and all but one of the values lie above the expected value. This hints at the fact that the assimilation was not optimal. Clearly, the assimilation of experiment C must have been suboptimal.

An investigation of the posterior data residuals yields roughly the same results as for Reference Experiment II. The residuals are white in time and the hypothesis of a normal distribution cannot be rejected in almost all cases. For experiments A and B, the residuals show no obvious spatial patterns. Only the residuals of experiment C exhibit a weak spatial structure (not shown). In summary, using shorter assimilation windows and a relatively naive reinitialization does not appear to have a significant negative effect on the optimality of the algorithm, provided the assimilation windows are chosen such as to allow adequate evolution of the error covariance before observation times.

Finally, we would like to note that the computational effort for experiments A, B, and C is substantially smaller than for Reference Experiment II. For a more detailed discussion please turn to Section 8.1.3.

## 7.2 Assimilation without Precipitation Data

Of all model inputs, precipitation is the one parameter which dominates soil moisture conditions. At the same time, precipitation is also the input associated with the highest uncertainty. Precipitation observations from rain gauges are point measurements, and the interpolation to larger areas is notoriously ill understood. On the other hand, large-scale precipitation measurements from radar sensors are equally imprecise. The quality of the soil moisture estimates stands and falls with the accuracy of the large-scale precipitation data. It is therefore desirable to take a closer look at the sensitivity of the assimilation algorithm to the precipitation inputs.

### 7.2.1 Experiment Design

We now present the results of an experiment where the precipitation data are withheld from the assimilation. The experiment is based on Reference Experiment I (Section 6.1), and all inputs are the same unless otherwise mentioned. For this experiment, instead of supplying the observed precipitation time series to the assimilation algorithm, we specify zero precipitation throughout the two-week period. In order to compensate for the lack of precipitation in the assimilation, we specify certain times at which the model error in the upper moisture boundary condition has a very high variance. The times at which such model error occurs are chosen to be times at which significant area average precipitation has been observed. Such times can be regarded as precipitation indicators. This is a realistic, albeit extreme, scenario, because it is fairly easy to detect whether or not there is precipitation,

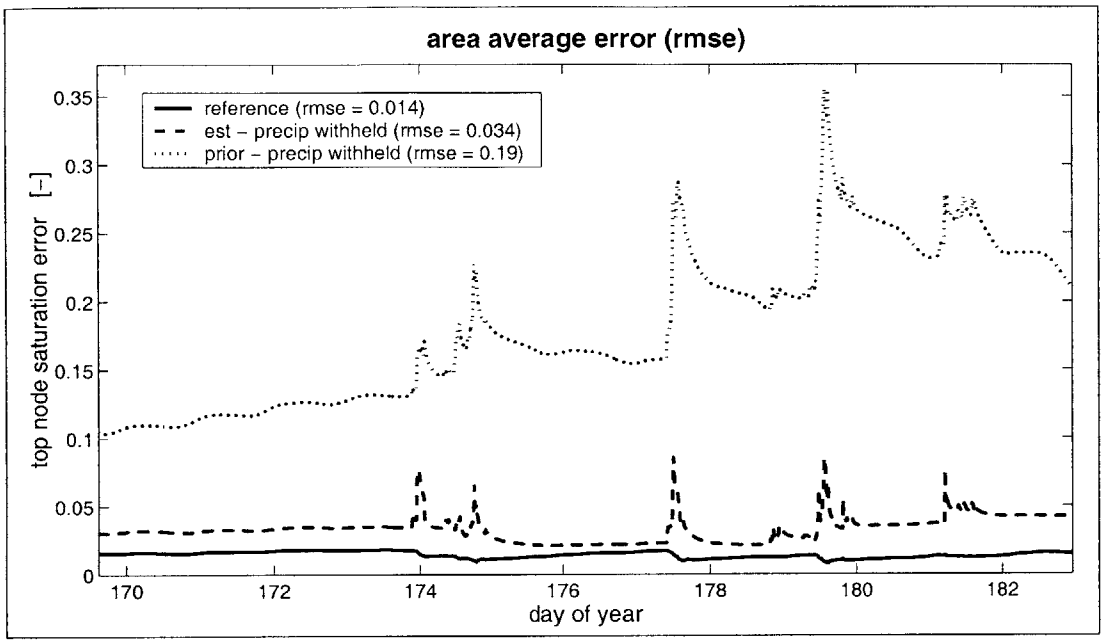


Figure 7.3: Area average errors when precipitation is withheld. The errors in the top node saturation of the estimate, the prior, and the estimate of Reference Experiment I (Figure 6.4) are shown. The errors are in the root-mean-square sense with respect to the (synthetic) true solution. In the legend we also indicate the temporal mean of the area average rmse. Note that the soil moisture errors are in terms of saturation (see Figure 6.4). Even when quantitative precipitation data are withheld entirely, the algorithm can estimate soil moisture to good accuracy.

but it is rather difficult to observe the rain event quantitatively. In practice, one would of course use whatever quantitative precipitation information is available.

Note that the intermittent model error is nonstationary, and in practice we then have to use a white noise error model to make the noise update computationally feasible (Section 4.8). When precipitation is indicated, we set the spatially uniform standard deviation of the model error in the moisture flux boundary condition equal to $86.4mm/d$, which is a typical area average rain rate at $15min$ resolution. This is the case for 70 out of 1280 time steps. At all other times, the white component of the model error has zero variance.

Note that we did not tune the error model to improve the performance of the estimation algorithm. Moreover, the above error model is of course not the only choice. Many other error models can be thought up to compensate for the lack of precipitation data in the assimilation. For example, the model error variance could be horizontally distributed, which may better capture the strong gradient in rainfall across the domain. Also note that with the above error model, "negative precipitation" could result from the estimation, that is the model error could lead to unrealistically high evapotranspiration rates. To prevent this from happening, the model error could be formulated from the start with a logarithmic transform.

### 7.2.2 Estimation of the True Fields

Figure 7.3 shows the area average top node saturation errors with respect to the (synthetic) true fields when quantitative precipitation data are withheld from the assimilation. For comparison, we also show the best estimate of the ideal assimilation run of Section 6.1 (Figure 6.4). Even when all quantitative precipitation information is withheld from the assimilation, we can estimate the top node saturation to within 3.4% in saturation terms, compared to 1.4% under ideal conditions. Unlike in the case of the ideal Reference Experiment I (Figure 6.4), the prior error in this experiment increases over the two-week assimilation window. Since the prior solution does not contain any rain events, the prior saturation is governed by a single, two-week long drydown. For longer assimilation windows, the prior saturation error reaches a plateau.

Around precipitation events, the average error in the top node saturation increases somewhat. This follows naturally from the fact that we cannot resolve the temporal structure of the events from brightness data that are available only once daily. To illustrate this point, we plot in Figure 7.4 the observed precipitation and the corresponding model error estimates for three different pixels during the three major precipitation events of the two-week assimilation window. The three pixels shown in Figure 7.4 are in the southwestern corner (pixel 100), the center (pixel 398), and the northeastern corner (pixel 412) of the domain. For the three pixels, the cumulative observed precipitation over the entire two-week period is $1.6cm$, $1.5cm$, and $4cm$, respectively. The cumulative model error estimates are $1.6cm$, $0.9cm$, and $2.8cm$. The area average cumulative precipitation is $2.8cm$, and the corresponding area average cumulative model error estimate is $1.6cm$.

By using the precipitation indicators we naturally get the overall timing of the storms right. In addition, we also get reasonable estimates of the volume of the storms from the brightness observations. It is clear, however, that the detailed temporal structure of the storms eludes us. Moreover, the model error estimates are generally lower than the observed precipitation. To understand this, recall that we assimilate brightness data only once a day. In this case, the nonlinearities in the infiltration and exfiltration processes defy a more accurate estimate of the volume of the storms. Since the true precipitation is heavier than the model error estimates, the soil actually gets wetter than estimated, but this also leads to stronger evaporation and possibly runoff. If the observations are available only some time after the storm, the difference between the true and the estimated soil saturation at the observation time is then much smaller than the difference between the model error estimate and the observed precipitation at the time of the storm, and we cannot distinguish between the two scenarios.

Note that the increase in the model error estimate with time during blocks of nonzero estimates is a direct consequence of the whiteness of the model error. If the temporal correlation of the model error is white, the model error estimates are in essence scaled version of the adjoint variables. The adjoint variables, in turn, decay backwards in time, as can be seen for example in Figure 6.10.

### 7.2.3 Reduced Objective Function and Computational Effort

Figure 7.5 shows the reduced objective function during the iteration. After convergence, the reduced objective function is 10,907, which is not compatible with an expected value of 6,144. Obviously, our choice of error model does not capture the real error introduced by withholding all quantitative precipitation information. In fact, the above experiment is only

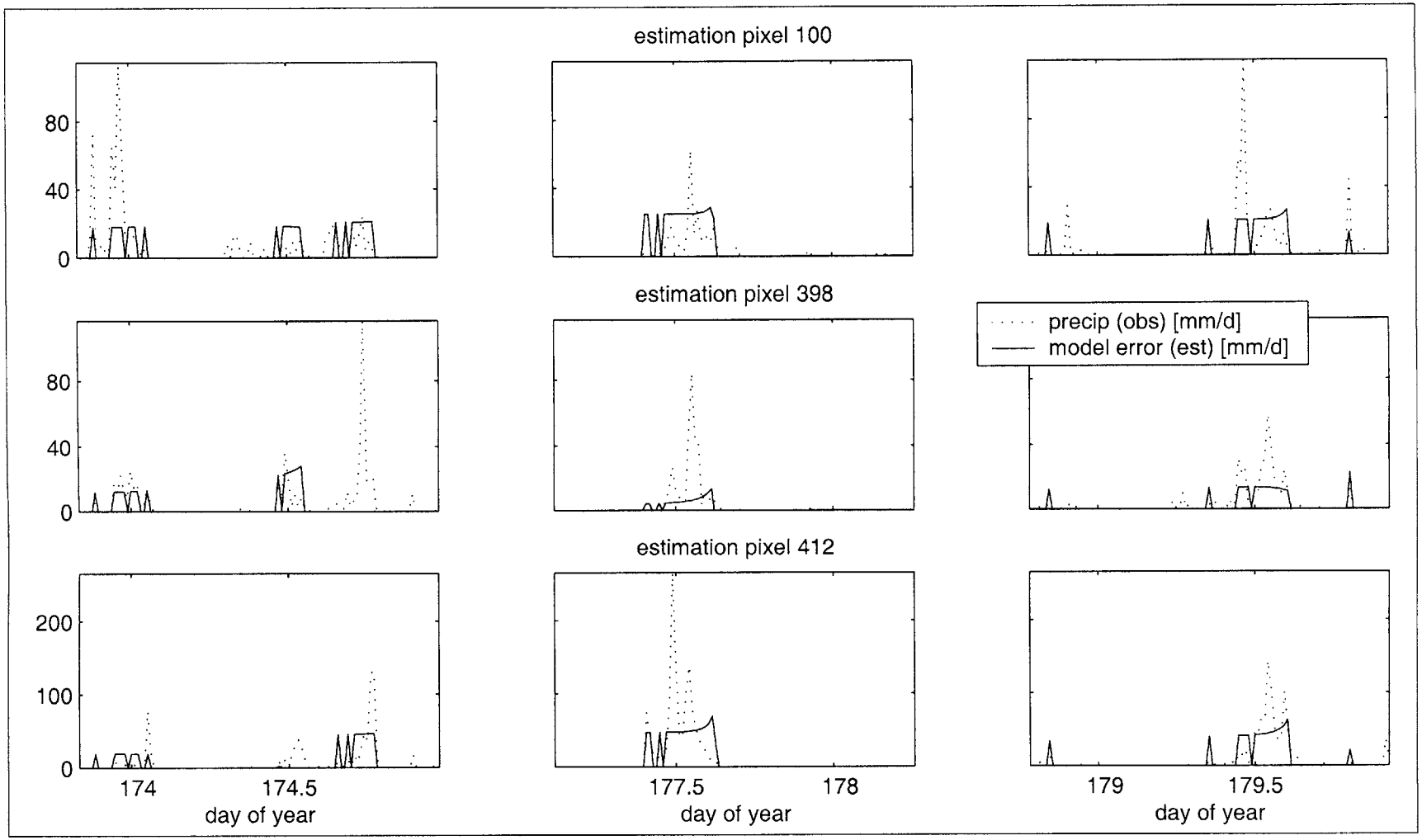


Figure 7.4: Model error estimates when precipitation is withheld. Model error estimates for three different pixels during the three major precipitation events. The pixels 100, 398, and 412 are located in the southwestern corner, the center, and the northeastern corner of the domain, respectively. By using precipitation indicators, we supply the overall timing of the storms to the estimation algorithm. By assimilating brightness observations, we get reasonable estimates of the volume of the storms. To better compare the model error estimates to the observed precipitation, which is defined as a positive quantity, we plot the negative of the model error estimate.

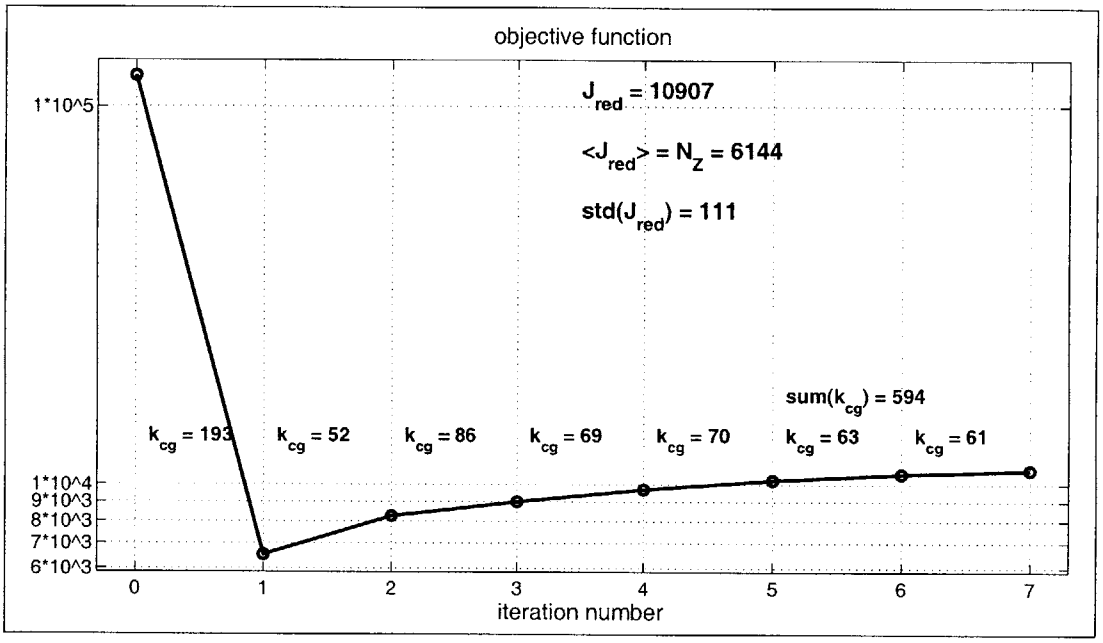


Figure 7.5: Objective function versus iteration number when precipitation is withheld. The reduced objective function after convergence is 10,907. The number of data points is 6144, which is also the expected value of the reduced objective function. Obviously, the estimates do not pass the hypothesis test on the value of the reduced objective function. The values of $k_{cg}$ indicate the number of linear combinations of representer functions that needed to be evaluated during the conjugate gradient iteration of the indirect representer approach (Chapter 8).

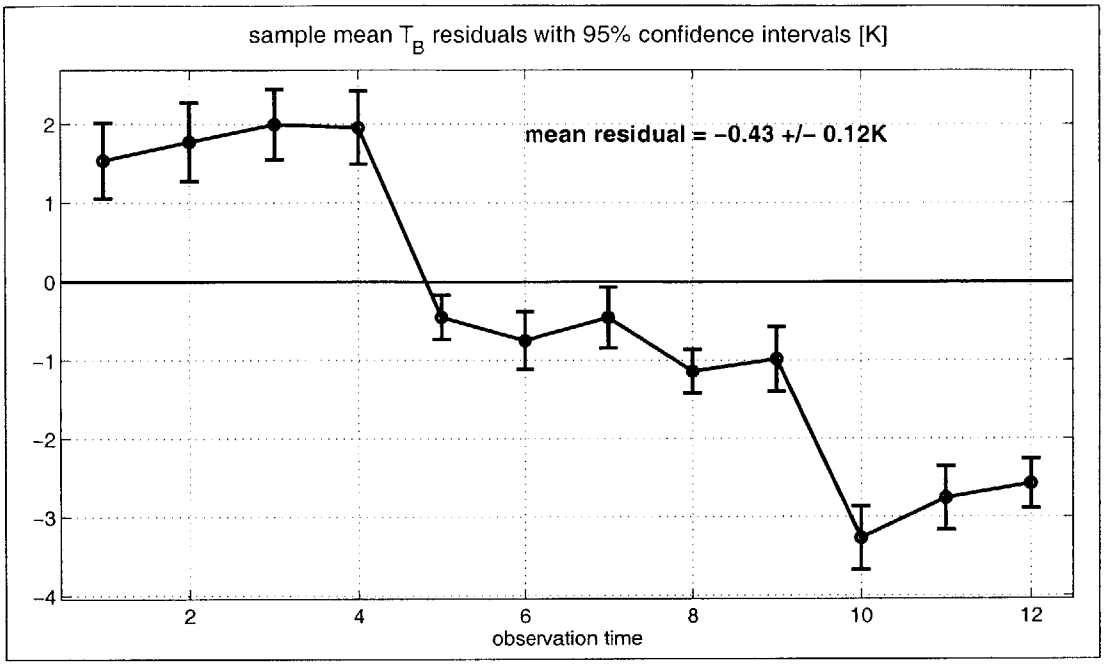


Figure 7.6: Sample mean of the posterior data residuals when precipitation is withheld. The mean residuals for the twelve images that have been assimilated are shown together with the 95% confidence intervals. Very obviously, the mean value of the posterior data residuals differs from zero.

a crude approximation of a much more sophisticated optimal assimilation procedure. Ideally, if the precipitation inputs are considered very uncertain, the true precipitation should be estimated by assimilating the rain data into a suitable model of precipitation processes. In the future, the current soil moisture assimilation algorithm may be augmented to also estimate precipitation.

Interestingly, the computational effort for this experiment is only about two thirds of the computational burden of Reference Experiment I. This stems from the fact that we changed the prior statistics in order to compensate for the lack of precipitation data. For a more detailed discussion see Section 8.1.2.

### 7.2.4 Posterior Data Residuals

A closer look at the data residuals sheds more light on the suboptimal nature of the assimilation in this experiment. We first examine the mean of the residuals (Figure 7.6). Unsurprisingly, the mean of all residuals with a 95% confidence interval is $-0.43 \pm 0.12K$, which does not include zero. Similarly, none of the individual residual brightness images has a mean whose 95% confidence interval includes zero. Figure 7.7 shows the standardized posterior data residuals for each of the twelve brightness images that have been assimilated. Obviously, some of the residual images show spatial structure. This is especially true for observation time 10, which happens to be during a major rain event in the northern half of the domain.

