Hydrologic Data Assimilation of Multi-resolution Microwave Radiometer and Radar Measurements using Ensemble Smoothing

by

Susan Catherine Dunne

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2006

Author .....
Department of Civil and Environmental Engineering
January 24th, 2006

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

Accepted by .....
Andrew Whittle
Chairman, Department Committee for Graduate Students

Hydrologic Data Assimilation of Multi-resolution Microwave Radiometer and Radar Measurements using Ensemble Smoothing

by

Susan Catherine Dunne

Submitted to the Department of Civil and Environmental Engineering
on January 24th, 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in the field of Hydrology

Abstract

Previously, the ensemble Kalman filter (EnKF) has been used to estimate soil moisture and related fluxes by merging noisy low frequency microwave observations with forecasts from a conventional though uncertain land surface model (LSM). Here it is argued that soil moisture estimation is a reanalysis-type problem and thus smoothing is more appropriate than filtering.

An ensemble moving batch smoother, an extension of the EnKF in which the state vector is distributed in time, is used to merge synthetic ESTAR observations with modeled soil moisture. Results demonstrate that smoothing can improve over filtering. However, augmentation of the state vector increases the computational cost significantly, rendering this approach unsuitable for spatially distributed problems.

The ensemble Kalman smoother (EnKS) is an inexpensive alternative as the costly computations are already performed in the EnKF which provides the initial guess. It is used to assimilate observed L-band radiobrightness temperatures during the Southern Great Plains Experiment 1997. Estimated surface and root zone soil moisture is evaluated using gravimetric measurements and flux tower observations. It is shown that the EnKS can be implemented as a fixed-lag smoother with the required lag determined by the memory in subsurface soil moisture.

In a synthetic experiment over the Arkansas-Red river basin, "true" soil moisture from the TOPLATS model is used to generate synthetic Hydros observations which are subsequently merged with modeled soil moisture from the Noah LSM using the EnKS. It is shown that the EnKS can be used in a large problem, with a spatially distributed state vector, and spatially-distributed multi-resolution observations. This EnKS-based framework is used to study the synergy between passive and active observations, which have different resolutions and error distributions.

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

Acknowledgments

This research was funded by a NASA Earth System Science Fellowship (NGT5-30479), and supported by NASA grants NNG05GE58G and NAG5-11602. Many thanks to Tom Jackson and the SGP team for the use of their data from SGP97, and Wade Crow for providing the Hydros OSSE data for Chapter 5.

I am indebted to my advisor, Professor Dara Entekhabi, for his guidance and patience during my PhD program. I have learned a great deal from him academically, professionally and personally, and will always appreciate that.

I would like to thank my committee members Jim Hansen, Dennis McLaughlin and Eni Njoku for their invaluable suggestions and ideas on this project. Thanks also to the previous and current students from the Entekhabi, McLaughlin and Bras research groups for the many inspiring debates and discussions. Special thanks to Steve Margulis for all the great help and advice.

I am grateful for the continuous support and encouragement from my friends at MIT, and my family and friends at home in Dublin. Thanks to Gary Steele for keeping me sane, fed and watered during my PhD program.

Most of all I am grateful to my mother, Mona, who has always inspired me to work hard and make the most of the wonderful opportunities I have been given.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Soil moisture in the climate system

The land surface plays an important role in the climatic system as a large fraction of incoming solar radiation passes through the atmosphere and is dissipated at the surface through turbulent and radiative fluxes. Soil moisture is a key state variable which integrates much of the land surface hydrology, and exerts considerable control on several land-atmosphere exchanges. Figure 1-1 illustrates the many interactions between the land surface and the overlying atmosphere which are influenced by soil moisture. Antecedent soil moisture controls the partitioning of incident precipitation between infiltration and run-off. Through its impact on soil albedo, soil moisture affects the net radiation absorbed at the surface. Together with soil temperature, soil moisture determines the partitioning of incident solar radiation between ground, sensible and latent heat fluxes. By controlling latent heat flux, soil moisture influences the available convective potential energy, and consequently precipitation. These interactions occurs across a wide range of temporal and spatial scales, rendering soil moisture a key variable of interest in hydroclimatology and hydrometeorology.

At finer scales, hydrometeorology is concerned with processes on the scale of convective storms and small basins such as flash-flooding. It has been shown that the formation and growth of clouds as well as the evolution of precipitating weather systems over land is influenced by surface fluxes and surface soil moisture [78]. Paegle et al.[67]

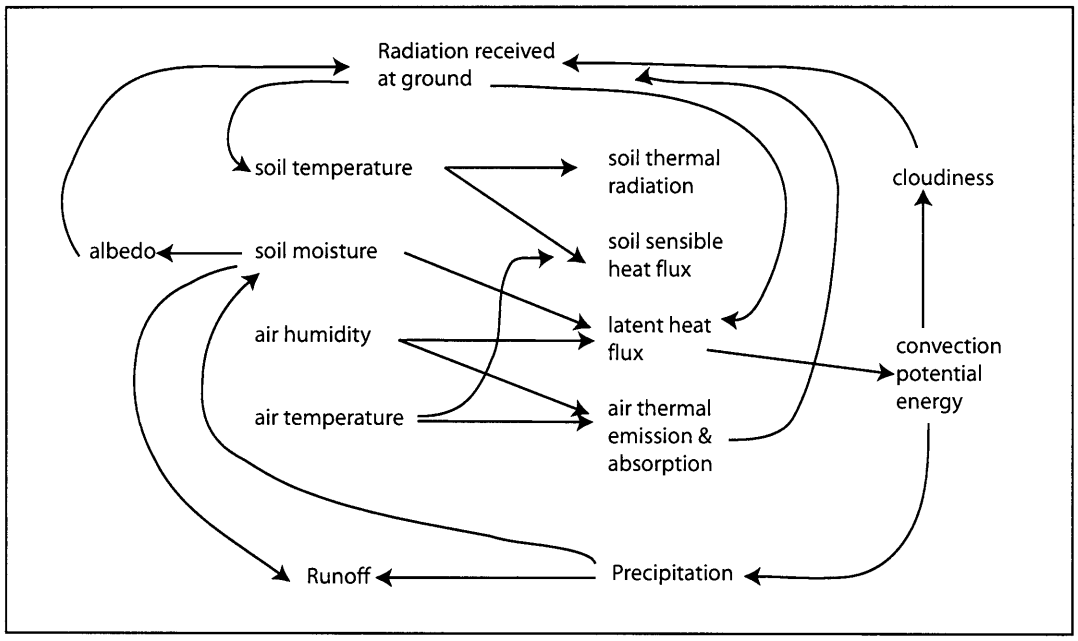

Figure 1-1: Conceptual diagram of the pathways through which soil moisture affects and is mutually influenced by the overlying atmosphere (from Entekhabi et al. [22]).

demonstrated that surface evaporation can influence rainfall production through the effect of the associated temperature field on buoyancy rather than through the water vapor distribution. The model study of Pan et al. [68] demonstrated that in flood conditions, the saturated surface made a significant contribution precipitation/water recycling. Chen et al. [6] demonstrated that knowledge of the temporal evolution of soil moisture improved the forecast of the convective rainfall event that caused flash flooding in Buffalo Creek, Colorado in 1996.

On larger scales, it has been shown that knowledge of soil moisture in some areas may be used to extend the lead times of seasonal forecasts. The atmosphere responds to anomalies in sea surface temperature (SST) in so-called "hot spots" of ocean-atmosphere coupling. Observations of SST in the eastern Pacific are used to predict the El Nino/La Nina cycle, thereby improving seasonal prediction. Koster et al. [54] conducted a multi-model experiment to identify analogous "hot spots" of land-atmosphere coupling, where persistence of an anomaly in soil moisture would impact precipitation on a seasonal timescale. As shown in Figure 1-2, major "hot spots"

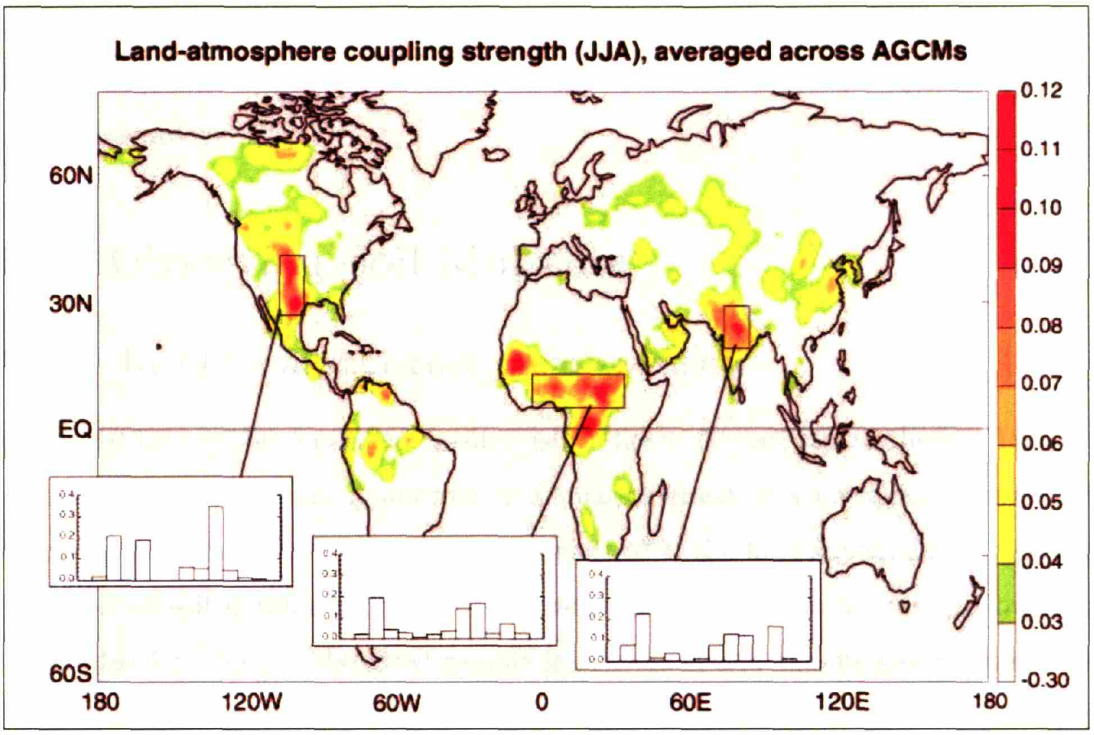

Figure 1-2: (From [54]) The land-atmosphere coupling strength diagnostic for boreal summer (the Ω difference, dimensionless, describing the impact of soil moisture on precipitation), averaged across the 12 models participating in GLACE. (Insets) Areally averaged coupling strengths for the 12 individual models over the outlined, representative hot spot regions. No signal appears in southern South America or at the southern tip of Africa.

occur in the central Plains of North America, the Sahel, equatorial Africa and India. Coupling is also significant in Canada, northern South America, central Asia and China. The strongest hot spots are generally found in areas where evaporation is limited by soil moisture, and where evaporation from the surface has a significant impact on moist convection, and consequently precipitation generation. Observations of soil moisture in these areas could improve the precipitation prediction skill during the Northern hemisphere summer. Furthermore, global satellite-based observations of soil moisture could be used to demonstrate the impact of soil moisture on precipitation using real data.

A consistent dataset of soil moisture, ground temperature and surface fluxes would enable a detailed study of land-atmosphere interactions and the role that they play

in climatic change. To address hydroclimatology questions, data on the scale of 30-50km is required. Higher resolution data, on the order of 10km is required to study hydrometeorology [23].

1.2 Observing Soil Moisture

1.2.1 In-situ soil moisture measurement

Soil moisture can be measured in-situ using direct or indirect methods. Indirect implies that soil moisture is inferred from measurement of some quantity which is affected by soil moisture. The gravimetric method is the best known direct method. A sample of soil is taken from the field, weighed and dried in an oven following a standardized protocol. The dried sample is reweighed and the moisture content (in terms of weight) is the difference between the wet and dry weights divided by the dry weight. Gravimetric measurements from the Southern Great Plains Experiment 1997 are used in Chapter 4 to validate estimated soil moisture. While this method is very simple, it is a destructive procedure, it is time-consuming and cannot be automated or used for continuous observation of soil moisture in a single location. Indirect methods such as Time or Frequency Domain Reflectometry (TDR or FDR) are often employed for continuous monitoring. These two methods exploit the relationship between the dielectric constant of the soil and its moisture content. The Global Soil Moisture Data Bank [75] has gathered soil moisture observations from research stations around the world. The earliest records date from 1927, from the former USSR and many data sets are 20-30 years in length. Most data are gravimetric observations at a point taken 3 times per month. Robock et al. [75] demonstrates how these data are useful for studying temporal and spatial scales of soil moisture variations, identifying trends in soil moisture and performing model inter-comparison studies. However, global or regional observations from in-situ measurements at the scales needed for hydroclimatological and hydrometeorological studies would be logistically and economically infeasible.

1.2.2 Remote-sensing of soil moisture

Remote sensing provides a way of obtaining frequent measurements of soil moisture over large areas. The suitability of low frequency microwave remote-sensing stems from the sharp contrast in dielectric constant between water ($\sim$ 80) and dry soil ($\sim$ 3) between 1-5GHz [81]. This difference in emissivity translates to a range of brightness temperature on the order of 100K for wet to dry soil at a given frequency. Passive microwave observations at low frequencies ($\sim$ 1.4GHz) are best suited to observation of soil moisture because at these frequencies microwave emissivity is primarily a function of soil moisture [88]. For any given soil texture, the sensitivity of the dielectric constant to soil moisture increases with decreasing frequency [36]. Soil texture largely dictates the relationship between moisture content and dielectric constant as it determines how bound water is held in the soil matrix. Soil texture is typically accounted for in dielectric mixing models ([18] and [90]), however difficulty can arise when soil texture is unknown or uncertain.

Soil moisture retrieval is complicated by surface roughness, although to a lesser extent at lower frequencies. The Fresnel equations assume that the air-soil interface is perfectly smooth within the area viewed by the radiometer. When the surface is rough, the area of the air-soil interface is greater, so there is a larger area to transmit upwelling energy. This results in a higher emissivity over a rough surface than that predicted by the Fresnel equations, which must be corrected for ([7],[89]).

The penetration depth, defined for emission from a soil as the e-folding depth (depth from above which 63% of the radiation emitted by the surface originates), increases with decreasing frequency [53]. This penetration depth is effectively the depth of the soil column being observed by the radiometer. While the final value of the penetration depth is a function of soil moisture, an example in Njoku and Entekhabi [66] illustrates that the penetration depth for L-band (say 1.41GHz) is on the order of centimeters, while in X-band (say 10GHz) it will be just millimeters. L-band remote-sensing of soil moisture is able to estimate volumetric water content in the top 0-5cm of the soil column with a precision of a few percent [47],[44],[45].

Even at microwave frequencies, retrieval is limited to bare soil or areas of low vegetation. Vegetation opacity is a function of frequency, the vegetation type (geometry and structure) and the vegetation water content [46]. Use of lower frequencies increases the area over which retrievals are possible.

As early as the late 1960's and early 1970's small studies were undertaken to determine the feasibility of using microwave brightness temperatures to estimate soil moistures. Figure 1-3 gives a brief history of passive microwave remote-sensing of soil moisture. It includes the main instruments, the type of experiments in which they were employed and a sub-section of the vast collection of subsequent publications. In the last five years, there have been several large scale field experiments; SMEX02 (e.g. [93]), SMEX03 (e.g. [43]) and SMEX04-NAME. With each experiment, retrieval algorithms are pushed to include more diverse conditions and vegetation.

Real and synthetic observations from ESTAR (SGP97) are used in Chapters 3 and 4 of this thesis. The radiative transfer model used to relate soil moisture to observed/simulated brightness temperature is given in Appendix B.

1.3 Hydrosphere State (Hydros) Mission

NASA's Hydrosphere State (Hydros) Satellite Mission is a pathfinder mission with the objective of providing global measurements of soil moisture at 10km resolution with a revisit time of three days [21].

Hydros will measure L-band microwave emission and backscatter from the Earth's surface using a combined passive and active sensor. A single feedhorn is shared by a 1.41 GHz passive channel and two active channels (1.26 H GHz and 1.29 V GHz). The feedhorn is positioned near the focus of a 6m diameter light-weight deployable mesh reflector (See Figure 1-4). This antenna diameter results in a radiometer footprint of 40km (root ellipsoid area), and a radar two-way 3-dB real aperture footprint of 30km. In order to image a 1000km wide swath, the antenna beam is rotated around the nadir direction resulting in an fixed incidence angle of 39.3^o with the Earth's surface (See Figure 1-5). Making a constant angle with the surface simplifies atmospheric

Figure 1-3: A Brief History of Developments in Passive Remote-Sensing of Soil Moisture (1960-2000)

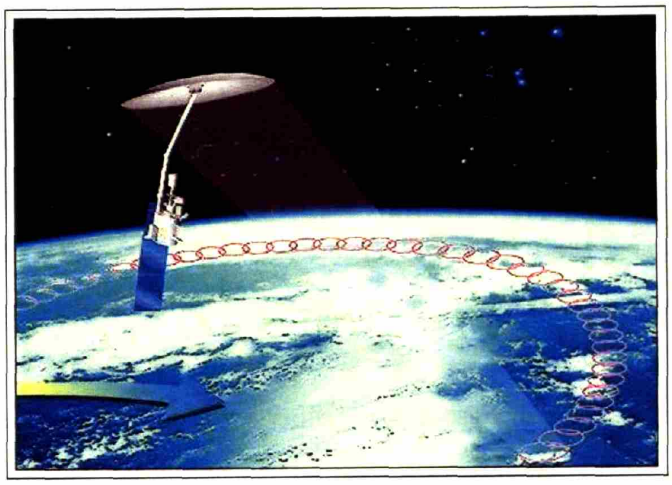

Figure 1-4: Artist's impression of NASA's Hydrosphere State (Hydros) Mission. (Source: http://hydros.nasa.gsfc.gov)

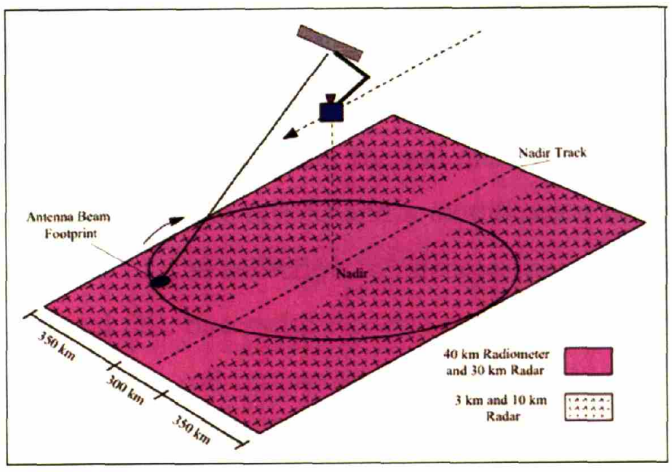

Figure 1-5: Hydros Radiometer and Radar Observation within 1000km swath. (Source:http://hydros.nasa.gsfc.gov)

corrections and facilitates accurate retrieval. The speed of rotation of the reflector relative to the instrument movement along the nadir track determines the sampling pattern on the ground. Rotating the reflector at 14.6rpm ensures contiguous coverage. Range and Doppler discrimination is used to subdivide the radar footprint to obtain 3km data. While radar resolution will be better than 3km over 70% of the swath, squint angle effects prevent high resolution observations at the center of the swath. Hydros will be in a low earth orbit at an altitude of 670km, ensuring whole-earth coverage with a revisit time of 2-3 days. It is in a sun-synchronous polar orbit, meaning that it precesses about the polar axis at one revolution per year. The advantage of this orbit is that the satellite will cross the same latitude at the same local time regardless of the longitude or the date. Hydros will have 6 a.m./6 p.m. equatorial crossings to provide ideal conditions for retrieval. Ionospheric effects are minimized at dawn, and soil moisture and temperature profiles are uniform.

Soil moisture may be obtained from the radiometer measurements through inversion of the following expression:

$$T_{Bp} = T_s e_p exp(-\tau_c) + T_c(1-\omega)[1-exp(-\tau_c)][1+r_p exp(-\tau_c)] \qquad (1.1)$$

where T_s and T_c are the soil and canopy temperatures (K), τ_c is the vegetation opacity along the slant path at look angle θ, ω is the single-scattering albedo, and r_p is the soil surface reflectivity which is obtained from the soil dielectric constant (a function of soil moisture) using the Fresnel equations and a modification to account for surface roughness. In Equation 1.1 the subscript "p" denotes polarization, which may be vertical (V) or horizontal (H). Successful retrieval requires ancillary data such as soil and vegetation temperatures, soil texture and surface roughness.

Total co-polarized (pp = HH or VV) radar backscatter from the surface, σ^t_{pp} is the sum of three contributions:

$$\sigma^t_{pp} = \sigma^s_{pp} exp(-2\tau_c) + \sigma^{vol}_{pp} + \sigma^{int}_{pp} \qquad (1.2)$$

The dominant contribution in bare soil or low-vegetation is the soil surface backscat-

ter, a function of soil moisture and RMS surface roughness. In the presence of vegetation, this signal is subject to two-way attenuation through the vegetation layer along the slant path. The other backscatter contributions are from the vegetation volume (σ_{pp}^{vol}) and the interaction between the vegetation and soil surface (σ_{pp}^{int}). The greatest obstacle in retrieval of soil moisture from backscatter data is vegetation. Backscatter is influenced by the geometry and orientation of vegetation components, as well as ground slope. Accurate retrievals are limited to regions where vegetation water content is less than 0.5 kg m^{-2} [80].

While radar observations have higher resolution than radiometer observations and contain valuable information on vegetation and roughness effects within the radiometer footprint, they are less accurate in vegetated areas than radiometer observations. One of the Hydros mission goals is to capitalize on the synergy between the active and passive observations to produce a 10km soil moisture product containing the best information from both observations.

The objective of this thesis is to develop a data assimilation framework in which uncertain brightness temperature and backscatter measurements from Hydros may be merged with uncertain modeled estimates of the land surface states to produce a value-added soil moisture data product. As synthetic Hydros data will be generated and used later in this thesis, a more thorough discussion of the Hydros microwave emission and backscatter model is provided in Appendix C.

1.4 Land Data Assimilation

While remote sensing offers the advantage of global coverage, the temporal resolution of observations is limited by the revisit time. The Hydros satellite will revisit a given location on the Earth's surface just once every 2-3 days. Furthermore, the L-band brightness temperature relates to the soil moisture at the surface (top 5cm) and yields no information on the root zone. Forcing a land surface model with meteorological data can produce soil moisture and temperature estimates, along with the associated fluxes at the temporal resolution of the model yielding information on the diurnal

cycle. However, such unconstrained simulations are subject to the errors in model structure and forcing uncertainty. Data assimilation offers a means to combine the advantages of modeling with those of remote sensing.

Data assimilation techniques have been used in meteorology and oceanography for decades. A comparison of the various techniques is provided by Ghil and Manalotte-Rizzoli [31]. Courtier [11] compiled a list of significant papers in the application of data assimilation techniques to meteorology problems. Data assimilation techniques can be roughly divided into two categories; variational techniques and those derived from the classic Kalman filter. Both methods have been applied to hydrological research in recent years.

Variational techniques were first introduced in meteorology by Y. Sasaki in the 1960s and 70s. The central concept in variational data assimilation is the adjoint model. This is obtained by linearizing the forward model along a trajectory producing the tangent-linear model, and obtaining the adjoint. Thus, variational techniques require that the forward model be differentiable. The adjoint model describes the evolution of sensitivity to initial conditions, boundary conditions and parameters of the model. [60] describes various variational techniques which have been applied in meteorology. Several applications in oceanography and meteorology are discussed by [31] and [91]. Variational techniques have been successfully applied to hydrological applications in recent years [4], [1], [71], [72], [73], [62]. 4DVAR, in which observations distributed in space and time are used with knowledge of temporal evolution of the state, is particularly suited to our problem as demonstrated by [71], but it requires development of the adjoint. While automatic adjoint compilers are available [32], they can prove difficult to use with large and intricate numerical models, and typically involve extensive tuning and sensitivity studies to validate the adjoint model generated. A means is sought by which temporally distributed observations may be used in a smoothing approach like 4DVAR without resorting to a simplified land surface model.

The classic Kalman filter as discussed by Gelb [30] provides the optimal state estimate for linear systems. It is therefore of little use in hydrological applications where

the physical model equations are often non-linear and contain thresholds. In the extended Kalman filter for non-linear systems ([30], [48]), approximate expressions are found for the propagation of the conditional mean and its associated covariance matrix. The structure of the propagation equations is similar to those of the classic Kalman filter for a linear system, as they are linearized about the conditional mean. The extended Kalman filter has been successfully applied to the land data assimilation problem ([20],[28], [87], [86], [12]), but its use in this application would require derivation of a tangent linear model to approximate the land surface model, as well as techniques to treat the instabilities which might arise from such an approximation. Ljung [58] performed a convergence analysis of the extended Kalman filter and demonstrated the potential for divergence or bias in estimates in non-linear systems. Nakamura et al. [65] encountered such instability in their application of the extended Kalman filter to soil moisture estimation.

An alternative sequential estimation technique for non-linear problems was proposed by Evensen [24]. In the ensemble Kalman filter (EnKF) an ensemble of model states is integrated forward in time using the non-linear forward model with replicates of system noise. At update times, the error covariance is calculated from the ensemble. The traditional update equation from the classical Kalman filter is used, with the Kalman gain calculated from the error covariances provided by the ensemble. The EnKF has been successfully implemented by Evensen and Van Leeuwen [27] and Houtekamer and Mitchell ([39], [40]) and has already been used to merge L-band observations with model output to estimate soil moisture ([74],[64],[13],[16]). Research in ensemble techniques has yielded innovative methods of improving estimates and reducing the computational burden ([77], [38], [84], [85]). The advantages and disadvantages of the EnKF are compared to those of variational techniques in Table 1.1.

In the past, soil moisture observations have typically been gathered during field experiments such as the Southern Great Plains Field Experiments (SGP97 and SGP99)

Table 1.1: The advantages and disadvantages of ensemble-based filters are compared to those of variational techniques.

	Ensemble-based Filters	Variational Techniques
Advantages	Any model can be used. Model does not need to be differentiable Noise can be placed anywhere, e.g. on uncertain parameters and forcing. Noise can be non-Gaussian and non-additive.	Uses all data in a batch window to estimate the state.
Disadvantages	Estimates are conditioned on past measurements only	Model must be differentiable to obtain tangent-linear model. Process noise can only be additive and Gaussian. Changes to model require that adjoint be obtained again.

and Soil Moisture Experiments in 2002 (SMEX02) and 2003 (SMEX03). Smoothing is ideal for analyzing historic data or data which are not available in real-time, as is the case with data from field experiments or exploratory missions such as Hydros and SMOS. Smoothing involves using all measurements in an interval $\mathbf{T} = [0, T]$, to estimate the state of the system at some time t where $0 \leq t \leq T$, so that the state estimate at a given time is determined by including information from subsequent observations. It will be argued that an ensemble-based smoothing (or batch estimation) approach is most suited to the soil moisture estimation problem.

Results from the EnKF experiment [64] suggest that the estimate could be improved through the inclusion of subsequent observations. Precipitation events divide the study interval into a series of dry-down events. In estimating soil moisture at a given time, one is estimating a single point value in a series. It is intuitive that the manner in which that series evolves in the future is related to the state at the estimation time. Future observations provide information on the shape of this series in the future and so contain useful information on the current state. Correlation between the states and the observations decreases with depth as the observations relate to the surface conditions. Consequently the impact of the observations is lessened with increasing depth. This means that it takes longer to correct for spurious initial conditions at depth than close to the surface. As the impact of the observations eventually penetrates the deeper layers, the latent heat flux estimate is seen to approach the observed values. Difficulty in estimating the root zone soil moisture results in poor initial estimates of the latent heat flux [64]. If including subsequent observations can improve on the initial conditions at depth, it would result in improved latent heat flux estimates.

Several ensemble smoothers exist in data assimilation literature e.g. the ensemble smoother (ES) of Van Leeuwen and Evensen [83] and the ensemble Kalman smoother (EnKS) of Evensen and Van Leeuwen [26]. They have been used in various applications such as ocean forecasting ([2], [82],[83]) and fish stock assessment [34] and the objective of this research is to determine their applicability to soil moisture estimation. Its performance is compared to the EnKF to determine if an improved estimate

of soil moisture with ensemble smoothing, and to identify issues which may be significant in the implementation of an ensemble smoother in a land data assimilation framework.

1.5 Thesis Outline

In Chapter 2, it will be shown that smoothing, the inclusion of observations prior to and after an estimation time, generally yields improved results over just filtering. The objective of Chapter 2 is to test a number of ensemble smoother algorithms to determine if they might be used to improve on the estimate obtained using the ensemble Kalman filter. For simplicity, the algorithms will be tested first for the linear gaussian problem. For this problem, the optimum filter is known to be the traditional Kalman filter, and the optimum smoother can be expressed in terms of the Rauch-Tung-Striebel (RTS) smoother algorithm. The solutions obtained using the optimal filter and smoother shall be used to evaluate the performance of the proposed smoothers. At the very least, the estimate obtained from the proposed smoother should be an improvement over that obtained using the Kalman filter. Ideally, a solution is sought which approximates that obtained using the RTS smoother.