Figure 7.8 shows the sample cumulative distribution function (cdf) for two representative

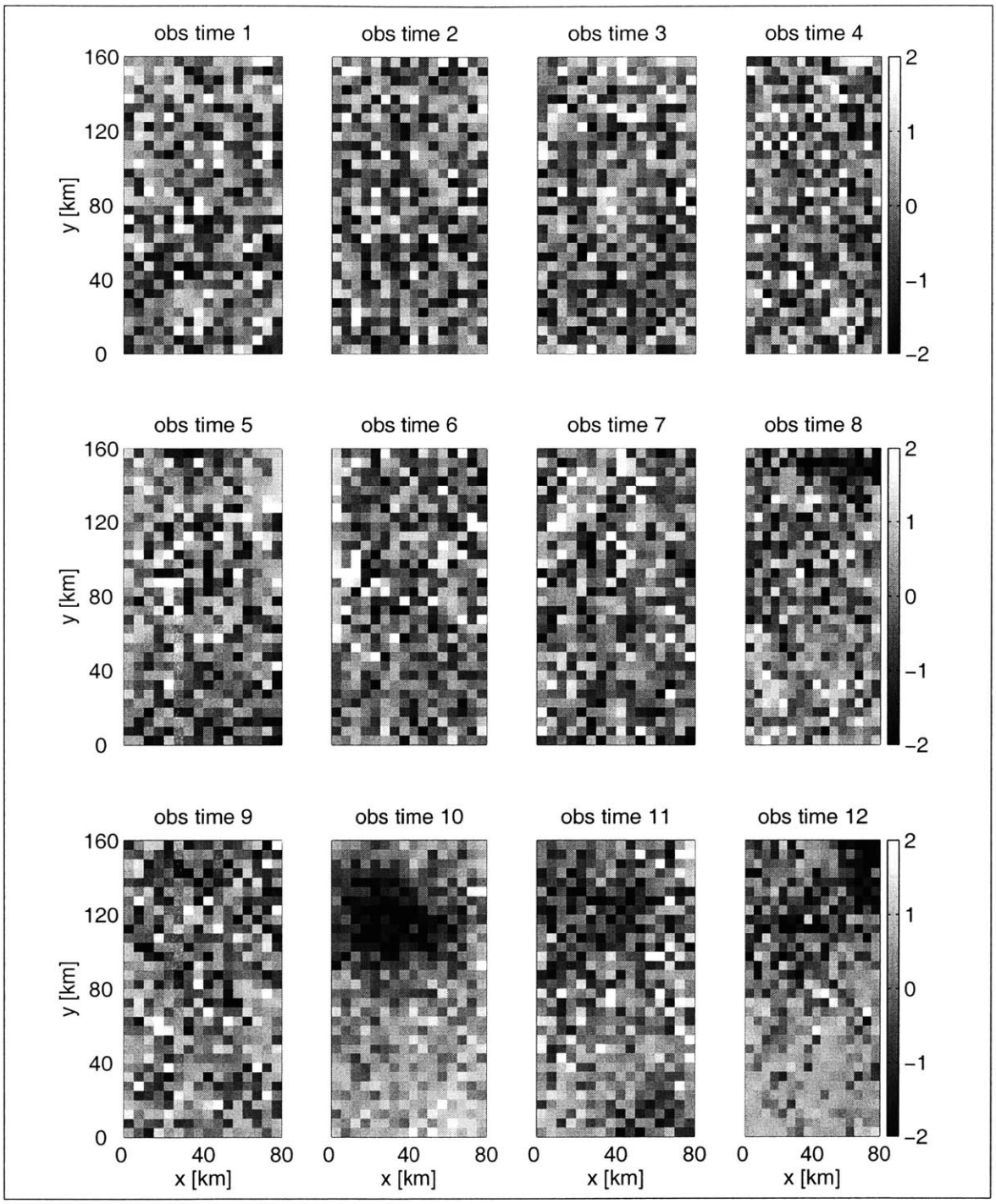


Figure 7.7: Standardized posterior data residuals when precipitation is withheld for the twelve brightness images that have been assimilated. The residuals of each image are standardized with the sample mean and standard deviation of the corresponding observation time. Some of the residual images show an obvious spatial structure, which indicates that the estimation algorithm does not work optimally.

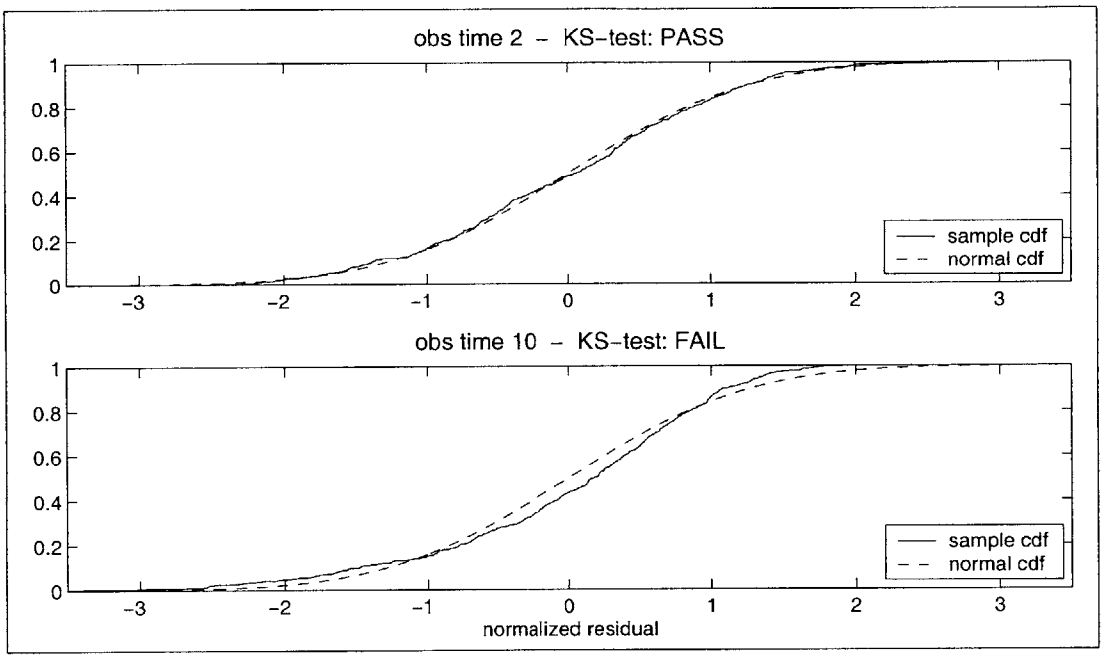


Figure 7.8: Sample cumulative distribution function (cdf) of the standardized posterior data residuals when precipitation is withheld for two of the twelve brightness images that have been assimilated (solid line). Also shown is the theoretical cumulative distribution function of the standard normal distribution (dashed line). For observation time 2 (upper panel), the residuals are close to normal and pass the Kolmogorov-Smirnov test for normality at a significance level of 5%. For observation time 10, however, the sample cdf is far from normal. The same is true for four other observation times.

residual images. For seven of the twelve observation times, the sample cdf is close to normal and passes the Kolmogorov-Smirnov test at a significance level of 5%. A representative example is observation time 2, which is shown in the upper panel of Figure 7.8. The lower panel shows an example of the five observation times for which the residuals are far from normally distributed. The examination of the brightness residuals thus corroborates that the assimilation was not optimal.

## 7.3 Assimilation with Poor Soil Hydraulic Parameters

Inevitably, the success of the estimation procedure depends on the quality of the model parameters that need to be specified. Many of these inputs, in particular the soil hydraulic parameters, are rather poorly known. Ideally, one would of course estimate the uncertain model parameters. Even though the data assimilation algorithm formulated in Chapter 2 is very general and provides for the estimation of model parameters, we have not implemented this feature in the synthetic experiments of this thesis for two reasons. First, the state estimation problem as implemented is already very complicated, and it is certainly wise to be conservative when specifying the uncertain inputs in a first application. Second, any hydrologic model that is to be used in an operational assimilation package had better be well calibrated. Estimating already calibrated parameters may make sense when one tries to improve the stability of the assimilation algorithm, but it is unlikely to be the most pressing problem in an operational context. Moreover, the parameter estimate adds significantly to the computational burden.

### 7.3.1 Experiment Design

Nevertheless, it is important to understand the sensitivity of the assimilation algorithm to poorly or wrongly specified model parameters. To address this issue, we have conducted a synthetic experiment in which we use different soil hydraulic parameters for the generation of the (synthetic) true fields and for the estimation. The setup of the experiment, including the (synthetic) true solution, is identical to Reference Experiment I of Section 6.1, with one exception. For the estimation, we change the soil hydraulic parameters of the land surface model by assigning the values from soil texture classes that have been randomly sampled from the existing soil texture classes. The resulting soil texture map is shown in Figure 7.9. Compared to the original map (Figure 5.3), 361 out of the 512 pixels differ in their texture classes.

We choose to randomly sample from existing soil texture classes for two reasons. First, the procedure guarantees that we only work with calibrated and tested input parameters. Unfortunately, the stability of the Richards' equation solver is fairly sensitive to the soil hydraulic parameters. Moreover, the scenario is realistic because in general we will have a good idea what soil texture classes occur in any given area, even though we may not accurately know their spatial distribution.

### 7.3.2 Estimation of the True Fields

Figure 7.10 shows the area average errors in the top node saturation of the prior and the estimate for the texture sensitivity experiment. Also shown is the error of the estimate from the ideal setup of Reference Experiment I (Figure 6.4). Even for wrongly specified soil

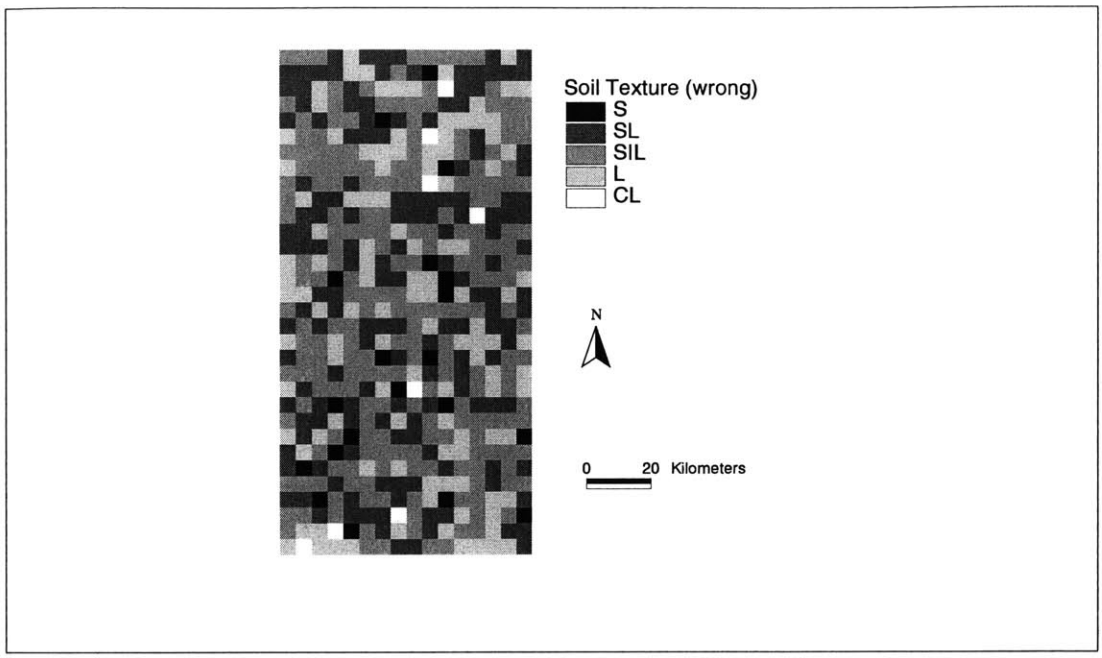


Figure 7.9: Soil texture classes for the texture sensitivity experiment. This "wrong" soil texture map has been derived by randomly subsampling from the original soil texture map (Figure 5.3). Compared to the original map, 361 out of the 512 pixels differ in their texture classes.

hydraulic parameters, the assimilation algorithm manages to estimate soil moisture satisfactorily. During precipitation events, however, the error increases. This is directly related to the fact that we supply wrong soil hydraulic parameters to the assimilation algorithm. Most of the infiltration happens during rain events, and exfiltration via evapotranspiration is highest just after the events. But with wrong soil hydraulic parameters, we cannot model the (synthetic) true moisture fronts accurately. This leads to higher errors during and after rain events.

Note that using the wrong soil hydraulic parameters leads to a prior error (Figure 7.10) which is only slightly higher than the error of Reference Experiment I (Figure 6.4), in which the true texture classes have been used. In Reference Experiment I, we already assume a very big uncertainty in the initial condition, and the prior rmse is about as big as it can be under the conditions of the experiment. Since we already know next to nothing about the initial condition distribution, the prior error does not increase much if we use the wrong soil hydraulic parameters.

### 7.3.3 Assessing the Optimality of the Estimates

Finally, Figure 7.11 shows the reduced objective function for the texture sensitivity experiment. The converged value of the reduced objective function is 7023, which is almost eight standard deviations above the expected value. This follows naturally from the fact that we use the same error statistics for the generation of the (synthetic) true solution and for the estimation, even though we changed the soil hydraulic parameters in the assimilation.

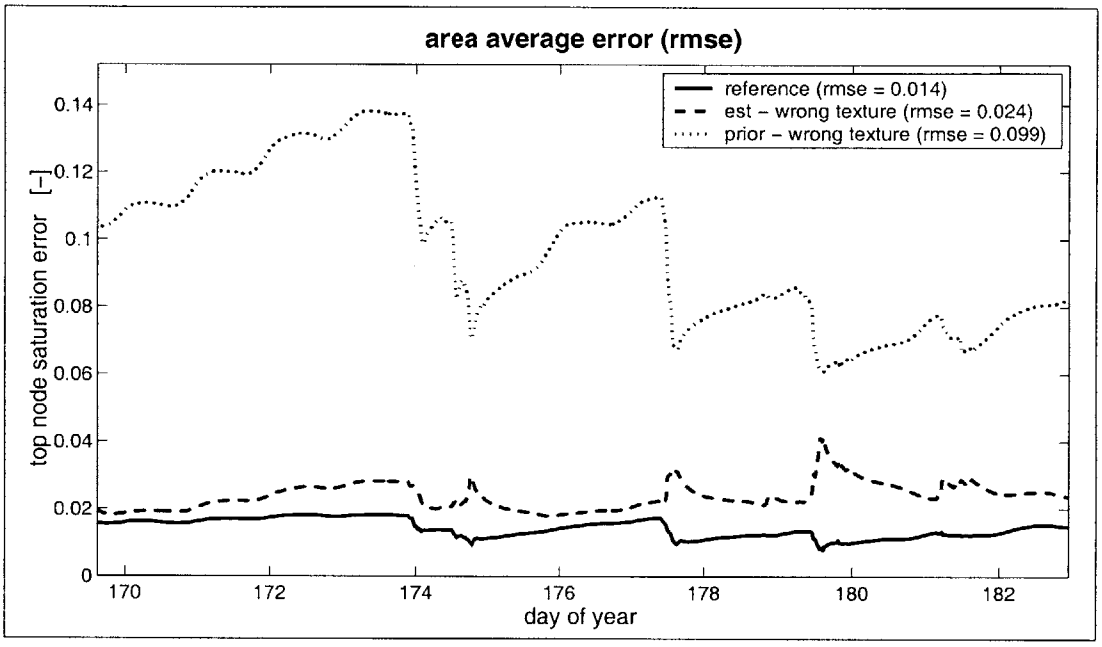


Figure 7.10: Area average errors for the texture sensitivity experiment. The errors in the top node saturation of the estimate, the prior, and the estimate of Reference Experiment I (Figure 6.4) are shown. The errors are in the root-mean-square sense with respect to the (synthetic) true solution. In the legend we also indicate the temporal mean of the area average rmse. Note that the soil moisture errors are in terms of saturation.

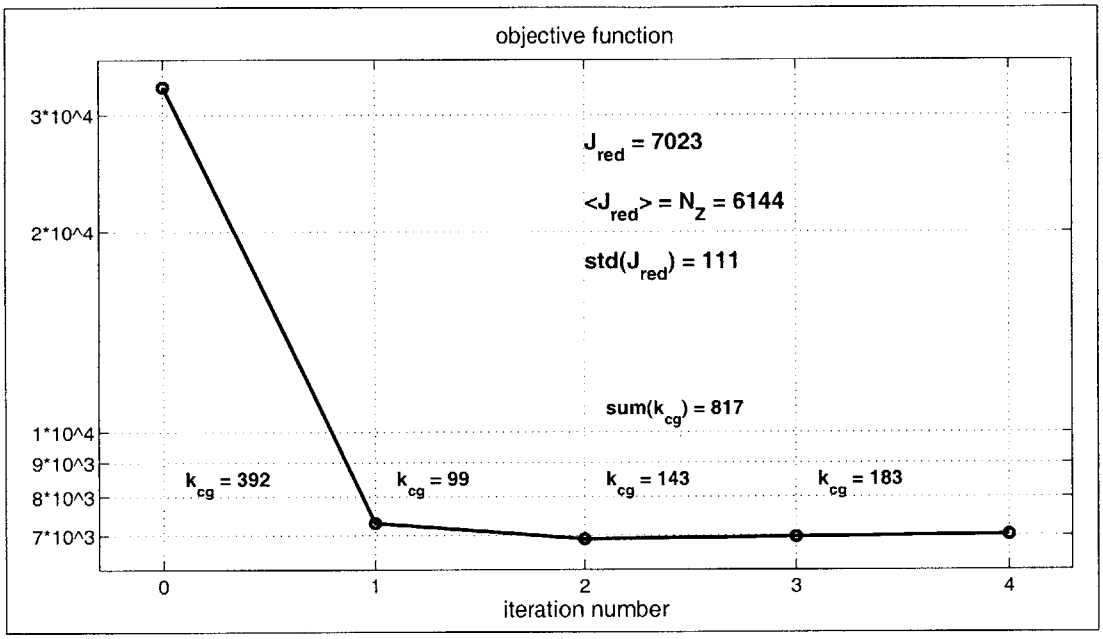


Figure 7.11: Objective function versus iteration number for the texture sensitivity experiment. The reduced objective function after convergence is 7023. The number of data points is 6144, which is also the expected value of the reduced objective function. Obviously, the estimates do not pass the hypothesis test on the value of the reduced objective function. The values of $k_{cg}$ indicate the number of linear combinations of representer functions that needed to be evaluated during the conjugate gradient iteration of the indirect representer approach (Chapter 8).

The change in model parameters leads to additional errors in the model, which are not adequately described with the original error statistics. This confirms that the assimilation is not operating optimally in this nonideal case.

On the other hand, a close look at the posterior data residuals yields qualitatively the same results as for Reference Experiment I (Figures 6.7, 6.8, and 6.9). The raw mean for all residuals with a 95% confidence interval is $0.13 \pm 0.13K$. For all but one of the residual brightness images we find a mean whose 95% confidence interval includes zero. There is no indication that the residuals are correlated in either space or time. Moreover, the sample cumulative distribution function for all but one of the residual images passes the Kolmogorov-Smirnov test for normality. In summary, this indicates that the estimation process is at least close to optimal.

# Chapter 8

# Computational Requirements

Optimal data assimilation has many advantages. Most importantly, we can derive truly best estimates, and we can assess whether the hydrologic model and our assumptions on its shortcomings are valid in the sense that they are statistically consistent with the assimilated data. Unfortunately, the computational burden for optimal data assimilation is also formidable. In this Chapter, we closely examine the computational requirements of the representer algorithm.

First, we review the computational demands of the synthetic experiments presented earlier and of a few more experiments that were not previously discussed (Section 8.1). Differences in the computational burden are explained and traced back to the design of the experiments. We then further investigate the computational requirements paying particular attention to the scalability of the approach (Section 8.2). This includes a review of the memory requirements and the possibility of parallel processing. Finally, we compare the computational requirements of the representer approach to other data assimilation methods (Section 8.3).

## 8.1 Computational Demands

In summary, the computational burden is influenced by three factors. First, the number of scalar data that are assimilated chiefly determines how many model integrations we need in order to derive the optimal estimates. Second, the number of nodes in the hydrologic model determines the CPU requirements for each model integration. Third, the computational effort is strongly influenced by the prior statistics and the number and length of the assimilation windows. The first two factors critically affect the scalability of the approach. We will postpone their discussion until Section 8.2. The last factors are the easiest to isolate, because it is straightforward to control the number of data and the size of the model. But they are also the hardest to understand (Section 8.1.2).