In Chapter 3 an ensemble smoother, described in Chapter 2, is used in which the state vector and measurement vector are distributed in time and updated as a batch. Its performance in a land data assimilation context is compared to that of the ensemble Kalman filter. Results will show that smoothing yields an improved estimate compared to filtering, reflected in the decreased deviation from truth and the reduction in uncertainty associated with the estimate. It will become clear that precipitation significantly impacts the performance of the smoother, acting as an information barrier between dry-down events. An adaptive hybrid filter/smoother is presented in which brightness temperature may be used to break the study interval into a series of drydown events. The smoother is used on dry-down events and the filter is used when precipitation is evident between estimation times. It will be shown that an improved estimate is obtained as all observations in a given dry-down

period are used to estimate soil moisture in that period, and backward propagation of information from subsequent precipitation events is avoided.

In Chapter 4 the ensemble Kalman smoother (EnKS) is employed to estimate surface and subsurface soil moisture and surface energy fluxes during the Southern Great Plains Experiment 1997 (SGP97) through the assimilation of observed L-band radiobrightness temperatures. While the EnKF uses observations as they become available to update the current state, the EnKS takes the EnKF estimate as its first guess. However, in addition to updating the current state, it also updates the best estimate at previous times. The performance of the EnKS is compared to the EnKF and the ensemble open loop (EnOL) in which no measurements are assimilated. Estimated surface soil moisture is compared to gravimetric observations at three locations. Root zone (5-100cm) soil moisture is evaluated by comparing the resultant latent heat flux to flux tower observations. In a fixed lag smoother, observations are used to update past estimates within a fixed time window. The EnKS can be implemented in a fixed lag formulation in problems with limited memory such as soil moisture estimation. It is shown that there is a trade-off to be made between the improved accuracy with longer lag and the increased computational cost incurred. It is demonstrated that the EnKS is a relatively inexpensive state estimation algorithm suited to operational data assimilation.

In Chapter 5, a synthetic experiment is carried out over the Arkansas-Red river basin, in which "true" soil moisture is obtained from the TOPLATS (TOPMODEL-based land-atmosphere transfer scheme) model at 1km and used to generate synthetic Hydros observations. The ensemble Kalman smoother (EnKS) is used to merge these multi-resolution observations with modeled soil moisture from the Noah Land Surface Model to estimate surface and subsurface soil moisture at 6km resolution. The EnKS is an extension of the ensemble Kalman filter (EnKF) in which observations are used to update states at previous times. It is shown that the EnKS can be used in a large problem, with a spatially distributed state vector, and spatially-distributed multi-resolution observations. The data assimilation framework built around the EnKS is used to study the synergy between passive and active observations which have

different resolutions and measurement error distributions. The extent to which the design parameters of the EnKS vary depending on the combination of observations assimilated is investigated.

In Chapter 6, a synthetic experiment is carried out to study the performance of the EnKF and EnKS in a soil moisture estimation problem when the observations are subject to a constant bias.

In Chapter 7, the original contributions and key findings of this thesis are summarized and future research directions are outlined.

Chapter 2

Smoothing Algorithms

2.1 Introduction

In this chapter, the goal is to test smoothing algorithms which will yield improved estimates (i.e. reduced root mean square error) over a Kalman filter. Gelb [30] developed an expression for the optimum smoother as a combination of a forward and a backward filter. It is shown that while this expression provides valuable insight into the power of smoothing rather than filtering, its implementation requires the tangent linear representation of the system of interest. An equivalent and more useful expression of the optimal smoother for linear systems is the Rauch-Tung-Striebel (RTS) formulation [70].

The objective here is to develop an ensemble-based algorithm to approximate the solution obtained using an RTS smoother for a simple linear model. Two algorithms are considered; the ensemble single batch smoother (EnSB) and the ensemble moving batch smoother (EnMB). These algorithms were tested on a linear auto-regressive model of order 1 (AR(1)), for which it is very easy to apply a linear RTS smoother. In later chapters, it will be determined whether the results are also applicable in a land surface data assimilation context.

The traditional Kalman filter equations are central to each of the algorithms developed in this chapter, and will be the starting point for the following discussion. The RTS smoother will be developed and will serve as the benchmark against which

the algorithms will be assessed. The EnSB and EnMB algorithms shall be developed and studied in the remainder of the chapter.

The results presented in this chapter were obtained by running an AR(1) (with $\Phi = 0.9$) over $N = 1000$ time steps. Unless otherwise stated, observations were available at every 10th time step. Model noise was $\sim N(0, 2)$ and noise in the observations was $\sim N(0, 1)$. The objective of this chapter is to find an ensemble smoothing algorithm which will yield results that are a considerable improvement over the Kalman filter results and tend towards those of the RTS smoother.

2.2 Kalman Filtering

The traditional Kalman filter is at the core of each of the algorithms discussed in this chapter. In the Rauch-Tung-Striebel algorithm, the Kalman filter is used to provide the filtered estimate for the state at each time step in the forward filtering step. In the EnSB smoother, the Kalman filter equations are used to update an augmented state vector consisting of the quantity of interest at all times in the interval. In the EnMB smoother, the Kalman filter equations are used to update an augmented state vector consisting of the quantity of interest at the times of interest in a given interval.

Developed first in the 1960s ([51],[50]), the traditional Kalman filter equations are derived in many standard text books in estimation theory (e.g. [30],[48],[49]).Consider the system model of equation (2.1). At time k, the states $\underline{x}$ are some linear combination of the states at time $(k-1)$, combined with zero mean white noise $\underline{w}_k$ with variance Q_k.

$$\underline{x}_k = \Phi_{k-1}\underline{x}_{k-1} + \underline{w}_{k-1} \tag{2.1}$$

We try to estimate $\underline{x}$ by merging this model with discrete observations. The measurement process is described in (2.2), where the set of m measurements at time k, $\underline{z}_k$, is some linear combination of the n elements of the state vector, $\underline{x}_k$, corrupted by

zero mean white noise $\underline{\epsilon}_k$ with variance R_k.

$$\underline{z}_k = H_k \underline{x}_k + \underline{\epsilon}_k \tag{2.2}$$

The system model is initialized by:

$$E[\underline{x}(0)] = \underline{\hat{x}}_0 \tag{2.3}$$

$$E[(\underline{x}(0) - \underline{\hat{x}}_0)(\underline{x}(0) - \underline{\hat{x}}_0)^T] = P_0 \tag{2.4}$$

It is assumed that noise is uncorrelated:

$$E[\underline{w}_k \underline{\epsilon}_j^T] = 0 \; \forall j, k \tag{2.5}$$

The conditional expectation or prior estimate at time k, and its error covariance are extrapolated from the updated values at the previous update step:

$$\underline{\hat{x}}_k(-) = \Phi_{k-1} \underline{\hat{x}}_{k-1}(+) \tag{2.6}$$

$$P_k(-) = \Phi_{k-1} P_{k-1}(+) \Phi_{k-1}^T + Q_{k-1} \tag{2.7}$$

The state estimate is updated using:

$$\underline{\hat{x}_k}(+) = \underline{\hat{x}}_k(-) + K_k[\underline{z}_k - H_k \underline{\hat{x}}_k(-)] \tag{2.8}$$

where the Kalman Gain is given by:

$$K_k = P_k(-) H_k^T [H_k P_k(-) H_k^T + R_k]^{-1} \tag{2.9}$$

The error covariance of the updated state is given by:

$$P_k(+) = [I - K_k H_k] P_k(-) \tag{2.10}$$

The equations summarized in this section are the traditional Kalman filter equations

used in each of the algorithms in this chapter.

2.3 Optimal Smoothing

2.3.1 Optimal Smoother as a Forward and Backward Filter

Consider the estimation problem in continuous time, in which we seek an estimate at some time t, using information from the entire interval T. Gelb [30] develops a framework for the optimum smoother ($\underline{\hat{x}}(t|T)$) in terms of a forward model ($\underline{\hat{x}}(t)$) and a backward filter ($\underline{\hat{x}}_b(t)$) as follows:

$$\underline{\hat{x}}(t|T) = A\underline{\hat{x}}(t) + A'\underline{\hat{x}}_b(t) \tag{2.11}$$

where A and A' are some weighting matrices. The forward filter operates on all the data prior to the time t, while the backward filter operates on the data at times later than t. The resultant smoothed estimate therefore includes information from all observations in the interval. If we write each estimate as the true value plus some estimation error, then we can write

$$\underline{x}(t) + \underline{\tilde{x}}(t|T) = [A\underline{x}(t) + A\underline{\tilde{x}}(t)] + [A'\underline{x}_b(t) + A'\underline{\tilde{x}}_b(t)] \tag{2.12}$$

$$\underline{\tilde{x}}(t|T) = [A + A' - I]\underline{x}(t) + A\underline{\tilde{x}}(t) + A'\underline{\tilde{x}}_b(t) \tag{2.13}$$

If our filter yields unbiased errors $\underline{\tilde{x}}(t)$ and $\underline{\tilde{x}}_b(t)$, we can guarantee an unbiased smoothing error $\underline{\tilde{x}}(t|T)$ if we set $[A+A'-I]$ to zero. The smoother error may therefore be written as

$$\underline{\tilde{x}}(t|T) = A\underline{\tilde{x}}(t) + [I - A]\underline{\tilde{x}}_b(t) \tag{2.14}$$

from which we can calculate the smoother error covariance, $P(t|T)$.

$$P(t|T) = E[\underline{\tilde{x}}(t|T)\underline{\tilde{x}}^T(t|T)] \quad (2.15)$$

$$= AP(t)A^T + [I - A]P_b(t)[I - A]^T \quad (2.16)$$

where $P(t)$ and $P_b(t)$ are the optimal forward and backward filter error covariance matrices.

We can find the covariance of the optimal smoother by minimizing (2.16) with respect to A. Differentiating yields:

$$0 = 2AP + 2[I - A]P_b[-I] \quad (2.17)$$

which may be rewritten

$$A = P_b(P + P_b)^{-1} \quad (2.18)$$

and implies that

$$I - A = P(P + P_b)^{-1}. \quad (2.19)$$

Simply inserting these expressions into (2.16) yields a rather cumbersome expression for $P(t|T)$.

$$P(t|T) = P_b(P + P_b)^{-1}P(P + P_b)^{-1}P_b + P(P + P_b)^{-1}P_b(P + P_b)^{-1}P \quad (2.20)$$

Algebraic manipulation simplifies this expression to a more useful form:

$$P^{-1}(t|T) = P^{-1}(t) + P_b^{-1}(t) \quad (2.21)$$

This expression demonstrates that the smoothed estimate is always better or as good as the filter estimate, as the covariance of the smoother estimate is always less than or equal to the filter covariance. This concept is illustrated in Figure (2-1).

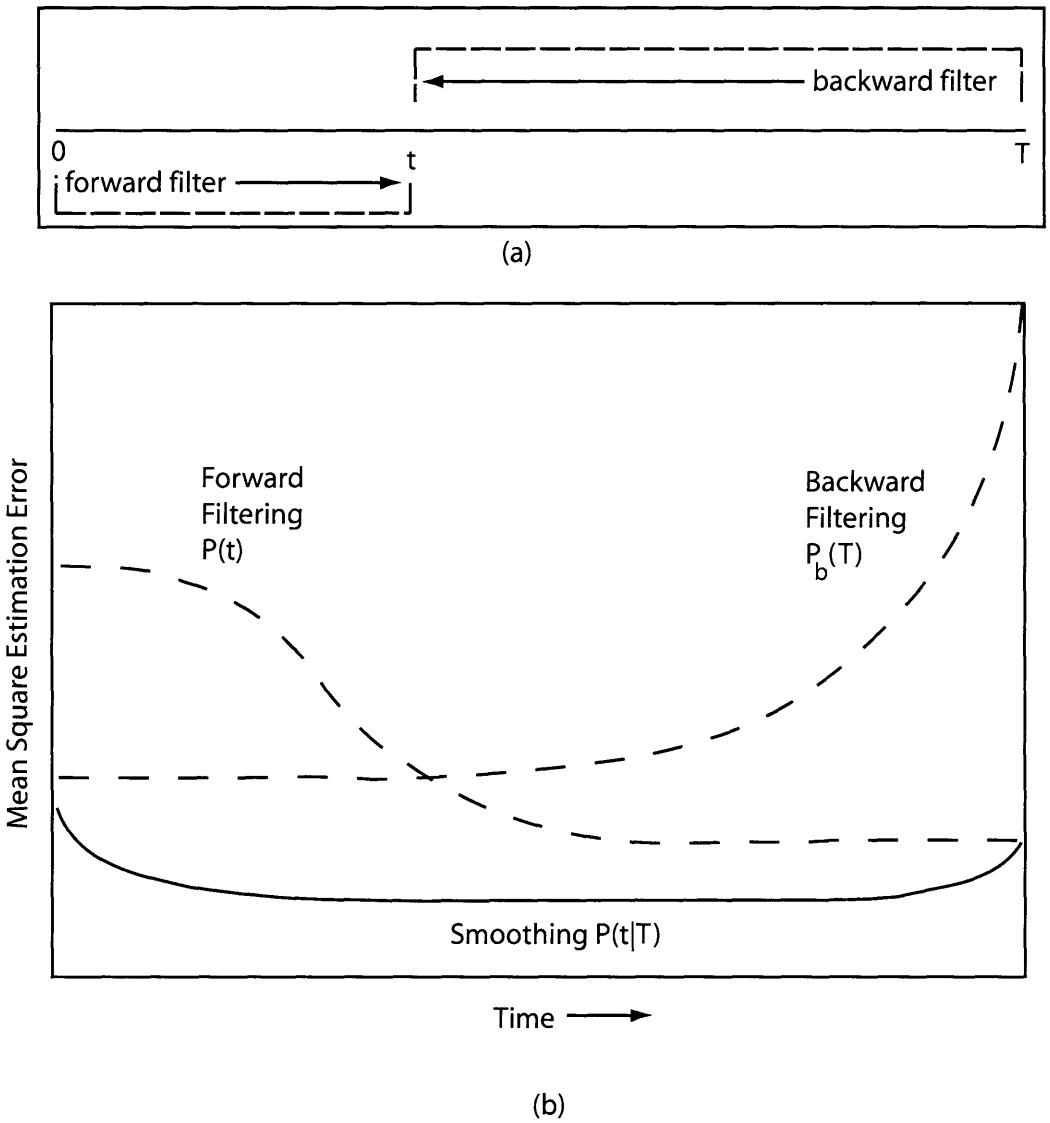

Figure 2-1: Advantage of performing optimal smoothing. Adapted from Gelb [30]

Inserting the same expressions into (2.11), gives an expression for the optimal smoother in terms of the forward and backward filters:

$$\hat{\underline{x}}(t|T) = P(t|T)[P^{-1}(t)\hat{\underline{x}}(t) + P_b^{-1}(t)\hat{\underline{x}}_b(t)] \tag{2.22}$$

The key benefit of expressing the optimal smoother in terms of a forward filter and a backward filter is the proof offered in equation (2.21) that smoothing yields an improved estimate over filtering. Its application in real problems, however, is limited by the need to obtain the tangent linear model to find $\hat{\underline{x}}_b(t)$.

2.3.2 Rauch-Tung-Striebel Formulation

Equations (2.21) and (2.22) provide expressions for the optimal smoother and its covariance in terms of those for the forward and backward filter. It is frequently more useful to work with expressions in which we do not need to literally filter backwards, but in which we can somehow propagate information backwards through the interval. The Rauch-Tung-Striebel smoother equations are one such set of equations for use with strictly linear systems. Rauch et al. [70] proposed a solution to the linear Gaussian smoothing problem based on the principle of the maximum likelihood estimator (MLE). Their solution takes the form of a backward recursive equation that relates the maximum likelihood estimate $\hat{\underline{x}}_{k|N}$ to the MLE of $\hat{\underline{x}}_{k+1|N}$ and the MLE of $\hat{\underline{x}}_{k|k}$.

From the principle of maximum likelihood estimation, the estimate of $\underline{x}_k$ given all observations in the interval [0,N] (Z_N), $\hat{\underline{x}}_{k|N}$ is that value of $\underline{x}_k$ which maximizes the function

$$L(\underline{x}_k, Z_N) = \log p(\underline{x}_k|Z_N) \tag{2.23}$$

Similarly, $\hat{\underline{x}}_{k|N}$ and $\hat{\underline{x}}_{k+1|N}$ are the values of $\underline{x}_k$ and $\underline{x}_{k+1}$ which maximize

$$L(\underline{x}_k, \underline{x}_{k+1}, Z_N) = \log p(\underline{x}_k, \underline{x}_{k+1}|Z_N) \tag{2.24}$$

Rauch et al. ([70]) equate this problem to that of minimizing the following with respect to $\underline{\hat{x}}_{k|N}$:

$$J = [\underline{\hat{x}}_{k+1|N} - \Phi_k \underline{x}_k] Q_k^{-1} [\underline{\hat{x}}_{k+1|N} - \Phi_k \underline{x}_k] + [\underline{x}_k - \underline{\hat{x}}_{k|k}] P_{k|k}^{-1} [\underline{x}_k - \underline{\hat{x}}_{k|k}] \quad (2.25)$$

This yields the solution

$$\underline{\hat{x}}_{k|N} = \underline{\hat{x}}(+) + A_k [\underline{\hat{x}}_{k+1|N} - \underline{\hat{x}}_{k+1}(-)] \quad (2.26)$$

where

$$A_k = P_k(+) \Phi_k^T P_{k+1}^{-1}(-) \quad (2.27)$$

The smoothing solution is obtained by computing backwards through the interval starting with:

$$\underline{\hat{x}}_{N|N} = \underline{\hat{x}}_N(+) \quad (2.28)$$

The error covariance is propagated backwards using:

$$P_{k|N} = P_k(+) + A_k [P_{k+1|N} - P_{k+1}(-)] A_k^T \quad (2.29)$$

starting with

$$P_{N|N} = P_N(+) \quad (2.30)$$

at $k = N - 1$. This algorithm requires storage of the filtered estimate and the associated covariances at all estimation times. This storage requirement may render this solution infeasible for large problems. It is also noteworthy that the solution was obtained for a linear Gaussian system, and its results are not generally applicable to non-linear problems.

Figures (2-2) to (2-5) show how the Rauch Tung Striebel smoother yields an

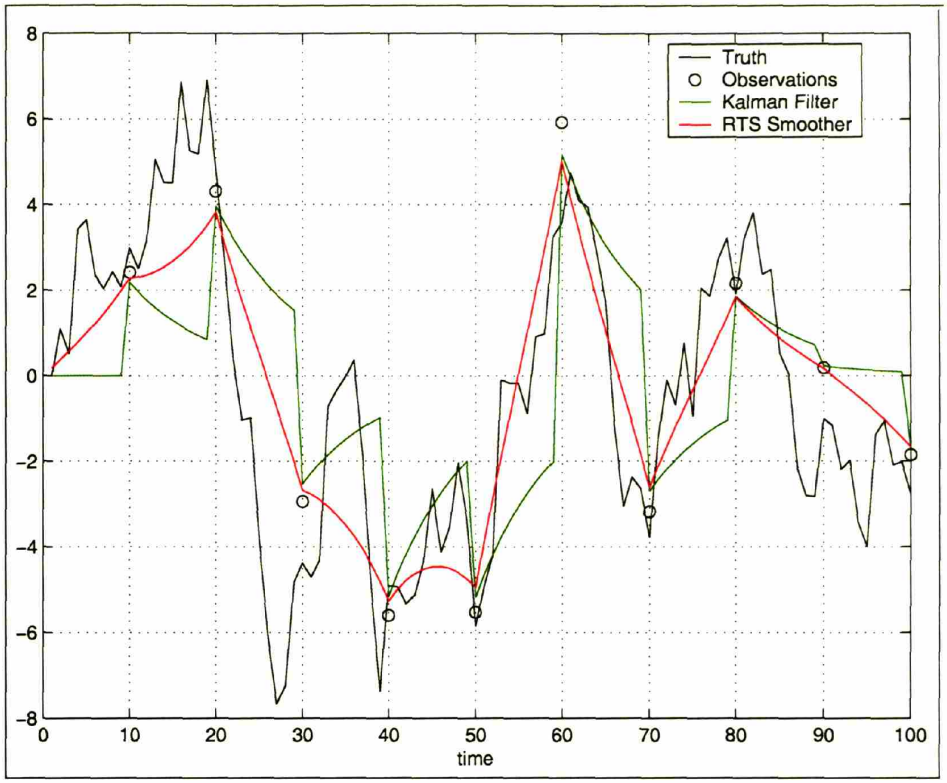

Figure 2-2: State Estimate compared to truth for the optimum filter and smoother. Noisy observations are available every 10 time steps. The Estimation interval is [0,100].

improved estimate over the Kalman filter. In Figure (2-2), a sample of the estimated time series is shown. The truth is shown in black, with observations at every 10th time step shown as circles. The Kalman filter estimate is shown in green, while that of the RTS smoother is shown in red. Clearly, the filter updates sharply towards the observations at update times, and is drawn slowly ($\Phi_k = \Phi = 0.9$) towards the process mean of zero between observations. The smoother, on the other hand, carves out a smooth transition between the observation values. We aim to find an estimator which will emulate this smooth curved path. In Figure (2-3), the covariances of the estimate is shown for the same interval of the time series. The covariance is a measure of the uncertainty in the estimate. The covariance of the filtered estimate is at a minimum at update times, and grows steadily between updates due to model error. The covariance of the smoother demonstrates that the smoothing algorithm

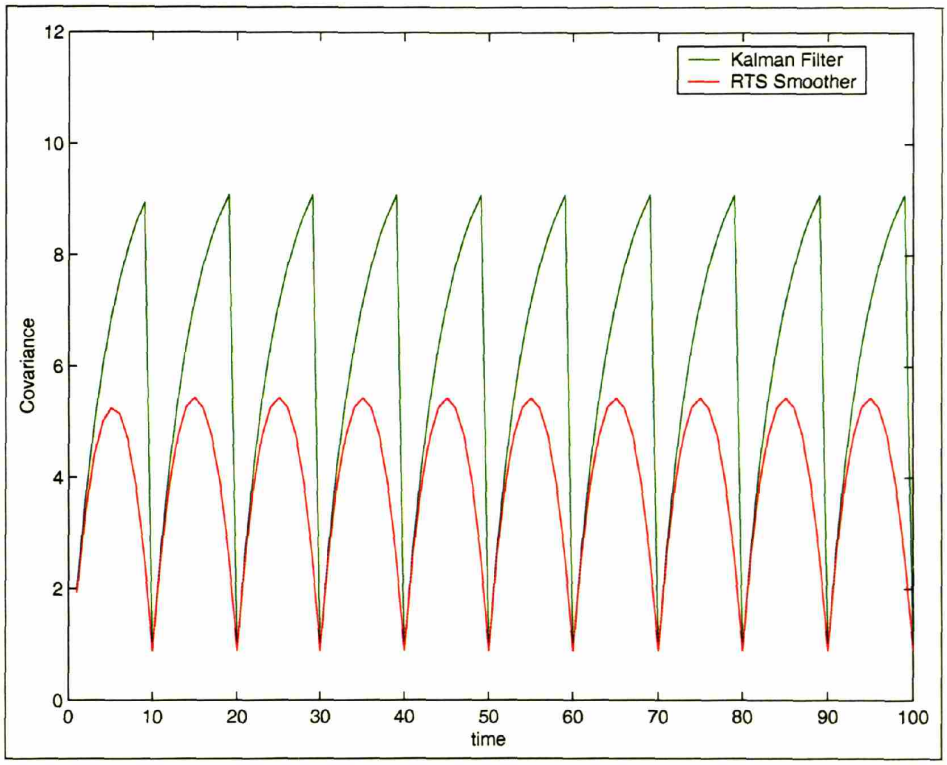

Figure 2-3: Time Series of the analysis covariance for the optimum filter and smoother. The estimation interval is [0,100].Observations are available every 10 time steps.

avails of future observations in estimating the state. The minimum covariance in the smoother case is lower than that of the filter, because updates at observation times are also influence by future observations through H. Furthermore, while the smoother covariance grows after updates at an identical rate to the filter, it falls again roughly halfway between observations as a result of the improvement in the estimate from the future observation. This decrease can be explained by recalling how the smoother sweeps backwards through the experiment interval.

Figures (2-4) and (2-5) show the normal probability plot of the analysis error normalized by the square root of the covariance for the Kalman filter and the RTS smoother respectively. For a perfect estimator, this quantity should be normally distributed with mean zero and a standard deviation of 1. In this example, it is clear that for values in the range [-2,2], this quantity is clearly normally distributed. The mean value is -0.0887 for the filter and -0.1263 for the smoother, and the corresponding standard deviations are 0.9534 and 0.9944 respectively.

2.4 Ensemble Kalman Filter and RTS Smoother

In the ensemble Kalman filter, an ensemble (size n_r) of realizations of the state vector is propagated forward in time. At some time k, the realizations can be gathered into a matrix X

$$X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{n_r}) \tag{2.31}$$

The covariance at time k is given by:

$$P = \frac{X'X'^T}{n_r - 1} \tag{2.32}$$

where:

$$X' = X - \bar{X} \tag{2.33}$$

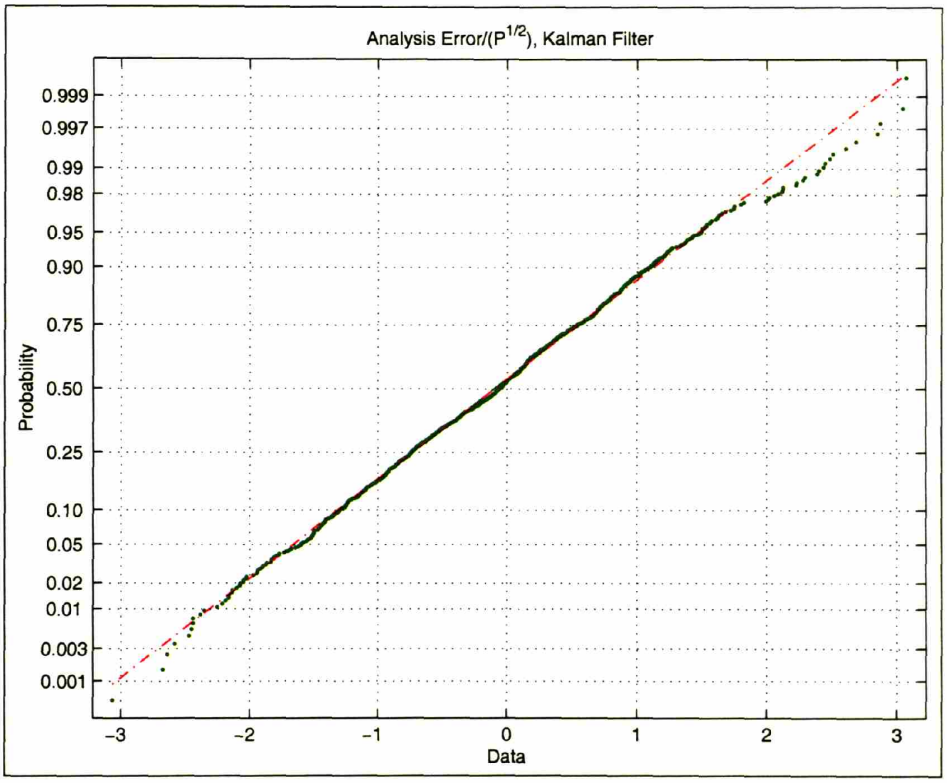

Figure 2-4: Normal Probability plot of Analysis Error normalized by $P^{1/2}$, for the Kalman filter. Observations are available every 10 time steps. Results are shown for an estimation interval [0,1000].

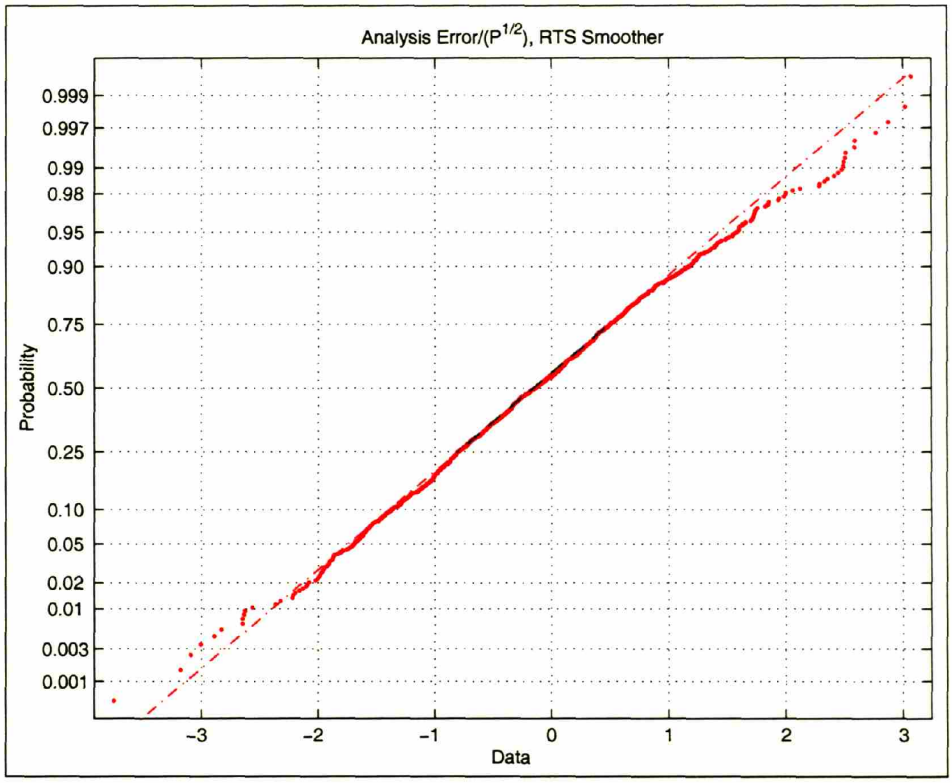

Figure 2-5: Normal Probability plot of analysis error normalized by $P^{1/2}$, for the RTS smoother. Observations are available every 10 time steps. Results are shown for an estimation interval [0,1000].

and

$$\bar{X} = X1_{n_r} \tag{2.34}$$

where 1_{n_r} is an n_r by n_r matrix, each element of which is equal to $1/n_r$. At an update time, a vector of measurements z of length m becomes available. Each ensemble member must be updated with respect to a perturbed observation to ensure that the spread of the updated ensemble is consistent with the true posterior estimation error covariance [3]. An ensemble of perturbed observations is generated at each update time:

$$z_j = z + \underline{\epsilon}_j \tag{2.35}$$

where j indicates the j-th ensemble member and $\underline{\epsilon}_j$. If the ensemble of perturbations are gathered into the matrix γ,

$$\gamma = (\underline{\epsilon}_1, \underline{\epsilon}_2, \ldots, \underline{\epsilon}_{n_r}) \tag{2.36}$$

the measurement error covariance, R is given by:

$$R = \frac{\gamma\gamma^T}{n_r - 1} \tag{2.37}$$

Finally, the ensemble of observations can be gathered into the matrix Z:

$$Z = (z_1, z_2, \ldots, z_{n_r}) \tag{2.38}$$

The update equation can therefore be written as:

$$X(+) = X(-) + PH^T(HPH^T + R)^{-1}(Z - HX(-)) \tag{2.39}$$

where $X(+)$ and $X(-)$ denote the updated and prior state respectively. The Kalman gain can be expressed as

$$K = PH^T(HPH^T + R)^{-1} \tag{2.40}$$

Analagous to the analytic case, the Kalman gain may be thought of as a weighting matrix which quantifies the relative confidence we can place in the model and the observations at an update time. The difference between the EnKF and the classic KF is that in the EnKF, the covariance terms of the Kalman gain term are calculated from the ensemble rather than determined analytically. If a sufficient number of replicates is used, the solution obtained using an ensemble Kalman filter converges to that obtained using the traditional Kalman filter.