### 8.1.1 Computational Effort for the Synthetic Experiments

Table 8.1 summarizes the computational requirements of the experiments of Chapters 6 and 7 and of a few more experiments which were not presented earlier. The experiments are partitioned into three groups, according to the reference experiment they are based on. For the experiments that were presented earlier, we give the Section where a detailed description of the setup can be found. The other experiments are straightforward variations

| Experiment | | | ideal | $N_{ep}$ | $N_{ot}$ | $N_Z$ | $k_{tl}$ | $\Sigma k_{cg}$ | $N_{\text{int}}^{\text{ind}}$ | CPU [$h$] | Noise upd. [% CPU] | $N_{\text{int}}^{\text{ind}}/N_{\text{int}}^{\text{dir}}$ [%] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **Reference Exp. I** | Sec. 6.1 | yes | 512 | 12 | 6144 | 5 | 902 | 1819 | 30.5 | 3 | 3.0 |
| 2 | (1:4) downscaling exp. | Sec. 6.3 | yes | 512 | 12 | 1536 | 3 | 334 | 677 | 11.8 | 3 | 7.3 |
| 3 | (1:16) downscaling exp. | Sec. 6.3 | yes | 512 | 12 | 384 | 3 | 148 | 305 | 5.4 | 3 | 13.2 |
| 4 | 3-day repeat cycle | † | yes | 512 | 4 | 2048 | 4 | 565 | 1142 | 20.4 | 3 | 7.0 |
| 5 | 6-day repeat cycle | † | yes | 512 | 2 | 1024 | 3 | 365 | 739 | 13.2 | 3 | 12.0 |
| 6 | Rain withheld | Sec. 7.2 | no | 512 | 12 | 6144 | 7 | 594 | 1209 | 20.9 | 3 | 1.4 |
| 7 | Wrong texture | Sec. 7.3 | no | 512 | 12 | 6144 | 4 | 817 | 1646 | 28.5 | 3 | 3.3 |
| 8 | **Reference Exp. II** | Sec. 6.2 | yes | 512 | 12 | 6144 | 4 | 439 | 890 | 18.0 | 25 | 1.8 |
| 9 | (1:4) downscaling exp. | † | yes | 512 | 12 | 1536 | 3 | 183 | 375 | 7.3 | 26 | 4.1 |
| 10 | (1:16) downscaling exp. | † | yes | 512 | 12 | 384 | 2 | 78 | 162 | 3.3 | 25 | 10.5 |
| 11 | (1:64) downscaling exp. | † | yes | 512 | 12 | 96 | 3 | 47 | 103 | 2.2 | 27 | 17.6 |
| 12 | 3-day repeat cycle | Sec. 6.4 | yes | 512 | 4 | 2048 | 4 | 290 | 592 | 11.5 | 26 | 3.6 |
| 13 | 6-day repeat cycle | Sec. 6.4 | yes | 512 | 2 | 1024 | 4 | 234 | 480 | 9.8 | 25 | 5.9 |
| 14 | 12-day repeat cycle | † | yes | 512 | 1 | 512 | 4 | 136 | 284 | 5.6 | 25 | 6.9 |
| 15 | 3 assim. windows A‡ | Sec. 7.1 | no | 512 | 12 | 6144 | 3.4 | 181 | 372 | 6.9 | 22 | 0.9 |
| 16 | 12 assim. windows B‡ | Sec. 7.1 | no | 512 | 12 | 6144 | 2.8 | 45 | 98 | 2.4 | 18 | 0.3 |
| 17 | 12 assim. windows C‡ | Sec. 7.1 | no | 512 | 12 | 6144 | 2.2 | 33 | 73 | 2.0 | 16 | 0.3 |
| 18 | **Reference Exp. IIa** | † | yes | 128 | 12 | 1536 | 3 | 197 | 403 | 1.9 | 20 | 4.4 |
| 19 | (1:4) downscaling exp. | † | yes | 128 | 12 | 384 | 3 | 99 | 207 | 0.9 | 22 | 8.9 |
| 20 | (1:16) downscaling exp. | † | yes | 128 | 12 | 96 | 2 | 36 | 78 | 0.4 | 21 | 20.0 |
| 21 | 3-day repeat cycle | † | yes | 128 | 4 | 512 | 4 | 162 | 336 | 1.7 | 18 | 8.2 |
| 22 | 6-day repeat cycle | † | yes | 128 | 2 | 256 | 3 | 94 | 197 | 0.9 | 22 | 12.8 |

†No other results from these experiments were presented. ‡Some entries are averages over the assim. windows.

Table 8.1: Computational effort for the synthetic experiments. The number of estimation pixels is denoted with $N_{ep}$, the number of observation times with $N_{ot}$, the number of scalar data with $N_Z$, the number of outer (tangent-linear) iterations with $k_{tl}$, and the total number of inner (conjugate-gradient) iterations with $\Sigma k_{cg}$. For the number of model integrations we count all forward and backward runs separately. The total number of model integrations for the indirect representer method is $N_{\text{int}}^{\text{ind}} = 3k_{tl} + 2\Sigma k_{cg}$. The CPU time is measured on a DEC Alpha workstation with a clock speed of $333MHz$ and does not include I/O and preprocessing. We also indicate how much of the CPU time can be attributed to the process noise update. In the last column we compare the computational effort of the indirect representer method with the burden of the direct approach, which requires $N_{\text{int}}^{\text{dir}} = k_{tl}(2N_Z + 3)$ model runs.

on the ones presented throughout the thesis. For example, the repeat cycle experiments of the first group are designed in analogy to the repeat cycle experiments of Section 6.4, except that they are based on Reference Experiment I. Likewise, the downscaling experiments of the second group are designed similar to the ones in the first group but based on Reference Experiment II.

The experiments of the third group are variants of the second group at a coarser scale. For Reference Experiment IIa, we divide the same computational domain into $8 \times 16 = 128$ estimation pixels of $10km \times 10km$ each. The prior statistics are the same as in Reference Experiment II, except that we increase the correlation lengths by a factor of two (Section 8.2.2). Moreover, instead of spatially aggregating the true solution of Reference Experiment II, we use a different random seed when generating the synthetic true solution. This adds one more realization to the sparse body of data and helps corroborate the results.

For all experiments of Table 8.1, we indicate whether or not the assimilation conditions were ideal (Chapter 6). We then list the number of estimation pixels $N_{ep}$, the number of observation times $N_{ot}$, the number of scalar data $N_Z$, the number of outer (tangent-linear) iterations $k_{tl}$, and the total number of inner (conjugate-gradient) iterations $\Sigma k_{cg}$, where the sum is over all outer iterations. We also show the number of model integrations $N_{\text{int}}^{\text{ind}}$ that are required in the indirect representer method. The last three columns contain the CPU time, the proportion of the CPU time attributed to the process noise update, and the percentage of the computational effort required by the indirect representer approach when compared to the direct method.

The total count for the indirect representer method is $N_{\text{int}}^{\text{ind}} = 3k_{tl} + 2\Sigma k_{cg}$ model integrations, where we count the forward and backward runs separately and equally. For each outer iteration, we have to compute the representer coefficients with the conjugate gradient solver in an inner loop (Section 2.3.4). During each of the iterations of the inner loop, we have to compute one linear combination of adjoint representer fields and one linear combination of state representer fields, which explains the factor of two in front of $\Sigma k_{cg}$. In addition, for each outer iteration we need to calculate the prior trajectory (one integration) and, after obtaining the representer coefficients, we must finally solve the tangent-linear Euler-Lagrange equations for the state estimates (two integrations). This is why we need an additional $3k_{tl}$ model integrations.

By contrast, for the direct representer approach we need $N_{\text{int}}^{\text{dir}} = k_{tl}(2N_Z + 3)$ model integrations. During each outer iteration, we calculate the representer matrix by computing one prior trajectory, $N_Z$ adjoint representers, and $N_Z$ state representers. Finally, after obtaining the representer coefficients, we must calculate the estimated adjoint and state trajectories.

The CPU time is measured on a DEC Alpha workstation with a clock speed of $333MHz$. Note that the CPU times given in Table 8.1 do not include the preprocessing steps and input/output (I/O) operations. In the current implementation, all inputs and the process noise are kept in RAM, but the state (or the state representers) must be stored on disk. Depending on the type of experiment, the I/O operations increase the time for the assimilation by $5 \ldots 20\%$, where the higher number is for the experiments with twelve assimilation windows.

### 8.1.2 Computational Effort and Prior Statistics

If we compare the computational requirements of the experiments in the first and in the second group (Table 8.1), one obvious difference is the proportion of CPU time that goes into the process noise update. In Reference Experiment I, the model error is weak and the correlation length is only a fraction of the size of one estimation pixel. Consequently, we only spend about 3% of the CPU time on the process noise update, or approximately $2s$ of CPU time per model integration. In Reference Experiment II, the process noise variance is much stronger relative to the uncertainty in the initial condition, and, more importantly, the correlation length of the model error is increased fourfold. We now have to spend about 25% of the CPU time on the process noise update, which is equivalent to spending $18.5s$ per model run.

For small correlation lengths, we expect the burden for the process noise update per model integration to grow quadratically with the increase in the correlation length, because the convolution integral of the process noise update scales with the number of estimation pixels that are within one correlation length from any given point. This is compatible with the numbers given above if we also take boundary effects into account.

Moreover, the actual CPU requirements for the process noise update depend strongly on how many components we deem uncertain in the problem at hand. Considering that only one out of four components of the model equations is directly affected by process noise, and that the horizontal correlation length in Reference Experiment II is only 1.2 times the side of an estimation pixel, 25% is a significant share of the total computational effort. The share is so high because the model itself is not horizontally coupled and therefore extremely computationally efficient. The process noise update, by contrast, provides the horizontal coupling of the domain.

Despite the increased burden for the process noise, the computational effort for Reference Experiment II is still much less than for Reference Experiment I, even though the number of pixels and data is the same for both experiments. Whereas we need more than 30 hours of CPU time for Reference Experiment I, we only need 18 hours for Reference Experiment II. This can be explained by considering the condition number of the representer matrix.

Although we never explicitly compute the representer matrix, its condition number does of course affect the number of iterations needed in the conjugate gradient solver. In brief, the representer matrix encapsulates the correlation between the measurement predictions and the states. If there is very little model error, as was the case in Reference Experiment I, earlier and later brightness observations carry approximately the same information on the initial condition. As a consequence, the representer matrix is poorly conditioned. Since in Reference Experiment II the model error is strong, earlier and later brightness observations contain different amounts of information on the initial condition and on the model error at a given time. By increasing the model error variance and simultaneously decreasing the initial condition variance, we effectively improve the condition number of the representer matrix.

Another interesting comparison is between Reference Experiment I and the assimilation without precipitation data. Intriguingly, the computational effort for the latter experiment is only about two thirds of the computational burden of Reference Experiment I, although we do not supply the rain data to the assimilation algorithm. But by using different error statistics to compensate for the lack of precipitation data, we effectively improve the condition number of the representer matrix. Of course, we pay for the withholding of input data

and the increased convergence speed with a poorer estimate.

In summary, we can say that the prior statistics strongly influence the computational requirements. Increasing the horizontal correlation length of the model error leads to an increased computational burden for the process noise update. On the other hand, increasing the variance of the model error relative to the initial condition variance improves the condition number of the representer matrix and therefore eases the computational requirements.

### 8.1.3 Computational Effort for Multiple Assimilation Windows

We also observe a substantial decrease in the computational requirements if we go to multiple but shorter assimilation windows. Cutting the experiment period into three assimilation windows decreases the CPU time from 18 hours in Reference Experiment II to only 6.9 hours for experiment A for the entire two-week period. The savings are even more dramatic if we use twelve assimilation windows. For experiment B we only need 2.4 hours of CPU time. This decrease in CPU requirements is due to two factors.

First, we need relatively fewer model integrations as we increase the number of assimilation windows. For Reference Experiment II, we need 890 model runs to get the estimate, whereas in experiments A and B we only need on average 372 and 98, respectively. Second, the percentage of CPU time that goes into the process noise update decreases from 25% to 22% and 18%, respectively. This reflects the smaller relative importance of the model error compared to the initial condition uncertainty, because in experiments A and B we estimate the initial condition for each of the three or twelve assimilation windows. Although the reinitialization of the error covariances follows a naive scheme and the assimilation in these experiments is not strictly optimal, the estimates are nevertheless quite close to the optimal estimates of Reference Experiment II (Section 7.1). For a small sacrifice in optimality, the savings in computational effort are substantial enough to make assimilation intervals of a few days attractive in future applications.

## 8.2 Computational Effort and Scalability

For operational applications, we will have to tackle bigger problems, that is we must assimilate more data into larger domains. It is therefore of utmost importance to understand the scaling of the computational requirements as we increase the size of the problem. The two critical factors which determine the scaling are the number of scalar data that are assimilated and the number of nodes in the hydrologic model. In summary, the computational effort scales roughly with the product of the number of data and the number of estimation pixels, although the effort appears to grow somewhat less than linearly with the number of data. What follows is a more detailed discussion of this finding.

### 8.2.1 Number of Data and Efficiency of the Indirect Representer Method

The scaling of the computational effort with the number of data is closely related to the efficiency of the indirect representer approach. When we compare the computational burden of the indirect to the direct representer approach (Table 8.1), it is striking by how much the computational load is reduced in the indirect method. Typically, the effort for the indirect method is only a small percentage of what we would need in the direct approach. Of course, these huge savings come at a cost. The price is the ready availability of the posterior

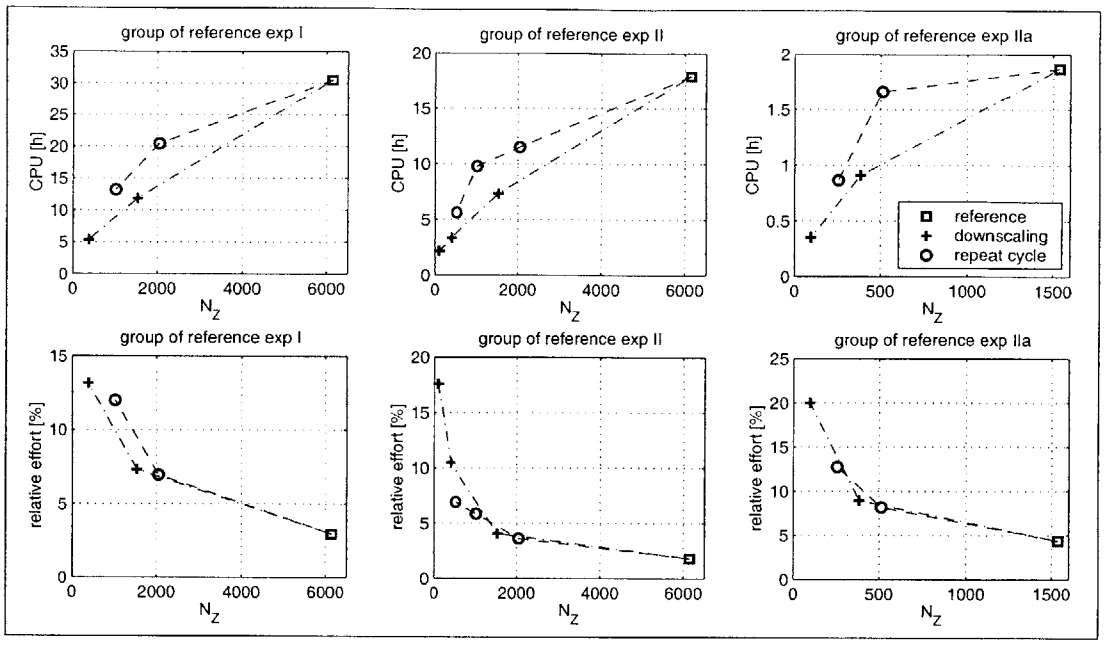


Figure 8.1: Computational effort for the ideal assimilation experiments. The top panels show the CPU requirements as a function of the number of scalar data that are assimilated. The lower panels show the computational effort of the indirect representer approach relative to the direct method. Obviously, there is redundancy in the data, and the computational burden grows less than linearly with the number of scalar data.

covariance information, which naturally comes with the direct method (Section 2.4). But this is a small price to pay, because it is certainly unnecessary to compute all of the posterior covariances at each outer iteration before the estimates are even dynamically consistent. We may still compute the individual representer fields after convergence has been achieved.

The efficiency of the indirect representer method does have an intuitive physical interpretation. In brief, there are many fewer degrees of freedom in the brightness images than there are scalar data, and the data are effectively compressed. To shed more light on this intuitive argument, it is helpful to investigate how the computational effort of the indirect representer method scales with the number of data. For reference, recall that the burden for the direct method scales linearly with the number of data (Section 8.1.1).

The top panels of Figure 8.1 show the CPU requirements for the three groups of synthetic experiments versus the number of scalar data. To avoid comparing apples and oranges, we only plot the ideal experiments of each group. For all three groups, the computational effort increases less than linearly with the number of data, suggesting that there is redundancy in the additional brightness observations. Another way to look at the same phenomenon is to plot the computational effort of the indirect method relative to the direct approach versus the number of data, which is done in the lower panels of Figure 8.1. We can see that the relative effort of the indirect method decreases as the number of data increases, again suggesting that there is relatively less information in the data as their number increases.

For the downscaling experiments, the result is obvious if we look at the spatial correlation scales of the dominant spatially distributed inputs, namely the soil texture and land cover

classes (Figure 5.3), the initial condition (Figure 6.3), and the precipitation. For all these inputs, the horizontal correlation scale is well above the scale of the estimation pixels, leading to redundant information in the brightness images. This finding confirms our hypothesis that the indirect representer approach and a priori data compression (Section 2.5) achieve related computational savings.

For the repeat cycle experiments, this result is not so obvious. From the 6-day repeat cycle experiment of Section 6.4 it is clear that an excellent estimate of the initial saturation can be obtained from only two brightness images at observation times 1 and 7. But the repeat cycle experiments based on Reference Experiment II have a strong model error which varies on a daily time scale. One would therefore expect that adding more information by assimilating more brightness images increases the computational effort proportionally. However, it appears that the brightness observations are effectively compressed in time as well as in space by the indirect representer method.

Unlike the spatial compression of individual images, such a compression in time is not straightforward, and it is not obvious whether it could be accommodated within the proposed data assimilation method by an a priori data compression scheme (Section 2.5). This means that the indirect representer approach is likely to be more efficient in reducing the computational effort than a priori data compression. But recall that a priori data compression greatly reduces the burden for calculating posterior error covariances (Section 2.5). Ideally, both schemes should be implemented.

Finally, it is interesting to take a quick look at the number of model integrations that are needed at each tangent-linear (outer loop) iteration to get the representer coefficients in the conjugate gradient solver. For Reference Experiment I, Figure 6.5 shows that $k_{cg} = 442$ (inner loop) iterations were necessary to derive the representer coefficients in the first (outer loop) iteration. In the remaining four tangent-linear (outer loop) iterations, on average only 115 such inner loop iterations were needed.

Similar results hold if we look at the other experiments (Figures 6.15, 6.18, 6.20, 7.5, and 7.11). On average, we only need about half the number of conjugate gradient iterations for the second and higher (outer) iterations than we do in the first (outer) iteration. This decrease in the computational burden is due to the much improved initial guess for the representer coefficients in the conjugate gradient solver. Obviously, we use the estimate for the representer coefficients from the previous iteration to initialize the conjugate gradient solver. This means that for nonlinear problems the indirect method is relatively more efficient (compared to the direct approach) than for linear problems.

### 8.2.2 Computational Effort, Resolution, and Size of the Domain

As we have seen above, the computational effort increases with the number of model runs that are required to solve for the estimate. One crucial factor determining the computational burden of the assimilation is therefore the time it takes to integrate the model once. Since the model is composed of a collection of independent vertical columns, the CPU requirements for each model integration will scale with the number of estimation pixels. Subtracting the effort for the model error update, each model integration covering the two-week experiment period takes approximately $60s$ of CPU time for the experiments of the first two groups with 512 estimation pixels. For the experiments of the third group with 128 estimation pixels, each model integration takes roughly $15s$. This translates into a CPU requirement of $0.01s$ per estimation pixel and per experiment day on the $333MHz$

DEC Alpha workstation.

Although the number of model integrations required for the estimate is mostly determined by the number of scalar data being assimilated, the number of nodes may affect the computational demand beyond merely increasing the CPU requirement for each model run. However, given all the other factors, in particular the complicated dependence of the computational burden on the prior statistics, it is hard to isolate the scaling of the computational demand with the number of nodes.

Suppose we increase the horizontal resolution of a given area, but we keep the horizontal correlation length for the model error fixed and we assimilate the same (coarse-scale) data. This will change the computational effort through at least three effects. First and trivially, each model integration will take longer at the fine resolution than at the coarse resolution. Second, one correlation length covers more estimation pixels at the fine resolution than at the coarse resolution, and the effort for the model error update will increase (Section 8.1.2). Third, the same(!) data may have different information content when viewed from the coarse and from the fine resolution of the estimation pixels, leading to different savings in the indirect representer method. This last effect is related to the increase in the relative computational effort of the indirect approach as we decrease the number of data by assimilating coarser brightness images (Section 8.2.1).

By comparing experiments with the same number of data from the second group and the third group (Table 8.1), we can check whether the computational effort scales only linearly with the number of estimation pixels. As mentioned above, we have increased the correlation length in Reference Experiment IIa and the related experiments with respect to Reference Experiment II. This makes the relative cost of the model error update approximately equal for both groups. If we now compare experiments of the two groups for which the same number of data are assimilated, we find that the CPU requirements increase approximately linearly with the number of estimation pixels. For example, if we divide the CPU time for experiment 9 by the CPU time for experiment 18, we get 3.9, while there are four times as many estimation pixels in experiment 9 than in experiment 18. For experiments 10 and 19, we get 3.7, and for experiments 14 and 21, we get 3.4. For experiments 11 and 20, on the other hand, we get 6.1, but this may be due to the extreme downscaling ratio of experiment 11. In summary, we have not found evidence that the computational effort increases more than linearly with the number of nodes.

### 8.2.3 Memory Requirements

The memory requirements for the assimilation algorithm are substantial. Ideally, all variables are kept in RAM (random-access memory) during the assimilation to keep the additional cost for I/O (input/output) operations low. As mentioned above, this was not possible with our computing equipment. But differences in speed aside, we can treat RAM and memory on the harddrive as being equivalent for the sake of this discussion. Also note that it is usually possible to trade off memory requirements against CPU time by recomputing dependent quantities each time they are needed rather than storing them.

We now discuss the memory requirements in detail. First, we have to store the model inputs. For the numerical parameters derived from the soil texture and land cover classes, the memory requirements are proportional to the number of estimation pixels and vertical nodes. For the micro-meteorologic inputs, the memory requirements are proportional to the length of the assimilation window. They are also proportional to the number pixels,

unless we opt for recomputing the interpolated fields whenever they are needed. Since during the assimilation the model must be run many times, it makes sense to keep the interpolated micro-meteorologic inputs at every pixel, together with some derived variables such as the land surface resistances, in RAM. In our implementation, the model inputs for the experiments of Chapters 6 and 7 require about $60MB$ in double precision.

Since the underlying hydrologic model is nonlinear, the adjoint operators depend on the state trajectories. Moreover, the variational method processes the entire assimilation window in a batch mode. It is therefore necessary to store the full state trajectory. In our example, the number of scalar states for all times is roughly $10^7$ (Sections 4.2 and 5.2). For double precision, this is equivalent to about $80MB$.

Moreover, we need to store the estimates of the uncertain inputs. The memory requirements for the initial condition parameters are small, but we must also store up to two copies of the model error fields. First, the adjoint operator depends on the model error, because the coefficient in front of the process noise term in the state equation depends on the state itself (Section 2.2). We must therefore store the model error estimate of the previous iteration. Second, we may have to store the model error that is under update in the current iteration. Depending on the implementation, we may be able to use the memory allocated for the state trajectories to store the current model error estimate. But since we do not keep the entire state trajectory in RAM due to memory limitations, this is not possible in our implementation.

The actual memory requirements for storing the process noise depend strongly on how many components of the model we deem uncertain in the problem at hand. However, the memory requirements will always scale with the length of the assimilation window. For the experiments of Chapters 6 and 7, we need about $26MB$ of memory for storing the model error.

### 8.2.4 Parallel Computing

By design, the direct representer method is ideally suited for parallel computing [Bennett and Baugh, 1992]. In theory, each of the individual representer fields can be integrated simultaneously on a separate processor. The indirect representer approach, by contrast, does not generally allow for such parallel processing. But in our case, the structure of the hydrologic model would allow us to integrate the individual estimation pixels in parallel. Ideally, with $N_{ep}$ processors, each model integration would only take as long as it takes to run a single one-dimensional column. Consequently, parallel computing offers great opportunities for the land surface data assimilation problem.

## 8.3 Comparison with Other Assimilation Techniques

In this Section, we briefly compare the computational requirements of the iterated indirect representer technique, the gradient-descent ("adjoint") method [Bennett, 1992], the full Kalman filter (KF) [Gelb, 1974], the Ensemble Kalman Filter (EnKF) (Section 1.2.2, [Evensen, 1994b]), error subspace statistical estimation (ESSE) (Section 1.2.2, [Lermusiaux and Robinson, 1999a]), and Optimal Interpolation (Section 1.2.1, [Daley, 1991]). Since we have only implemented the representer algorithm, the comparison with the other methods remains theoretical in nature. Nevertheless, an approximate count of the floating point operations provides sufficient insight into the characteristic requirements of each method.

Among the techniques mentioned above, the Kalman filter, the representer method, and the gradient-descent approach are optimal algorithms which fully propagate the error covariances. The big difference between these methods is that the Kalman filter propagates the error covariances explicitly, whereas the representer approach and the gradient-descent technique accomplish the error covariance propagation implicitly through the adjoint equations. The Ensemble Kalman Filter, ESSE, and Optimal Interpolation are suboptimal assimilation algorithms which rely on approximations of the dynamic evolution of the error covariances.

We now discuss the approximate operations count of each method for the experiments of Chapters 6 and 7. Recall from Sections 4.2 and 5.2 that the state vector at each time step is of length $N_x N_y (N_z + 5) = 6144$. With $N_t = 1280$ time steps, the total number of scalar states is $N_x N_y (N_z + 5) N_t \approx 10^7$. For each model integration, we simply count $\mathcal{O}(10^7)$ floating point operations. The dimension of the process noise in the example is $3 N_x N_y = 1536$ at every time step or a total of $3 N_x N_y N_t \approx 2 \cdot 10^6$ scalar variables. The number of data is 512 in each of the 12 images, and the total number of scalar data is $N_Z = 6144$. Note that we only discuss the leading order of the operations count.

### 8.3.1 Optimal Assimilation Algorithms

The explicit propagation of the error covariances in the Kalman filter requires manipulating matrices of the size of the state vector at each time step. The operations count for such matrix multiplications scales with the size of the matrix to the third power. Since this has to be done at every time step, the total count for the Kalman filter is $\mathcal{O}(10^{14})$. The actual operations count is even higher, because in the examples the process noise is colored, and the state must first be augmented. Note that solving the smoothing problem with the Kalman smoother approximately doubles the effort. The computational burden is so heavy because the Kalman filter comes with the full posterior covariance information.

As discussed above, the iterated indirect representer method scales with the time it takes to integrate the model once times the number of model integrations. Using 1000 model integrations and doubling the effort to accommodate the process noise update, the total operations count is $\mathcal{O}(10^{10})$. It is important to note that this does not include the posterior covariance information. We also stress that here we make use of the particular structure of the land surface model and the nature of the brightness observations, which implies that the results of this Section are valid only for the land surface data assimilation problem discussed in this thesis.

At first glance, the gradient-descent method requires only two model integrations at each iteration for the evaluation of the gradient. But the number of iterations needed for convergence strongly depends on the technique that is used for stepping down the gradient. The steepest-descent algorithm, which does not require any additional effort, is notoriously inefficient and should never be used [Press et al., 1992]. One alternative is to determine a "conjugate" gradient descent direction and to search for the minimum along this direction. For the line search, the model needs to be integrated multiple times and the objective function must be determined. If there is process noise, the latter step makes the method unattractive in practice, because the model error covariance is of size $10^6 \times 10^6$ and must be inverted. The problem becomes especially severe if the process noise is colored, as it is in our example. Regardless of the particular variant employed, the gradient-descent method is always impractical in our example because the search happens in the space of the uncertain

inputs, of which there are many more than there are scalar data.

### 8.3.2 Suboptimal Assimilation Algorithms

The effort for the Ensemble Kalman filter or related techniques, such as for example the error subspace statistical estimation (ESSE) scheme of Lermusiaux and Robinson [1999a], depends critically on the number of realizations or ensemble members that are used to approximate the error covariances. For a problem with 2800 data, a total of $10^8$ states, and no model error, Lermusiaux and Robinson [1999b] need approximately 200 ensemble members. Moreover, the decomposition of the forecast error covariance and the actual update step require additional matrix manipulations. If we use their expression for the operations count (Table 2 of [Lermusiaux and Robinson, 1999b]) and substitute our numbers, the ESSE scheme and the EnKF require roughly $\mathcal{O}(10^{11})$ floating point operations. Since these techniques approximate the error covariance propagation, the estimates are not optimal, but the methods provide valuable information about the posterior error covariances.

Finally, Optimal Interpolation as the most suboptimal scheme essentially requires one model integration and the manipulation of matrices which are of the size of the data vector at each update time. The operations count for Optimal Interpolation is roughly $\mathcal{O}(10^9)$.

### 8.3.3 Summary

In conclusion, the full Kalman filter or smoother cannot be used for large-scale land data assimilation. Likewise, the optimal adjoint-based gradient-descent method is not computationally feasible if model errors are present. From a computational point of view, the indirect iterated representer technique is competitive with suboptimal sequential Monte Carlo methods like the Ensemble Kalman Filter and ESSE. The great advantage of the representer technique lies with the optimality of the estimates. If posterior covariance information is required, the sequential Monte Carlo methods may be a better choice. Optimal Interpolation is of course the cheapest alternative, but the estimates are poorer and no posterior covariance information is provided.

# Chapter 9

# Conclusions

In Section 9.1 we first summarize the major original contributions of this thesis. Finally, in Section 9.2 we point out the limitations of the approach, the problems that still need to be addressed, and some possible solutions.

## 9.1 Summary of Original Contributions

In the introductory Section 1.3 we tried to assess the state of the art of soil moisture data assimilation. Having described our approach in detail in the previous Chapters, we now summarize the major original contributions of this work. In a nutshell, the five major contributions of this thesis are (1) the implementation of the optimal representer algorithm for a hydrologic data assimilation problem, (2) the development of a land surface model suitable for data assimilation, (3) the formulation of a general downscaling methodology for L-band passive microwave images, (4) the investigation of a number of topics which are crucial to the design of an operational soil moisture data assimilation system, and (5) the detailed assessment of the computational requirements of the algorithm for the soil moisture assimilation problem. Here is a more detailed summary of our contributions:

1. The optimal indirect iterated representer algorithm has been applied for the first time to a hydrologic data assimilation problem.

   (a) The assimilation algorithm is fully four-dimensional. We account explicitly for vertical and horizontal correlations.

   (b) The hydrologic model enters the assimilation as a weak constraint. Model error (or process noise) is taken into account.

   (c) The iterated representer method has been extended to allow for a nonlinear measurement operator, for a nonlinear dependence of the state on the parameters, and for a state-dependent coefficient multiplying the process noise.

   (d) Satellite radiances (or brightness temperatures) are assimilated directly. No off-line inversions of the remote sensing observations are necessary.

   (e) The formulation of the algorithm allows for the assimilation of various other data types, not only L-band passive microwaves.

   (f) Posterior error covariance calculations have been formulated for land surface hydrologic applications.

2. A land surface model suitable for hydrologic data assimilation has been developed.

   (a) Our model captures the key physical processes of the land-atmosphere boundary, and at the same time it is very computationally efficient.

   (b) The computational efficiency has been achieved foremost by dividing the model domain into laterally uncoupled one-dimensional vertical columns, which we call estimation pixels. Horizontal coupling has been incorporated through correlations of the inputs and of the model and parameter errors.

   (c) Our model fulfills the basic differentiability requirement for use in a variational assimilation algorithm.

   (d) For the soil moisture dynamics, any number of vertical nodes can be specified. It is easy to adjust the resolution of the soil moisture profile according to the problem at hand.

   (e) Soil and canopy temperatures are modeled together with soil moisture. This obviates the need to independently specify soil and canopy temperatures for the Radiative Transfer model. It also opens up the possibility of assimilating remotely sensed soil and canopy temperatures.

   (f) We have developed the adjoint of our land surface model.

3. A general downscaling capability has been incorporated into the assimilation algorithm, making it possible to effectively increase the resolution of the remote sensing brightness images, or equivalently, to estimate the land surface states at a scale finer than the resolution of the brightness images (Section 4.7).

4. We have conducted a series of synthetic experiments to test our land surface assimilation algorithm. These experiments demonstrate the usefulness of L-band radiobrightness data for soil moisture estimation under realistic conditions. We have assessed the optimality of the estimates for all experiments.

   (a) For a series of synthetic experiments under ideal and nonideal conditions, the assimilation algorithm can estimate the (synthetic) true land surface states to a high degree of accuracy.

   (b) From our results we conclude that large-scale soil moisture estimation from L-band passive microwave data is feasible.

   (c) In two reference experiments, we have successfully demonstrated the ability of the estimation algorithm to estimate the initial soil moisture conditions and the moisture flux boundary condition at the land surface (Sections 6.1 and 6.2).

   (d) We have investigated two typical downscaling scenarios which combine coarse-scale remote sensing data with fine-scale information from the model inputs (Section 6.3). Even for downscaling ratios of one to sixteen (each observation pixel contains sixteen estimation pixels), the estimate captures many of the fine-scale features of the true fields. This implies that brightness images with a resolution of $50km$ may be used to infer soil moisture on the scale of ten to twenty kilometers, provided sufficiently accurate model inputs are available at the finer scale.

(e) Repeat frequencies of up to three days for brightness images allow for satisfactory estimation of soil moisture conditions (Section 6.4). Generally, the repeat frequency should not be smaller than the typical frequency for rainstorms. If brightness data are available less frequently, the soil moisture estimates deteriorate rapidly.

(f) We have assessed the influence of the length of the data assimilation interval and the problem of reinitializing the windows in an operational fashion (Section 7.1). For an operational setup, it seems best to match the assimilation window approximately to an interstorm period. The assimilation intervals must be chosen such that there is sufficient time for the initial error covariance to evolve before the first observation time. This mitigates the negative effects of naively reinitializing the assimilation windows.

(g) We have demonstrated the ability of the assimilation algorithm to satisfactorily estimate soil moisture even if quantitative precipitation data are not available (Section 7.2).

(h) The assimilation algorithm is also capable of satisfactorily estimating soil moisture even if the soil hydraulic parameters are only poorly known (Section 7.3).

(i) For all experiments, we have thoroughly assessed the optimality of the estimates by examining the posterior data residuals and the value of the reduced objective function.

Even under ideal assimilation conditions, the nonlinear structure of the hydrologic model and the measurement operator leads to deviations of the residuals' sample cumulative distribution function from a normal distribution.

The residuals and the reduced objective function consistently show when the assimilation conditions were not ideal and the estimates were therefore suboptimal.

5. The computational requirements of the assimilation algorithm have been assessed in detail.

(a) For a typical application, the computational effort of the iterated indirect representer method grows less than linearly with the number of scalar data (Section 8.2.1).

(b) The computational effort grows only linearly with the number of pixels (Section 8.2.2).

(c) The prior statistics and the length of the assimilation interval strongly influence the computational requirements (Section 8.1.2 and 8.1.3). Assimilation windows of a few days are computationally very attractive while providing estimates that are close to optimal.

(d) For the land surface data assimilation problem discussed in this thesis, the iterated indirect representer approach is competitive from a computational perspective with suboptimal sequential Monte Carlo methods such as the Ensemble Kalman Filter (EnKF) and error subspace statistical estimation (ESSE) (Section 8.3). For a given computational cost, the trade-off is between the optimality of the estimates (representer method) and the availability of posterior covariance information (EnKF, ESSE).

## 9.2 Limitations and Problems to be Addressed

We now provide some comments on the research presented throughout this thesis. These comments are meant to serve three main purposes. First, we hope to raise the reader's awareness of any difficulties that may arise when implementing the methods described in this work. Second, we suggest areas that require more research, and perhaps entirely new directions of approaching the problem. Last but not least, we point at possible solutions that we think are appropriate but did not investigate for lack of time.

Given the highly nonlinear structure of land-atmosphere processes, and given the high complexity of real world applications, any large-scale land surface data assimilation algorithm will necessarily be a compromise between realistic physical representations and computational feasibility. A lot more research still needs to be done before an operational soil moisture data assimilation package can be set up. Here is a list of caveats, possible solutions, and suggestions for future research:

1. So far, the assimilation algorithm has only been verified with a few synthetic experiments. A field test and further synthetic experiments must be carried out.

   (a) All results have been obtained with a few synthetic experiments. Even though we made an effort to mimic realistic conditions, we have only discussed a total of three different realizations. There can be no guarantee that the results remain completely unchanged if more synthetic experiments are added which rely on different random seeds.

   (b) The results may depend on the particular choices of the error covariances. As we have seen Section 8.1.2, the computational demand is already quite sensitive to the prior statistics. To corroborate the findings, these intuitive choices will have to be validated in a field application, or, alternatively, the same outcome would have to be found for synthetic experiments with many different setups.

   (c) Before the assimilation algorithm can be applied to any particular field experiment, a robust calibration of the hydrologic model must be carried out for the field site in question.

   (d) The next, indispensable step is to conduct a field test of the algorithm by assimilating the SGP97 ESTAR observations. This will bring up difficult issues of model bias. Using the hypothesis tests described in Sections 2.3.6 and 2.4.1, we will then be able to thoroughly check whether the model and the statistical assumptions are appropriate in the field. In other words, the field test will help to identify realistic prior error statistics for land surface data assimilation.

2. The computation of the posterior error covariances must be implemented.

   (a) Ideally, when running the estimator in an operational mode, the prior error covariance of the initial condition should be specified or at least approximated respecting the posterior error covariances of the previous assimilation window. If this turns out to be too expensive to do operationally, the computation of the posterior error covariances in research studies will still provide valuable insights into the accuracy of the estimates and the operating conditions of the assimilation algorithm.

(b) It is certainly unnecessary in any case to compute the full posterior error covariances of the state and the measurements. Suitable approximations for the calculation of the posterior error covariances must be identified.

(c) A priori data compression should be implemented to decrease the cost of calculating the posterior covariances. There may also be additional savings from a priori data compression that are not realized with the indirect representer method. This topic is worth a detailed investigation because of the practical relevance of such savings.

3. The application and interpretation of Richards' equation at the regional scale clearly deserve more attention.

   (a) In our approach, the second-order Richards' equation is but a means to model moisture fronts that move nonlinearly in both directions. Any simple formulation capable of producing such upward and downward movement will in practice resemble Richards' equation.

   (b) Consequently, we do not claim that the small-scale soil hydraulic properties published in the literature can necessarily be interpreted as the parameters that we should use for our large-scale Richards' equation. Moreover, we do not interpret the soil moisture values in different vertical layers as a soil moisture profile that could be verified with ground-based point observations.

   (c) In our interpretation, Richards' equation can be viewed as governing a nonlinear multilayer reservoir of soil moisture that is reasonably consistent with the true land surface fluxes and with the groundwater recharge. The verification of this hypothesis and of the validity of the profile estimates will have to come from data assimilation experiments which are conducted over long experiment periods in a quasi-operational manner (Section 6.1.3).

4. In the current implementation, the assimilation algorithm depends heavily on the information needed to run the hydrologic model. To relax this dependence, other data types must be assimilated.

   (a) L-band brightness temperatures are determined not only by soil moisture, but also by soil temperature. Therefore our ability to estimate soil moisture from observations of the L-band brightness temperature critically depends on how accurately we know soil temperature. The experiments of Chapters 6 and 7 are not designed to test the capability of the algorithm to estimate the soil temperature. Instead, the soil temperature was rather well known by design. In an operational setting, such good knowledge of soil temperature can only be achieved by assimilating supplemental data, for example infrared remote sensing observations.