One limitation of the EnKF is the inherent assumption that at the update time, the second order moments are sufficient to characterize the probability distribution of the prior ensemble. In this simple case with a linear model and gaussian model error, this difficulty does not arise as the p.d.f. of the prior ensemble is Gaussian. However in soil moisture estimation this will not necessarily be the case.

An ensemble approach can be taken with Rauch-Tung-Striebel smoothing, whereby the RTS equations are used to update the state and the covariances are calculated using the ensemble. For the linear gaussian problem, the results converge to those obtained using the classic RTS. For the AR(1) problem under consideration, ensemble Kalman filtering and subsequent ensemble RTS smoothing yielded results identical to those in Figures (2-2) to (2-5).

Figure 2-6 shows the cross-covariance between the updated states at each time step in the interval [0,100]. The variance (diagonal terms) is clearly at a minimum at observation times, and increases between observations. The extent of non-zero off-diagonal terms is indicative of memory in the system. The strength of the correlation decreases with distance from the diagonal. At update times (i.e. every 10th step), the variance decreases suddenly to ~ 1 and there is little or no correlation with the state at previous times. Just before update steps, there is a significant correlation

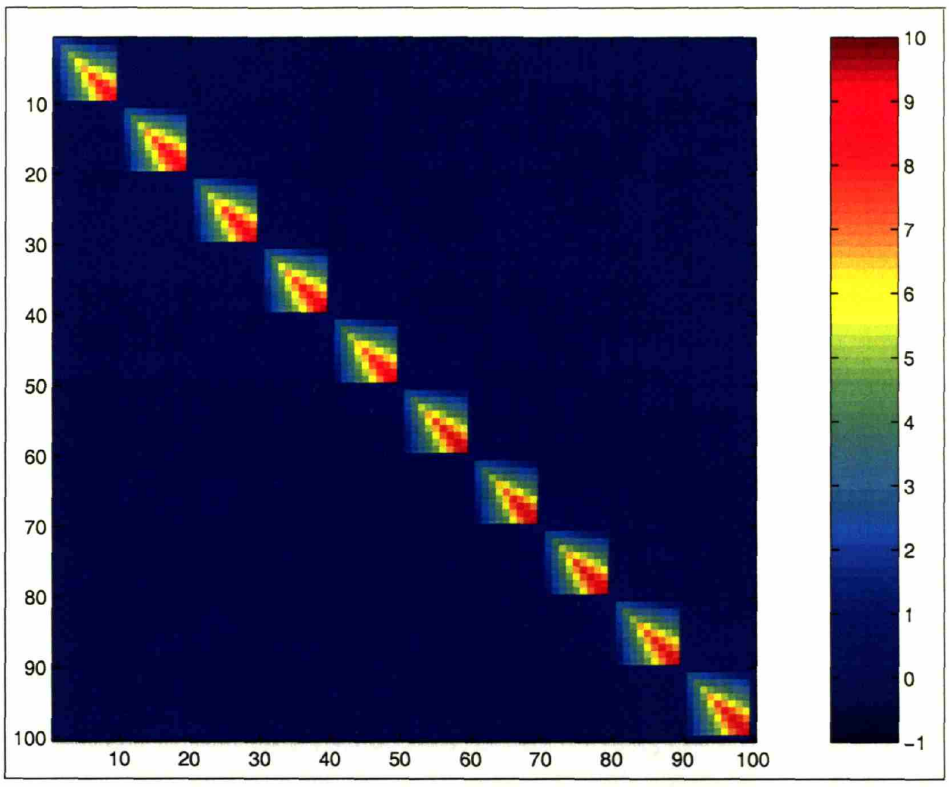

Figure 2-6: Cross-covariance between updated state estimates from the ensemble Kalman filter in the interval [0,100]. Observations were available every 10 time steps.

between the state and the state up to 8 time steps beforehand. This is due to the high value of Φ. In Figure 2-7, the same quantity is shown after the RTS smoother has been applied. In the smoother case, there is a beaded structure between observation times. The covariance is at a minimum at observation times and a maximum halfway between observations. This indicates the propagation of information backwards from future observations to improve the estimate at intermediate times. While there was significant correlation between states up to 8 time steps apart in the filter case, this is much reduced in the smoother case. In testing the proposed smoothers, we hope to see the beaded structure of Figure (2-7).

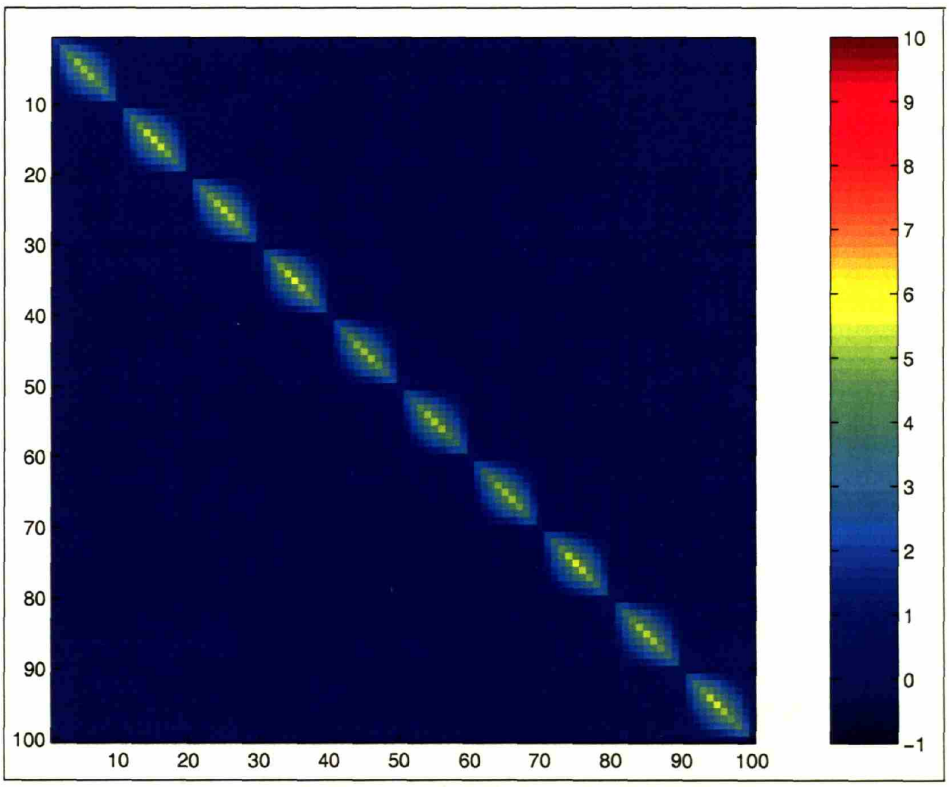

Figure 2-7: Cross-covariance between updated state estimates from the ensemble RTS smoother in the interval [0,100]. Observations were available every 10 time steps.

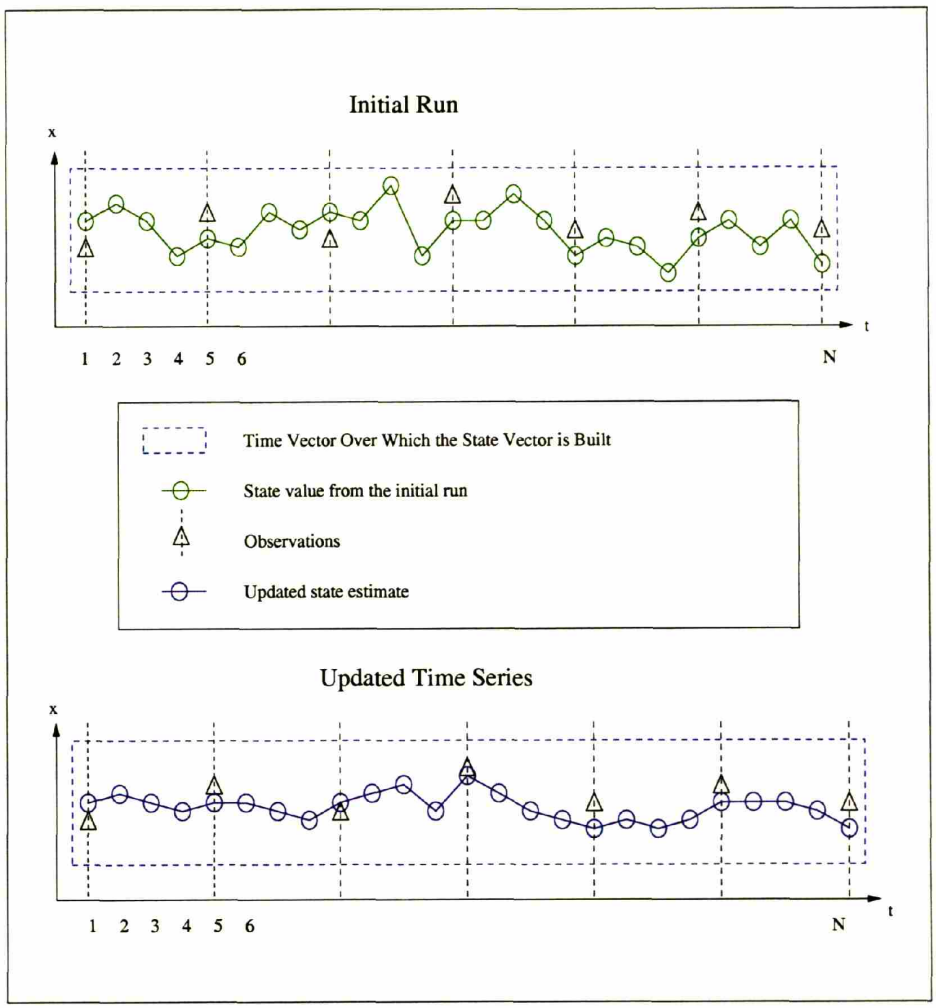

Figure 2-8: Schematic Diagram of the Ensemble Single Batch Smoother

2.5 Ensemble Single Batch Smoother

This is the simplest, albeit expensive, way of estimating the state at each estimation time step in the interval using all of the observations in the interval. Figure (2-8) is a simple illustration of the algorithm. Previously, the state vector at time t, $\underline{x}(t)$, has been defined as the variables of interest at that time. This state vector is of dimension n by 1. In the ensemble single batch smoother, we want to update the variables of interest at all estimation time steps in the interval in a single batch. This is achieved by having an "augmented" state vector $\underline{\mathcal{X}}$ of dimension $(n_a \times n, 1)$ for each replicate

in the ensemble, such that:

$$\underline{\mathcal{X}} = [\underline{x}_1 \ \ldots \ \underline{x}_{n_a}]^T \tag{2.41}$$

This vector contains the states at all n_a estimation times in the interval. From Figure (2-8), the blue dashed box indicating the smoother window extends to the Nth time step in the interval. Instead of proceeding sequentially through the interval, this augmented state vector will be updated once, using all of the observations in a batch. The vector of observations must also therefore be augmented:

$$\underline{\mathcal{Z}} = [\underline{z}_{t_{z(1)}} \ \underline{z}_{t_{z(2)}} \ \ldots \ \underline{z}_{t_{z(n_z)}}]^T \tag{2.42}$$

For clarity, Figure (2-8) shows just one replicate of the ensemble. In practice, there are $n_r = 2000$ such replicates to ensure convergence. Each of these n_r realizations of the process consists of an initial value at the first time step propagated through the entire interval. The prior covariance is calculated using this ensemble. Each replicate is then updated using this prior covariance. The mean across the updated ensemble of realizations is the updated and final estimate.

The ensemble single batch smoother was evaluated by comparing its results to those obtained from the Kalman filter and the RTS smoother. In Figure (2-9), it is clear that the estimate obtained is closer to that of the smoother. However, the estimate does not vary as smoothly with time as the RTS smoother estimate. Figure (2-10) demonstrates that the algorithm is using future observations to update the estimate, as the trace of the covariance is symmetric between observations, as distinct from the saw-tooth covariance associated with optimal filtering. So, the quality of the EnSB estimate may be considered closer to that of the RTS smoother than to that of the Kalman filter. This is the expected result, as the estimate at a time between two observations uses information from the future observation as well as the previous observation. We expect that halfway between the observations, the covariance will be at a maximum as we are furthest from both observations. The performance of the EnSB as a good estimator is further illustrated in Figure (2-11). Just like the

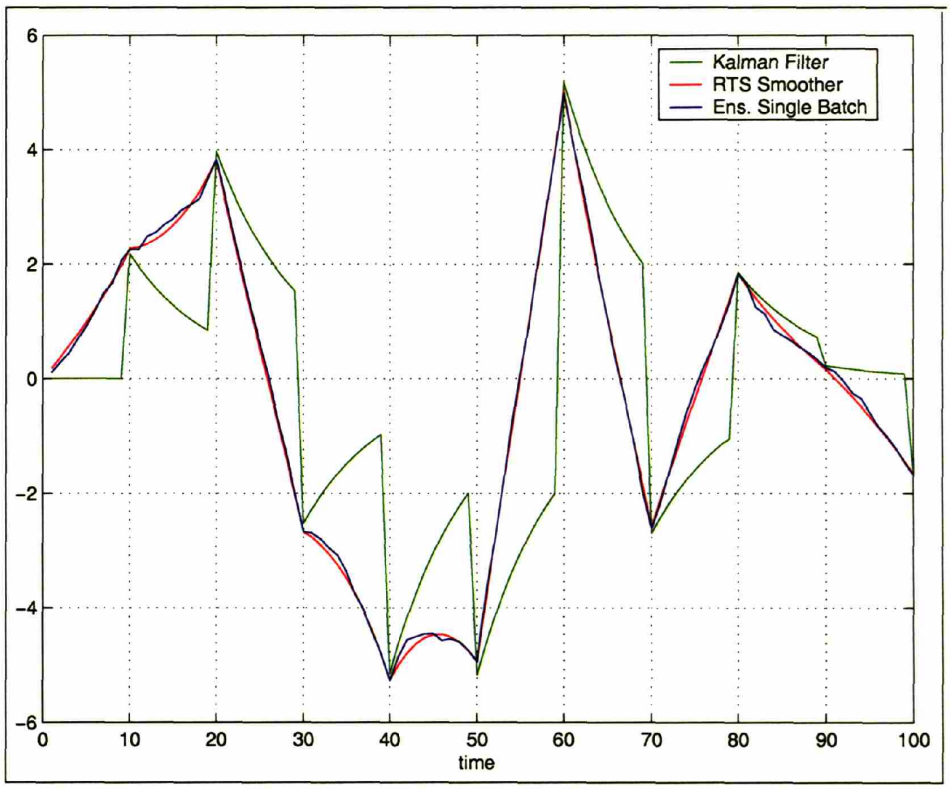

Figure 2-9: State Estimate from the ensemble single batch smoother compared to truth, and solutions from the optimal filter and smoother for the interval [0,100]. Observations were available every 10 time steps.

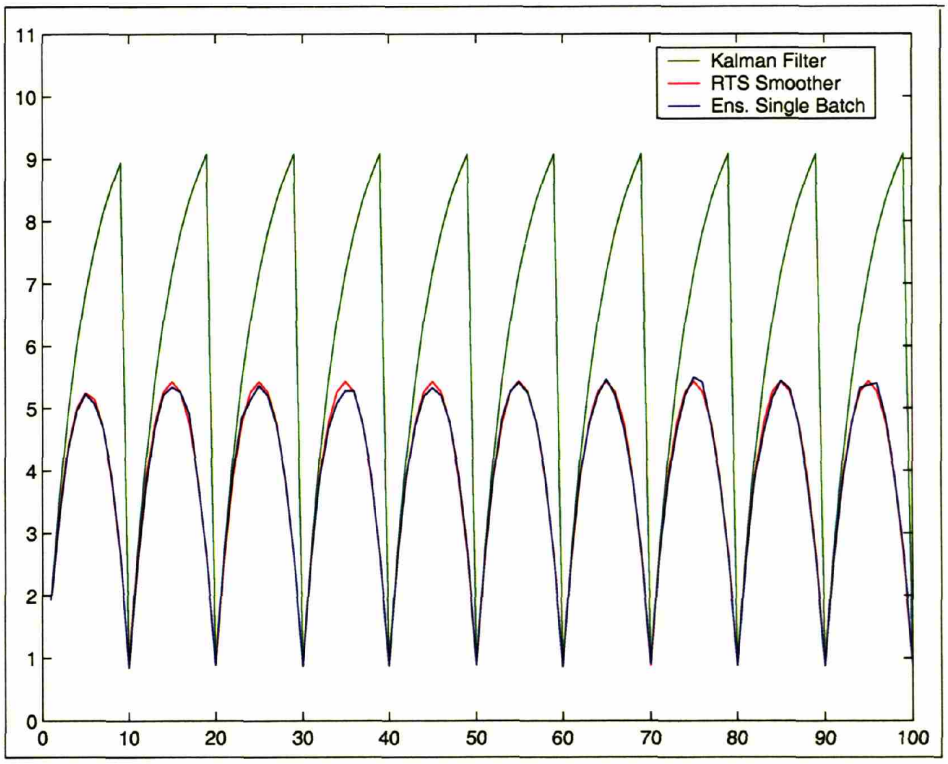

Figure 2-10: Time series of the analysis covariance from the ensemble single batch smoother is compared to results from the Kalman filter and RTS smoother in the interval [0,100]. Observations were available every 10 time steps.

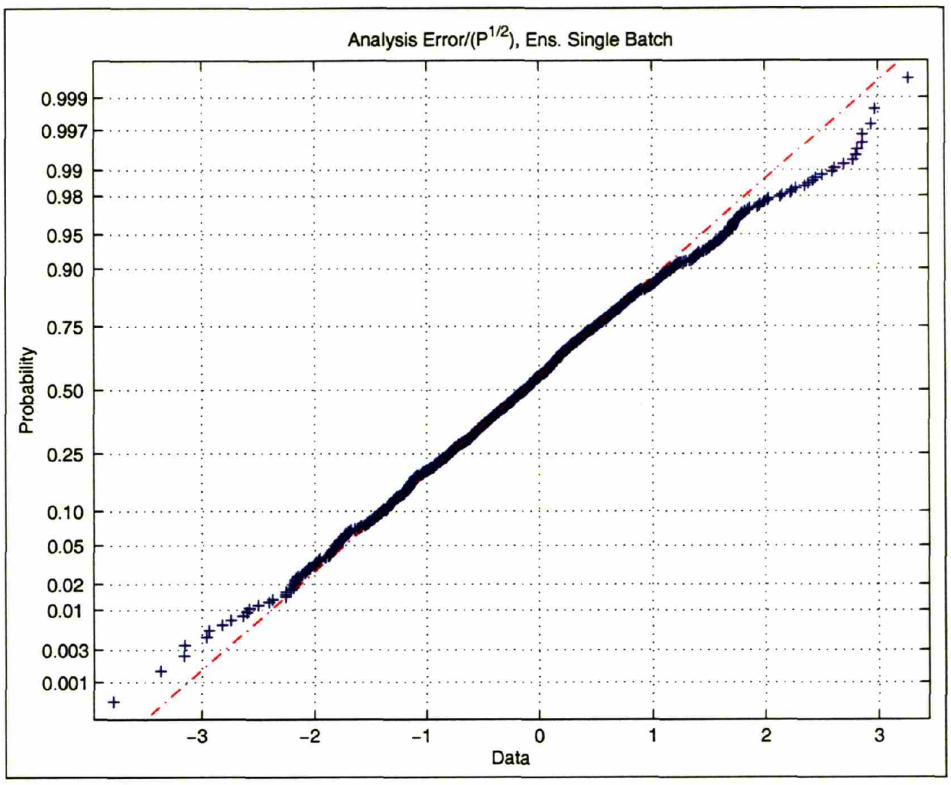

Figure 2-11: Normal probability plot of analysis error normalized by $P^{1/2}$ for the ensemble single batch smoother in the interval [0,1000].

case of the RTS smoother, the analysis error normalized by the square root of the covariance is clearly normally distributed in the range [-2,2]. The mean is -0.1327, and the standard deviation is 1.0331, so this is reasonably close to $\sim N(0.1)$. There are more outliers beyond the range [-2,2], as the ensemble single batch is obviously not as good as the optimal smoother.

Figures 2-13 and 2-12 show the cross-covariance between the states in the interval [0,100] before and after the update. Figure 2-12 may be thought of as the ensemble open loop cross-covariance, i.e. the case where no observations are assimilated. The extent and magnitude of the off-diagonal terms are indicative of memory in the system. When no observations are assimilated, the state is correlated with states up to 15 time steps away. The cross-covariance shown in Figure 2-12 is in fact part of the prior covariance matrix used to update the temporally distributed state vector.

The corresponding part of the updated covariance matrix is shown in Figure (2-

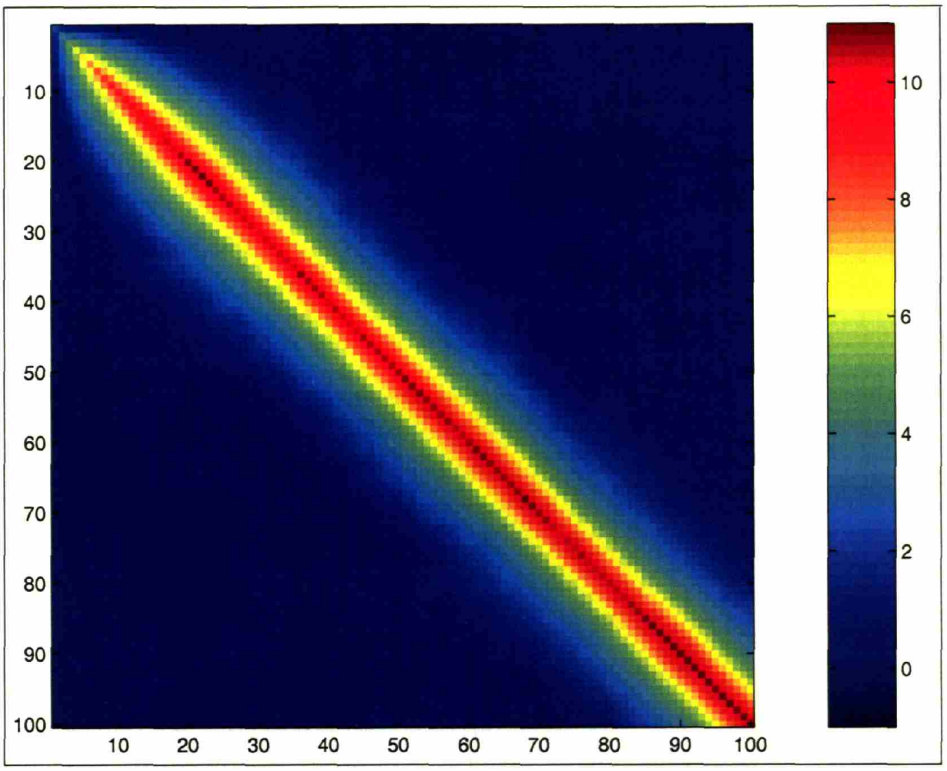

Figure 2-12: Cross-covariance between the a priori state estimate from the ensemble single batch smoother, shown in the interval [0,100]. Observations are available every 10 time steps

13). Note that the scale is identical to that used in Figure (2-12), so the magnitude of the cross-covariances are visibly reduced. The diagonal terms have been reduced from ~ 11 to ~ 5.5. Furthermore, the beaded covariance structure obtained using the RTS smoother is also produced using this algorithm. The covariance is at a maximum halfway between between observations, and at a minimum at observation times, as expected. This beaded structure demonstrates that information from future observations is used to update the state at previous times. The covariance between states more than 2-3 timesteps apart is drastically reduced. This is probably particular to the AR(1) model.

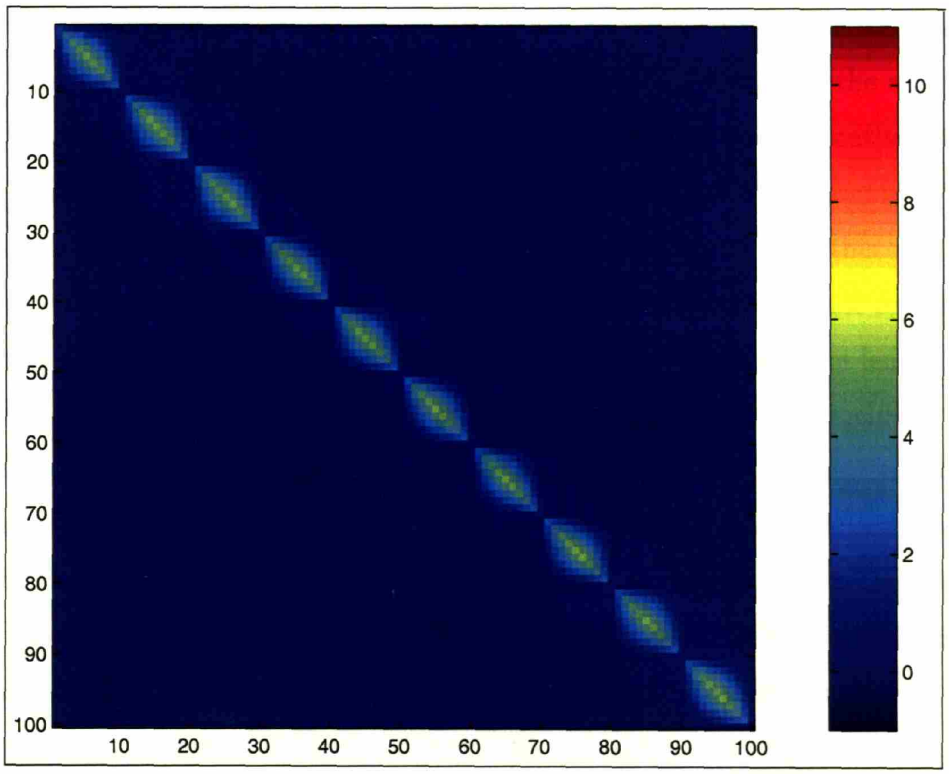

Figure 2-13: Cross-covariance between the updated state estimate from the ensemble single batch smoother, shown in the interval [0,100]. Observations are available every 10 time steps

2.6 Ensemble Moving Batch smoother

This approach is similar to the ensemble single batch smoother in that it uses augmented state and measurement vectors like those of (2.41) and (2.42). However, the augmented vector is not comprised of the state variables at all N time steps but is instead limited to the timesteps within LSW observations. LSW is used to denote the **L**ength of the **S**moother **W**indow. The moving window should be bounded by observations as it is the observations which introduce new information into the algorithm, thereby improving the estimate. The minimum value of LSW is 2, so the augmented state vector would consist of the state at all times between two consecutive observations. If the LSW is set to N_z, so that all observations in the estimation interval are included in the smoother window, the ensemble moving batch smoother and single batch smoother perform the same calculation and should yield identical results.

Figure (2-14) illustrates how the algorithm works. In this case, the length of the smoother window is 3, so the red box containing all the time steps included in the augmented state vector encompasses three observations and all the intermediate time steps at which we want to know the state variables. Note that first and last time steps in the box are observation times as explained previously. As in the ensemble single batch, the prior estimate consists of n_r realizations of the process run through the smoother window. For clarity, just one such realization is shown in Figure (2-14). Using the prior covariance calculated from this ensemble, each replicate is updated to yield the updated estimate. In the second panel shows the same replicate of the augmented state vector has been updated to the values colored in red. The next time step is shown in Panel 3, where the moving smoother window now begins at the next observation time. The prior estimate is obtained by running the model forward from the updated value at time 5 through 13 for each of the n_z replicates. The augmented state vector is then updated using the covariance across the ensemble. The state between times 5 and 9 has now been updated twice, so this is the final estimate for these times. Times 9 to 13 have been updated for the first time. In the next time step, the moving smoother window moves along to the next observation time (9).