   (b) If satellite derived soil skin temperatures are assimilated, the one-layer soil temperature force-restore approximation may have to be extended to a two layer model. This would, however, conflict with the desire to keep the model as computationally efficient as possible for the transition towards continental-scale operational assimilation.

(c) The precipitation inputs to the land surface model are crucial for the quality of the soil moisture estimates. In the current implementation, errors in the micrometeorologic forcings including precipitation are lumped into the model error terms. In other words, we treat precipitation as a model parameter which is not explicitly estimated.

If the precipitation data are poor, it may be beneficial to assimilate these data into a coupled land surface and precipitation model. Note, however, that a simple scheme would not offer much information on the complicated nature of the precipitation processes. A sophisticated precipitation model, on the other hand, may prove hard to invert or too computationally expensive.

5. The experiments of Chapters 6 and 7 represent only a small fraction of the possibilities to study land surface data assimilation problems with our algorithm. Many other useful synthetic experiments can be thought up.

   (a) Together with the soil properties, the land cover parameters are critical for estimating soil moisture and other land surface variables. Therefore, the impact of the vegetation on the quality of the estimates must be investigated in more detail.

   (b) We have not fully explored the potential to estimate land surface variables other than soil moisture. Suitable experiments must be designed to assess the capability of the algorithm to infer soil temperature, canopy temperature, and the land surface fluxes from L-band passive microwaves and possibly other remote sensing observations.

   (c) Operational brightness observations at higher frequencies are already available or will soon become available. The algorithm could be used to investigate the trade-off of assimilating many data of poorer quality with respect to soil moisture, such as C-band ($5.3GHz$) observations, versus fewer data of higher quality, such as L-band ($1.4GHz$) observations.

6. The land surface model may require modifications for certain applications.

   (a) The measurement operator for the general downscaling methodology must be re-evaluated when only C-band brightness temperatures are available for assimilation (instead of L-band brightness). The use of the arithmetic mean over the estimation pixels within each observation pixel becomes questionable at higher microwave frequencies.

   (b) The rainfall interception model proved very difficult to include in the assimilation algorithm. This is not surprising given the disparity in the time scales of canopy interception and of the other land surface states. Since the quantity of water stored in the canopy is very small, the influence of the interception process on the soil moisture estimates is limited. On the other hand, canopy interception provides invaluable information about brightness data. Due to complicated reflections at and within the canopy, brightness images are essentially useless when the canopy is wet.

   Simply neglecting canopy interception altogether, as we have done so far, is one way out of the dilemma, but this works only when the brightness images are

subject to careful quality control. In future implementations, it might be beneficial to run the interception model off-line and in forward mode only. Brightness images taken at times when the canopy is wet can then be discarded in a quality control step.

(c) For continental-scale applications, our land surface model will have to be re-evaluated. The physics of the model may not be appropriate at much larger scales, although no obvious alternatives spring to mind. In addition, more computational savings may be required before a continental-scale application becomes feasible.

(d) It is desirable to couple the land surface model with a runoff and a groundwater model in order to more accurately represent land surface hydrologic processes.

(e) Any change in the land surface model must also be reflected in its adjoint.

7. Since the iterated indirect representer algorithm is the cheapest optimal approach, it constitutes an invaluable benchmark. The actual computational requirements, however, may defy its use for operational applications.

(a) The results on the computational demands and the scalability of the algorithm (Chapter 8) must be corroborated or modified through numerical experiments conducted with more pixels and longer experiment periods.

(b) With its CPU and memory requirements, the iterated indirect representer method may be too expensive for an operational assimilation package for some time still. Moreover, there is little hope of calculating accurate and detailed posterior error covariances even in research studies.

(c) Nonetheless, the iterated indirect representer approach is an efficient way to derive the optimal estimates, something which is not possible with a Kalman filter or an adjoint-based gradient search at the same computational expense. The representer method can therefore be used as a benchmark against which cheaper but less optimal algorithms can be evaluated.

# Appendix A

# Data Assimilation

## A.1 Derivation of the Euler-Lagrange Equations

This Section outlines the derivation of the Euler-Lagrange equations presented in Section 2.1.3. In order to find the minimum of the objective function, we perform a first variation on $J$.

$$\delta J = \frac{\delta J}{\delta X}\delta X + \frac{\delta J}{\delta Y}\delta Y + \frac{\delta J}{\delta \nu}\delta \nu + \frac{\delta J}{\delta \omega}\delta \omega + \frac{\partial J}{\partial \alpha}\delta \alpha + \frac{\partial J}{\partial \beta}\delta \beta + \frac{\delta J}{\delta \mu}\delta \mu + \frac{\delta J}{\delta \lambda}\delta \lambda + \frac{\partial J}{\partial \lambda_0}\delta \lambda_0 \quad \text{(A.1)}$$

At the minimum we have $\delta J = 0$. Since all variations are considered arbitrary and independent, each of the individual partial derivatives of the objective function must vanish.

Obviously, variation of (2.7) with respect to the adjoint parameters $\mu$, $\lambda$ and $\lambda_0$ simply returns the state equation, i.e. the forward equation (2.8) and its initial condition (2.8*b*) in the set of Euler-Lagrange equations.

From the variation of (2.7) with respect to the parameters $\alpha$ and $\beta$, we get

$$\left[ 2(\alpha - \overline{\alpha})^T C_\alpha^{-1} - 2\int_0^{t_f} \mu^T \frac{\partial \phi}{\partial \alpha} dt - 2\int_0^{t_f} \lambda^T \frac{\partial \varphi}{\partial \alpha} dt \right] \delta\alpha = 0$$

and

$$\left[ 2(\beta - \overline{\beta})^T C_\beta^{-1} - 2\lambda_0^T \frac{\partial Y_0}{\partial \beta} \right] \delta\beta = 0$$

Using $\lambda_0 = \lambda|_{t=0}$ (see below), these are the parameter update equations (2.10).

Next, variation of (2.7) with respect to $\omega$ yields

$$2\int_0^{t_f} \left[ \int_0^{t_f} \omega(t')^T C_\omega^{-1}(t', t'')dt' - \lambda(t'')^T D_\omega(t'')P_\omega \right] \delta\omega(t'')dt'' = 0$$

and therefore

$$\int_0^{t_f} \omega(t')^T C_\omega^{-1}(t', t'')dt' = \lambda(t'')^T D_\omega(t'')P_\omega$$

Postmultiplying by $C_\omega(t'',t)$, integrating over $t''$, and using (2.6), we find

$$\omega(t)^T = \int_0^{t_f} \lambda(t'')^T D_\omega(t'') P_\omega C_\omega(t'',t) dt''$$

which is the process noise update (2.11$a$). Similarly, we get the update equation for $\nu$.

Variation with respect to $X$ yields

$$-2\left(Z - M[X,Y]\right)^T C_v^{-1} \frac{\partial M[X,Y]}{\partial X}\delta X - 2\int_0^{t_f} \mu^T \frac{\partial \phi}{\partial X}\delta X dt - 2\int_0^{t_f} \lambda^T \frac{\partial \varphi}{\partial X}\delta X dt = 0$$

In this expression, we substitute from (2.3)

$$\frac{\partial M[X,Y]}{\partial X} = \int_0^{t_f} [\delta] \frac{\partial f(X(t),Y(t))}{\partial X} dt$$

and get the first part of the backward equation after collecting all terms under a single integral and setting the integrand to zero.

In order to perform the variation with respect to the state $Y$, we start from (2.7) and integrate by parts the term containing $\frac{\partial Y}{\partial t}$.

$$\begin{aligned}
J = &\left(Z - M[X,Y]\right)^T C_v^{-1} \left(Z - M[X,Y]\right) \\
&+ \left(\alpha - \overline{\alpha}\right)^T C_\alpha^{-1} \left(\alpha - \overline{\alpha}\right) + \left(\beta - \overline{\beta}\right)^T C_\beta^{-1} \left(\beta - \overline{\beta}\right) \\
&+ \int_0^{t_f}\int_0^{t_f} \nu(t')^T C_\nu^{-1}(t',t'')\nu(t'')dt'dt'' + \int_0^{t_f}\int_0^{t_f} \omega(t')^T C_\omega^{-1}(t',t'')\omega(t'')dt'dt'' \\
&- 2\int_0^{t_f} \mu^T \left(\phi(X,Y;\alpha) + D_\nu P_\nu \nu\right) dt \\
&+ 2\left(\lambda^T Y\right)|_{t=t_f} - 2\left(\lambda^T Y\right)|_{t=0} - 2\int_0^{t_f} \left(\frac{\partial \lambda}{\partial t}^T Y + \lambda^T \varphi(X,Y;\alpha) + \lambda^T D_\omega(Y) P_\omega \omega\right) dt \\
&+ 2\lambda_0^T \left(Y|_{t=0} - Y_0(\beta)\right)
\end{aligned}$$

From the variation with respect to $Y|_{t=t_f}$ we immediately get the terminal condition (2.9$a$) for the backward equation. (Recall that $t_m \in (0,t_f)$, i.e. there are no measurements at the final time $t_f$). Similarly, variation with respect to $Y|_{t=0}$ yields

$$\lambda|_{t=0} = \lambda_0$$

Finally, variation with respect to the state $Y(t)$, $t \in (0,t_f)$, yields

$$\begin{aligned}
&-2\left(Z - M[X,Y]\right)^T C_v^{-1} \frac{\partial M[X,Y]}{\partial Y}\delta Y \\
&\qquad - 2\int_0^{t_f} \mu^T \frac{\partial \phi}{\partial Y}\delta Y dt - 2\int_0^{t_f} \left[\frac{\partial \lambda}{\partial t}^T + \lambda^T \frac{\partial \varphi}{\partial Y} + \lambda^T \frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right] \delta Y dt = 0
\end{aligned}$$

In this expression, we substitute from (2.3)

$$\frac{\partial M[X,Y]}{\partial Y} = \int_0^{t_f} [\delta] \frac{\partial f(X(t),Y(t))}{\partial Y} dt$$

and get the backward equation after collecting all terms under a single integral and setting the integrand to zero.

## A.2 Derivation of the Posterior Covariance Equations

This Section outlines the derivation of the posterior covariance equations of Section 2.4. Recall that the problem at hand is nonlinear, and that the linearized posterior covariances derived below are at best approximations of the true posterior covariances. In particular, we treat the previous estimate as a fixed deterministic input, although strictly speaking the previous estimate depends on the data and therefore on the measurement error.

### A.2.1 Equivalence of Representers and Prior Cross-Covariances

We first prove the fact that the state representers are equal to the linearized (prior) cross-covariances of the measurement predictions and the states.

$$\overline{L_k[X',Y']X'(t)} = \Xi^k(t) \qquad \overline{L_k[X',Y']Y'(t)} = \Upsilon^k(t) \tag{2.34}$$

The idea is to show that $\overline{L_k[X',Y']X'(t)}$ and $\overline{L_k[X',Y']Y'(t)}$ obey the same differential equations as $\Xi^k(t)$ and $\Upsilon^k(t)$, which are of course the state representer equations (2.25).

From the tangent-linear state equation (2.13) and the equation for the prior state (2.22), we obtain an equation for the perturbation of the state $X' \equiv X - \overline{X}^{\eta+1}$ and $Y' \equiv Y - \overline{Y}^{\eta+1}$.

$$\begin{aligned} 0 &= \left.\frac{\partial \phi}{\partial X}\right|_\eta X' + \left.\frac{\partial \phi}{\partial Y}\right|_\eta Y' + \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta \alpha' + D_\nu P_\nu \nu \\ \frac{\partial Y'}{\partial t} &= \left.\frac{\partial \varphi}{\partial X}\right|_\eta X' + \left.\frac{\partial \varphi}{\partial Y}\right|_\eta Y' + \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta \alpha' + D_\omega(Y^\eta) P_\omega \omega + \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta Y' \end{aligned} \tag{A.2}$$

$$Y'|_{t=0} = \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta \beta' \tag{A.2a}$$

For the perturbations of the parameters we write $\alpha' \equiv \alpha - \overline{\alpha}$ and $\beta' \equiv \beta - \overline{\beta}$.

Next, we multiply (A.2) with the scalar $L_k[X',Y']$ and take the expectation. We get

$$\begin{aligned} 0 &= \left.\frac{\partial \phi}{\partial X}\right|_\eta \overline{L_k[X',Y']X'(t)} + \left.\frac{\partial \phi}{\partial Y}\right|_\eta \overline{L_k[X',Y']Y'(t)} + \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta \overline{L_k[X',Y']\alpha'} \\ &\quad + \overline{L_k[X',Y'] D_\nu P_\nu \nu} \\ \frac{\partial}{\partial t}\overline{L_k[X',Y']Y'(t)} &= \left.\frac{\partial \varphi}{\partial X}\right|_\eta \overline{L_k[X',Y']X'(t)} + \left.\frac{\partial \varphi}{\partial Y}\right|_\eta \overline{L_k[X',Y']Y'(t)} + \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta \overline{L_k[X',Y']\alpha'} \\ &\quad + \overline{L_k[X',Y'] D_\omega(Y^\eta) P_\omega \omega} + \overline{L_k[X',Y'] \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta Y'} \end{aligned}$$

$$\overline{L_k[X',Y']Y'|_{t=0}} = \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta \overline{L_k[X',Y']\beta'}$$

Comparing these equations to the state representer equations (2.25), we see that the fact (2.34) holds if the following four identities hold.

$$\overline{L_k[X',Y']\alpha'} = C_\alpha \int_0^{t_f} \left( \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta^T \Omega^k(t') + \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta^T \Lambda^k(t') \right) dt' \tag{A.3}$$

$$\overline{L_k[X',Y']D_\nu P_\nu \nu} = \int_0^{t_f} D_\nu P_\nu C_\nu(t,t') P_\nu^T D_\nu^T \Omega^k(t') dt' \tag{A.4}$$

$$\overline{L_k[X',Y']D_\omega(Y^\eta)P_\omega \omega} = \int_0^{t_f} D_\omega(Y^\eta(t)) P_\omega C_\omega(t,t') P_\omega^T [D_\omega(Y^\eta(t'))]^T \Lambda^k(t') dt' \tag{A.5}$$

$$\overline{L_k[X',Y']\beta'} = C_\beta \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta^T \Lambda^k|_{t=0} \tag{A.6}$$

For the proof of (A.3)-(A.6) we seek an expression of the form

$$L_k[X',Y'] \approx \delta M_k[X,Y] = \frac{\delta M_k}{\delta X}\delta X + \frac{\delta M_k}{\delta Y}\delta Y = \frac{\delta M_k}{\delta \nu}\delta\nu + \frac{\delta M_k}{\delta \omega}\delta\omega + \frac{\delta M_k}{\delta \alpha}\delta\alpha + \frac{\delta M_k}{\delta \beta}\delta\beta$$

To this end, we apply an adjoint technique which avoids the explicit computation of $\frac{\delta M_k}{\delta X}$ and $\frac{\delta M_k}{\delta Y}$ and yields the derivatives with respect to $\nu$, $\omega$, $\alpha$, and $\beta$ directly. We first define as an objective the function for which we need the derivatives.

$$\tilde{J}^k = M_k[\overline{X}^{\eta+1}, \overline{Y}^{\eta+1}] + L_k[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}] \tag{A.7}$$

In order to satisfy the relation between the states $X$, $Y$ and the inputs $\nu$, $\omega$, $\alpha$, and $\beta$, we adjoin the tangent-linear state equation (2.13) to the objective.

$$\begin{aligned} J^k = {} & M_k[\overline{X}^{\eta+1}, \overline{Y}^{\eta+1}] + L_k[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}] \\ & + \int_0^{t_f} (\tilde{\mu}^k)^T \left( \phi(X^\eta, Y^\eta; \alpha^\eta) + \left.\frac{\partial \phi}{\partial X}\right|_\eta (X - X^\eta) + \left.\frac{\partial \phi}{\partial Y}\right|_\eta (Y - Y^\eta) \right. \\ & \qquad \left. + \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta (\alpha - \alpha^\eta) + D_\nu P_\nu \nu \right) dt \\ & - \int_0^{t_f} (\tilde{\lambda}^k)^T \left( \frac{\partial Y}{\partial t} - \varphi(X^\eta, Y^\eta; \alpha^\eta) - \left.\frac{\partial \varphi}{\partial X}\right|_\eta (X - X^\eta) - \left.\frac{\partial \varphi}{\partial Y}\right|_\eta (Y - Y^\eta) - \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta (\alpha - \alpha^\eta) \right. \\ & \qquad \left. - D_\omega(Y^\eta) P_\omega \omega - \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta (Y - Y^\eta) \right) dt \\ & \qquad - (\tilde{\lambda}_0^k)^T \left( Y|_{t=0} - Y_0(\beta^\eta) - \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta (\beta - \beta^\eta) \right) \end{aligned} \tag{A.8}$$

Using partial integration we substitute

$$-\int_0^{t_f} (\tilde{\lambda}^k)^T \frac{\partial Y}{\partial t} dt = +\int_0^{t_f} \frac{\partial \tilde{\lambda}^k}{\partial t}^T Y dt - \left((\tilde{\lambda}^k)^T Y\right)|_{t=t_f} + \left((\tilde{\lambda}^k)^T Y\right)|_{t=0}$$

in (A.8) and then use the resulting expression to compute the first variation of $J^k$.

$$\delta J^k = \frac{\delta J^k}{\delta X}\delta X + \frac{\delta J^k}{\delta Y}\delta Y + \frac{\delta J^k}{\delta \nu}\delta\nu + \frac{\delta J^k}{\delta \omega}\delta\omega + \frac{\delta J^k}{\delta \alpha}\delta\alpha + \frac{\delta J^k}{\delta \beta}\delta\beta \tag{A.9}$$

We now choose $\tilde{\mu}^k$, $\tilde{\lambda}^k$ and $\tilde{\lambda}_0^k$ such that $\frac{\delta J^k}{\delta X}\delta X \equiv \frac{\delta J^k}{\delta Y}\delta Y \equiv 0$, because then we have

$$\delta J^k = \delta \tilde{J}^k \equiv \delta L_k[X, Y] = L_k[X', Y'] \tag{A.10}$$

First, the variation of $J^k$ with respect to $X$ and $Y$ yields

$$\begin{aligned}
\frac{\delta J^k}{\delta X}\delta X &= \int_0^{t_f} \left( \delta(t - t_k) \left.\frac{\partial f_k}{\partial X}\right|_\eta + (\tilde{\mu}^k)^T \left.\frac{\partial \phi}{\partial X}\right|_\eta + (\tilde{\lambda}^k)^T \left.\frac{\partial \varphi}{\partial X}\right|_\eta \right) \delta X dt = 0 \\
\frac{\delta J^k}{\delta Y}\delta Y &= \int_0^{t_f} \left( \delta(t - t_k) \left.\frac{\partial f_k}{\partial Y}\right|_\eta + (\tilde{\mu}^k)^T \left.\frac{\partial \phi}{\partial Y}\right|_\eta + \frac{\partial \tilde{\lambda}^k{}^T}{\partial t} + (\tilde{\lambda}^k)^T \left.\frac{\partial \varphi}{\partial Y}\right|_\eta \right. \\
&\qquad \left. + (\tilde{\lambda}^k)^T \left.\frac{\partial [D_\omega(Y) P_\omega \omega]}{\partial Y}\right|_\eta \right) \delta Y dt = 0
\end{aligned} \tag{A.11}$$

as well as $\tilde{\lambda}_0^k = \tilde{\lambda}^k|_{t=0}$ and $\tilde{\lambda}^k|_{t=t_f} = 0$. Comparing (A.11) with the adjoint representer equations (2.24), we see that $\tilde{\mu}^k$ and $\tilde{\lambda}^k$ obey the same equations as $\Omega^k$ and $\Lambda^k$ and therefore $\tilde{\mu}^k \equiv \Omega^k$ and $\tilde{\lambda}^k \equiv \Lambda^k$.

The other variations of $J^k$ yield

$$\begin{aligned}
\frac{\delta J^k}{\delta \nu}\delta\nu &= \int_0^{t_f} (\tilde{\mu}^k)^T D_\nu P_\nu \delta\nu dt \\
\frac{\delta J^k}{\delta \omega}\delta\omega &= \int_0^{t_f} (\tilde{\lambda}^k)^T D_\omega(Y^\eta) P_\omega \delta\omega dt \\
\frac{\delta J^k}{\delta \alpha}\delta\alpha &= \int_0^{t_f} \left( (\tilde{\mu}^k)^T \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta + (\tilde{\lambda}^k)^T \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta \right) \delta\alpha dt \\
\frac{\delta J^k}{\delta \beta}\delta\beta &= (\tilde{\lambda}_0^k)^T \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta \delta\beta
\end{aligned}$$

Using $\tilde{\mu}^k \equiv \Omega^k$, $\tilde{\lambda}^k \equiv \Lambda^k$ and the fact that $L_k[X', Y']$ is a scalar (for example $\frac{\delta J^k}{\delta \omega}\delta\omega = \delta\omega^T \frac{\delta J^k}{\delta \omega}^T$ ), we can now write $L_k[X', Y']$ according to (A.9) and (A.10) as

$$\begin{aligned}
L_k[X', Y'] = &\int_0^{t_f} \delta\nu^T P_\nu^T D_\nu^T \Omega^k dt + \int_0^{t_f} \delta\omega^T P_\omega^T [D_\omega(Y^\eta)]^T \Lambda^k dt \\
&+ \int_0^{t_f} \delta\alpha^T \left( \left.\frac{\partial \phi}{\partial \alpha}\right|_\eta^T \Omega^k + \left.\frac{\partial \varphi}{\partial \alpha}\right|_\eta^T \Lambda^k \right) dt + \delta\beta^T \left.\frac{\partial Y_0}{\partial \beta}\right|_\eta^T \Lambda^k|_{t=0}
\end{aligned}$$

If we left-multiply this expression with $\delta\alpha \equiv \alpha'$, take the expectation, and keep in mind that the cross-covariances (2.4*a*) between $\alpha$ and $\nu$, $\omega$ or $\beta$ vanish, we find (A.3). Similarly, multiplication with $\nu$, $\omega$, and $\beta'$ yield (A.5), (A.4), and (A.6). This completes our proof of (2.34).

### A.2.2 Derivation of the Posterior Covariance Equations

Before we start with the derivation of the posterior covariances, we prove the following result. Whereas the representer functions are deterministic, the representer coefficients are random variables. Their covariance is

$$\overline{b_k b_l} = [U^{-1}]_{kl} \tag{A.12}$$

where we defined earlier

$$U \equiv C_v + R \qquad \text{and} \qquad [R]_{kl} \equiv L_k[\Xi^l, \Upsilon^l] \tag{2.27}$$

This result is easily derived from (2.26). We have

$$\begin{aligned}
\overline{b_k b_l} &= \sum_{rs} [U^{-1}]_{kr} \overline{(Z_r - M_r[X^\eta, Y^\eta] - L_r[\overline{X}^{\eta+1} - X^\eta, \overline{Y}^{\eta+1} - Y^\eta]) \cdot} \\
&\qquad \overline{\cdot (Z_s - M_s[X^\eta, Y^\eta] - L_s[\overline{X}^{\eta+1} - X^\eta, \overline{Y}^{\eta+1} - Y^\eta])} [U^{-1}]_{sl} \\
&= \sum_{rs} [U^{-1}]_{kr} \overline{(v_r + L_r[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}])(v_s + L_s[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}])} [U^{-1}]_{sl} \\
&= \sum_{rs} [U^{-1}]_{kr} \left( [C_v]_{rs} + \overline{L_r[X', Y'] L_s[X', Y']} \right) [U^{-1}]_{sl}
\end{aligned}$$

With $\overline{L_r[X', Y'] L_s[X', Y']} = L_r[\overline{L_s[X', Y']X'}, \overline{L_s[X', Y']Y'}] = L_r[\Xi^s, \Upsilon^s] = R_{rs}$, the desired result follows immediately.

We are now finally ready to derive the equations (2.35), (2.36), and (2.37) for the posterior covariances. Using the previous results, the derivations are straightforward. First, we expand the expression for the posterior state covariance $C_{\tilde{Y}\tilde{Y}}$.

$$\begin{aligned}
[C_{\tilde{Y}\tilde{Y}}(t_1, t_2)]_{ij} &\equiv \overline{(Y_i(t_1) - Y_i^{\eta+1}(t_1))(Y_j(t_2) - Y_j^{\eta+1}(t_2))} \\
&= \overline{\left(Y_i(t_1) - \overline{Y}_i^{\eta+1}(t_1) - \sum_k b_k \Upsilon_i^k(t_1)\right)\left(Y_j(t_2) - \overline{Y}_j^{\eta+1}(t_2) - \sum_l b_l \Upsilon_j^l(t_2)\right)} \\
&= [C_{Y'Y'}(t_1, t_2)]_{ij} - \sum_k \Upsilon_i^k(t_1) \overline{b_k (Y_j(t_2) - \overline{Y}_j^{\eta+1}(t_2))} \\
&\quad - \sum_l \Upsilon_j^l(t_2) \overline{b_l (Y_i(t_1) - \overline{Y}_i^{\eta+1}(t_1))} + \sum_{kl} \Upsilon_i^k(t_1) \overline{b_k b_l} \Upsilon_j^l(t_2)
\end{aligned}$$

Using (A.12) and

$$\begin{aligned}
&\overline{b_k(Y_j(t_2) - \overline{Y}_j^{\eta+1}(t_2))} \\
&= \sum_r [U^{-1}]_{kr} \overline{(Z_r - M_r[X^\eta, Y^\eta] - L_r[\overline{X}^{\eta+1} - X^\eta, \overline{X}^{\eta+1} - X^\eta])(Y_j(t_2) - \overline{Y}_j^{\eta+1}(t_2))} \\
&= \sum_r [U^{-1}]_{kr} \overline{(v_r + L_r[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}])(Y_j(t_2) - \overline{Y}_j^{\eta+1}(t_2))} \\
&= \sum_r [U^{-1}]_{kr} \overline{L_r[X', Y'] Y_j'(t_2)} \\
&= \sum_r [U^{-1}]_{kr} \Upsilon_j^r(t_2)
\end{aligned}$$

we immediately get the corresponding equation in (2.35). The other posterior state covariance equations of can be derived analogously.

In order to derive the posterior covariance of the measurement predictions (2.36), we expand

$$\begin{aligned}
[C_{\tilde{v}}]_{mn} &\equiv \overline{L_m[X - X^{\eta+1}, Y - Y^{\eta+1}] L_n[X - X^{\eta+1}, Y - Y^{\eta+1}]} \\
&= \overline{\big(L_m[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}] - \sum_k b_k L_m[\Xi^k, \Upsilon^k]\big) \cdot} \\
&\quad \overline{\cdot \big(L_n[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}] - \sum_l b_l L_n[\Xi^l, \Upsilon^l]\big)} \\
&= \overline{L_m[X', Y'] L_n[X', Y']} - \sum_k L_m[\Xi^k, \Upsilon^k] \overline{b_k L_n[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}]} \\
&\quad - \sum_l L_n[\Xi^l, \Upsilon^l] \overline{b_l L_m[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}]} + \sum_{kl} L_m[\Xi^k, \Upsilon^k] \overline{b_k b_l} L_n[\Xi^l, \Upsilon^l]
\end{aligned}$$

With $L_m[\Xi^k, \Upsilon^k] = R_{mk}$ and

$$\begin{aligned}
&\overline{b_k L_n[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}]} \\
&= \sum_l [U^{-1}]_{kl} \overline{(Z_l - M_l[X^\eta, Y^\eta] - L_l[\overline{X}^{\eta+1} - X^\eta, \overline{Y}^{\eta+1} - Y^\eta]) L_n[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}]} \\
&= \sum_l [U^{-1}]_{kl} \overline{(v_l + L_l[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}]) L_n[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}]} \\
&= \sum_l [U^{-1}]_{kl} \overline{L_l[X', Y'] L_n[X', Y']} \\
&= \sum_l [U^{-1}]_{kl} R_{ln}
\end{aligned}$$

we immediately obtain (2.36).

Finally, for the covariance of the posterior data residuals (2.37) we expand

$$\begin{aligned}
[C_{\hat{v}}]_{mn} =& \overline{(Z_m - M_m[X^\eta, Y^\eta] - L_m[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta])\cdot} \\
& \overline{\cdot(Z_n - M_n[X^\eta, Y^\eta] - L_n[X^{\eta+1} - X^\eta, Y^{\eta+1} - Y^\eta])} \\
=& \overline{(v_m + L_m[X - X^{\eta+1}, Y - Y^{\eta+1}])(v_n + L_n[X - X^{\eta+1}, Y - Y^{\eta+1}])} \\
=& \overline{\big(v_m + L_m[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}] - \sum_k b_k L_m[\Xi^k, \Upsilon^k]\big)\cdot} \\
& \overline{\cdot\big(v_n + L_n[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}] - \sum_l b_l L_n[\Xi^l, \Upsilon^l]\big)} \\
=& \overline{v_m v_n} - \sum_l L_n[\Xi^l, \Upsilon^l]\overline{b_l v_m} - \sum_k L_m[\Xi^k, \Upsilon^k]\overline{b_k v_n} + [C_{\tilde{v}}]_{mn}
\end{aligned}$$

Using $L_n[\Xi^l, \Upsilon^l] = R_{nl}$ and

$$\begin{aligned}
\overline{b_l v_m} &= \sum_r [U^{-1}]_{lr} \overline{(Z_r - M_r[X^\eta, Y^\eta] - L_r[\overline{X}^{\eta+1} - X^\eta, \overline{Y}^{\eta+1} - Y^\eta])v_m} \\
&= \sum_r [U^{-1}]_{lr} \overline{(v_r + L_r[X - \overline{X}^{\eta+1}, Y - \overline{Y}^{\eta+1}])v_m} \\
&= \sum_r [U^{-1}]_{lr}[C_v]_{rm}
\end{aligned}$$

we get

$$\begin{aligned}
[C_{\hat{v}}]_{mn} &= [C_v]_{mn} - \sum_{lr} R_{nl}[U^{-1}]_{lr}[C_v]_{rm} - \sum_{ks} R_{mk}[U^{-1}]_{ks}[C_v]_{sn} + [C_{\tilde{v}}]_{mn} \\
&= [C_v]_{mn} - [RU^{-1}C_v]_{nm} - [RU^{-1}C_v]_{mn} + [C_{\tilde{v}}]_{mn}
\end{aligned}$$

which is obviously (2.37).

# Appendix B

# Land Surface Model

## B.1 List of Symbols

Tables B.1 to B.8 provide a list of all symbols used in the land surface model. The last column generally indicates in which spatial dimensions the variables or parameters vary and whether they are time-dependent. First, a list of the state and observation variables is shown in Table B.1. The three state variables for soil moisture can be used interchangeably. They are connected through the Clapp-Hornberger relations (3.4). The soil moisture and temperature states and the interception water are governed by ODE's, therefore initial conditions must be specified.

Next, Table B.2 lists the meteorologic inputs to the model. Tables B.3 and B.4 compile all the time-dependent variables and parameters. The functional dependence is indicated. Note that empirical and physical constants are not listed in this functional dependence. Tables B.5 and B.6 list the time-independent parameters, most of which must be specified as model inputs. Table B.7 contains all the scalar empirical constants with their values or appropriate references. Finally, Table B.8 shows all the physical constants in the model. Those numbers are fixed and never used for calibration.

Recall the notational convention to label most of the empirical constants in the various parameterizations with $\kappa$ for scalar constants and with $\beta$ for distributed parameters (which for example depend on texture or vegetation). The empirical parameters are superscripted with the variable which is being parameterized and subscripted with a number in case more than one empirical constant is needed.

Moreover, the subscripts $r, a, c, g$ refer to **r**eference (or screen) height, **a**ir (within the canopy), **c**anopy (plant material), and **g**round, respectively. Note that all variables at screen height are inputs that are directly measured or derived from meteorologic observations. The subscripts $s$ and $l$ are used for **s**hortwave and **l**ongwave, $s$ and $u$ are used for **s**aturated and **u**nsaturated, depending on context. The symbol $f$ always denotes a fraction varying from 0 to 1.

Lastly, in our convention the matric head $\psi_g$ is negative for unsaturated conditions. The vertical coordinate $z$ is positive upward, and the numbering of the layers increases upward.

| Symbol | Units | Description | Dimension |
|---|---|---|---|
| $W_g$ | $[-]$ | soil wetness/saturation | $x, y, z, t$ |
| $\theta_g$ | $[m^3/m^3]$ | volumetric soil moisture content | |
| $\psi_g$ | $[m]$ | matric head | |
| $T_g$ | $[K]$ | soil surface temperature | $x, y, t$ |
| $W_c$ | $[m]$ | canopy interception water | $x, y, t$ |
| $T_c$ | $[K]$ | canopy temperature | $x, y, t$ |
| $e_a$ | $[mb]$ | canopy air vapor pressure | $x, y, t$ |
| $T_a$ | $[K]$ | canopy air temperature | $x, y, t$ |
| $T_B$ | $[K]$ | radiobrightness temperature | $x, y, t$ |

Table B.1: State variables of the land surface model.

| Symbol | Units | Description | Dimension |
|---|---|---|---|
| $P_r$ | $[m/s]$ | precipitation at ref. height | $x, y, t$ |
| $R_{rs}$ | $[W/m^2]$ | incoming shortw. radiation at ref. height | $x, y, t$ |
| $T_r$ | $[K]$ | atmospheric temperature at ref. height | $x, y, t$ |
| $e_r$ | $[mb]$ | vapor pressure at ref. height | $x, y, t$ |
| $u_r$ | $[m/s]$ | wind speed at ref. height | $x, y, t$ |
| $T_d$ | $[K]$ | depth average soil temperature | $x, y, (t)$ |

Table B.2: Meteorologic inputs. The depth average soil temperature changes on a seasonal time-scale only.

| Sym. | Units | Description | Dependency | Dim. |
|---|---|---|---|---|
| $P_t$ | $[m/s]$ | throughfall rate of precipitation | $P_t(P_r, P_i, D_c, f_c)$ | $x,y,t$ |
| $P_i$ | $[m/s]$ | interception rate of precipitation | $P_i(P_r, W_c, W_c^{\max})$ | $x,y,t$ |
| $D_c$ | $[m/s]$ | dripping rate from interception storage | $D_c(W_c, t_c^{\text{drip}})$ | $x,y,t$ |
| $S_g$ | $[1/s]$ | root sink term for transp. loss | $S_g(E_{ct}^{\text{pot}}, \chi_g(W_g))$ | $x,y,z,t$ |
| $G_g$ | $[W/m^2]$ | ground heat flux | $G_g(R_{gs}^{\text{net}}, R_{gl}^{\text{net}}, E_g, H_g)$ | $x,y,t$ |
| $R_{rl}$ | $[W/m^2]$ | incoming longw. radiation at ref. height | $R_{rl}(T_r, \epsilon_r)$ | $x,y,t$ |
| $R_{gs}^{\text{net}}$ | $[W/m^2]$ | net shortw. radiation at ground surface | $R_{gs}^{\text{net}}(R_{rs}, a_g, f_c)$ | $x,y,t$ |
| $R_{gl}^{\text{net}}$ | $[W/m^2]$ | net longw. radiation at ground surface | $R_{gl}^{\text{net}}(R_{rl}, T_g, T_c, \epsilon_g, \epsilon_c, f_c)$ | $x,y,t$ |
| $R_{cs}^{\text{net}}$ | $[W/m^2]$ | net shortw. radiation at canopy | $R_{cs}^{\text{net}}(R_{rs}, a_c, f_c)$ | $x,y,t$ |
| $R_{cl}^{\text{net}}$ | $[W/m^2]$ | net longw. radiation at canopy | $R_{cl}^{\text{net}}(R_{rl}, T_g, T_c, \epsilon_g, \epsilon_c, f_c)$ | $x,y,t$ |
| $a_g$ | $[-]$ | ground surface albedo | $a_g(W_g, \theta_s)$ | $x,y,t$ |
| $\epsilon_g$ | $[-]$ | longw. soil surface emissivity | $\epsilon_g(W_g)$ | $x,y,t$ |
| $\epsilon_r$ | $[-]$ | longw. atmospheric emissivity | $\epsilon_r(e_r)$ | $x,y,t$ |
| $E_a$ | $[kg/m^2/s]$ | evapotransp. rate to atmosphere | $E_a(e_a, e_r, r_a)$ | $x,y,t$ |
| $E_g$ | $[kg/m^2/s]$ | ground surface evap. rate | $E_g(e_s(T_g), e_a, r_g, r_d, f_c)$ | $x,y,t$ |
| $E_{ce}$ | $[kg/m^2/s]$ | (wet) canopy evap. rate | $E_{ce}(e_s(T_c), e_a, r_c, \text{LSAI}, f_c, f_{ce})$ | $x,y,t$ |
| $E_{ct}^{\text{pot}}$ | $[kg/m^2/s]$ | potential (dry) canopy transp. rate | $E_{ct}^{\text{pot}}(e_s(T_c), e_a, r_c, r_s, \text{LAI}, f_c, f_{ce})$ | $x,y,t$ |
| $E_{ct}$ | $[kg/m^2/s]$ | actual (dry) canopy transp. rate | $E_{ct}(E_{ct}^{\text{pot}}, \chi_g(W_g))$ | $x,y,t$ |
| $H_a$ | $[W/m^2]$ | sensible heat flux to atmosphere | $H_a(T_r, T_a, r_a)$ | $x,y,t$ |
| $H_g$ | $[W/m^2]$ | sensible heat flux at ground surface | $H_g(T_g, T_a, r_d, f_c)$ | $x,y,t$ |
| $H_c$ | $[W/m^2]$ | sensible heat flux from canopy | $H_c(T_c, T_a, r_c, \text{LSAI}, f_c)$ | $x,y,t$ |
| $e_s$ | $[mb]$ | saturation vapor pressure | $e_s(T)$ | $x,y,t$ |
| $u_c$ | $[m/s]$ | wind speed at canopy height | $u_c(u_r, z_r, z_0, d_c, h_c)$ | $x,y,t$ |
| $u_*$ | $[m/s]$ | friction velocity | $u_*(u_r, z_r, z_0, d_c)$ | $x,y,t$ |
| $r_a$ | $[s/m]$ | atmospheric resistance | $r_a(u_r, z_r, z_0, d_c)$ | $x,y,t$ |
| $r_c$ | $[s/m]$ | bulk canopy resistance | $r_c(u_c)$ | $x,y,t$ |
| $r_d$ | $[s/m]$ | aerodynamic resistance within canopy | $r_d(u_c)$ | $x,y,t$ |
| $r_g$ | $[s/m]$ | surface resistance of bare soil | $r_g(W_g, \theta_s)$ | $x,y,t$ |
| $r_s$ | $[s/m]$ | stomatal resistance | $r_s(r_s^{\min}, \chi_c(R_{rs}))$ | $x,y,t$ |
| $C_g$ | $[J/m^3/K]$ | vol. heat capacity of wet top soil layer | $C_g(\rho_{gb}, c_g, W_g, \theta_s)$ | $x,y,t$ |
| $\lambda_g$ | $[W/m/K]$ | thermal conductivity of wet top soil layer | $\lambda_g(W_g, \theta_s)$ | $x,y,t$ |
| $q_t$ | $[m/s]$ | moisture flux b.c. at top | $q_t(P_t, E_g)$ | $x,y,t$ |
| $q_b$ | $[m/s]$ | moisture flux b.c. at bottom | $q(K_u)$ | $x,y,t$ |
| $K_u$ | $[m/s]$ | unsaturated hydraulic conductivity | $K_u(K_s, W_g)$ | $x,y,z,t$ |
| $f_{ce}$ | $[-]$ | wet canopy fraction (for canopy evap.) | $f_{ce}(W_c, W_c^{\max})$ | $x,y,t$ |
| $\chi_g$ | $[-]$ | stress function for water-limited transp. | $\chi_g(W_g, \psi^{\text{wilt}}, f_R)$ | $x,y,z,t$ |
| $\chi_c$ | $[-]$ | solar radiation influence on transp. | $\chi_c(R_{rs})$ | $x,y,t$ |