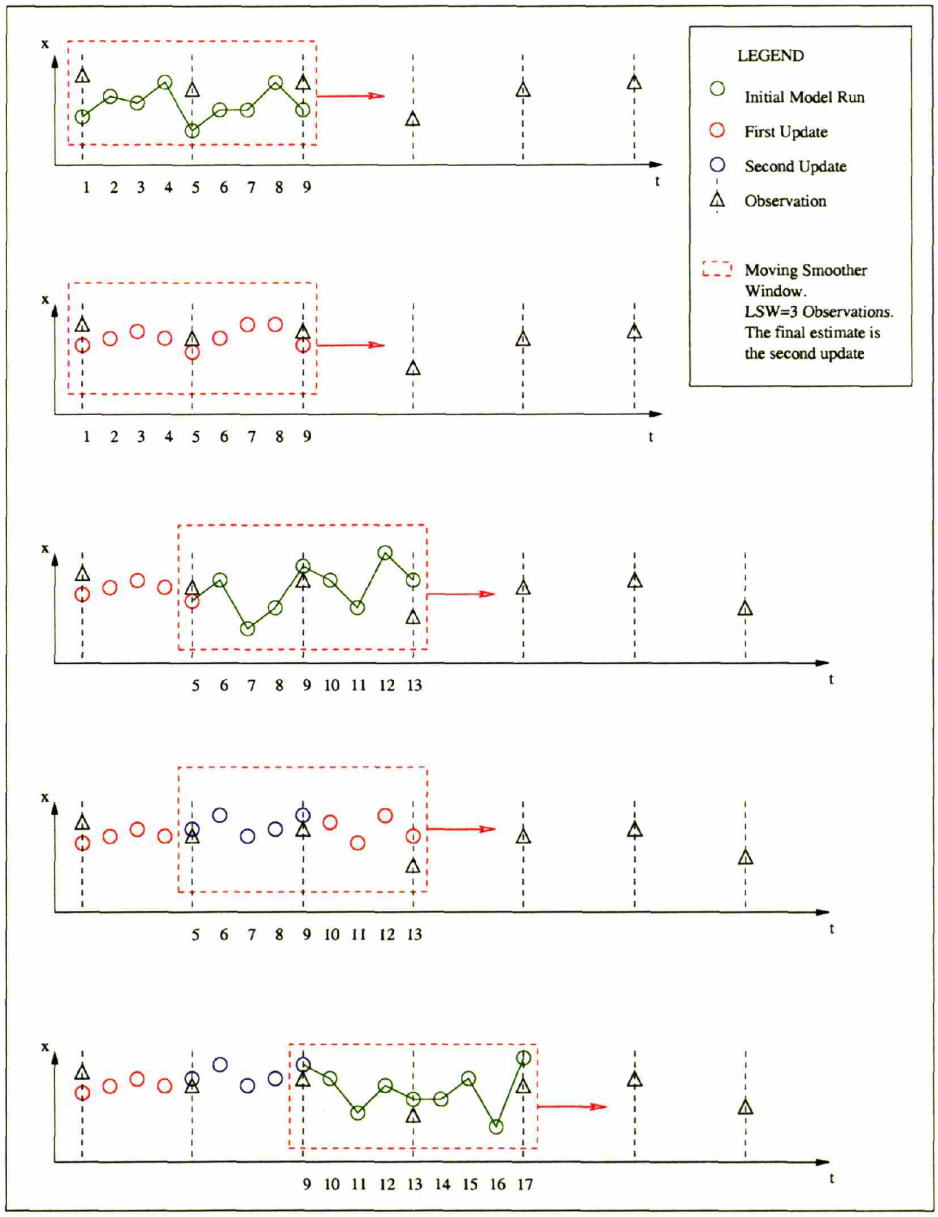

Figure 2-14: Schematic diagram of the ensemble moving batch smoother

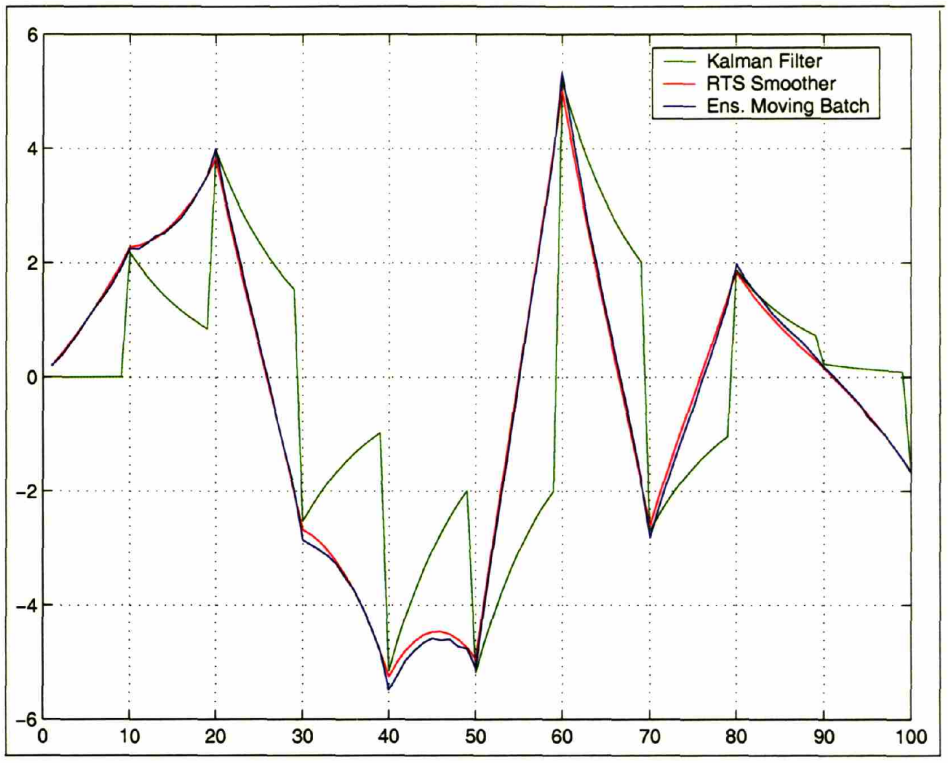

Figure 2-15: State estimate from the ensemble moving batch smoother is compared to the estimate from the optimal filter and smoother in the interval [0,100]. Observations are available at every 10th time step, and LSW=5.

The prior estimate is obtained by running the model forward for each replicate from its updated value at 9. This process continues until the end of the smoother algorithm reaches the end of the study interval.

The results obtained using a moving batch smoother with LSW=5 for the problem outlined in Section 1 are shown in Figures (2-15) to (2-17).

Figure 2-15 shows that the EnMB provides a good approximation to the solution obtained using the RTS smoother. The noise seen in the estimate obtained using the ensemble single batch smoother has clearly been removed by the moving smoother window.

The series of covariance values for the same interval , shown in Figure (2-16) also follows that of the RTS smoother more consistently than those of the ensemble single batch smoother.

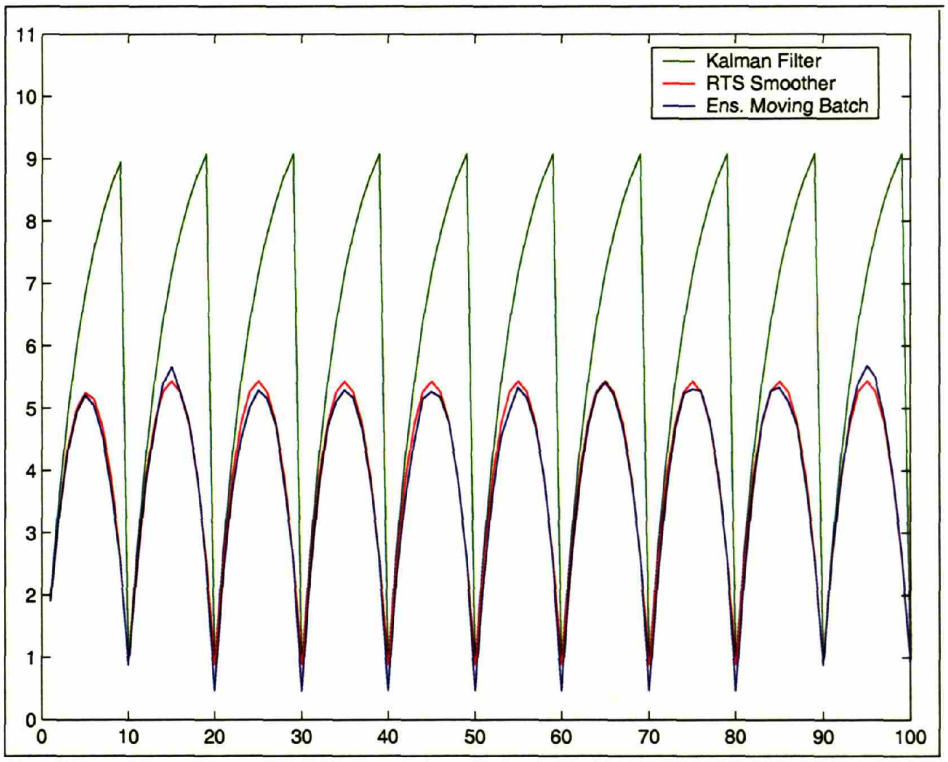

Figure 2-16: Analysis covariance from the ensemble moving batch smoother is compared to that from the optimal filter and smoother in the interval [0,100]. Observations are available at every 10th time step, and LSW=5.

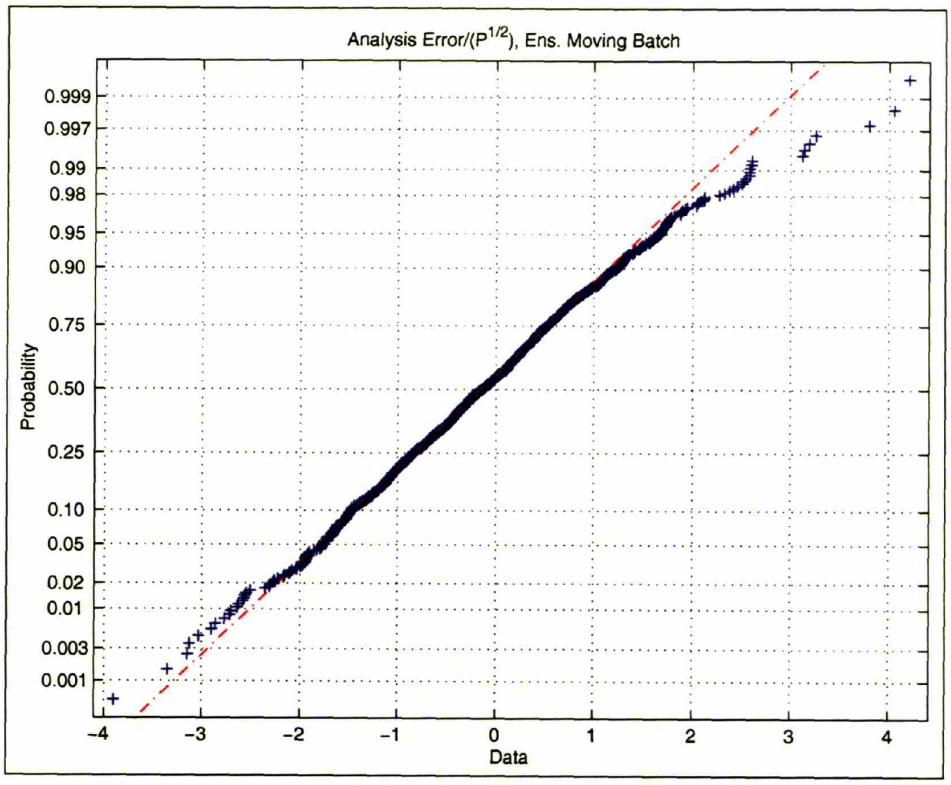

Figure 2-17: Normal probability plot of analysis error normalized by $P^{1/2}$, for the ensemble moving batch smoother in the interval [0,1000]. LSW=5 and observations were available at every 10th time step.

The normal probability plot of the analysis error normalized by the square root of the covariance is given in Figure (2-17). Within the same range as the RTS smoother, [-2, 2], this quantity is very visibly normally distributed. The mean, for the ensemble moving batch smoother is -0.1271 and the standard deviation is 1.0654. So it is a pretty good estimator. As in the case of the ensemble single batch smoother, there are a few more outliers beyond ± 3. Figure 2-18 shows the cross-covariance between the updated states in the interval [0,100], obtained from the ensemble moving batch smoother with LSW=5. It shows similar beaded structure to results obtained using the RTS smoother and the EnSB, indicating that information has been propagated backwards from the observations to update previous states.

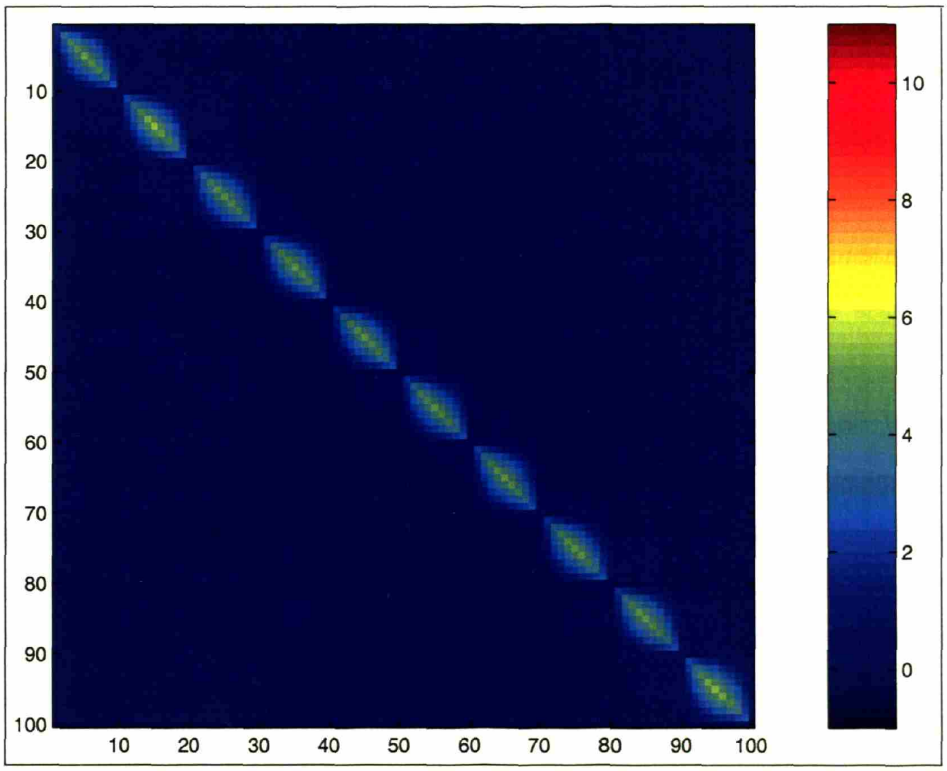

Figure 2-18: Cross-covariance between updated state estimates from the ensemble moving batch smoother in the interval [0,100]. Observations are available every 10 time steps and LSW=5.

Table 2.1: Comparison of nRMSE for analytic and ensemble Kalman filter and RTS smoother.

	Analytic	Ensemble
Kalman Filter	0.7171	0.7182
RTS Smoother	0.6158	0.6179

2.7 Root Mean Square Errors

It is clear from the results shown that both the proposed algorithms yield improved estimates over the Kalman filter. Both algorithms result in updated covariance matrices similar in structure to that produced by the RTS smoother, demonstrating that they successfully used future and prior observations to estimate the state at each estimation time.

The Root Mean Square Error associated with each algorithm can be used to examine the overall success relative to the Kalman filter and RTS smoother, to investigate the impact of less frequent observations, and to find the optimum length of the smoother window for the EnMB. We shall use a normalized Root Mean Square Error (nRMSE), obtained by normalizing the Root Mean Square Error by the known system noise. The key findings are as follows:

- Provided a sufficiently large number of replicates is used, the ensemble Kalman filter and ensemble RTS smoother converge to the results of their analytical equivalents.

- It follows that since the analytic RTS smoother improves on the estimate obtained using the Kalman filter, the ensemble RTS smoother clearly yields improved estimates compared to the ensemble Kalman filter.

- From Figure 2-19, the nRMSE for both the Kalman filter and the RTS smoother increase steadily as the interval between observations increases. This result is intuitive as both algorithms operate sequentially, so the error covariance has

a chance to grow more between observations. With fewer observations, the modeled estimate can deviate further from the truth between observations.

- The difference in nRMSE between the Kalman filter estimate and the RTS smoother estimate appears to increase as the interval between observations increases. This is particularly apparent where the interval between observations is less than 10 time steps. One possible explanation is that the memory of the system was seen earlier to be on the order of 10 time steps. Perhaps the change is due to the fact that the observations are further apart than the length of the memory of the process.

- It is also clear from Figure 2-19 that the ensemble single batch smoother yields an nRMSE remarkably close to that obtained using the RTS. This result is expected as both algorithms use all observations to obtain the estimate at each estimation time. It is clear, however, that the nRMSEs are not absolutely equal. It will be shown later that the nRMSE associated with the EnSB can take one of a range of values due to the difficulty in estimating such a large number of states all at once.

As the interval between observations is cleary significant, three experiments were undertaken to study the relative performance of the EnSB and the EnMB as the length of the smoother window is increased. Results are shown for LSW=1, 5 and 10 in Figures 2-20, 2-21 and 2-22 respectively.

- In Figure 2-20, the difference in nRMSE between the Kalman filter and the RTS smoother is relatively small. This is because the error covariance has no chance to grow if there are observations at every time step. One expected result is that for the EnMB the minimum nRMSE occurs when the length of the smoother window is the same as the memory of the process i.e. about 10-20 time steps. Oddly though, the EnMB appears to outperform the RTS smoother, which is supposed to be the optimal smoother. When the length of the smoother window increases such that all observations are included, the EnMB and EnSB

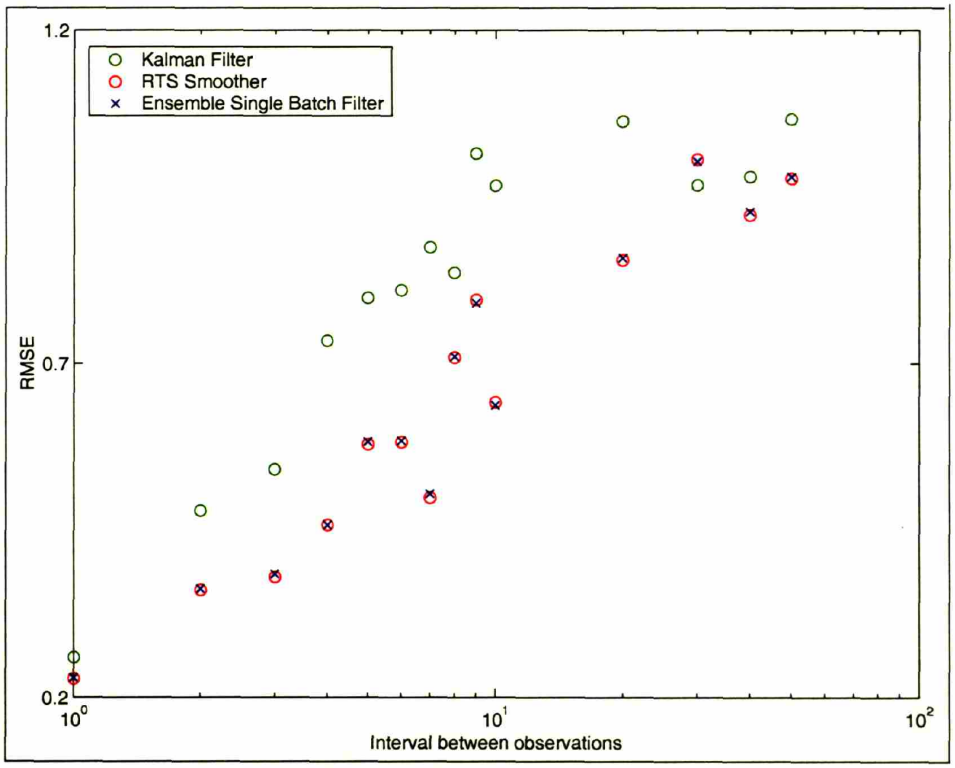

Figure 2-19: Comparison of RMSE from ensemble single batch smoother compared to results from the Kalman filter and RTS smoother as the interval between observations is increased. Note that this figure is plotted on a semi-logarithmic scale for clarity.

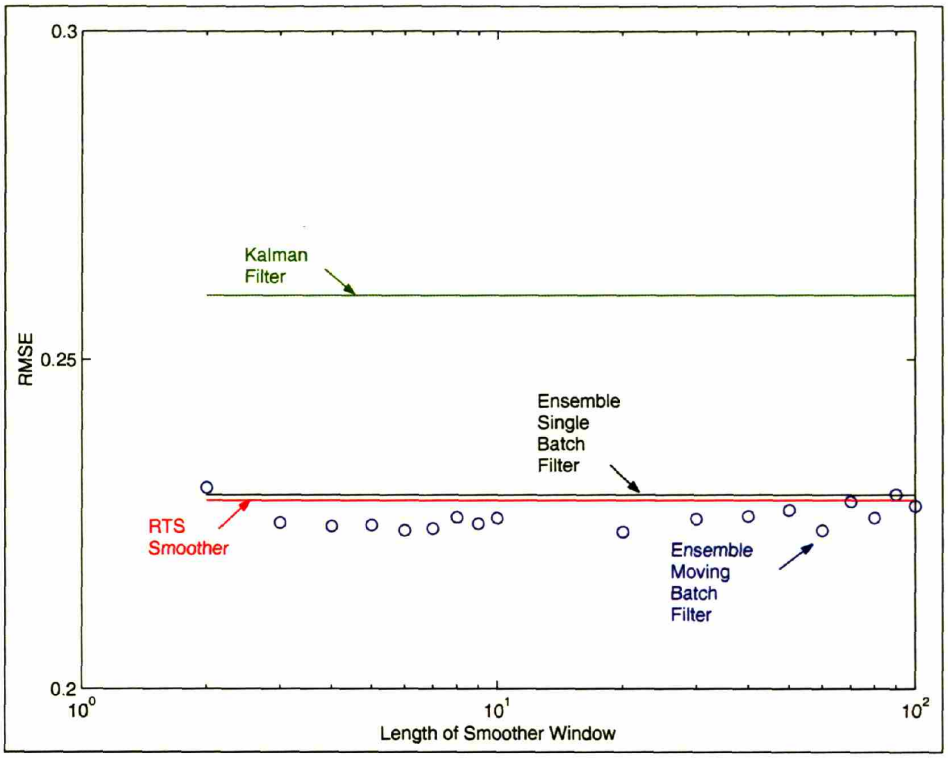

Figure 2-20: RMSE from the EnMB as the length of the smoother window is increased is compared to the RMSE from the Kalman filter, RTS smoother and EnSB smoother. Observations are available at every time step. This figure is plotted on a semi-logarithmic scale for clarity.

algorithms are identical. The nRMSEs for both algorithms are therefore pretty close where LSW=90 and 100 in this case. The variability in nRMSE in this range suggests that both algorithms may have difficulty estimating so many states at once, and that a very large number of replicates might be necessary.

- Figure 2-21 shows the nRMSEs for the case where observations were available at every fifth time step. Recall that the length of the smoother window is defined in terms of observations included, so LSW=5 in this case describes a window of 40 time steps. In this case the difference in nRMSE between the Kalman filter and the RTS smoother is larger than when observations were available at every time step. This is because the error covariance now has a chance to grow between observations.

In order to investigate the performance of the algorithm at high LSWs more closely, the EnMB experiments were repeated 15 times to observe the mean and standard deviation of the nRMSE. The blue circles denote the mean nRMSE obtained for each LSW, and the bar denotes the standard deviation in nRMSE.

Recall that when the LSW is increased to include all observations the EnSB and the EnMB are identical. So, the mean and standard deviation of the nRMSE for the EnMB at LSW=20 is also that of the nRMSE for the EnSB. Clearly then, the black line denoting the nRMSE for the EnSB is really just one realization of a range of possible values in this range. Interestingly, the mean of the nRMSE for the ensemble single batch coincides with the nRMSE value for the RTS smoother. One would hope therefore that if sufficient replicates could be used, the nRMSE for both of the proposed smoothers would converge to that of the RTS smoother.

It is also interesting to note that when the LSW=2 or 3 (corresponding to an interval of 10 or 20 time steps) the standard deviation in the nRMSE is close to zero. Beyond this range the standard deviation is clearly non-zero. This suggests that if the length of the smoother window is long enough to just capture the memory in the system, the result will always be the same. If the length of the smoother window is longer than the memory of the system, it appears to result in a range (albeit small) of possible solutions. In order to guarantee convergence, either the experiment must be repeated over and over again or the number of replicates must be further increased to allow for the large ratio of states to observations.

While it seems that the EnMB is outperforming the RTS smoother, this is clearly not the case as the RTS smoother is known to be the optimal solution. I expect that this is merely a sampling issue, and if the experiment was repeated a larger number of times the results would converge to those of the RTS smoother.

- The results obtained when observations were available at every 10th time step

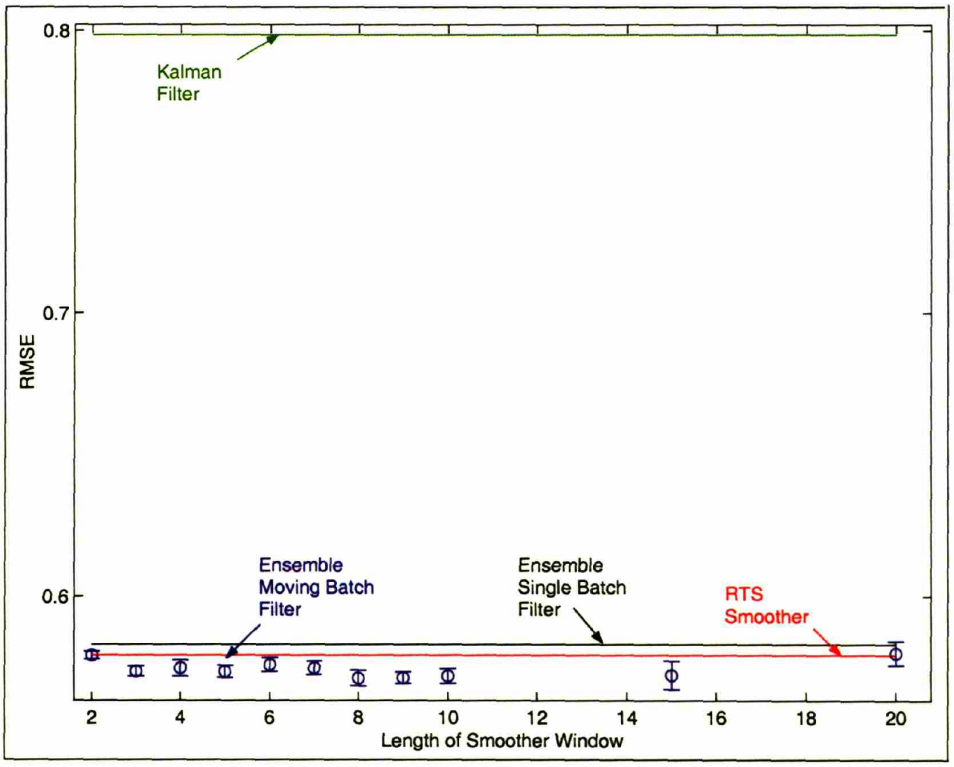

Figure 2-21: RMSE from the EnMB as the length of the smoother window is increased is compared to the RMSE from the Kalman filter, RTS smoother and EnSB smoother. Observations are available at every 5th time step.

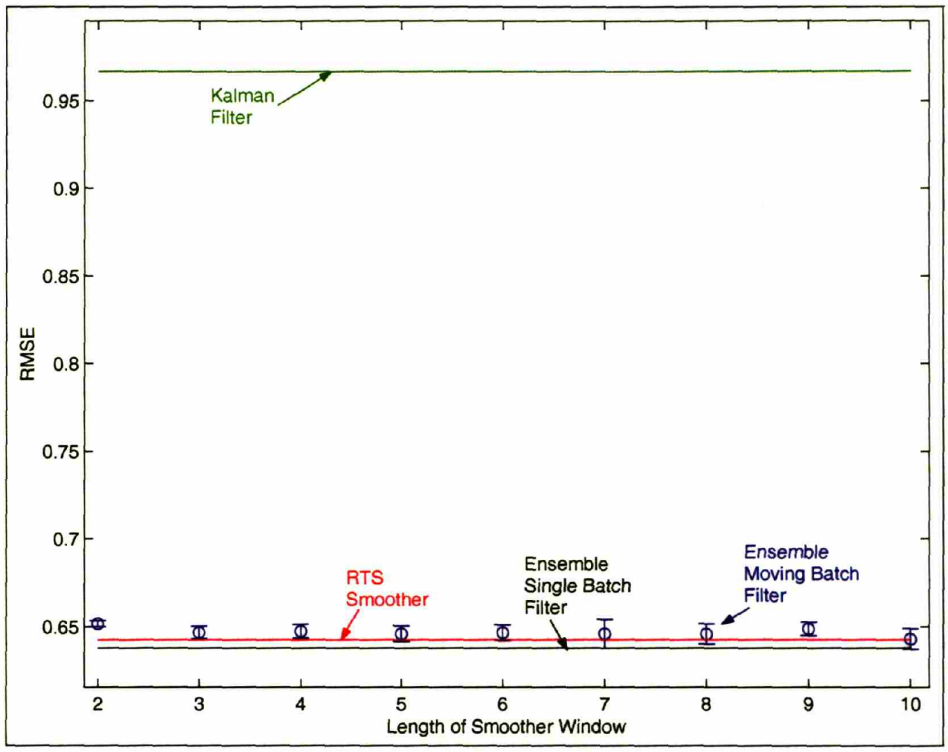

Figure 2-22: RMSE from the EnMB as the length of the smoother window is increased is compared to the RMSE from the Kalman filter, RTS smoother and EnSB smoother. Observations are available at every 10th time step.

are shown in Figure 2-22. The difference between the nRMSEs obtained using the Kalman filter and RTS smoother is slightly greater than the difference in Figure 2-21. This is because the error covariance grows more slowly when it approaches the maximum.

Recall again that the black line representing the nRMSE associated with the EnSB is just one realization of the possible range of solutions, and that the mean and standard deviation of this range are equal to the mean and standard deviation of the nRMSE for the EnMB for LSW=10 in this case. The ensemble single batch yields an RMSE close to that obtained using the RTS smoother. In fact, here the nRMSE for the RTS smoother coincides with the mean value of the nRMSE for the EnSB, and for most LSWs the nRMSE for the EnMB is close to the range of values for the EnSB.

To some extent we can see the expected decline in nRMSE for the EnMB as the LSW increases. We would expect to see the nRMSE continue to decrease to the value associated with LSW=10 (i.e. the EnSB). We can surmise from the increasing standard deviation in nRMSE that trying to estimate so many states with so few observations is difficult.

2.8 Computational Burden

The choice of algorithm is not purely dependent on the quality of the estimate obtained. It may be be controlled by memory limitations. In particular, calculation and storage of large covariance matrices may render an otherwise effective estimation technique unfeasible. Table 2.2 contains the definitions of some key variables needed to quantify the computational burden of the algorithms under consideration.

Table 2.3 contains the limiting quantities and their sizes for each of the algorithms considered here.

The issue of computational burden is very significant in these algorithms. One would expect that the EnSB would yield results comparable to those of the RTS smoother as both algorithms use all of the observations in the interval. However while the RTS is a sequential algorithm, the EnSB involves estimating the states at all times simultaneously. Preliminary results suggest that a very large number of replicates is required to do this with any degree of reproducibility. This has been verified in Figures 2-21 and 2-22; when the EnSB was implemented 15 times, it produced a range of results. Storage of such a large state vector for an even larger number of replicates requires enormous amounts of memory and renders this approach infeasible for lengthy study intervals. If this algorithm is to be used with any success, some consideration must be given to minimizing the size of the state vector or the estimation interval.

Clearly, as the EnSB is effectively a special case of the EnMB, the issue of convergence is also significant in applying the EnMB. Efforts must be made to ensure that the length of the smoother window used is no greater than the apparent mem-

Table 2.2: Notation used in the ensemble single and moving batch smoother algorithms

N	Total number of timesteps in the study interval. In this study, N=1000
N_z	Total number of observation times in the study interval. In this study, observations were at every 10th timestep, so $N_z = 100$.
n	Number of states at each estimation time, i.e. number of quantities of interest at each estimation time. In this example $n = 1$.
n_r	Number of replicates in the ensemble. In this study, $n_r = 2000$ to ensure convergence.
LSW	**L**ength of the **S**moother **W**indow. This quantity is defined only for the ensemble moving batch smoother. The window length is defined in terms of the number of observations it encompasses. The window begins and ends at an observation time, as new information is only included when a new observation is encountered.
n_a	Number of estimation times included in the augmented state vector. In the ensemble Kalman filter, $n_a = 1$, as it is a recursive algorithm, updating one timestep at a time. For the ensemble single batch smoother, $N_z \leq n_a \leq N$. In the ensemble moving batch smoother, n_a could be just LSW, corresponding to the case where an estimate is only required at observation times. The largest value of n_a for this algorithm is when an estimate is also required at all intermediate timesteps in the study interval.

Table 2.3: Limiting Memory Requirements for Several Smoothing Algorithms

Algorithm	Limiting Quantity	Memory Requirement
Kalman Filter	state	$n \times 1$
	covariance	$n \times n$
Rauch-Tung-Striebel Smoother	state	$n \times 1$
	covariance	$n_a(n \times n)$
Ensemble Kalman Filter	state	$n \times n_r$
	covariance	$n \times n$
Ensemble Single Batch Smoother	augmented state	$N(n \times n_r)$
	covariance	$(n \times N) \times (n \times N)$
Ensemble Moving Batch Smoother	augmented state	$n_a(n_r \times n)$
	covariance	$(n \times n_a) \times (n \times n_a)$

ory of the system. This ensures that future observations relevant to the state are used while those in the distant future are not. Keeping the number of states to be estimated simultaneously to a minimum permits the use of fewer replicates, thereby minimizing the computational burden. This point is also illustrated in Figures 2-21 and 2-22 where nRMSEs at low values of LSW had standard deviations close to zero. This indicates that repeated experiments using different initial seeds converged on the same result.

Chapter 3

A synthetic experiment using the EnMB smoother during SGP97

In Chapter 2 it was shown that for a test linear model with gaussian model and error observations, an ensemble moving batch smoother (EnMB) could yield an improved state estimate compared to the ensemble Kalman filter (EnKF). This EnMB is a simple extension of the EnKF in which the state and observation vectors are distributed in time and updated in a batch. If the memory in the system is sufficiently short, the batch may span just a few observations. In this chapter, the feasibility of using ensemble-based smoothers for soil moisture estimation using non-linear land surface models is studied. The EnMB is used to merge synthetic ESTAR observations with modeled soil moisture during the Southern Great Plains Experiment 1997. The performance of the EnMB is compared to the EnKF to determine if an improved estimate of soil moisture can be obtained using ensemble smoothing, and to identify issues which may be significant in the implementation of an ensemble-based smoother in a land data assimilation framework.

3.1 Ensemble Smoother Algorithm

In this section the EnKF and EnMB algorithms are revisited with emphasis on their implementation in a soil moisture estimation problem.

3.1.1 Ensemble Kalman Filter Equations

In the EnKF an ensemble of model states, $y(t)$ is integrated forward in time using the full non-linear model, $A[\cdot]$.

$$\begin{aligned} y(t) &= A[y(\tau), \alpha, u(\tau), w(t), t, \tau] \\ y(t_0) &= y_0 \end{aligned} \tag{3.1}$$

The state at time t depends on the state at a previous time τ, the time invariant parameters α of the model, the forcing applied to the model $u(\tau)$ and system error $w(t)$. Here, y(t) contains the soil moisture in six layers of the soil column and A[$\cdot$] is the NOAH Land Surface Model (LSM) [5]. Time-invariant parameters include descriptors of the soil texture and vegetation cover. The model is initialized with random initial conditions y_0. The observations z are related to the state y through the measurement operator $M[\cdot]$ and have additive gaussian error $\omega(t)$.

$$z = M[y(t)] + \omega \tag{3.2}$$

Here, the Radiative Transfer Model (RTM) is the measurement operator, relating the volumetric soil moisture values in y(t) to the observed L-band brightness temperature (Section 3.2). The EnKF is a sequential processor, updating the state through (3.3) when observations become available. Each ensemble member $y^j(-)$ is updated individually using the Kalman Gain, K, which is calculated from the ensemble statistics in (3.4).

$$y^j(+) = y^j(-) + K\{z + \omega^j - M[y^j(-)]\} \tag{3.3}$$

$$K = C_{yz}(C_{zz} + C_\nu)^{-1} \tag{3.4}$$

$$C_{yz} = \frac{1}{N_R}\tilde{\mathbf{y}}(-)\tilde{\mathbf{z}}^T \tag{3.5}$$

$$C_{zz} = \frac{1}{N_R}\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T \tag{3.6}$$

K weighs the relative uncertainty in the modeled estimate to that associated with the observation. C_{yz} is the cross covariance between the prior state and its transformed value in observations space, C_{zz} is the covariance of the transformed prior states in observation space and C_ν is the known variance of the observations (here $C_\nu = (3K)^2$ for L-band observations). In equations (3.5) and (3.6), $\tilde{()}$ denotes the perturbation matrix, N_Y is the number of states(6), N_Z is the number of observations (1) and N_R is the number of ensemble members (2000). For each ensemble member random noise ω^j is added to the observation z to account for the contribution of observation error to the posterior covariance [3]. In the soil moisture estimation problem, the model is highly non-linear, and uncertainty in parameters and forcing can result in non-Gaussian distributions of the states. By updating each ensemble member individually, this algorithm is particularly advantageous as it does not force a Gaussian posterior distribution. A thorough description of the Ensemble Kalman Filter and its implementation is provided by Evensen [25].

3.1.2 Ensemble Moving Batch Smoother

The EnKF described above has been used to estimate soil moisture during SGP97 [64]. The smoothing algorithm used here is a simple extension of the EnKF in which the states are distributed in time and updated in a "batch". Recall from Chapter 2 that the number of observations included determines the length of the observation vector, the state vector and consequently the covariance matrices. Including observations too far into the future would increase the computational burden without adding any useful information. Fortuitously, the memory in soil moisture is limited by the occurrence of precipitation which disrupts the dry-down and effectively reinitializes the problem. In this implementation of the EnMB observations will be available every three days, while an estimate of soil moisture is desired four times daily, based on the data assimilation product requirement of the Hydros mission [21]. The batch includes just two observations, to demonstrate that the inclusion of any information on the future state would yield an improved estimate. Consequently, the state vector will consist of the volumetric soil moisture in six layers at 12 time-steps, and the measurement

vector will contain two brightness temperatures.

Computational burden is a concern in employing batch smoothing techniques and ensemble techniques. As the length of the augmented vectors grow, larger memory will be required to make estimates conditioned on all measurements in the batch window. A concern is that including spatial correlation would increase the computational burden indefinitely. However, estimation variables can be a combination of model states. The standard Hydros data product is 0-5cm and 5-100cm soil moisture, so the dimension of the state vector can be significantly reduced even though the land surface model may have more layers for computational stability. There are computationally more efficient ways of implementing ensemble smoothing [25]. These techniques will be used in subsequent chapters when the the spatial dimension is added to the problem.

3.2 Data Assimilation Framework

Here, the Ensemble Moving Batch (EnMB) smoothing algorithm was evaluated using data from the Southern Great Plains Experiment 1997 (SGP97) to facilitate comparison with results from [64]. Experiments focused on two points in the SGP97 domain, namely El Reno and Little Washita.

Data from the Oklahoma Mesonet were used to create an Observing System Simulation Experiment (OSSE). The land surface model (Section 3.2.1) was forced using meteorological data to create a synthetic truth. Synthetic observations were generated from this "truth" using the radiative transfer model (Section 3.2.2). Additive zero-mean Gaussian noise with standard deviation of 3K was added to the synthetic observations to account for observation error. Using synthetic rather than real observations offers the following advantages:

- The estimation technique can be evaluated since the synthetic truth is known. Furthermore, this obviates the need to compare the estimate from data assimilation to ground observations which are prone to added sampling error.

- The availability of observations is not constrained by adverse weather or instrument troubles.
- Observations can be made at any time. Here, they were taken at 6am every 3 days to simulate the revisit time of the Hydros mission.
- Meteorological data from the Oklahoma Mesonet was used to generate the truth from the land surface model, so the experiment duration could be extended to run from 1st May to 1st September 1997.

3.2.1 Model

The NOAH Land Surface Model is used to propagate the ensemble of states forward between observations. This 1-D model of the soil column provides estimates of soil moisture and temperature profiles in addition to the mass and energy terms of the surface water and energy balances. It is a widely-used and freely available community land surface model which has been extensively validated and is currently used in the NASA Land Data Assimilation System ([59]). It was used by Margulis et al.[64] to estimate soil moisture using the EnKF. A more detailed discussion of the NOAH LSM is provided in Appendix A.

3.2.2 Radiative Transfer Model

A Radiative Transfer Model (RTM) is required to transform the states from state space to observation space. The RTM used here is identical to that used by Margulis et al.[64]. It is based on the retrieval algorithm used by Jackson et al.[45] to retrieve soil moisture from ESTAR observations during SGP97, but using the mixing model of Wang and Schmugge [90]. Surface roughness and vegetation effects are also accounted for ([7],[46]). Detailed descriptions of the ESTAR instrument and radiative transfer model are provided in Appendix B.

3.2.3 Model Error and Uncertainty

Model error was implicitly added to the data assimilation framework by allowing key parameters to assume different values in an expected range for each ensemble member. Uncertainty was imposed on four key soil and vegetation parameters, namely the saturated hydraulic conductivity, the minimum canopy resistance, the porosity and the wilting point. Varying the saturated hydraulic conductivity effectively varies the rate at which water can move through the soil column. Allowing the porosity and wilting point to vary means that each replicate has a distinct possible range of soil moisture values. Each replicate having a different minimum canopy resistance means that the rate of evaporation will be different for each ensemble member.

The values for these parameters afforded by the model based on land class or soil class were used as nominal values. The time-invariant parameter value for each ensemble member consists of the nominal value multiplied by a random variable of mean one and a coefficient of variation of 1.0 for both the saturated hydraulic conductivity and minimum canopy resistance, and 0.05 for the porosity and wilting point. Log-normal multiplicative Gaussian noise was added to yield a large range of values while ensuring that negative values did not occur. The relative frequency distributions of the parameters for the El Reno pixel are shown in Figure (3-1). Uncertainty was also included in the initial condition. Nominal relative saturation at the surface was set to 0.5, with the nominal values at depth determined by assuming a hydrostatic profile. Uncertainty was including by adding Gaussian noise of mean 0.0 and standard deviation decaying exponentially with depth from 0.2 at the surface.

Unpublished experiments found that the most effective way to introduce ensemble spread is through uncertainty in precipitation. Further discussion of uncertainty in precipitation is included in sections 3.3 and 3.4.

3.2.4 Algorithm Evaluation

The ensemble open loop (EnOL) provides the model estimate and associated model error in the absence of data assimilation, a valuable benchmark by which to measure

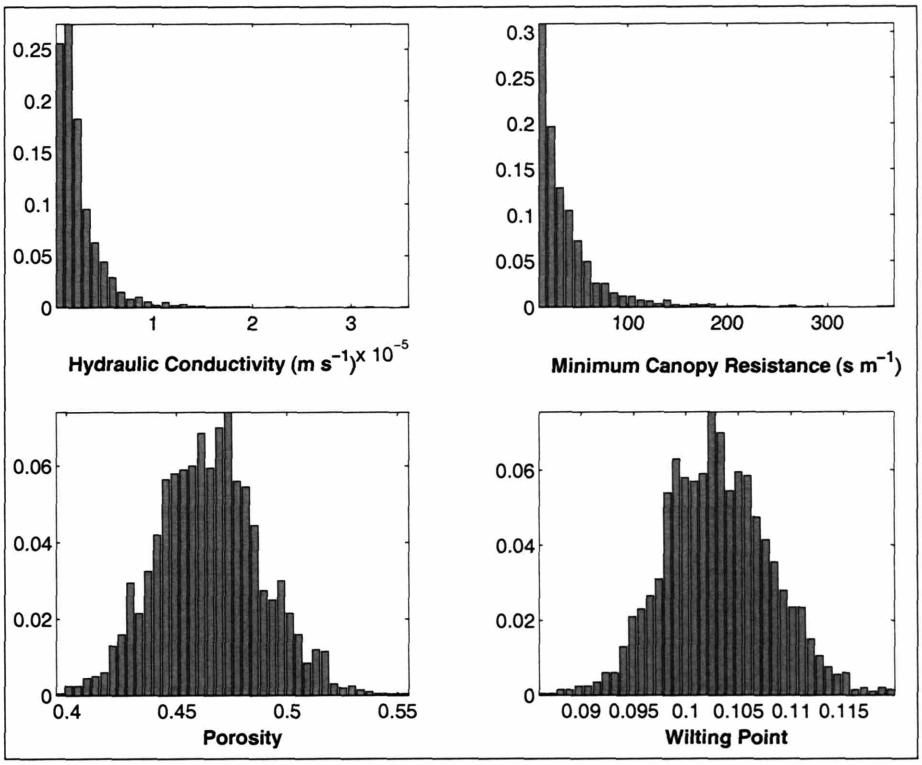

Figure 3-1: Relative frequency distribution of saturated hydraulic conductivity (top left), minimum stomatal resistance (top right), porosity (bottom left) and wilting point (bottom right) at the El Reno pixel.

the improvement after filtering or smoothing. To evaluate ensemble algorithms, the quantities of interest are the ensemble mean, which will be compared to the "true" state, and the standard deviation across the ensemble. Two summary statistics will also be used to assess the smoother algorithms performance relative to the EnKF and EnOL:

- The Root-Mean-Square-Error (RMSE) provides an average measure of the deviation of the ensemble mean from the true state over all estimation times. Clearly, the data assimilation algorithm is performing well if the ensemble mean is close to the truth.

- The Estimation Error Standard Deviation (EESD) is the average standard deviation across the ensemble calculated over all estimation times. The ensemble spread is a measure of the confidence which should be placed in the estimate.

Observations were available every 3 days, and estimates were required four times daily at 6am, 2pm, 6pm and 12am (2pm is used rather than 12pm as it is closer to the peak in soil surface temperature). For the four month experiment duration, this yielded a sample of 493 estimation times with which to calculate the RMSE and EESD.

3.3 Experiment 1: Precipitation Forcing Derived from Monthly Total Information

In a global land data assimilation application, precipitation data will likely be derived from satellite data such as Global Precipitation Climatology Project (GPCP). Daily, pentad and monthly total precipitation products are available from GPCP. Daily totals provide higher frequency information than pentad or monthly observations but due to temporal sampling and algorithm uncertainty the monthly total is more reliable. This temporal resolution is too coarse to characterize storm events for the purposes of land surface modeling which requires hourly data or better. The spatial resolution of observations ($2.5^o \times 2.5^o$) is orders of magnitude greater than that of the

Table 3.1: Rectangular Pulses Model parameters for Oklahoma City

Month	$E[t_r]$	$E[i_r]$	$E[t_B]$	T	$E[m]$	$\hat{m}$	$E[t_B]'$
May	7	2.616	**83**	744	164.15	149.09	**91**
June	5	3.220	**92**	720	126.00	192.81	**60**
July	8	2.264	**137**	744	98.36	103.9	**130**
August	4	2.374	**116**	744	60.91	113.24	**62**

estimation pixel (typically km), so information on spatial variability of precipitation is lost. Use of such data requires spatial and temporal disaggregation to the resolution of the model. Consequently, use of satellite-based data implies uncertainty in the timing, amount, and spatial distribution of precipitation.

3.3.1 Ensemble Precipitation using the Rectangular Pulses Model (RPM) to disaggregate the monthly total

The objective is to generate an ensemble of precipitation forcing which is constrained only by the monthly total precipitation. Using the Rectangular Pulses Model of Rodriguez-Iturbe et al.[76], it is assumed for each ensemble member that precipitation occurs as distinct rectangular pulses with random parameters. The expected arrival time, duration and intensity of a storm are exponentially distributed with mean values $E[T_B]$, $E[T_r]$ and $E[I_r]$ respectively.

Using historical meteorological data, Hawk and Eagleson [37] derived these climatological parameters for many stations across the United States. The Hawk and Eagleson parameters for the months of interest are shown in Table 3.1.

The method of Margulis and Entekhabi [63] is used here to derive a modified $E[t_B]$,$E[t_B]'$ which takes into account the observed monthly precipitation. The total monthly precipitation was derived from Oklahoma Mesonet precipitation records at El Reno. Using these "monthly observations", $E[t_B]'$ was calculated for the four months of interest in 1997 (Table 3.1). Further value can be derived from the monthly measured rainfall, by using it to discriminate between realizations. Here, realizations were

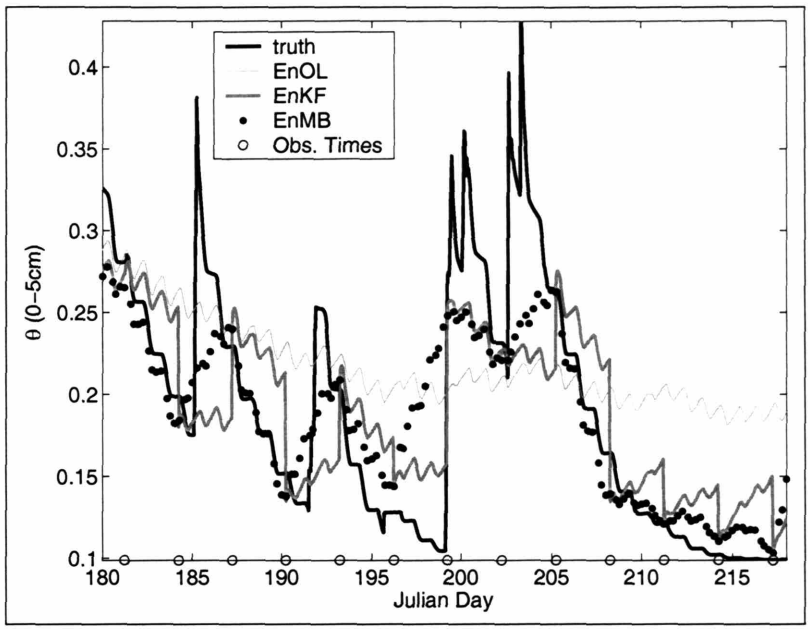

Figure 3-2: Ensemble mean volumetric soil moisture (θ) in the top 5cm of the soil column at El Reno, compared to the synthetic truth. Results are shown for the period between Julian days 180 and 218.

rejected if they were beyond 25% of the total observed precipitation at the end of the four month period.

3.3.2 Surface Soil Moisture at El Reno

Figure 3-2 compares the estimated surface soil moisture from the EnOL, EnKF and EnMB to the truth. In the absence of information on the timing and magnitude of precipitation events, the EnOL soil moisture is distributed across the dynamic range. Both the EnKF and EnMB are drawn towards to the truth at observation times. While the EnKF drifts uncorrected towards the EnOL between observations, the backward propagation of subsequent observations yields a smooth transition between observations in the EnMB. This is particularly advantageous during dry-down periods (e.g. Julian Day 205 to 219). The relative timing of observations and precipitation significantly impacts the performance of the smoother. Backward propagation of the increased soil moisture following a storm results in spuriously moist estimates in

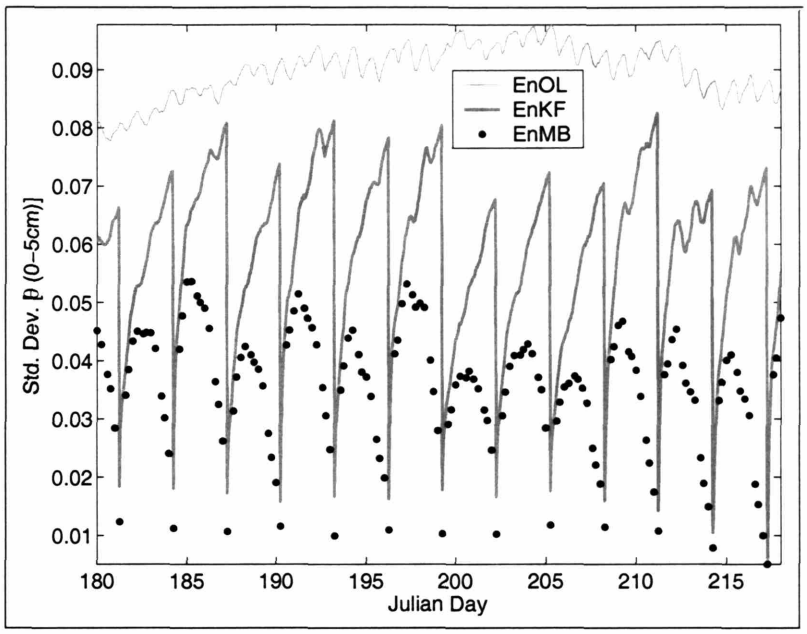

Figure 3-3: Estimation Error Standard Deviation (EESD) in the estimate of surface (0-5cm) volumetric soil moisture (θ) at El Reno for the period between Julian day 180 and 218.

the EnMB. The effect is most detrimental if the observation immediately precedes an observation (days 196-199), and less harmful if the precipitation is early in the interval (days 184-187, and 202-205). As smoothing is most effective on dry-down curves, it would be useful if we could identify dry-down curves over which to smooth. This issue is discussed further in Section 3.5.

Figure 3-3 shows the reduced EESD obtained from the EnMB compared to the EnKF and EnOL. The EESD in the EnOL is relatively constant at 0.09, about 25% of the dynamic range of soil moisture. In the filter case, the ensemble spread exhibits a characteristic sawtooth shape, growing rapidly betwen observations. The symmetry in the EnMB standard deviation indicates that the backward propagation of information through the covariance matrix is improving the estimate. The reduced standard deviation indicates that we should have increased confidence in the smoothed result compared to the filter. Figure 3-4 shows the reduction in ensemble spread after filtering/smoothing as a function of timing within the 3-day inter-observation period. At each estimation time the EESD for the smoother and filter were normalized by that of

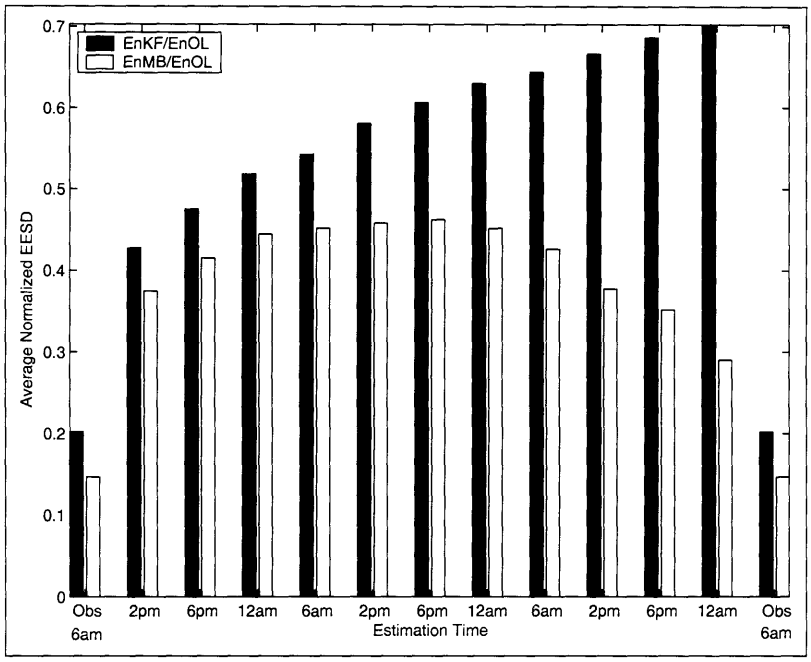

Figure 3-4: Average Normalized EESD in surface volumetric soil moisture as a function of timing within the inter-observation period at El Reno

the open loop. The results were then averaged for each point in the inter-observation period. The EESD in the ensemble filter grew to 0.7 times that of the open loop case as observations were available every 3 days. Shortening/lengthening the observation interval would reduce/increase this value. The maximum ratio in EESD between the smoother and the open loop is around 0.45, two-thirds of the maximum from the filter. The greatest improvement due to smoothing is immediately prior to the later observation. The correlation between states and the future observation is highest immediately prior to the observation and is diminished as the difference between the estimation time and the future observation increases. This is counterbalanced by the fact that EESD is at a minimum at the observation time and grows with time. The combination of the two effects is a symmetric rather than sawtooth evolution of the EESD between observations in the smoothed case.

3.3.3 Subsurface Soil Moisture Estimation at El Reno

Figure 3-5 shows the deviation between the estimated and true soil moisture at depth. With increasing depth, the length of time required by the EnOL to recover from spurious initial conditions increases. While the EnKF improves on the open loop, the EnMB reduces the deviation by over 50% as soon as the first observations become available. In the deepest layer, the EnKF takes 20-30 days to catch up with the smoother.

3.3.4 Summary Statistics at El Reno

Figure 3-6 demonstrates that smoothing improves over filtering at all depths in terms of both RMSE and EESD. The filtered estimate is quickly drawn towards the EnOL between observations, limiting the reduction in RMSE to 25% at the surface. The smoother leads to a further 20% improvement over the filter. At depth, smoothing alleviates the impact of initial conditions much faster than the filter. In layers 4 and 5 (20-45cm, and 45-100cm respectively), the RMSE is close to half that of the filter.

There is almost a 50% reduction in average EESD due to the filter compared to the EnOL. There is a further 33% reduction at the surface when the EnMB is employed. Although ensemble growth is slower at depth due to the dampened response to incident precipitation, there is a persistent reduction in EESD due to smoothing.