Table B.3: Forcing variables and time-dependent parameters for the land surface model. Note that we assume the soil thermal diffusivity $K_T$ to be constant in time (Section 3.1.7).

| Symbol | Units | Description | Dependency | Dim. |
|---|---|---|---|---|
| $k_g$ | $[-]$ | microw. dielectric constant of wet soil | $k_g(W_g, \theta_s, k_w, k_{gd}, f_S, f_C)$ | $x, y, t$ |
| $k_w$ | $[-]$ | microw. dielectric constant of water | $k_w(k_{w0}, k_{w\infty}, \nu_r, \tau_w)$ | $x, y, t$ |
| $k_{w0}$ | $[-]$ | static dielectric constant of water | $k_{w0}(T_g)$ | $x, y, t$ |
| $\tau_w$ | $[s]$ | relaxation time of water | $\tau_w(T_g)$ | $x, y, t$ |
| $\varepsilon_{gp}$ | $[-]$ | rough surface microw. emissivity for polariz. $p$ | $\varepsilon_{gp}(k_g, \phi_r)$ | $x, y, t$ |
| $\varepsilon_{gp}^{\text{smooth}}$ | $[-]$ | smooth surface microw. emissivity for polariz. $p$ | $\varepsilon_{gp}^{\text{smooth}}(k_g, \phi_r)$ | $x, y, t$ |
| $T_g^{\text{eff}}$ | $[K]$ | eff. soil temp. for microw. emission | $T_g^{\text{eff}}(T_g, k_g)$ | $x, y, t$ |
| $\phi_g$ | $[rad]$ | in-soil propagation angle | $\phi_g(k_g, \phi_r)$ | $x, y, t$ |
| $z_{\text{grad}}$ | $[m]$ | gradient RT effective depth | $z_{\text{grad}}(\alpha_e, \phi_g)$ | $x, y, t$ |
| $\alpha_e$ | $[1/m]$ | microw. attenuation coefficient | $\alpha_e(k_g, \nu_r)$ | $x, y, t$ |

Table B.4: Time-dependent variables and parameters for the Radiative Transfer model. Note that $\phi_g$, $z_{\text{grad}}$, and $\alpha_e$ are part of the Gradient RT approximation and not used.

| Symbol | Units | Description/Value | Model Input ? | Dim. |
|---|---|---|---|---|
| $\delta_g$ | $[m]$ | depth of surface temp. layer | yes | - |
| $\Delta_i$ | $[m]$ | depth of finite difference soil moisture layer $i$ | yes | $z$ |
| $\bar{\Delta}_i$ | $[m]$ | depth of layer "around" soil moisture node $i$ | $\bar{\Delta}(\Delta)$ | $z$ |
| $\theta_s$ | $[m^3/m^3]$ | porosity | yes | $x, y, z$ |
| $K_s$ | $[m/s]$ | saturated hydraulic conductivity | yes | $x, y, z$ |
| $\psi^{\mathrm{CH}}$ | $[m]$ | [Clapp and Hornberger, 1978] | yes | $x, y, z$ |
| $B^{\mathrm{CH}}$ | $[-]$ | [Clapp and Hornberger, 1978] | yes | $x, y, z$ |
| $W^{r_g}$ | $[-]$ | surface resistance param. (0.25 ... 0.6) | yes | $x, y$ |
| $r_g^{\max}$ | $[s/m]$ | surface resistance param. (3000 ... 7000) | yes | $x, y$ |
| $\beta^{r_g}$ | $[-]$ | surface resistance param. | $\beta^{r_g}(W^{r_g}, r_g^{\max})$ | $x, y$ |
| $K_T$ | $[m^2/s]$ | thermal diffusivity of soil surface layer | yes | $x, y$ |
| $d_g$ | $[m]$ | damping depth of temp. forcing | $d_g(K_T)$ | $x, y$ |
| $\alpha_g$ | $[-]$ | param. of force-restore approx. | $\alpha_g(d_g, \delta_g)$ | $x, y$ |
| $\Gamma_g$ | $[1/s]$ | coefficient in force-restore approx. | $\Gamma_g(\alpha_g)$ | $x, y$ |
| $\Gamma'_g$ | $[s/m]$ | coefficient in force-restore approx. | $\Gamma'_g(K_T)$ | $x, y$ |
| $\rho_{gb}$ | $[kg/m^3]$ | bulk density of (dry) soil in layer $\delta_g$ | $\rho_{gb}(\theta_s, \rho_g)$ | $x, y$ |
| $c_g$ | $[J/kg/K]$ | specific heat of dry soil in layer $\delta_g$ | yes | $x, y$ |
| $z_r$ | $[m]$ | reference (or screen) height | yes | - |
| $h_c$ | $[m]$ | vegetation height (also for bare soil $z_0$) | yes | $x, y, (t)$ |
| $d_c$ | $[m]$ | zero plane displacement height | $d_c(h_c)$ | $x, y, (t)$ |
| $z_0$ | $[m]$ | roughness length | $z_0(h)$ | $x, y, (t)$ |
| $\beta^{r_c}$ | $[\sqrt{s/m}]$ | [Sellers and Dorman, 1987] | yes | $x, y, (t)$ |
| $\beta^{r_d}$ | $[-]$ | [Sellers and Dorman, 1987] | yes | $x, y, (t)$ |
| $r_s^{\min}$ | $[s/m]$ | minimum stomatal resistance | yes | $x, y, (t)$ |
| $a_c$ | $[-]$ | shortw. canopy albedo | yes | $x, y, (t)$ |
| $\epsilon_c$ | $[-]$ | longw. canopy emissivity | yes | $x, y, (t)$ |
| $W_c^{\max}$ | $[m]$ | max intercepted water | $W_c^{\max}$(LSAI) | $x, y, (t)$ |
| $t_c^{\mathrm{drip}}$ | $[s]$ | dripping time for canopy interception storage | yes | $x, y, (t)$ |
| $f_c$ | $[-]$ | area fraction shaded by vegetation canopy | yes | $x, y, (t)$ |
| $\psi^{\mathrm{wilt}}$ | $[m]$ | wilting point | yes | $x, y, (t)$ |
| $d_R$ | $[m]$ | typical rooting depth | yes | $x, y, (t)$ |
| $\rho_R$ | $[1/m]$ | root density distribution | $\rho_R(d_R)$ | $x, y, z, (t)$ |
| $f_R$ | $[-]$ | root distribution factor | $f_R(d_R)$ | $x, y, z, (t)$ |
| LAI | $[-]$ | (green) leaf area index | yes | $x, y, (t)$ |
| SAI | $[-]$ | stem (and dead leaf) area index | yes | $x, y, (t)$ |
| LSAI | $[-]$ | leaf and stem area index | LSAI( LAI, SAI) | $x, y, (t)$ |
| $\beta_1^{\lambda_g}$ | $[W/m/K]$ | $-0.197$ (clay), 0.243 (loam), 0.228 (sand) | (yes) | $x, y$ |
| $\beta_2^{\lambda_g}$ | $[W/m/K]$ | $-0.962$ (clay), 0.393 (loam), $-2.406$ (sand) | (yes) | $x, y$ |
| $\beta_3^{\lambda_g}$ | $[W/m/K]$ | 2.521 (clay), 1.534 (loam), 4.909 (sand) | (yes) | $x, y$ |

Table B.5: Time-independent parameters for the land surface model. Note that we assume the soil thermal diffusivity $K_T$ to be a constant in time (Section 3.1.7), that is the $\beta_i^{\lambda_g}$ are not used. Note also that vegetation parameters are time-dependent on the time-scale of plant growth, which is indicated by $(t)$.

| Symbol | Units | Description | Model Input ? | Dim. |
|---|---|---|---|---|
| $\nu_r$ | $[Hz]$ | microw. observation frequency | yes | - |
| $\phi_r$ | $[rad]$ | (in-air) look-angle from nadir at ref. height | yes | - |
| $f_S$ | $[-]$ | sand fraction in layer $\delta_g$ | yes | $x, y$ |
| $f_C$ | $[-]$ | clay fraction in layer $\delta_g$ | yes | $x, y$ |
| $\beta^{k_g}$ | $[-]$ | Dobson dielectric mixing model param. | $\beta^{k_g}(f_S, f_C)$ | $x, y$ |
| $\beta^{\varepsilon_g}$ | $[-]$ | surface roughness param. [Choudhury et al., 1979] | yes | $x, y$ |
| $W_v$ | $[kg/m^2]$ | vegetation water content | yes | $x, y, (t)$ |
| $\beta^{\delta_c}$ | $[m^2/kg]$ | "vegetation b param." [Jackson and Schmugge, 1991] | yes | $x, y, (t)$ |
| $\delta_c$ | $[-]$ | canopy optical depth | $\delta_c(W_v, \phi_r)$ | $x, y, (t)$ |
| $\alpha_c$ | $[-]$ | canopy microwave attenuation | $\alpha_c(\delta_c)$ | $x, y, (t)$ |
| $z_{\text{grey}}$ | $[m]$ | grey body RT param. | yes | - |

Table B.6: Time-independent parameters for the Radiative Transfer model. Note that vegetation parameters are time-dependent on the time-scale of plant growth, which is indicated by $(t)$.

| Symbol | Value | Units |
|---|---|---|
| $\kappa_1^{\alpha_g}$ | 0.25 | $[-]$ |
| $\kappa_2^{\alpha_g}$ | 0.125 | $[-]$ |
| $\kappa_1^{\epsilon_r}$ | 0.74 | $[-]$ |
| $\kappa_2^{\epsilon_r}$ | 0.0049 | $[-]$ |
| $\kappa_1^{\epsilon_g}$ | 0.9 | $[-]$ |
| $\kappa_2^{\epsilon_g}$ | 0.18 | $[-]$ |
| $\kappa^{r_g}$ | 10 | $[s/m]$ |
| $\kappa_1^{\alpha_g}$ | 0.943 | $[-]$ |
| $\kappa_2^{\alpha_g}$ | 0.223 | $[-]$ |
| $\kappa_3^{\alpha_g}$ | $1.68 \cdot 10^{-2}$ | $[-]$ |
| $\kappa_4^{\alpha_g}$ | $-5.27 \cdot 10^{-3}$ | $[-]$ |
| $\kappa^{d_c}$ | 0.63 | $[-]$ |
| $\kappa^{z_0}$ | 0.13 | $[-]$ |
| $\kappa^{W_c}$ | $10^{-4} \ldots 10^{-3}$ | $[m]$ |
| $\kappa_1^{\chi_c}$ | $50 \ldots 100$ | $[-]$ |
| $\kappa_2^{\chi_c}$ | $100 \ldots 200$ | $[W/m^2]$ |
| $\kappa^{f_{cc}}$ | 2/3 | $[-]$ |
| $\kappa_1^{k_g}$ | 0.65 | $[-]$ |
| $\kappa_2^{k_g}$ | 1.09 | $[-]$ |
| $\kappa_3^{k_g}$ | 0.11 | $[-]$ |
| $\kappa_4^{k_g}$ | -0.18 | $[-]$ |
| $\kappa_1^{k_{w0}}$ | 88.045 | $[-]$ |
| $\kappa_2^{k_{w0}}$ | $-0.4147$ | $[1/K]$ |
| $\kappa_3^{k_{w0}}$ | $6.295 \cdot 10^{-4}$ | $[1/K^2]$ |
| $\kappa_4^{k_{w0}}$ | $1.075 \cdot 10^{-5}$ | $[1/K^3]$ |
| $\kappa_1^{\tau_w}$ | $1.1109 \cdot 10^{-10}$ | $[s]$ |
| $\kappa_2^{\tau_w}$ | $-3.824 \cdot 10^{-12}$ | $[s/K]$ |
| $\kappa_3^{\tau_w}$ | $6.938 \cdot 10^{-14}$ | $[s/K^2]$ |
| $\kappa_4^{\tau_w}$ | $-5.096 \cdot 10^{-16}$ | $[s/K^3]$ |
| $c_{\text{grad}}$ | 1.03 | $[-]$ |

Table B.7: Scalar empirical constants.

| Symbol | Value | Units | Description |
|---|---|---|---|
| $\rho_w$ | 1000 | $[kg/m^3]$ | density of water |
| $\rho_a$ | 1.20 | $[kg/m^3]$ | density of air |
| $\rho_g$ | $2.65 \cdot 10^3$ | $[kg/m^3]$ | density of soil particles in layer $\delta_g$ |
| $\omega_d$ | $2\pi/86400$ | $[1/s]$ | angular frequency of diurnal cycle |
| $L$ | $2.5 \cdot 10^6$ | $[J/kg]$ | latent heat of vaporization |
| $\sigma$ | $5.57 \cdot 10^{-8}$ | $[W/m^2/K^4]$ | Stefan-Boltzmann constant |
| $\gamma$ | 0.65 | $[mb/K]$ | psychrometric constant |
| $c_a$ | 1004 | $[J/kg/K]$ | specific heat of air at constant pressure |
| $c_w$ | 4187 | $[J/kg/K]$ | specific heat of water |
| $K$ | 0.4 | $[-]$ | von Karman constant |
| $T_0$ | 273.15 | $[K]$ | reference temperature |
| $k_{gd}$ | 4.67 | $[-]$ | microw. dielectric constant of dry soil |
| $k_{w\infty}$ | 4.9 | $[-]$ | high frequency dielectric constant of water |
| $g$ | 9.81 | $[m/s^2]$ | acceleration due to gravity |
| $c_{\text{light}}$ | $3 \cdot 10^8$ | $[m/s]$ | speed of light |

Table B.8: Physical constants

# Bibliography


Abramopoulos, F., Rosenzweig, C. and Choudhury, B. [1988]. Improved ground hydrology calculations for global climate models (GCMs): Soil water movement and evapotranspiration, *J. Climate* **1**: 921–941.

Ács, F., Mihailović, D. T. and Rajković, B. [1991]. A coupled soil moisture and surface temperature prediction model, *J. Appl. Meteorol.* **30**(6): 812–822.

Asif, A. and Moura, J. M. F. [1999]. Data assimilation in large time-varying multidimensional fields, *IEEE Trans. Image Process.* **8**(11): 1593–1607.

Basara, J. B., Brotzge, J. A. and Crawford, K. C. [1999]. Investigating land-atmosphere interactions using the Oklahoma Mesonet, *Preprint Volume of the 3rd Symposium on Integrated Observing Systems, Dallas, TX*, American Meteorological Society, Boston, MA.

Bengtsson, L., Ghil, M. and Källén, E. (eds) [1981]. *Dynamic Methodology: Data Assimilation Methods*, Springer Verlag, New York.

Benjamin, J. R. and Cornell, C. A. [1970]. *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, New York, NY.

Bennett, A. F. [1992]. *Inverse Methods in Physical Oceanography*, Cambridge University Press, New York, NY.

Bennett, A. F. [1999]. Inverse methods and data assimilation, Lecture Notes of the 1999 Summer School at the College of Oceanic and Atmospheric Sciences at Oregon State University, Corvallis, Oregon.

Bennett, A. F. and Baugh, J. R. [1992]. A parallel algorithm for variational assimilation in oceanography and meteorology, *J. Atmos. Ocean. Technol.* **9**: 426–433.

Bennett, A. F., Chua, B. S., Harrison, D. E. and McPhaden, M. J. [1998]. Generalized inversion of tropical atmosphere-ocean data and a coupled model of the tropical Pacific, *J. Climate* **11**(7): 1768–1792.

Bennett, A. F., Chua, B. S. and Leslie, L. M. [1996]. Generalized inversion of a global numerical weather prediction model, *Meteorol. Atmos. Phys.* **60**: 165–178. See also [Bennett et al., 1997].

Bennett, A. F., Chua, B. S. and Leslie, L. M. [1997]. Generalized inversion of a global numerical weather prediction model, II: Analysis and implementation, *Meteorol. Atmos. Phys.* **62**: 129–140.

Bennett, A. F., Leslie, L. M., Hagelberg, C. R. and Powers, P. E. [1993]. Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method, *Mon. Weather Rev.* **121**: 1714–1729.

Bhumralkar, C. M. [1975]. Numerical experiments on the computation of ground surface temperature in an atmospheric global circulation model, *J. Appl. Meteorol.* **14**: 1246–1258.

Birchak, J. R., Gardner, C. G., Hipp, J. E. and Victor, J. M. [1974]. High dielectric constant microwave probes for sensing soil moisture, *Proceedings of the IEEE* **62**(1): 93–98.

Bouttier, F., Mahfouf, J.-F. and Noilhan, J. [1993a]. Sequential assimilation of soil moisture from atmospheric low-level parameters. Part I: Sensitivity and calibration studies, *J. Appl. Meteorol.* **32**(8): 1335–1351.

Bouttier, F., Mahfouf, J.-F. and Noilhan, J. [1993b]. Sequential assimilation of soil moisture from atmospheric low-level parameters. Part II: Implementation in a mesoscale model, *J. Appl. Meteorol.* **32**(8): 1352–1364.

Bouyssel, F., Cassé, V. and Pailleux, J. [1999]. Variational surface analysis from screen level atmospheric parameters, *Tellus* **51**(A): 453–468.

Brock, F. V., Crawford, K. C., Elliott, R. L., Cuperus, G. W., Stadler, S. J., Johnson, H. L. and Eilts, M. D. [1995]. The Oklahoma Mesonet: A Technical Overview, *J. Atmos. Ocean. Technol.* **12**: 5–19.

Brooks, R. H. and Corey, A. T. [1964]. Hydraulic properties of porous media, *Hydrology Paper 3*, Colorado State Univ., Civil Engineering Dept., Fort Collins, Colorado.

Burgers, G., van Leeuwen, P. J. and Evensen, G. [1998]. Analysis scheme in the Ensemble Kalman Filter, *Mon. Weather Rev.* **126**: 1719–1724.

Callies, U., Rhodin, A. and Eppel, D. P. [1998]. A case study on variational soil moisture analysis from atmospheric observations, *J. Hydrol* **212–213**: 95–108.

Calvet, J.-C., Noilhan, J. and Bessemoulin, P. [1998]. Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: A feasibility study based on field measurements, *J. Appl. Meteorol.* **37**(4): 371–386.

Camillo, P. J., Gurney, R. J. and Schmugge, T. J. [1983]. A soil and atmospheric boundary layer model for evapotranspiration and soil moisture studies, *Water Resour. Res.* **19**: 371–380.

Castelli, F., Entekhabi, D. and Caporali, E. [1999]. Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance, *Water Resour. Res.* **35**(10): 3115–3125.

Celia, M. A., Bouloutas, E. T. and Zarba, R. L. [1990]. A general mass-conservative numerical solution for the unsaturated flow equation, *Water Resour. Res.* **26**(7): 1483–1496.

Chen, M., Rood, R. B. and Joiner, J. [1999]. Assimilating TOVS humidity into the GEOS-2 data assimilation system, *J. Climate* **12**: 2983–2995.

Choudhury, B. J., Schmugge, T. J., Chang, A. and Newton, R. W. [1979]. Effect of surface roughness on the microwave emission from soils, *J. Geophys. Res.* **84**(C9): 5699–5706.

Choudhury, B. J., Schmugge, T. J. and Mo, T. [1982]. A parameterization of effective soil temperature for microwave emission, *J. Geophys. Res.* **87**(C2): 1301–1304.

Chung, S.-O. and Horton, R. [1987]. Soil heat and water flow with a partial surface mulch, *Water Resour. Res.* **23**: 2175–2186.

Clapp, R. B. and Hornberger, G. M. [1978]. Empirical equations for some soil hydraulic properties, *Water Resour. Res.* **14**: 601–604.

Cohn, S. E., Da Silva, A., Guo, J., Sienkiewicz, M. and Lamich, D. [1998]. Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System, *Mon. Weather Rev.* **126**(11): 2913–2926.