3.4 Experiment 2: Precipitation forcing from rain-gauge data

3.4.1 Derivation of precipitation forcing data

The objective of this experiment is to evaluate the performance of the EnMB in a data assimilation framework where precipitation data are from rain gauges. While gauge data is a useful indicator of when precipitation occurs, the amount is uncertain as the measurement is at a point, and is prone to errors due to spatial variability and under-

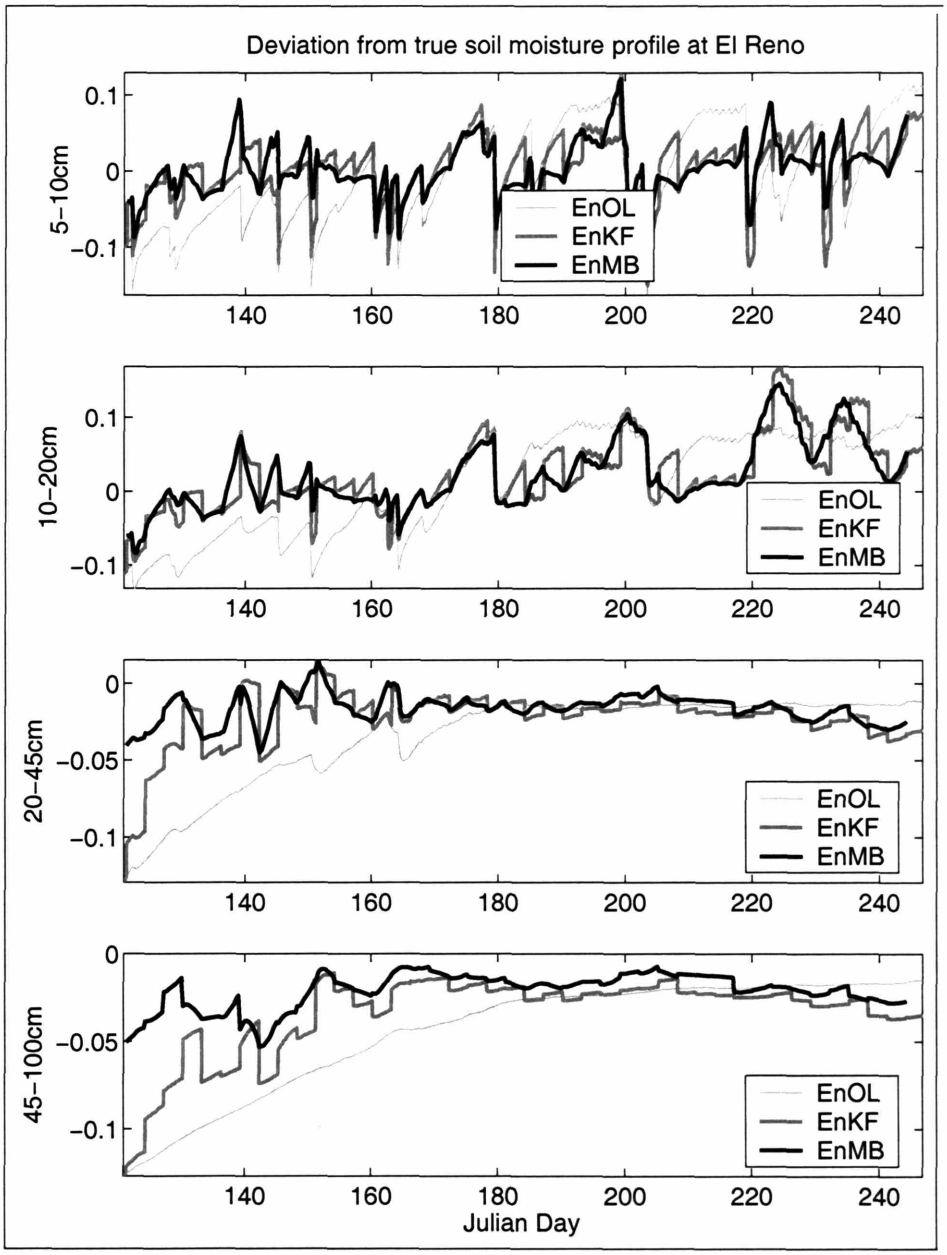

Figure 3-5: Deviation from "true" soil moisture at El Reno is shown at various depths. The smoothed estimate (EnMB) is compared to the filtered estimate (EnKF) and the ensemble open loop (EnOL).

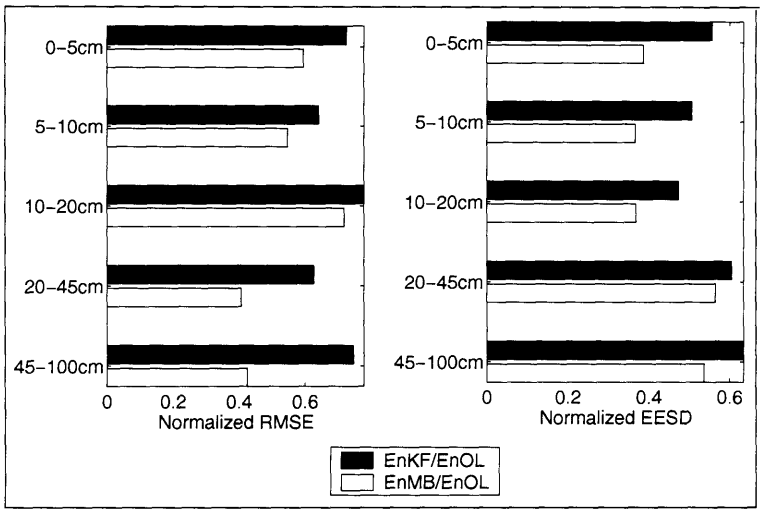

Figure 3-6: Normalized RMSE and Average Normalized EESD of volumetric soil moisture at depth (θ) at El Reno.

reach. An ensemble of precipitation forcing was generated to reflect this uncertainty. Nominal precipitation was multiplied by a lognormally-distributed random factor of mean 1.0 and standard deviation set equal to 50% of the nominal precipitation. The performance of the EnMB was evaluated at two locations:

- At El Reno (the gauge location) the timing of precipitation is known. A single realization of the precipitation forcing was used to generate "truth".

- At Little Washita it was assumed that the best available data is that recorded at El Reno. Gauge density in the SGP97 region is considerably higher than the rest of the world. This experiment evaluates the performance of the EnMB under the incorrect assumption that storm timing is perfectly known. Figure 3-7 shows that the amount and timing of precipitation are considerably different at Little Washita and El Reno.

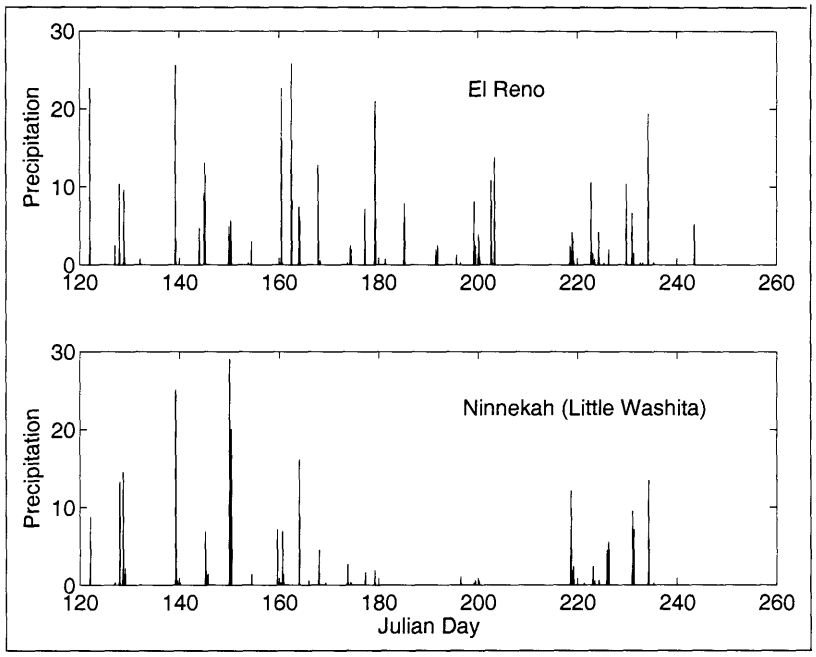

Figure 3-7: Observed precipitation (in mm hour^{-1} time series for El Reno and Ninnekah (Little Washita). This illustrates the difference in the timing and quantity of precipitation at the two stations.

3.4.2 Estimating Surface Soil Moisture at El Reno with precipitation forcing from gauge data at El Reno

Figure 3-8 compares the true soil moisture to that estimated using the EnOL, EnKF and EnMB. The benefit of smoothing is particularly noticeable between Julian days 226 and 229. Elsewhere, the improvement over filtering is relatively modest. This may be due to the limited growth of uncertainty in this experiment due to the assumption that the timing of precipitation is perfectly known. The growth of uncertainty between observations is limited to the uncertainty associated with the unknown parameters. From Figure 3-9, the reduction in standard deviation due to smoothing exceeds that achieved by filtering. The uncertainty introduced in this experiment is very small; the standard deviation across the filtered ensemble is on the order of 0.02, about 5% of the dynamic range of volumetric soil moisture. The limited improvement due to data assimilation between days 219 and 225 suggests that the uncertainty in the modeled estimate is comparable with the observation error.

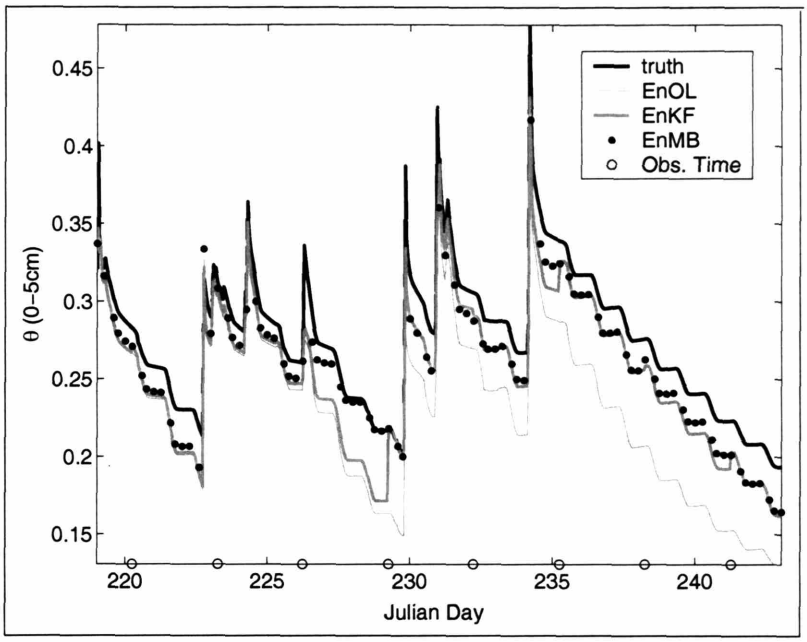

Figure 3-8: Ensemble mean volumetric soil moisture (θ) in the top 5cm of the soil column at El Reno, compared to the synthetic truth. Results are shown for the period between Julian day 219 and 243.

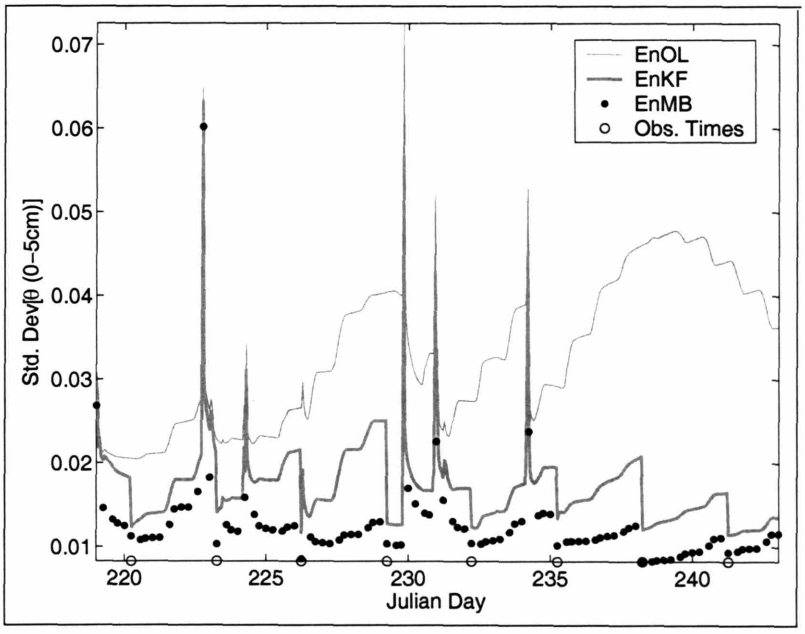

Figure 3-9: Estimation Error Standard Deviation in the estimate of surface (0-5cm) volumetric soil moisture (θ) at El Reno. Results are shown for the period between Julian day 219 and 243.

3.4.3 Estimating Soil Moisture at Depth at El Reno using El Reno forcing data

Figure 3-10 shows the deviation between the true soil moisture at depth, and the the estimate from the ensemble algorithms. Spurious dry initial conditions persist longer at depth, as illustrated by the EnOL estimate in layers 4 and 5. In the filter and smoother, the states at depth are updated through their correlation with the surface state and the observations. The filter improves more slowly than the smoother as it processes the observations sequentially. The smoother updates using observations in a batch, therby tying the estimate closer to the truth between observations.

3.4.4 Estimating Surface Soil Moisture at Little Washita with precipitation forcing from gauge data at El Reno

Figure 3-11 shows the estimated surface soil moisture at Little Washita. The EnMB improves over the EnKF and EnOL, but is unable to correct entirely for the fact that storms occurred at El Reno while Little Washita was dry. When precipitation occurs at El Reno all ensemble members receive precipitation, thereby reducing ensemble spread. Due to this apparent certainty that the soil at Little Washita is wet, the filter and smoother fail to update the ensemble mean towards the true value. This demonstrates the importance of correctly characterizing the sources of error and uncertainty in land data assimilation.

3.4.5 Summary Statistics at El Reno and Little Washita

From Figure 3-12, filtering yields about a 50% reduction in RMSE compared to the EnOL at the surface. Ensemble smoothing yields a further 20% reduction on average. At depth, the greater improvement in smoothing over filtering is largely due to smoothing's ability to correct for erroneous initial conditions. With longer experiments, this effect would be reduced. The EnKF yields a 50% reduction in EESD over the EnOL, but the EnMB yields a further 20% improvement over the

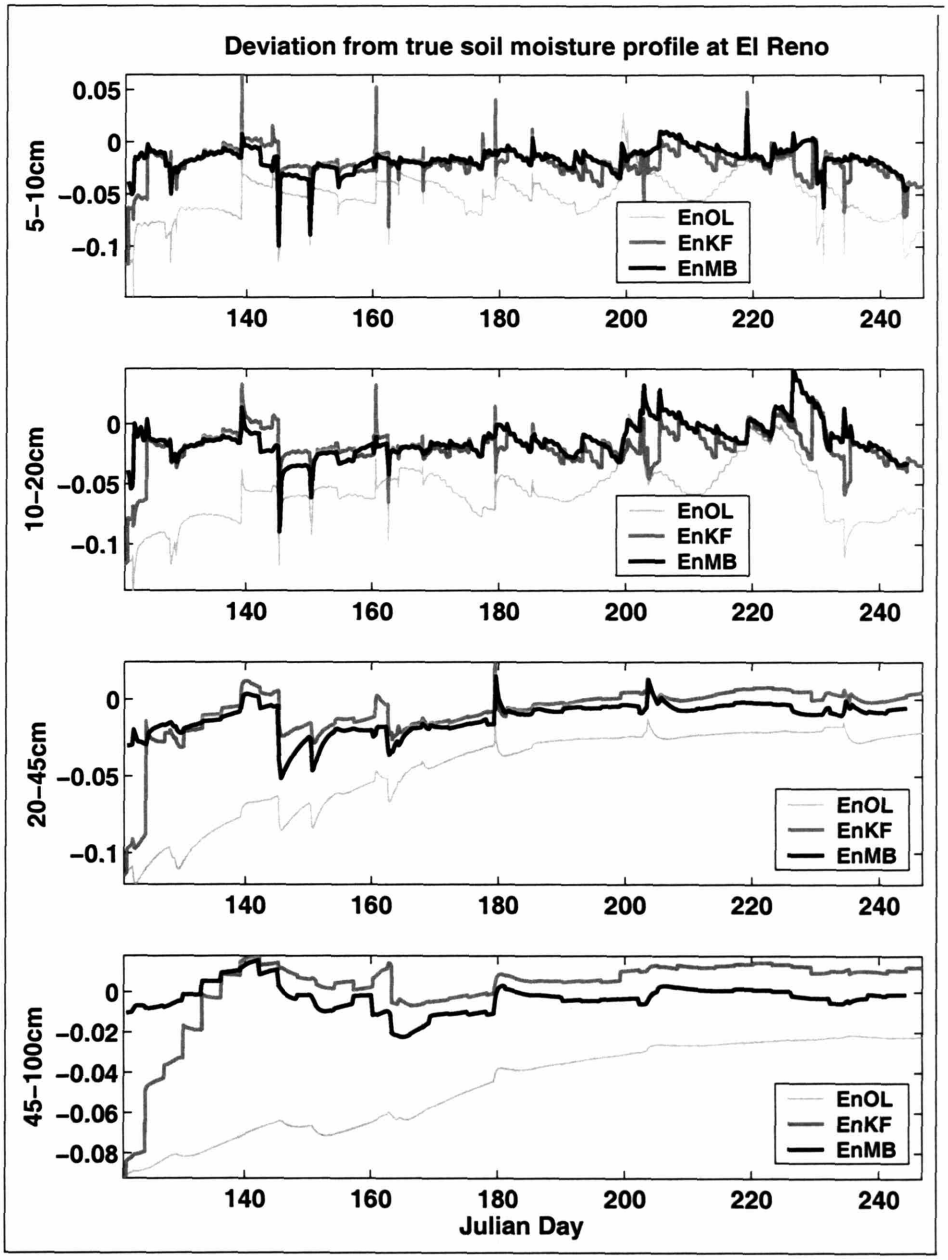

Figure 3-10: Deviation from "true" soil moisture at El Reno is shown for layers 2 to 5. The results from the moving batch smoother (EnMB) are compared to that of the EnKF and the ensemble open loop (EnOL).

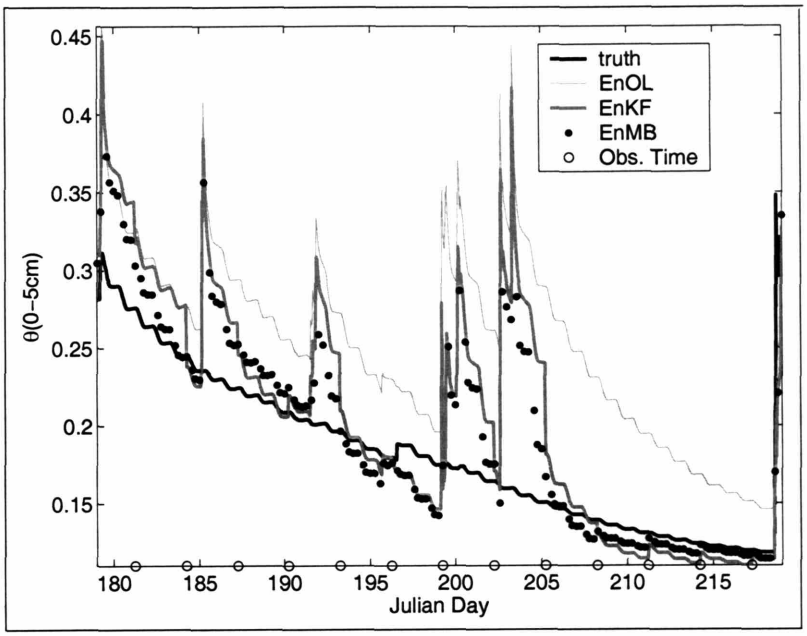

Figure 3-11: Ensemble mean volumetric soil moisture θ in the top 5cm of the soil column at Little Washita is compared to the synthetic truth for the interval from Julian Day 179 to 219.

sequential filter. The improvement is apparent at both El Reno and Little Washita.

3.5 Hybrid Filter/Smoother Approach

Recall from Section 3.3 that while the EnMB yielded improved results compared to filtering, the backward propagation of information pertaining to the soil's response to subsequent precipitation led to spuriously moist estimates. From Figure 3-2, it is evident that smoothing is most advantageous where the objective is to measure a particular dry-down series. Conversely, the smoother is least beneficial when the smoothing interval is disrupted by intermittent precipitation.

Here, a method is proposed to objectively divide the study interval into a series of dry-down events over which to smooth. It would be undesirable to use precipitation data for this purpose, as the objective is to estimate soil moisture with uncertain precipitation data and satellite data only. Fortuitously, the L-band brightness temperature observations can be used to make a first order assessment of when in the

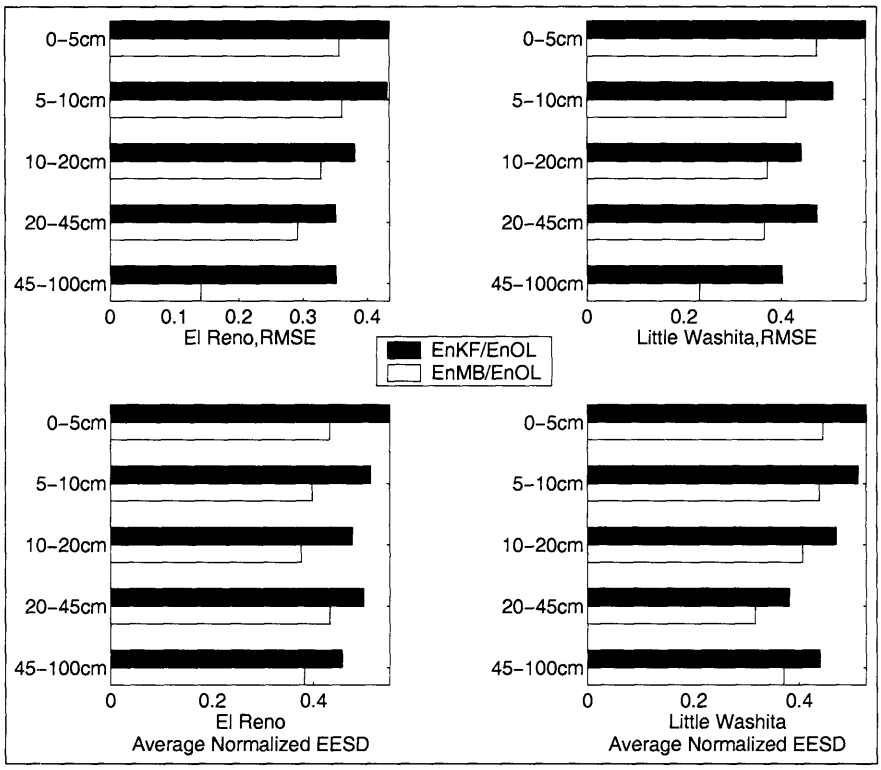

Figure 3-12: Normalized RMSE and average normalized EESD in the smoothed estimate of volumetric soil moisture θ compared to that of the filter and open loop.

study interval wetting has occurred.

Figure 3-13 shows the precipitation recorded at El Reno over the four-month long synthetic experiment (top). The resultant modeled soil moisture at the surface at El Reno is shown in the middle panel. L-band brightness temperature T_B observations were simulated from this soil moisture using the Radiative Transfer Model (bottom). T_B is a function of soil moisture, soil temperature, and many soil and vegetation parameters. In general, however, a decrease in T_B indicates the interim occurrence of precipitation so smoothing would offer no improvement over filtering. Because the observations have an error of 3K (1σ), decreases of less than 6K are disregarded. Provided the brightness temperature is increasing the soil is drying down, and smoothing should improve over filtering. Instead of prescribing the fixed length of a moving smoother window, in this approach the length of the smoother window is dynamic such that the augmented state vector consists of all estimation times on a given dry-down curve. Soil moisture estimation using this technique should yield improved estimates through two mechanisms:

- Preventing backward propagation of information from a subsequent dry-down
- Lengthening the smoother window to encompass all observations on the dry-down curve of interest guarantees that the state is estimated using all relevant observations.

3.5.1 Results

In Figure 3-14, the "Hybrid Smoother/Filter" performance is compared to that of the EnKF and the EnMB. The key benefit of this hybrid algorithm is seen, for example, on the dry-down beginning on day 205. As the brightness temperatures are increasing for 12 days, the smoother window encompasses five observations. Using all of these observations in a single batch to estimate soil moisture at all estimation times in that interval yields an improvement over using the EnMB (LSW=2). The hybrid algorithm also improves the estimate during wetting periods where the filter is used instead of the EnMB. Precipitation occurs immediately prior to the observation on

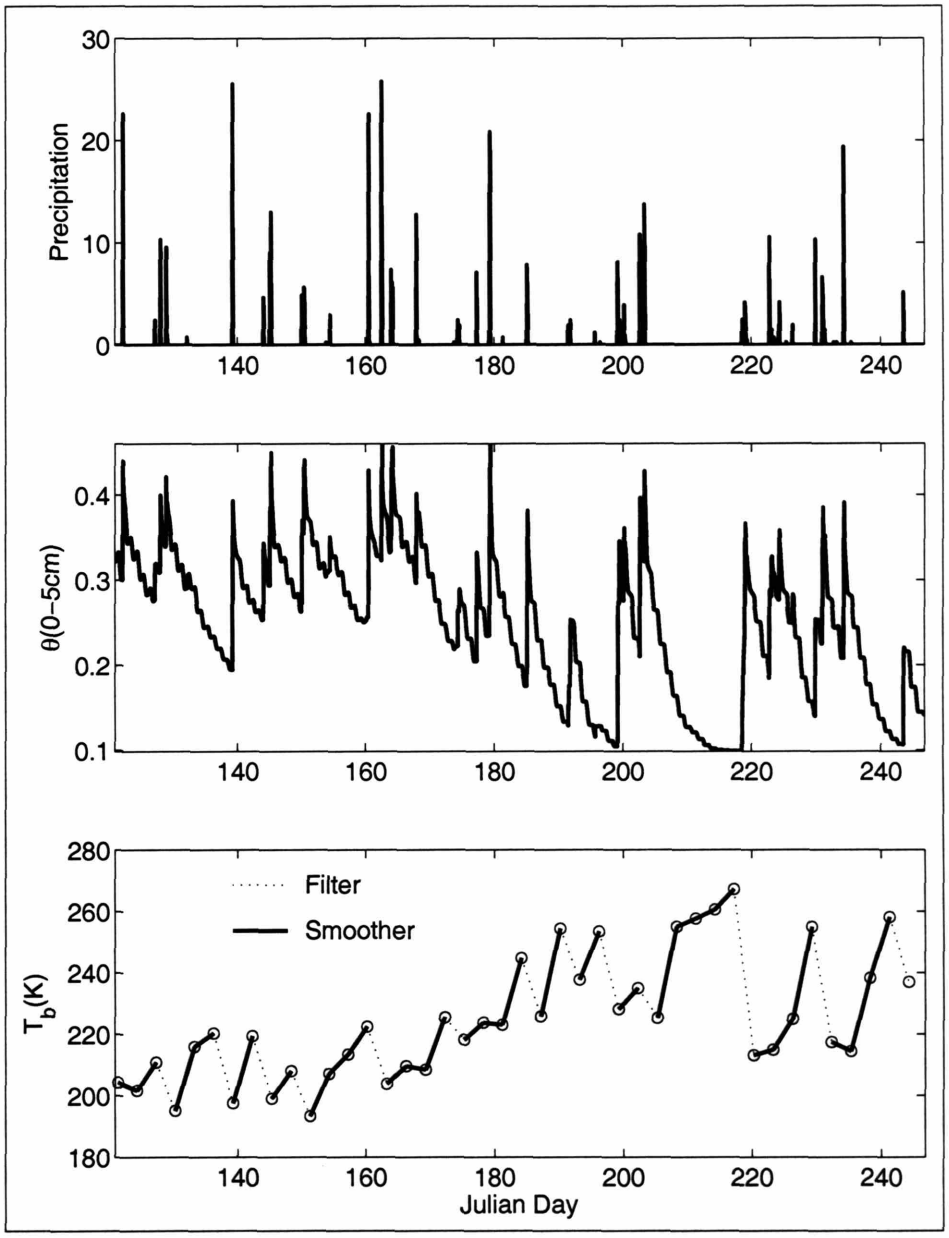

Figure 3-13: The incident precipitation (in mm hour^{-1}) at El Reno is shown in the top panel. The resultant modeled volumetric soil moisture θ at El Reno is shown in the middle panel. In the bottom panel, the simulated brightness temperatures T_B associated with these soil moisture values are plotted. The solid lines indicated smoothing intervals which are separated by filtered intervals.

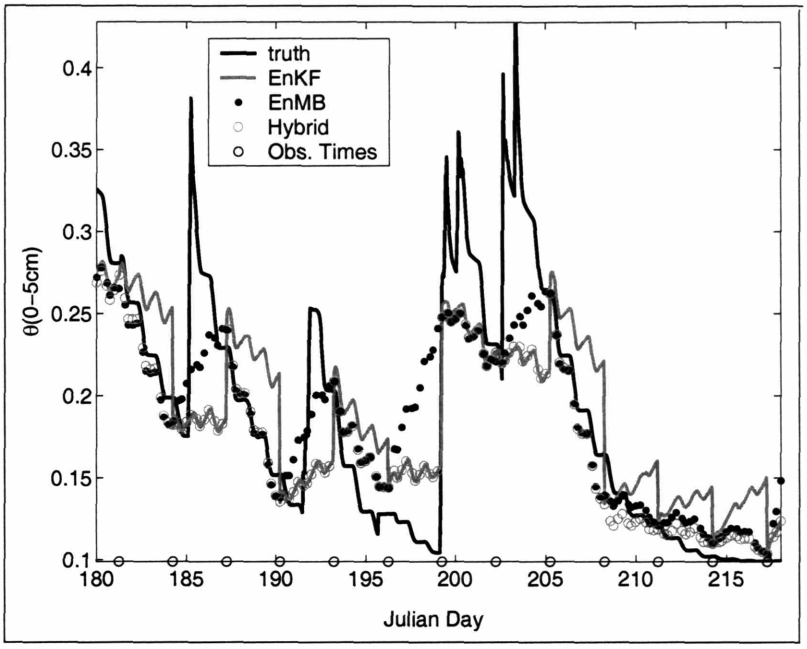

Figure 3-14: Ensemble mean volumetric soil moisture θ in the top 5cm of the soil column at El Reno, compared to the synthetic truth for the period between Julian Days 180 and 218.

day 199. The EnMB propagates information on the wet condition back in time yielding a moister estimate between days 196 and 199. In the hybrid algorithm the filter is used in this interval, preserving the drier soil moisture condition. There are cases where the hybrid algorithm can result in a poorer estimate than the EnMB (days 184-187 and 202-205). Here, the precipitation occurs just after an observation filtering underestimates the soil moisture, and the hybrid is therefore too dry in this interval. When the precipitation occurs halfway between observations (Julian Day 190-193), the hybrid algorithm has no net effect. These results demonstrate the difficulties of estimating soil moisture in intermittent precipitation using temporally sparse observations.