Cohn, S. E. and Todling, R. [1996]. Approximate data assimilation schemes for stable and unstable dynamics, *J. Meteorol. Soc. Japan* **74**: 63–75.

Courant, R. and Hilbert, D. [1953]. *Methods of Mathematical Physics*, Vol. 1, Wiley Interscience, New York, NY.

Courtier, P., Derber, J., Errico, R., Louis, J.-F. and Vukićević, T. [1993]. Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology, *Tellus* **45A**: 342–357.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. [1994]. A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol. Soc.* **120**: 1367–1387.

Daley, R. [1991]. *Atmospheric Data Analysis*, Cambridge University Press, New York, NY.

Daniel, M. M. [1997]. *Multiresolution Statistical Modeling with Application to Modeling Groundwater Flow*, PhD dissertation, Massachusett Institute of Technology, Dept. of Electrical Engineering and Computer Science.

Daniel, M. M., Willsky, A. S. and McLaughlin, D. [1999]. A multiscale approach for estimating solute travel time distributions, *Adv. Water Resour.* . Submitted Aug 9, 1999, see also [Daniel, 1997].

Deardroff, J. W. [1978]. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation, *J. Geophys. Res.* **82**(C4): 1889–1903.

Dee, D. P. [1991]. Simplification of the Kalman filter for meteorological data assimilation, *Q. J. R. Meteorol. Soc.* **117**: 365–384.

Dickinson, R. E. [1988]. The force-restore model for surface temperatures and its generalizations, *J. Climate* **1**(11): 1086–1097.

Dickinson, R. E., Henderson-Sellers, A. and Kennedy, P. J. [1993]. Biosphere-Atmosphere Transfer Scheme (BATS) as coupled to the NCAR Community Climate Model, *Technical Note 387*, National Center for Atmospheric Research (NCAR).

Dobson, M. C., Ulaby, F. T., Hallikainen, M. T. and El-Rayes, M. A. [1985]. Microwave dielectric behavior of wet soil – Part II: Dielectric mixing models, *IEEE Trans. Geosci. Remote Sens.* **GE-23**(1): 35–46.

Dorman, J. L. and Sellers, P. J. [1989]. A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the Simple Biosphere Model (SiB), *J. Appl. Meteorol.* **28**: 833–855.

Drusch, M., Wood, E. F. and Simmer, C. [1999]. Up-scaling effects in passive microwave remote sensing: ESTAR 1.4GHz measurements during SGP '97, *Geophys. Res. Lett.* **26**(7): 879.

Eagleson, P. S. [1978]. Climate, soil, and vegetation: 3. A simplified model of soil moisture movement in the liquid phase, *Water Resour. Res.* **14**(5): 722–730.

Egbert, G. D., Bennett, A. F. and Foreman, M. G. G. [1994]. TOPEX/POSEIDON tides estimated using a global inverse model, *J. Geophys. Res.* **99**(C12): 24821–24852.

Eknes, M. and Evensen, G. [1997]. Parameter estimation solving a weak constraint variational formulation for an Ekman model, *J. Geophys. Res.* **102**(C6): 12479–12491.

Eltahir, E. A. B. [1998]. A soil moisture-rainfall feedback mechanism 1. Theory and observations, *Water Resour. Res.* **34**(4): 765–776. See also [Zheng and Eltahir, 1998] and [Findell and Eltahir, 1997].

Entekhabi, D., Nakamura, H. and Njoku, E. G. [1994]. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations, *IEEE Trans. Geosci. Remote Sens.* **32**(2): 438–448.

Errico, R. M. [1997]. What is an adjoint model?, *Bull. Amer. Meteorol. Soc.* **78**(11): 2577–2591.

Errico, R. M. [1999]. Workshop on assimilation of satellite data, *Bull. Amer. Meteorol. Soc.* **80**(3): 463–471.

Evensen, G. [1994a]. Inverse methods and data assimilation in nonlinear ocean models, *Physica D* **77**: 108–129.

Evensen, G. [1994b]. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.* **99**(C5): 10143–10162. Note that [Burgers et al., 1998] contains an important correction to the original Ensemble Kalman Filter.

Evensen, G. and van Leeuwen, P. J. [1996]. Assimilation of Geosat altimeter data for the Agulhas Current using the Ensemble Kalman Filter with a quasigeostrophic model, *Mon. Weather Rev.* **124**: 85–96.

Famiglietti, J. S., Devereaux, J. A., Laymon, C. A., Tsegaye, T., Houser, P. R., Jackson, T. J., Graham, S. T., Rodell, M. and van Oevelen, P. J. [1999]. Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment, *Water Resour. Res.* **35**(6): 1839–1851.

Famiglietti, J. S. and Wood, E. F. [1994a]. Application of multiscale water and energy balance models on a tallgrass prairie, *Water Resour. Res.* **30**(11): 3061–3078.

Famiglietti, J. S. and Wood, E. F. [1994b]. Multiscale modeling of spatially variable water and energy balance processes, *Water Resour. Res.* **30**(11): 3079–3093.

Fieguth, P., Karl, W. C., Willsky, A. S. and Wunsch, C. [1995]. Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry, *IEEE Trans. Geosci. Remote Sens.* **33**(2): 280–292.

Fieguth, P., Menemenlis, D., Ho, T., Willsky, A. S. and Wunsch, C. [1998]. Mapping Mediterranean altimeter data with a multiresolution optimal interpolation algorithm, *J. Atmos. Ocean. Technol.* **15**: 535–546.

Findell, K. L. and Eltahir, E. A. B. [1997]. An analysis of the soil moisture-rainfall feedback, based on direct observations from Illinois, *Water Resour. Res.* **33**(4): 725–735.

Galantowicz, J. F., Entekhabi, D. and Njoku, E. G. [1999]. Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness, *IEEE Trans. Geosci. Remote Sens.* **37**(4): 1860.

Gelb, A. (ed.) [1974]. *Applied Optimal Estimation*, M.I.T. Press, Cambridge, MA. Written by the Technical Staff, The Analytic Sciences Corporation.

Ghil, M. [1989]. Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework, *Dyn. Atmos. Oceans* **13**: 171–218.

Ghil, M. and Malanotte-Rizzoli, P. [1991]. Data assimilation in meteorology and oceanography, *Adv. Geophys.* **33**: 141–266.

Giering, R. and Kaminski, T. [1998]. Recipes for adjoint code construction, *ACM Trans. Math. Software* **24**(4): 437–474.

Houser, P. R. [1996]. *Remote-Sensing Soil Moisture Using Four-Dimensional Data Assimilation*, PhD dissertation, University of Arizona, Dept. of Hydrology and Water Resources.

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H. and Goodrich, D. C. [1998]. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resour. Res.* **34**(12): 3405–3420. See also [Houser, 1996].

Hu, Y., Gao, X., Shuttleworth, W. J., Gupta, H., Francois Mahfouf, J. and Viterbo, P. [1999]. Soil-moisture nudging experiments with a single-column version of the ECMWF model, *Q. J. R. Meteorol. Soc.* **125**: 1879–1902.

Hu, Z. and Islam, S. [1995]. Prediction of ground surface temperature and soil moisture content by the force-restore method, *Water Resour. Res.* **31**: 2531–2539.

Idso, S., Jackson, R., Reginato, R., Kimball, B. and Nakayama, F. [1975]. The dependence of bare-soil albedo on soil water content, *J. Appl. Meteorol.* **14**: 109–113.

Jackson, T. J. [1997]. *Southern Great Plains 1997 (SGP97) Hydrology Experiment Plan*, USDA-ARS, Hydrology Lab, U. S. Department of Agriculture, Beltsville, MD.

Jackson, T. J., Le Vine, D. M., Hsu, A. Y., Oldak, A., Starks, P. J., Swift, C. T., Isham, J. D. and Haken, M. [1999]. Soil moisture mapping at regional scales using microwave radiometry: The Southern Great Plains Hydrology Experiment, *IEEE Trans. Geosci. Remote Sens.* **37**(5): 2136–2151.

Jackson, T. J., O'Neill, P. E. and Swift, C. T. [1997]. Passive microwave observation of diurnal surface soil moisture, *IEEE Trans. Geosci. Remote Sens.* **35**(5): 1210–1222.

Jackson, T. J. and Schmugge, T. J. [1991]. Vegetation effects on the microwave emission of soils, *Remote Sens. Environ.* **36**: 203–212.

Jenkins, G. M. and Watts, D. G. [1968]. *Spectral analysis and its applications*, Holden-Day, San Francisco, CA.

Katul, G. G., Wendroth, O., Parlange, M. B., Puente, C. E., Folegatti, M. V. and Nielsen, D. R. [1993]. Estimation of in situ hydraulic conductivity function from nonlinear filtering theory, *Water Resour. Res.* **29**(4): 1063–1070.

Klinker, E., Rabier, F., Kelly, G. and Mahfouf, J.-F. [1999]. The ECMWF operational implementation of four dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration, *Technical Memorandum TM273*, European Centre for Medium Range Weather Forecasts, Reading, UK.

Kondo, J., Saigusa, N. and Sato, T. [1990]. A parameterization of evaporation from bare soil, *J. Appl. Meteorol.* **29**: 385–389.

Koster, R. D. and Suarez, M. J. [1992a]. A comparative analysis of two land surface heterogeneity representations, *J. Climate* **5**: 1379–1390.

Koster, R. D. and Suarez, M. J. [1992b]. Modeling the land surface boundary in climate models as a composite of independent vegetation stands, *J. Geophys. Res.* **97**(D3): 2697–2715.

Kustas, W. P. and Jackson, T. J. [1999]. The impact on area-averaged heat fluxes from using remotely sensed data at different resolutions: A case study with Washita '92 data, *Water Resour. Res.* **35**(5): 1539–1550.

Lanczos, C. [1966]. *The Variational Principles of Mechanics*, third edn, University of Toronto, Toronto, Canada.

Le Dimet, F.-X. and Talagrand, O. [1986]. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, *Tellus* **38A**: 97–110.

Lermusiaux, P. F. J. and Robinson, A. R. [1999a]. Data assimilation via error subspace statistical estimation. Part I: Theory and schemes, *Mon. Weather Rev.* **127**: 1385–1407.

Lermusiaux, P. F. J. and Robinson, A. R. [1999b]. Data assimilation via error subspace statistical estimation. Part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation, *Mon. Weather Rev.* **127**: 1408–1432.

Lhomme, J.-P., Elguero, E., Chehbouni, A. and Boulet, G. [1998]. Stomatal control of transpiration: Examination of Monteith's formulation of canopy resistance, *Water Resour. Res.* **34**: 2301–2308.

Lin, J. D. [1980]. On the force-restore method for prediction of ground surface temperature, *J. Geophys. Res.* **85**(C6): 3251–3254.

Liou, Y. A., Kim, E. J. and England, A. W. [1998]. Radiobrightness of prairie soil and grassland during dry-down simulations, *Radio Sci.* **33**(2): 259–265.

Lorenc, A. C. [1986]. Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.* **112**: 1177–1194.

Mahfouf, J.-F. [1991]. Analysis of soil moisture from near-surface parameters: A feasibility study, *J. Appl. Meteorol.* **30**: 1534–1547.

Mahfouf, J. F. and Noilhan, J. [1991]. Comparative study of various formulations of evaporation from bare soil using in situ data, *J. Appl. Meteorol.* **30**: 1354–1365.

Malanotte-Rizzoli, P. (ed.) [1996]. *Modern Approaches to Data Assimilation in Ocean Modeling*, Vol. 61 of *Oceanography Series*, Elsevier, New York, NY.

McLaughlin, D. [1995]. Recent developments in hydrologic data assimilation, *Rev. Geophys.* pp. 977–984. U.S. National Report to the IUGG (1991-1994), published as Supplement to Rev. Geophys.

McLaughlin, D. and Townley, L. R. [1996]. A reassessment of the groundwater inverse problem, *Water Resour. Res.* **32**(5): 1131–1161.

Menemenlis, D., Fieguth, P., Wunsch, C. and Willsky, A. [1997]. Adaptation of a fast optimal interpolation algorithm to the mapping of oceanographic data, *J. Geophys. Res.* **102**(C5): 10573–10584.

Miller, C. T., Williams, G. A., Kelley, C. T. and Tocci, M. D. [1998]. Robust solution of Richards' equation for nonuniform porous media, *Water Resour. Res.* **34**: 2599–2610.

Milly, P. C. D. [1986]. Integrated remote sensing and modeling of soil moisture: Sampling frequency, response time, and accuracy of estimates, *in* M. E. Moss (ed.), *Integrated Design of Hydrological Networks, Proceedings of the Budapest Symposium*, number 158 in *IAHS Redbooks*, IAHS, pp. 201–211.

Njoku, E. G. and Kong, J.-A. [1977]. Theory for passive microwave remote sensing of near-surface soil moisture, *J. Geophys. Res.* **82**(20): 3108–3118.

Noilhan, J. and Planton, S. [1989]. A simple parameterization of land surface processes for meteorological models, *Mon. Weather Rev.* **117**(3): 536–549.

Owe, M., van de Griend, A. A., de Jeu, R., de Vries, J. J., Seyhan, E. and Engman, E. T. [1999]. Estimating soil moisture from satellite microwave observations: Past and ongoing projects, and relevance to GCIP, *J. Geophys. Res.* **104**(D16): 19735–19742.

Parlange, M. B., Katul, G. G., Folegatti, M. V. and Nielsen, D. R. [1993]. Evaporation and the field scale soil water diffusivity function, *Water Resour. Res.* **29**(4): 1279–1286.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. [1992]. *Numerical Recipes in C*, second edn, Cambridge University Press, New York, NY.

Rabier, F., Courtier, P., Pailleux, J., Talagrand, O. and Vasiljevic, D. [1993]. A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensional variational analysis, *Q. J. R. Meteorol. Soc.* **119**: 845–880.

Raju, S., Chanzy, A., Wigneron, J.-P., Calvet, J.-C., Kerr, Y. and Laguerre, L. [1995]. Soil moisture and temperature profile effects on microwave emission at low frequencies, *Remote Sens. Environ.* **54**: 85–97.

Reichle, R., McLaughlin, D. and Entekhabi, D. [1999]. Four-dimensional data assimilation and down-scaling of remote sensing SGP97 observations for the estimation of profile soil moisture and soil temperature, *Preprint Volume of the 14th Conference on Hydrology, Dallas, TX*, American Meteorological Society, Boston, MA, pp. 133–136.

Reichle, R., McLaughlin, D. and Entekhabi, D. [2000]. Variational data assimilation of soil moisture and temperature from remote sensing observations, *IAHS Red Book on Calibration and Reliability in Groundwater Modelling – Coping with Uncertainty. Proceedings of the ModelCARE 99 Conference, Zürich, September 1999*, IAHS, Wallingford, Oxfordshire OX10 8BB, UK. Accepted for publication.

Reid, L. B. [1996]. *A Functional Inverse Approach for Three-Dimensional Characterization of Subsurface Contamination*, PhD dissertation, Massachusett Institute of Technology, Dept. of Civil and Environmental Engineering.

Rhodin, A., Kucharski, F., Callies, U., Eppel, D. P. and Wergen, W. [1999]. Variational analysis of effective soil moisture from screen-level atmospheric parameters: Application to a short-range weather forecast model, *Q. J. R. Meteorol. Soc.* **125**: 2427–2448.

Schmugge, T. J. and Choudhury, B. J. [1981]. A comparison of radiative transfer models for predicting the microwave emission from soil, *Radio Sci.* **16**(5): 927–938.

Schmugge, T. J. and Jackson, T. J. [1992]. A dielectric model of the vegetation effects on the microwave emission from soils, *IEEE Trans. Geosci. Remote Sens.* **30**(4): 757–760.

Sellers, P. J. and Dorman, J. L. [1987]. Testing the Simple Biosphere Model (SiB) using point micrometeorological and biophysical data, *J. Climate Appl. Meteorol.* **26**: 622–651.

Sellers, P. J., Los, S. O., Tucker, C. J., Justice, C. O., Dazlich, D. A., Collatz, G. J. and Randall, D. A. [1996]. A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: The generalization of global fields of terrestrial biophysical parameters from satellite data, *J. Climate* **9**(4): 706–737. See also [Sellers, Randall, Collatz, Berry, Field, Dazlich, Zhang, Collelo and Bounoua, 1996].

Sellers, P. J., Mintz, Y., Sud, Y. C. and Dalcher, A. [1986]. A simple biosphere model (SiB) for use within general circulation models, *J. Atmos. Sci.* **43**(7): 505–531.

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., Zhang, C., Collelo, G. D. and Bounoua, L. [1996]. A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation, *J. Climate* **9**(4): 676–705.

Simunek, J., Huang, K., Sejna, M. and van Genuchten, M. T. [1997]. *The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media*, U. S. Salinity Laboratory, Agricultural Research Service, U. S. Department of Agriculture, Riverside, CA. Draft Version 1.0.

Sun, C.-C. [1997]. *A Stochastic Approach for Characterizing Soil and Groundwater Contamination at Heterogeneous Field Sites*, PhD dissertation, Massachusett Institute of Technology, Dept. of Civil and Environmental Engineering.

Thépaut, J.-N. and Courtier, P. [1991]. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model, *Q. J. R. Meteorol. Soc.* **117**: 1225–1254.

Thépaut, J.-N., Vasiljevic, D., Courtier, P. and Pailleux, J. [1993]. Variational assimilation of conventional meteorological observations with a multilevel primitive-equation model, *Q. J. R. Meteorol. Soc.* **119**: 153–186.

Thiebaux, H. J. [1976]. Anisotropic correlation functions for objective analysis, *Mon. Weather Rev.* **104**: 994–1002.

Thompson, S. L. and Pollard, D. [1995]. A global climate model (GENESIS) with a land-surface transfer scheme (LSX): 1. Present climate simulation, *J. Climate* **8**(4): 732–761.

Todling, R. and Cohn, S. E. [1994]. Suboptimal schemes for atmospheric data assimilation based on the Kalman filter, *Mon. Weather Rev.* **122**: 2530–2557.

Ulaby, F. T., Moore, R. K. and Fung, A. K. [1986]. *Microwave Remote Sensing*, Vol. I–III, Artech House Inc., Norwood, MA.

van de Griend, A. A. and Owe, M. [1994]. Bare soil surface resistance to evaporation by vapor diffusion under semi-arid conditions, *Water Resour. Res.* **30**(2): 181–188.

Verlaan, M. and Heemink, A. W. [1997]. Tidal flow forecasting using reduced rank square root filters, *Stoch. Hydrol. Hydraul.* **11**(5): 349–368.

Wetzel, P. J. and Chang, Y.-T. [1988]. Evapotranspiration from nonuniform surfaces: A first approach for short-term numerical weather prediction, *Mon. Weather Rev.* **116**(3): 600–621.

Wilheit, T. T. [1978]. Radiative transfer in a plane stratified dielectric, *IEEE Trans. on Geosci. Electronics* **GE-16**(2): 138–143.

Willsky, A. S. [1995]. Recursive estimation, 6.433 Course Notes, Dept. of Electrical Eng. and Computer Sciences, Mass. Inst. of Technol., Cambridge, MA.

Wunsch, C. [1996]. *The Ocean Circulation Inverse Problem*, Cambridge University Press, New York, NY.

Xue, Y., Sellers, P. J., Kinter, J. L. and Shukla, J. [1991]. A simplified biosphere model for global climate studies, *J. Climate* **4**: 345–364.

Yang, Z.-L., Dickinson, R. E., Shuttleworth, W. J. and Shaikh, M. [1998]. Treatment of soil, vegetation and snow in land surface models: A test of the Biosphere-Atmosphere Transfer Scheme with the HAPEX-MOBILHY, ABRACOS and Russian data, *J. Hydrol* **212–213**: 109–127.

Zheng, X. and Eltahir, E. A. B. [1998]. A soil moisture-rainfall feedback mechanism 2. Numerical experiments, *Water Resour. Res.* **34**(4): 777–785.

Zimmerman, D. A., de Marsily, G., Gotway, C. A., Marietta, M. G. et al. [1998]. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.* **34**(6): 1373 1413.