The impact of using the hybrid algorithm is also apparent in the reduction in ensemble standard deviation (Figure 3-15) compared to the EnKF and the EnMB alone. When there is intermittent precipitation, the algorithm switches constantly between filtering and smoothing. When the hybrid selects the filter, ensemble spread

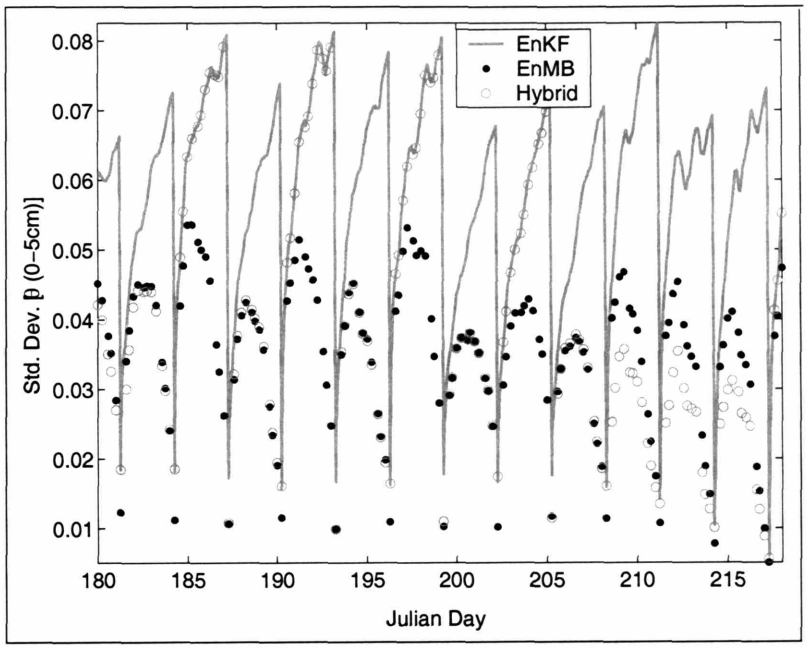

Figure 3-15: Estimation Error Standard Deviation in the estimate of surface (0-5cm) volumetric soil moisture at El Reno. The EESD from the Hybrid Smoother/Filter approach is compared to that obtained using the EnKF and EnMB alone. Results are shown for the period between Julian day 180 and 218.

grows like that in the filter, unconstrained by subsequent observations. Similarly, when the smoother is used for an interval of length 2 (i.e. two observations), the standard deviation across the ensemble is comparable to that of the EnMB. However, when the hybrid recognizes a lengthy dry-down and estimates the soil moisture over the entire interval as a long batch, the impact of additional future observations reduces the ensemble standard deviation below that of the moving batch smoother. The issue of the relative timing of precipitation and observations merits further attention. Nonetheless, this approach makes tentative steps to address the apparent pitfalls in using ensemble smoothing techniques in soil moisture estimation.

3.6 Conclusion and Discussion

It is argued soil moisture estimation is a reanalysis-type problem rather than a control-type or forecast problem and consequently a smoothing approach is more appropriate

than filtering. An ensemble-based smoothing algorithm was presented in which all observations within a prescribed window are used in a batch estimator to determine soil moisture at the surface and at depth. The algorithm was compared to the ensemble Kalman filter in two experiments with different precipitation data, and smoothing improved the estimated soil moisture at the surface and at depth. Smoothing was particularly effective in correcting for erroneous initial conditions at depth. This improvement is significant as it may lead to improved surface flux estimation through the dependence of the latent heat flux on root zone soil moisture. The smoother incorporates more observations than the filter to obtain the estimate, and thus is characterized by significantly reduced estimation estimation errors and increased confidence in the estimate.

The use of smoothing in land data assimilation is complicated by the occurrence of precipitation. A hybrid smoother/filter approach was presented to address this by breaking the study interval into a series of smoothing windows. The smoother window length is dynamic rather than prescribed, including all observations in a single dry-down period. The soil moisture for the whole dry-down is determined in one batch. This method improves the estimate by preventing the backward propagation of information from precipitation events after an observation at the end of a smoothing window. Here, the hybrid assumes precipitation has occurred if the decrease in brightness temperature is greater than 2σ, i.e. twice the standard deviation in the observation. While this is simplistic, it demonstrates the feasibility of using brightness temperature to break the interval into drydown events. Further experiments will address the issue of spatial variability in brightness temperature. The impact of the relative timing of precipitation and observations merits further attention as the performance of the hybrid depends on when in the inter-observation period the precipitation occurred.

So far, the performance of the smoother has been evaluated on independent uncorrelated pixels. In Chapter 5, ensemble smoothing techniques will be used to estimate soil moisture over a grid of spatially correlated pixels to estimate soil moisture from combined active and passive (multiscale) microwave-based observations like those ex-

pected from Hydros.

Chapter 4

State and Flux Estimation using the EnKS with real data during SGP97

4.1 Introduction

In Chapter 3, soil moisture was estimated using an ensemble moving batch (EnMB) smoother in which the state vector was augmented to include all states within a prescribed time window. All observations within this window were used to update all of the soil moisture states in a batch. Using this simple smoother rather than a filter resulted in an improved estimate of surface and root zone soil moisture and was particularly effective in correcting erroneous initial conditions at depth. A major disadvantage of the EnMB was that the state and observation vectors were augmented to be distributed in time, resulting in a computationally expensive smoother. Having demonstrated that smoothing could yield improved results over filtering, the objective in this chapter is to find a less computationally expensive approach to ensemble smoothing.

The ensemble Kalman smoother was introduced by Evensen and Van Leeuwen [26] as an improvement over the original ensemble smoother (ES) of Van Leeuwen and

Evensen [83]. Evensen and Van Leeuwen [26] compared its performance to that of the EnKF and the ES in a problem with the Lorenz equation. It was demonstrated that for linear dynamics it is identical to the ES, but for non-linear dynamics its use of the EnKF solution as its first guess resulted in a considerable improvement over the ES. Gronnevik and Evensen [34] demonstrated the feasibility of using the EnKF, ES and EnKS in a fish stock assessment problem. As the model was only weakly non-linear, the EnKS and ES yielded similar results. Brusdal et al. [2] compared the performance of the EnKF, EnKS and Singular Evolutive Extended Kalman (SEEK) filter in an operational ocean forecasting problem. The EnKS yielded improved results of sea level anomaly over the EnKF. It is noteworthy that the EnKS actually resulted in a somewhat poorer estimate of sea-surface temperature than the EnKF due to the introduction of artificial noise through updating with observations too far into the future. This arose due to the short decorrelation scale of sea-surface temperature.

There is limited memory in soil moisture in the unsaturated zone. Rather than use each observation to update all past estimates, the EnKS can be implemented as a fixed lag smoother in which the observation is only used to update past estimates within a fixed time window. In section 4.5, it is shown how a suitable lag (or length of this time window) might be determined. It is argued that there is a trade-off to be made between the improved accuracy achieved by increasing the lag and the increased computational burden.

To some extent, this paper may be thought of as revisiting the experiment of [64] that demonstrated the feasibility of using the EnKF for land data assimilation using the same forward model, observations and validation data used here. Assimilation of brightness temperature data improved the estimated soil moisture at the surface and at depth compared to a traditional "open loop" simulation using nominal forcing and parameters. The root zone soil moisture, and consequently the latent heat flux estimates were poorest at the beginning of the experiment as it took time for the information from the observations to significantly impact the states at depth. The hypothesis of this study is that using the EnKS rather the EnKF will address this shortcoming; through assimilation of future observations we expect to yield improved

estimates of surface and root zone soil moisture and latent heat flux.

A key difference between this experiment and that of Margulis et al. [64] is in the definition of the "open loop" run. In [64] the open loop is a single run of the model using nominal forcing and parameters. In this experiment, we use an ensemble open loop (EnOL) in which an ensemble of realizations is propagated forward without assimilating any observations. The resulting ensemble can be compared to the EnKF and EnKS to see how the ensemble statistics evolve in the absence of data assimilation.

4.2 Data Assimilation Approaches

4.2.1 Ensemble Kalman Filter

The objective is to estimate the value of the states of the system, that are stored in the state vector y. Here, the state vector contains the volumetric soil moisture in six layers of the soil column. The state y is propagated forward in time using the model, A, which has parameters in the vector α, forcing data in the vector $u(\tau)$, and vector of system uncertainty $\omega(t)$.

$$y(t) = A[y(\tau), \alpha, u(\tau), t, \tau, \omega(t)] \tag{4.1}$$

where $t > \tau > 0$. In this experiment, A is the Noah land surface model discussed in Section 4.3.1. In the ensemble Kalman filter, an ensemble of N realizations of this state vector is propagated forward in time. The ensemble size should be large enough to ensure repeated experiments converge on the same result. Tests (not reported) demonstrated that an ensemble size of N=400 is sufficient. Each realization has random initial conditions, model parameters and forcing as discussed in Sections 4.3.4 and 4.3.5.

At each update time t, a vector of observations (z) becomes available. A non-linear operator, $\mathcal{M}$, relates the true state to the measured variable. Here, $\mathcal{M}$ is the radiative transfer model discussed in Section 4.3.3, which relates the surface (0-5cm)

soil moisture to the observed L-band radiobrightness temperature in z.

$$z(t) = \mathcal{M}[y(t)] + e(t) \tag{4.2}$$

The uncertainty in the observation is given in the vector e, which is assumed to be zero-mean with covariance matrix R_e. From [3], each replicate is updated with respect to a perturbed observation to ensure that the spread of the updated ensemble is consistent with the true posterior estimation error covariance. So, for a given vector of measurements at time t, an ensemble of perturbed observations is generated:

$$z_j(t) = z(t) + \epsilon_j(t) \tag{4.3}$$

where j refers to the j-th ensemble member. If the ensemble of perturbations is gathered into the matrix $\gamma = (\epsilon_1, \epsilon_2, ...\epsilon_N)$, the measurement error covariance can be written as:

$$R_e = \frac{\gamma\gamma^T}{N-1} \tag{4.4}$$

From [24], the so-called analysis or update (^{a}) is obtained by updating each replicate individually:

$$y_j^a(t) = y_j(t) + K(t)(z_j(t) - \mathcal{M}[y_j(t)]) \tag{4.5}$$

where $K(t)$ is the Kalman gain matrix:

$$K(t) = C_{YM}(C_M + R_e)^{-1} \tag{4.6}$$

C_{YM} is the forecast cross covariance between the state $y(t)$ and the measurement predictions $\mathcal{M}[y(t)]$. C_M is the forecast error covariance of the measurement predictions.

The states, perturbed observations and predicted measurements can be gathered into the matrices Y, Z and M respectively, such that each column contains a single realization of the relevant variable. The terms of the Kalman gain matrix can be

written as follows:

$$C_{YM} = \frac{1}{N-1} Y' M'^T \tag{4.7}$$

$$C_M = \frac{1}{N-1} M' M'^T \tag{4.8}$$

where primed matrices indicate that the ensemble mean has been removed from each column. Equation 4.5 can therefore be expressed as:

$$Y^a(t) = Y(t) + Y'(t)M'^T(t)(M'(t)M'^T(t) + \gamma\gamma^T)^{-1}(Z(t) - M(t)) \tag{4.9}$$

Although the L-band brightness temperature is directly related to surface soil moisture, subsurface soil moisture states are updated through their covariance with the surface soil moisture and the observations. They will also respond to the change in surface soil moisture as water is transported within the soil column in the land surface model.

4.2.2 Ensemble Kalman Smoother

The EnKS is often described as an extension of the EnKF in which information from the observation at update time t is used to update, not just the state estimate at that update time, but also at previous times, t' using:

$$Y^a(t') = Y(t') + Y'(t')M'(t)^T(M'(t)M'(t)^T + \gamma\gamma^T)^{-1}(Z(t) - M(t)) \tag{4.10}$$

If the terms pertaining to time (t) are grouped together in $B(t)$, this can be written as:

$$Y^a(t') = Y(t') + Y'(t')B(t) \tag{4.11}$$

The EnKS is a sequential algorithm, requiring only forward model runs. There are no additional model runs beyond those required by the EnKF. Furthermore at an update time t, the matrix $B(t)$ is computed to update the current state in the EnKF. The

computationally expensive inversion of $(M'(t)M'(t)^T + \gamma\gamma^T)$ has therefore already been performed. Implementation of the EnKS requires that the ensemble at the prior times must be stored and available to be updated each time new observations become available. The computational burden can be controlled by limiting the number of times at which a smoothed estimate is required. In this application the model time step is half hourly. While applications such as flood-forecasting would require hourly or sub-hourly estimates, we are primarily interested in the components of the the surface energy balance and so estimates four times daily are sufficient.

The EnKF solution is the "first guess" of the EnKS. Each update with a subsequent set of observations results in a change in ensemble mean and a reduction in ensemble variance. The EnKS should therefore always produce an estimate which is at least as good as the EnKF. As observations further into the future are used, the improvements become negligible, indicating that they are beyond the decorrelation time [25].

The EnKS is subject to the same Gaussian assumption as the EnKF. At update times, it is assumed that the probability distribution of the state across the ensemble is Gaussian and therefore the ensemble mean and covariance are adequate descriptors. The ensemble of soil moisture realizations is rarely perfectly Gaussian and during extremely wet or dry conditions it can become particularly skewed. Consequently the EnKF and EnKS are suboptimal. In the EnKS, it is assumed that observations are independent in time as each observation is used when it becomes available to update the current and past state estimates. Any useful information that may be gleaned from the temporal correlation of satellite observations would therefore be lost.

4.3 Estimation of Soil Moisture during SGP97 with the Ensemble Kalman Smoother

The EnKS and EnKF will be used to estimate surface and root zone soil moisture as well as surface energy fluxes during the Southern Great Plains Experiment 1997.

In this paper, we are using the same model domain, land surface model, radiative transfer model, land cover and soil texture data, radiobrightness observations and validation data as Margulis et al. [64]. The key differences in the experiment set-up are in model error and precipitation forcing.

4.3.1 System Model: Noah Land Surface Model

The forward model in this data assimilation framework is the Noah land surface model, the user's guide for which may be found online (http://www.emc.ncep.noaa.gov/mmb/gcp/noahlsm/README_2.2.htm). It is a 1-D soil-vegetation-atmosphere transfer (SVAT) model that simulates the soil moisture and temperature profile in addition to the mass and energy terms of the surface water and energy balances [19][5]. Additional information on the Noah LSM is provided in Appendix A. The forcing data required by the model include precipitation, air temperature and humidity, surface pressure, wind speed, and downward longwave and solar radiation at the surface. In this application, the soil column was modeled as six layers centered at 0.025m, 0.075m, 0.15, 0.325m, 0.725m and 1.475m. Each cell is classified as one of 13 vegetation types and one of 9 soil types that determine the nominal values of the vegetation and soil parameters. Maps of the vegetation and soil classifications used on the model domain are provided by Margulis et al. [64]. The 4km $\times$ 4km grid cells on the model domain are modeled independently. The assumption that there is no lateral moisture or heat flux between the soil columns is reasonable due to the low relief landscape.

4.3.2 Southern Great Plains Experiment 1997

The Southern Great Plains Experiment 1997 was a soil moisture mapping experiment which took place between 18th June and 17th July 1997 in the central plains of Oklahoma. The objective of the experiment was to evaluate the retrieval algorithm developed for the electronically scanned thinned array radiometer (ESTAR) at coarse resolution [45]. Airborne ESTAR observations from SGP97 are used here with the

EnKS and EnKF to estimate the soil moisture profile.

While our whole domain study is on a 4km grid, the EnKF and EnKS are first validated using in-situ observations from the long-term monitoring sites at El Reno, Central Facility and Little Washita. In-situ gravimetric observations were made in several 800m $\times$ 800m field sites at each location to validate the soil moisture retrieval from 800m radiobrightness data. As in [64], soil moisture validation was carried out at El Reno site 5 (ER05), Little Washita site 13 (LW13) and Central Facility site 8 (CF08). The latent heat flux is validated against data from the ARM-CART station at Central Facility 1 (CF01). These sites were selected to reflect the variation in precipitation forcing across the domain. While meteorological stations at El Reno and Central Facility recorded comparable total amounts of precipitation, the intensity and timing of storm events were different at the two sites. Little Washita was characterized by lengthy dry periods.

4.3.3 Radiative Transfer Model

The assimilated observations are from ESTAR, a synthetic aperture, passive L-band radiometer operated at a central frequency of 1.413 GHz. The radiative transfer model is based on the retrieval algorithm of Jackson et al. [45] that was used to retrieve soil moisture from ESTAR radiobrightness temperature observations during SGP97. The effect of surface soil moisture on the soil dielectric properties is determined using the mixing model of Wang and Schmugge [90]. Surface roughness and vegetation effects are accounted for [7],[46]. Additional detail on this radiative transfer model is provided in Appendix B. This is identical to the RTM used by Margulis et al. [64] and in Chapter 3. Observation error is assumed to be normally distributed with mean zero and standard deviation 3K.

4.3.4 Model Error and Uncertainty

An advantage of ensemble techniques is that they allow great flexibility in the representation of model error. Errors in modeled soil moisture arise due to parame-

terization and representation of physical processes in the model itself, and because forcing variables are not measured at every grid cell. Model error was accounted for in this data assimilation experiment by allowing the parameters and forcing variables in Table 4.1 to vary within a physically reasonable range, and through uncertain initial conditions and precipitation. Nominal values are those given by the model for the prescribed vegetation and soil classes. Uncertain initial conditions were obtained by setting the nominal relative saturation at the surface to 0.5, and assuming a hydrostatic profile to determine the nominal saturation at depth. Variability was introduced by adding zero-mean Gaussian noise with standard deviation decaying exponentially with depth from 0.05 at the surface. The impact of perturbing forcing data and model parameters on the soil hydrology and evaporation scheme of the Noah land surface model is discussed further in Appendix A.

4.3.5 Precipitation Forcing

The objective is to simulate a data assimilation framework in which the available precipitation data are from a satellite-based global precipitation product. Here, the Global Precipitation Climatology Project (GPCP) merged analysis of pentad precipitation is used which is available on a $2.5^o \times 2.5^o$ grid [92]. As the model time step is half-hourly and the spatial resolution is 4km, the precipitation data must be temporally and spatially disaggregated.

Figure 4-1 shows the SGP97 domain on the right, which intersects three cells of the 2.5^o GPCP grid. For each of these three grid cells, the total precipitation in each of the five day periods during SGP97 is shown on the left. The total precipitation during the 35 day period in the grid cells denoted 1, 2 and 3 varied considerably; 122.76mm, 173.16mm and 80.90mm respectively.

The data are disaggregated based on the approach used by Margulis et al. [61]. For each pentad in which precipitation occurred, it was assumed that the total precipitation occurred in a single event with uniform intensity. The duration and intensity are obtained from the Rectangular Pulses Model [76], using the parameters for Oklahoma City [37]. The timing of the start of the storm is uniformly distributed in

Table 4.1: Sources of Uncertainty in Model Parameters and Forcing

Variable		Distribution	Standard Deviation (σ) or Coeff. of Variation (CV)
Parameters			
Saturated Hydraulic Conductivity	K_s	Lognormal	CV=1.0
Minimum Stomatal Resistance	R_{cmin}	Lognormal	CV=1.0
Porosity	θ_s	Lognormal	CV=0.05
Wilting Point	θ_w	Lognormal	CV=0.05
Roughness Length	z_0	Lognormal	CV=0.75
Leaf Area Index	LAI	Additive Gaussian	σ=0.25
Forcing			
Air Temperature	T_a	Additive Gaussian	σ=3K
Shortwave Radiation	R_g	Multiplicative Gaussian	σ=0.1
Relative Humidity	f	Additive Gaussian	σ=1.0%
Wind Speed	u	Additive Gaussian	σ=0.2 ms^{-1}
Precipitation	P	See Section 4.3.5	

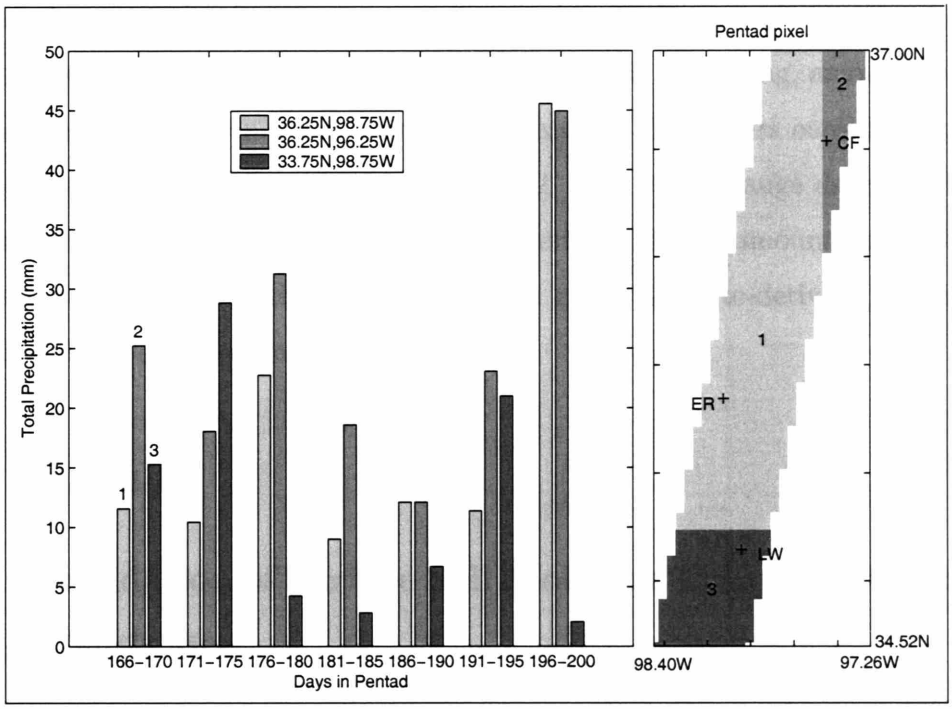

Figure 4-1: Precipitation totals for each pentad during SGP97, over each of the three $2.5^o \times 2.5^o$ pixels intersecting the SGP97 domain. The locations of El Reno (ER), Central Facility (CF) and Little Washita (LW) are indicated by crosses.

the five day interval. The total precipitation from the storm is then conditioned on the pentad total assuming that the observed precipitation is lognormally distributed with a mean of the nominal value and a coefficient of variation of 0.6. This provides a half-hourly precipitation time series for each ensemble member for each 2.5^o grid cell. For each ensemble member, the coarse resolution precipitation data are then disaggregated from a 2.5^o resolution grid to the 4km model resolution using a 7-level random cascade model [33]. It is assumed that the spatial structure of the storm is constant for the storm duration.

This disaggregation scheme results in uncertainty in the timing, amount and spatial distribution of precipitation across the domain. This differs considerably from the precipitation forcing used by Margulis et al. [64], where gauge data at El Reno was used throughout the domain, assuming uncertainty in the amount only. The new approach is motivated by the likely reliance on global satellite-derived precipitation data in a global data assimilation problem.

4.4 Results

4.4.1 Validation of Surface Soil Moisture Estimate at Ground Truth Sites

Surface soil moisture at three locations (Central Facility (CF08), El Reno (ER05) and Little Washita (LW13)) is compared to "ground truth" gravimetric measurements in Figure 4-2. At each location, multiple soil samples were taken to measure variability within the site. The grey circles represent the mean volumetric soil moisture across the site, and the error bars indicate $\pm\sigma$, where σ is the standard deviation in volumetric soil moisture across that site. The soil moisture estimate from the EnOL, EnKF and EnKS in the following figures is the ensemble mean. It is critical to recognize that the ensemble mean is an average across the ensemble, and not a physical realization from the model.

As precipitation forcing in this experiment was derived from GPCP Pentad data,

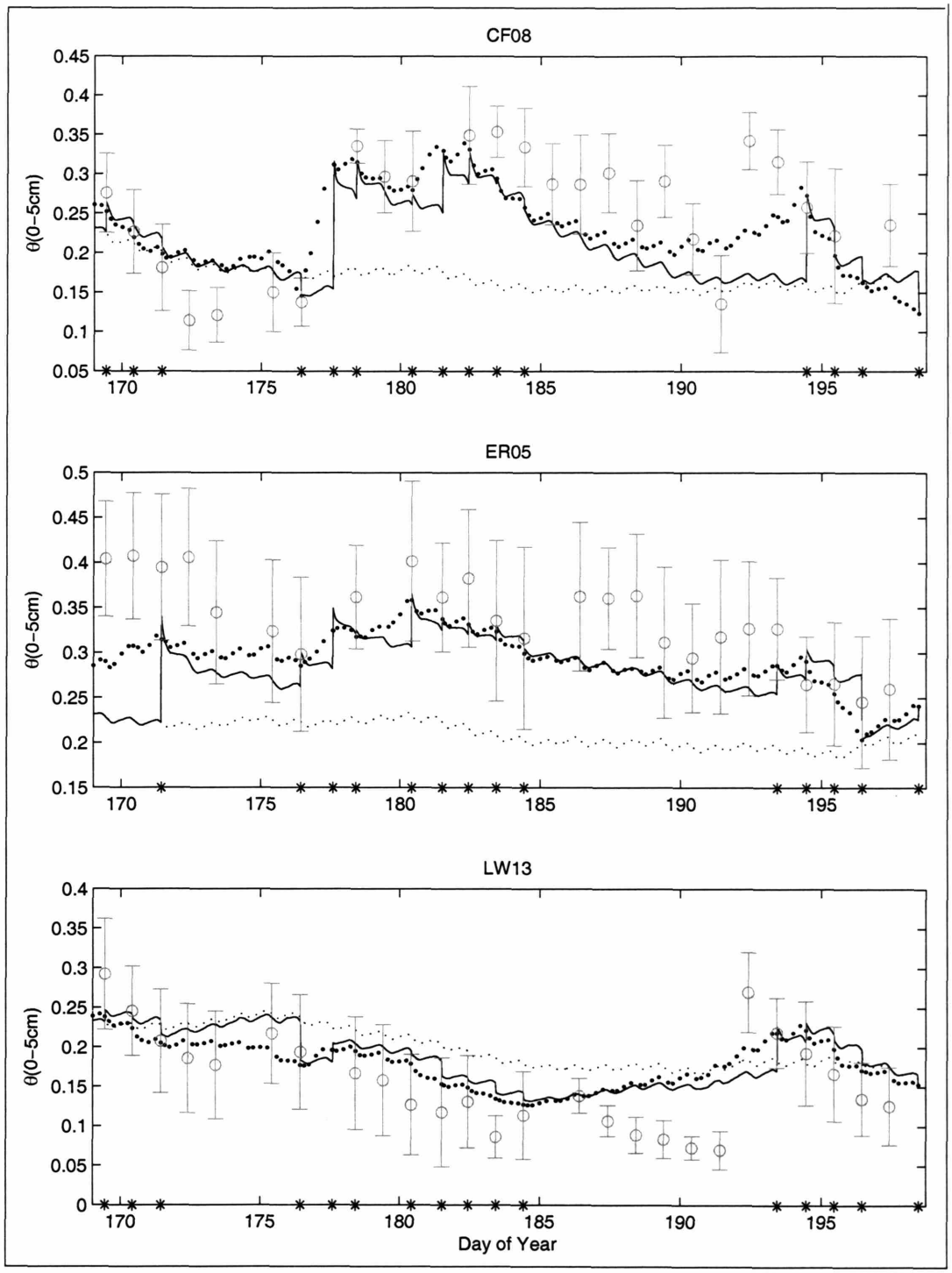

Figure 4-2: Estimated surface volumetric soil moisture from the EnOL (dotted), EnKF (solid) and EnKS (solid circles) are compared to the observed soil moisture from gravimetric measurements. Grey circles and vertical bars indicate the measurement mean $\pm$ one standard deviation. ESTAR measurement times are indicated by asterisks on the time axis.

the derived ensembles of hourly precipitation forcing data reflect uncertainty in the timing, duration and intensity of storm events. Soil moisture in each of the N ensemble members will therefore increase at different times, to varying moisture levels and dry down at different rates. As an average across this ensemble, the EnOL estimated surface soil moisture will lack the familiar temporal features seen in true soil moisture. At any time step, some ensemble members will be very moist, some will be drying down, while some will be completely dry. The average effect is an estimate in the middle of the dynamic range of possible values. In general, L-band brightness temperature will drop suddenly when precipitation occurs, and increase as the soil column dries down. As the EnKF and EnKS estimates are constrained by the brightness temperature observations, they both capture the variability in soil moisture due to the occurrence of precipitation.

While the EnKF extracts information from observations as they become available, the EnKS also propagates the information backwards in time, updating states before the update time. Consequently, at each of the validation sites the EnKS corrects for the spurious initial conditions before the first observation. This is particularly beneficial for situations such as at ER05 where there were no observations on Days 169 or 170.

In the EnKF there are abrupt changes in ensemble mean at update times and the estimate drifts towards the EnOL between observations. At first glance, it may seem that the EnKF estimate is simulating an infiltration front when there is a sudden increase in soil moisture. However, this is really the EnKF updating to a moister state upon assimilation of a brightness temperature observation. The precipitation event and true increase in soil moisture occurred some time since the previous observation. The EnKS produces a smooth transition between observations. This is particularly advantageous when the soil is drying down (see LW13, Day 169-185), and where observations are missing (Day 172-176). When observations are available daily, the EnKF and EnKS surface soil moisture estimates are very similar as the filter does not have time to drift significantly.

The smooth transition between observations in the EnKS can produce a poorer

estimate when observations are missing for lengthy periods. As there are no observations available for about 10 days beginning Day 184, both the EnKF and EnKS perform relatively poorly. At all three sites, the EnKF drifts toward the EnOL value. In the EnKS, information from the observation on Day 193 (CF08) or 194 (ER05 and LW13) is propagated backwards in time drawing the estimate for the whole period to a moister condition at CF08, and impacting the last 5 days at ER05 and LW13.

Ensemble spread is indicative of uncertainty in the estimate, which reflects the uncertainty introduced through the forcing, parameters and initial conditions. In Figure 4-3, uncertainty is highest in the EnOL estimate, as it is unconstrained by any state measurements. Between observations, ensemble spread in the EnKF estimate grows towards that of the EnOL while the spread in the EnKS estimate grows and then falls before the next observation, indicating that information from subsequent observations is being propagated backwards and reducing the uncertainty in the smoothed estimate. The ensemble standard deviation from the EnKS is always less than that of the EnKF which is in turn generally less than that of the EnOL. A notable exception is between Days 189 and 194 at CF08. During this period the soil is dry, and the range of values is constrained by the lower limit on soil moisture. As the EnOL estimate is drier than the EnKF estimate, its range is more confined, so the standard deviation is less than that of the filter. At the final time step, the standard deviation across the ensemble from the EnKS is identical to that of the EnKF by definition as both estimates have used all of the available observations.

4.4.2 Evaluation of Root Zone Soil Moisture Estimate Using Observed Latent Heat Fluxes

Our interest in root zone soil moisture is driven by its role in determining the rate of evapotranspiration. Transpiration can only occur when there is sufficient water available to plant roots, so our ability to estimate latent heat flux is inextricably linked to our ability to estimate root zone soil moisture. Our use of latent heat flux to evaluate root zone soil moisture is valid in this experiment, as this field study is

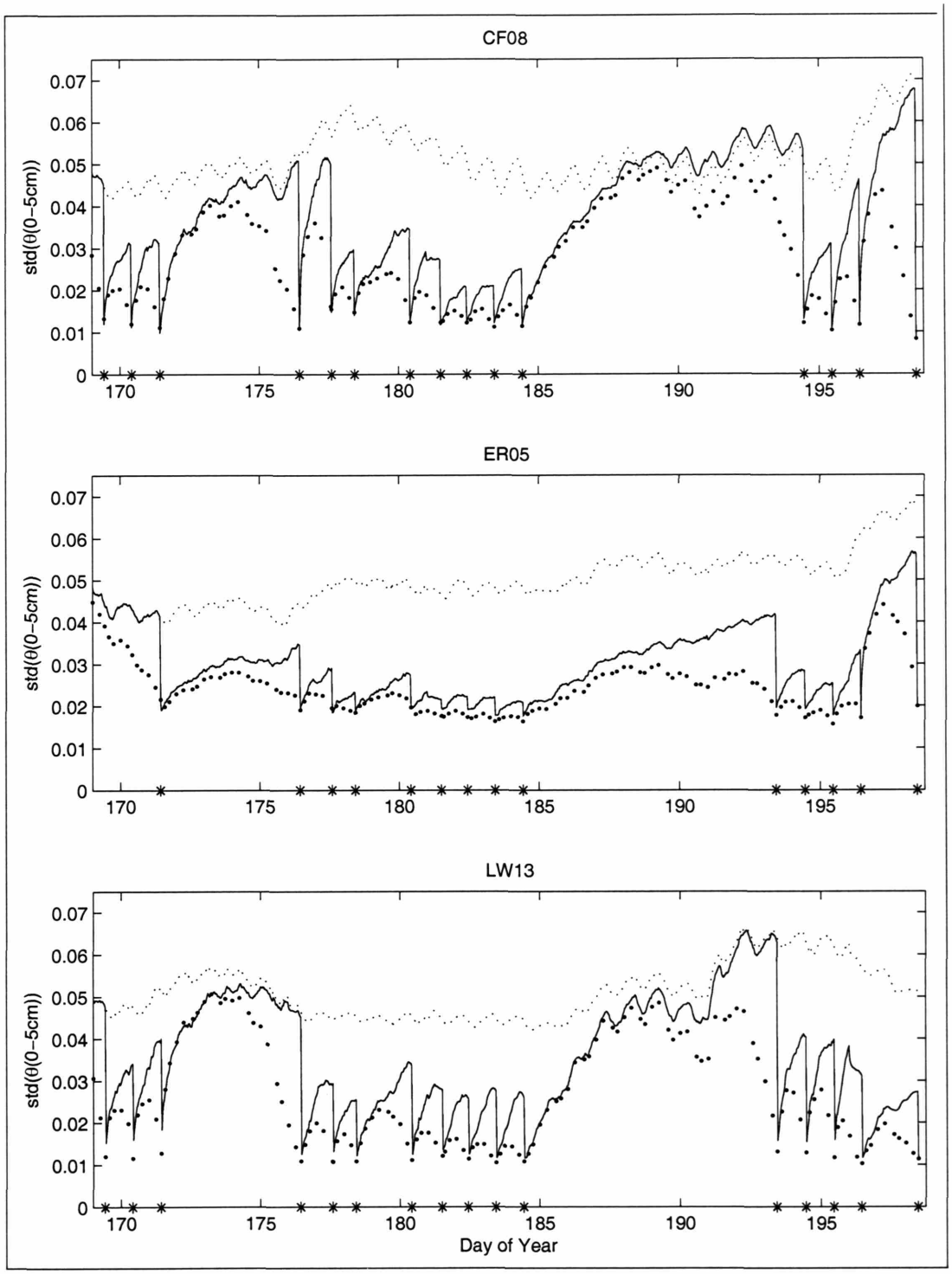

Figure 4-3: Analysis Error Standard Deviations in surface soil moisture for the EnOL (dotted), EnKF(solid) and EnKS (solid circles) at three validation sites. ESTAR measurement times are indicated by asterisks on the time axis.

characterized by lengthy periods with water-limited conditions.

The state vector has been defined to contain the soil moisture in each of the six soil layers of the model. While observations relate only to the top 0-5cm, the subsurface soil moisture is also updated through its cross-covariance with the observations and surface soil moisture, and more slowly as the impact of adjusted surface soil moisture is redistributed through the soil column. Through updating the soil moisture profile, soil temperature as well as the energy fluxes are impacted by the assimilation of radiobrightness observations.

In this section, the latent heat fluxes from the EnOL, EnKF and EnKS are compared to observations from the modified Energy Balance Bowen Ratio (EBBR) data obtained at the Department of Energy's Atmosphere and Radiation Measurement Cloud and Radiation Testbed (ARM-CART) Central Facility. This EBBR system has been operational since 1992, and has measurement uncertainty of 10% associated with it [9]. The data here are from the modified data set, which have been quality-checked and modified to provide reasonable values where the Bowen ratio is close to zero or -1, or data is missing.

The estimated surface and root zone soil moistures at this location and their corresponding standard deviations are plotted in Figure 4-4. While the EnKF estimate includes several sharp jumps in root zone soil moisture, the EnKS results in a smoother transition towards the same ultimate result. The estimate of root zone soil moisture reflects conditions at the surface. The standard deviation across the ensemble of root zone soil moisture for the EnKS is less than that for the EnKF which is in turn less than that for the EnOL, indicating that the inclusion of subsequent observations also reduces the uncertainty in the estimate of the root zone soil moisture.

The lack of event-driven soil moisture features from the EnOL is manifested in the latent heat flux estimate shown in Figure 4-5. Surface soil moisture dries from an initial value of 0.24 to a persistent value around 0.18 with diurnal variation while root zone soil moisture remains at about 0.23. This dry bias in the EnOL soil moisture results in underestimation of the latent heat flux. From Day 169 to 177, the EnKF and EnKS also underestimate the daily total latent heat flux. However, as the surface

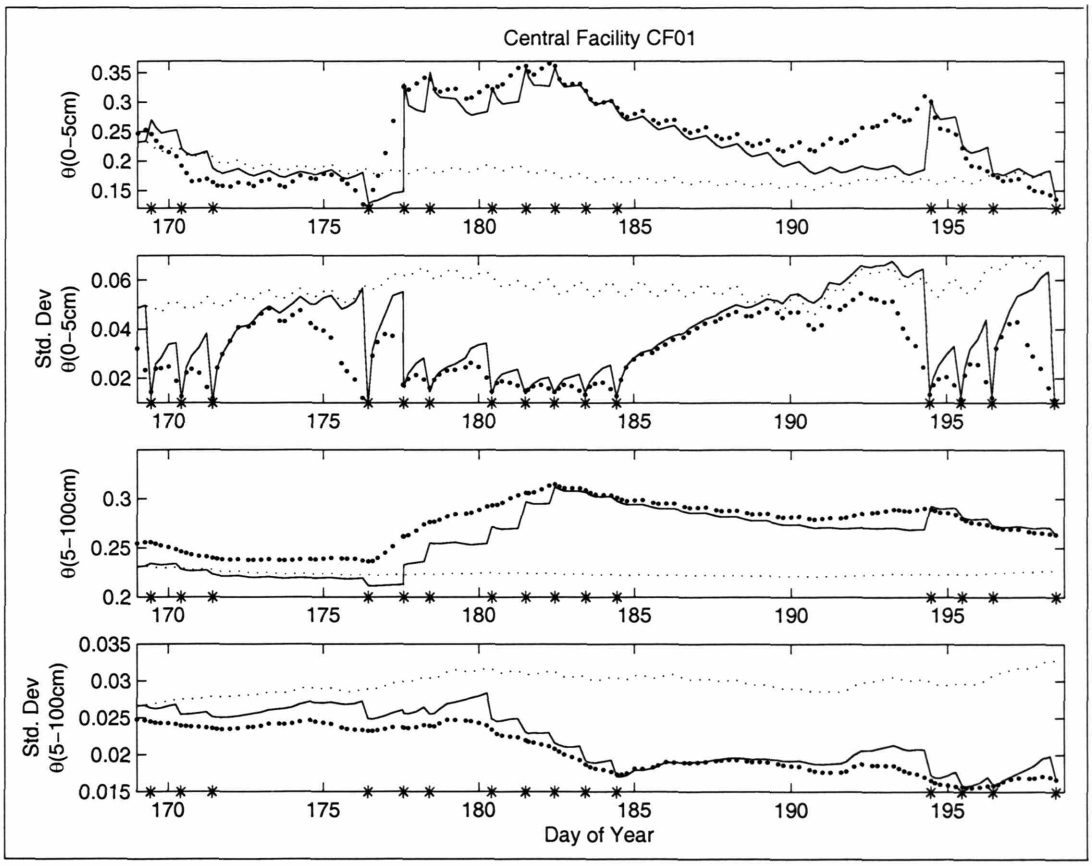

Figure 4-4: Surface soil moisture, root zone soil moisture and their corresponding ensemble standard deviations at Central Facility (CF01). The corresponding latent heat fluxes are compared to observations in Figure 4-5

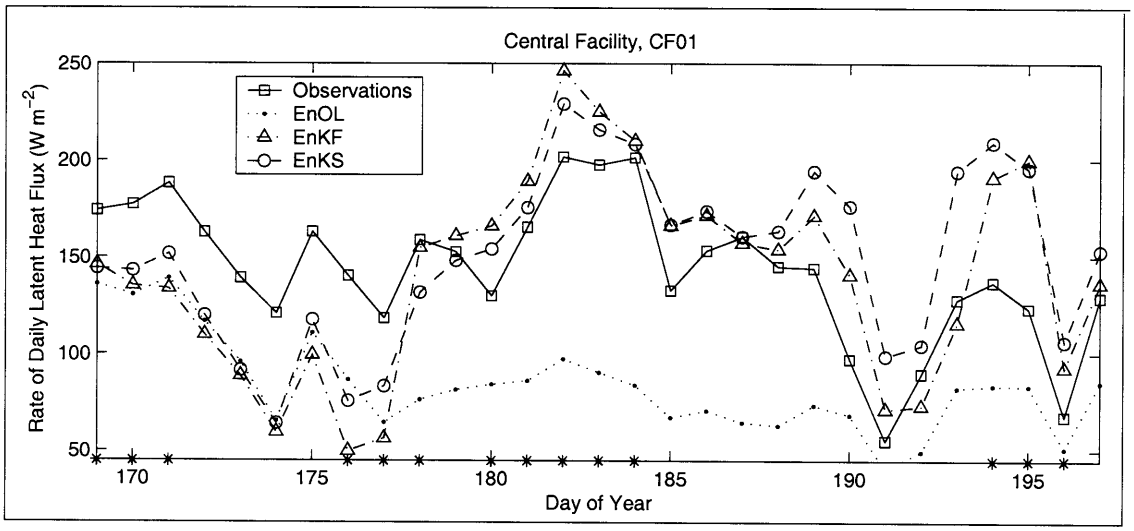

Figure 4-5: Estimated Daily Rate of Latent Heat Flux from the EnOL (dot), EnKF (triangle) and EnKS (circle) are compared to observations (square) from the ARM-CART flux tower at Central Facility CF01. Asterisks indicate days on which brightness temperature observations were available.

and subsurface soil moisture have been updated to a moister state in the EnKS, the EnKS yields an estimate closer to the observations. From Day 178 to Day 184, both algorithms overestimate the latent heat flux, but again the EnKS produces an estimate closer to the observed.

There are no observations between Days 184 and 194. From Figure 4-4, the surface and subsurface soil moisture ensembles from Day 185 to 187 are almost identical, and consequently the latent heat fluxes in Figure 4-5 are virtually indistinguishable. However, the estimates from the two algorithms soon diverge as the EnKS estimate is influenced by information propagated back from Day 194. The EnKF estimated soil moisture dries down toward the EnOL estimate, but the EnKS estimate is drawn to a moister condition by the subsequent observations, causing the EnKF to produce an estimate closer to the observations in this period. A notable exception is Day 192, when the EnKS is closer to the observations than the EnKF because there was some precipitation on this day. The effect of the EnKS updating the soil moisture to a spurious moist condition is still apparent on Days 196 and 197. The elevated latent heat flux in the EnKS from Days 188 to 195 cools the surface compared to the EnKF.

When the EnKS and EnKF update to a moister state on Day 194, one would expect the latent heat flux in the EnKS to be less than the EnKF estimate because the surface and root zone soil moisture is lower. However, because the surface is cooler in the EnKS, sensible heat flux is reduced and latent heat flux is increased. Consequently, the latent heat flux in the EnKF is closer to that observed. In a data assimilation framework based on Hydros observations, observations would be available every three days, so a scenario in which observations are missing for ten days is unlikely.

4.4.3 Surface Soil Moisture Estimation over the SGP97 Domain

The estimated volumetric soil moisture in the top 5cm of the soil column across the SGP97 domain is shown in Figure 4-6, for the first 10 estimation times in the experiment. Observations are available on Day 169.44 and 170.42. Data is missing in the northwest corner and around El Reno on Day 169.44, and around El Reno on Day 170.42.

The EnOL dries slowly from the fairly homogeneous initial condition. At the first measurement time (Day 169.44), the EnKF updates the north and center of the domain to a moister condition, while the area with sandy soil updates to a drier condition. A similar pattern is seen in the EnKS estimate, but in the EnKS information from the observation is propagated back to the first two estimation times as well (Days 169.00 and 169.25). In the northwest of the domain where observations are missing, the EnKF retains the EnOL value, while the EnKS updates to the moister state as it extracts information from subsequent observations. Between observations, the EnKF drifts towards the EnOL solution. The EnKS estimate varies smoothly between the observations across the whole domain. Spatial variability arises due to variation in soil texture and land cover. Soils with high sand content just north of the center of the domain are driest. There is more spatial variability in the southern half of the domain as soil and land cover are more variable than in the north.

Figure 4-7 shows the ensemble standard deviation in surface volumetric soil mois-

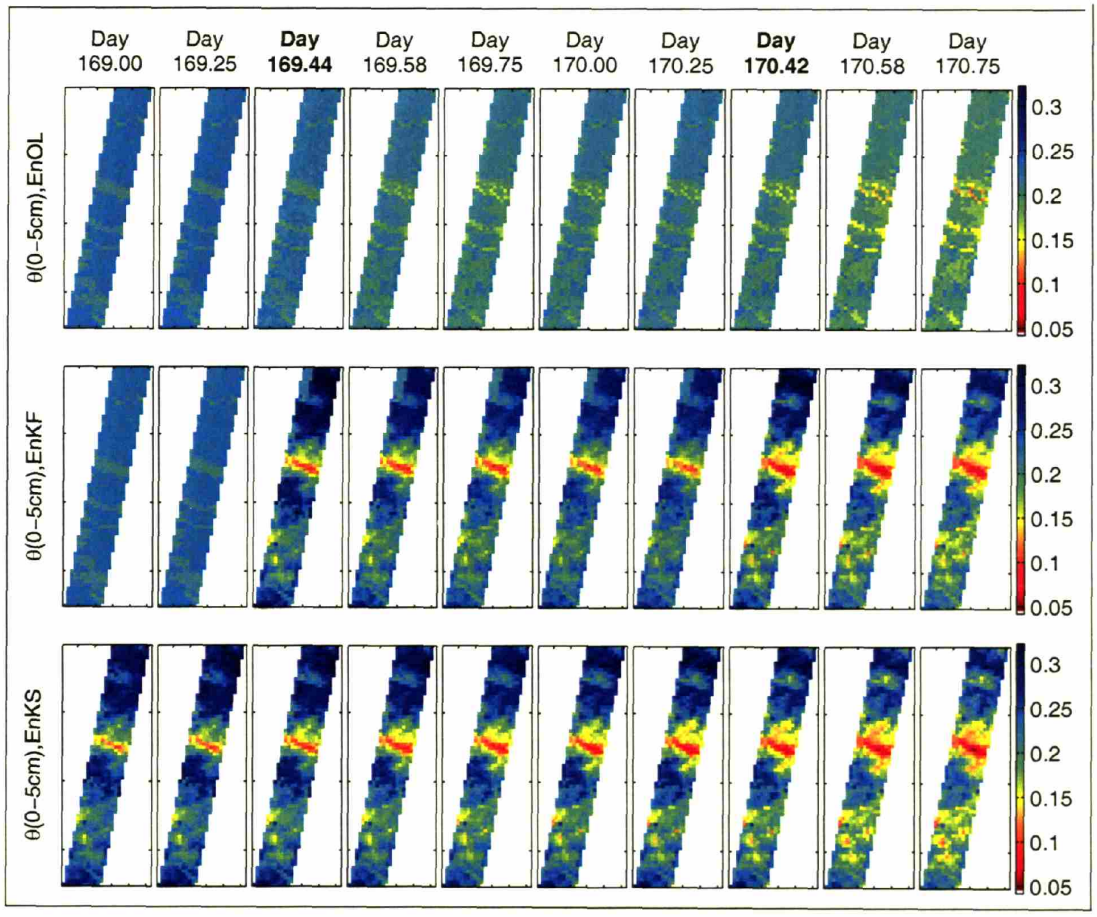

Figure 4-6: Estimated volumetric soil moisture across the SGP97 domain (Day 169.00 - Day 170.75) from the EnOL (top), EnKF (middle) and EnKS (bottom). Airborne radiobrightness temperatures were available at the times indicated in bold.

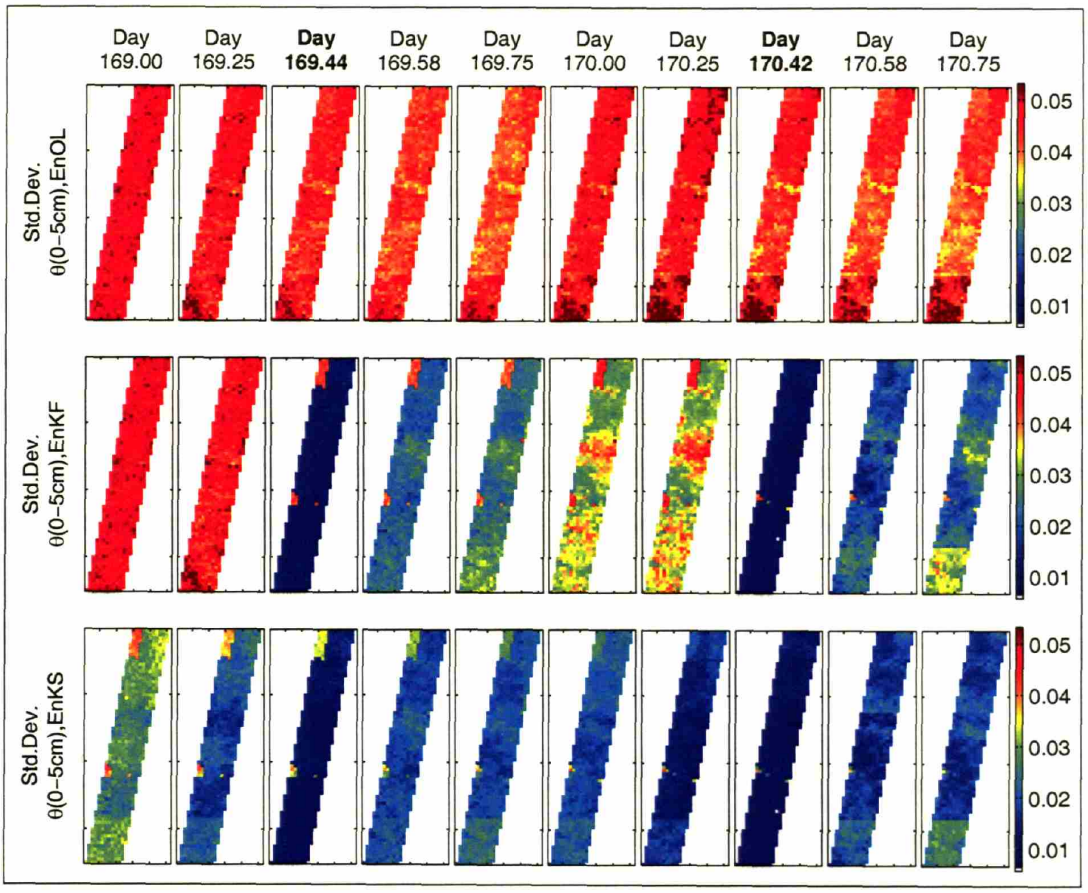

Figure 4-7: Standard deviation across ensemble of surface volumetric soil moisture over the SGP97 domain (Day 169.00- Day 170.75) from the EnOL(top), EnKF (middle) and EnKS (bottom). Airborne radiobrightness temperatures were available at the times indicated in bold.

ture for the same period as Figure 4-6. In the EnOL, the uncertainty exhibits only a slight variation during the day, but remains high across the domain. Spatial variability is limited; the outline of the pentad pixels is apparent in the south and northeast of the domain, and patterns in soil texture are just discernible. In the EnKF, the observations at Day 169.44 and 170.42 cause an abrupt reduction in ensemble spread. As observations were missing in the northwest corner on Day 169.44, the standard deviation is the same as the EnOL until the observation on Day 170.42. Between the observations, the standard deviation can be seen increasing, and the spatial variability becomes more pronounced. Ensemble growth rate varies with soil moisture, soil texture, land cover and precipitation forcing. Ensemble spread grows most where the

EnKF updated to a drier state on Day 169.44. As ensemble members continue to experience precipitation, the soil moisture in this area has a greater range of attainable values. The converse is true at the north of the domain where the filter updated to a moist condition, so there is less ensemble growth before the next observation.

In the EnKS, the ensemble spread at update times is close to that of the EnKF. Between observations, the spread grows at first, but is then reduced as information from the next observation is sent backwards. In the northwest corner where observations were missing on the first day, the standard deviation is less than that in the EnOL and EnKF because the state was updated with subsequent observations. This is also true for the first two estimation times, which preceded the first observation.

4.4.4 Root Zone Soil Moisture Estimation over the SGP97 Domain

The change in root zone soil moisture is much slower than at the surface. Figure 4-8 shows snapshots of the root zone soil moisture at five times at which observations were available. Note that the observations are 1, 6, 5 and 12 days apart. The corresponding ensemble standard deviations are shown on the right. In the EnOL, the root zone soil moisture varies little over the experiment duration. By Day 193, sandy areas have dried out considerably more than their surroundings. The ensemble spread has increased as the ensemble is not constrained by any observations. The edges of the pentad pixels are apparent, as variability in precipitation is proportional to the total precipitation. In the pentad period which includes Day 193, pentad grid cell 1 receives the least amount of precipitation and therefore variability in precipitation and soil moisture is lower than the other two pixels. In the EnKF, the root zone soil moisture has been updated through the cross-covariance between the root zone soil moisture and the observed state at the surface. Spatial features, largely due to soil texture, become increasingly apparent with time. The ensemble spread decreases slowly over the experiment as the root zone soil moisture is constrained by the inclusion of additional observations.

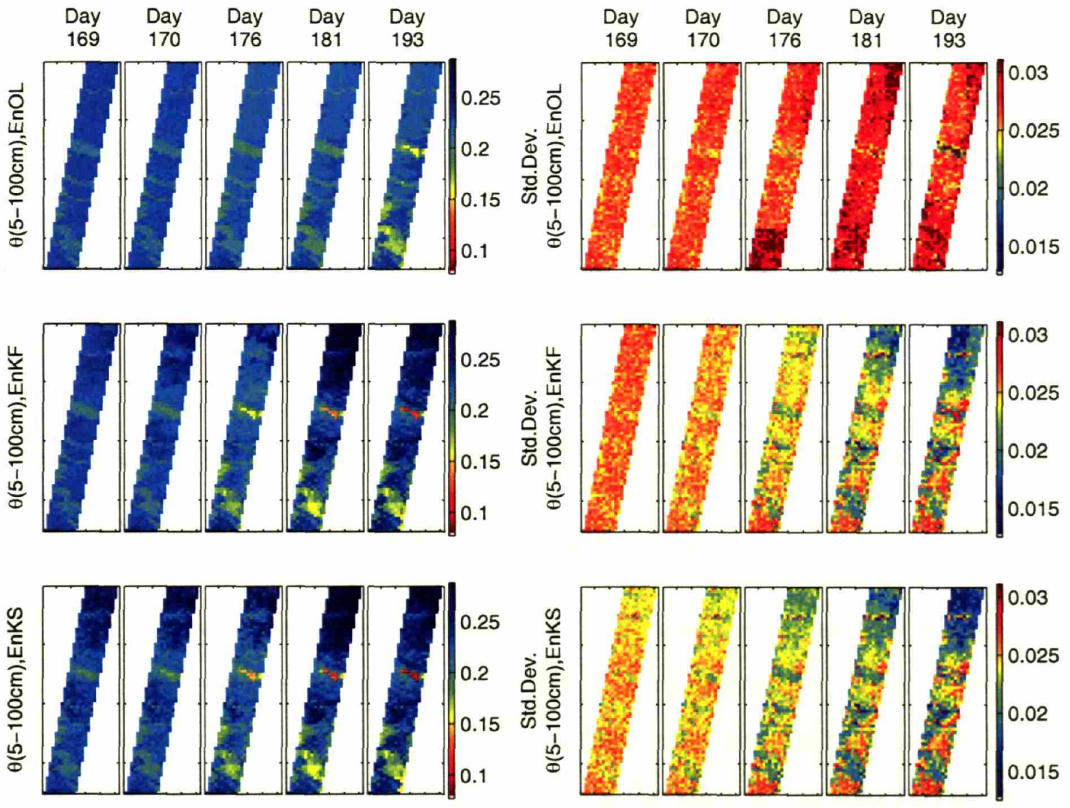

Figure 4-8: Estimated root zone volumetric soil moisture (left) and associated ensemble standard deviation (right) are shown for five measurement times during SGP97.

By the end of the experiment the EnKF and EnKS estimates are virtually identical as they have both used the same set of observations. The same is true of the standard deviation. Earlier in the experiment on Days 169 and 170, the spatial features yet to develop in the EnKF estimated root zone soil moisture are already visible in the EnKS estimate due to the backward propagation of information from observations later in the experiment. The uncertainty in the EnKS estimate is consistently less than that of the EnKF which is in turn less than that of the EnOL.

One of the most interesting features of these plots is a small area of needle-leaf evergreen trees just above the center of the domain. On Days 181 and 193, both the EnKF and EnKS show that due to a combination of sandy soils and deep roots, these cells have the lowest root zone soil moisture across the domain. This feature is discernible in the EnKS estimate as early as day 170, while it develops later in the EnKF and EnOL.

4.5 A Fixed Lag Ensemble Kalman Smoother for Operational Implementation

The EnKS can be implemented in a real-time data assimilation system alongside the EnKF. As observations become available, B(t) is calculated to perform the EnKF update to obtain the best estimate of the current state given all the available information. B(t) is then used in the EnKS to update the state in the previous estimation times of interest. The additional computational demand of the EnKS arises due to the storage or retrieval of the prior ensemble and its post-multiplication with B(t). The computational demand can be controlled to some extent by limiting the number of prior estimation times at which a smoothed estimate is obtained.

Successive corrections with B(t) produce a slight change in mean and a reduction in variance until the observations are beyond the decorrelation time of the state. Previously, results have compared the EnKF to the EnKS where each observation is used to update all of the previous states ("infinite lag"). Figure 4-9 shows the

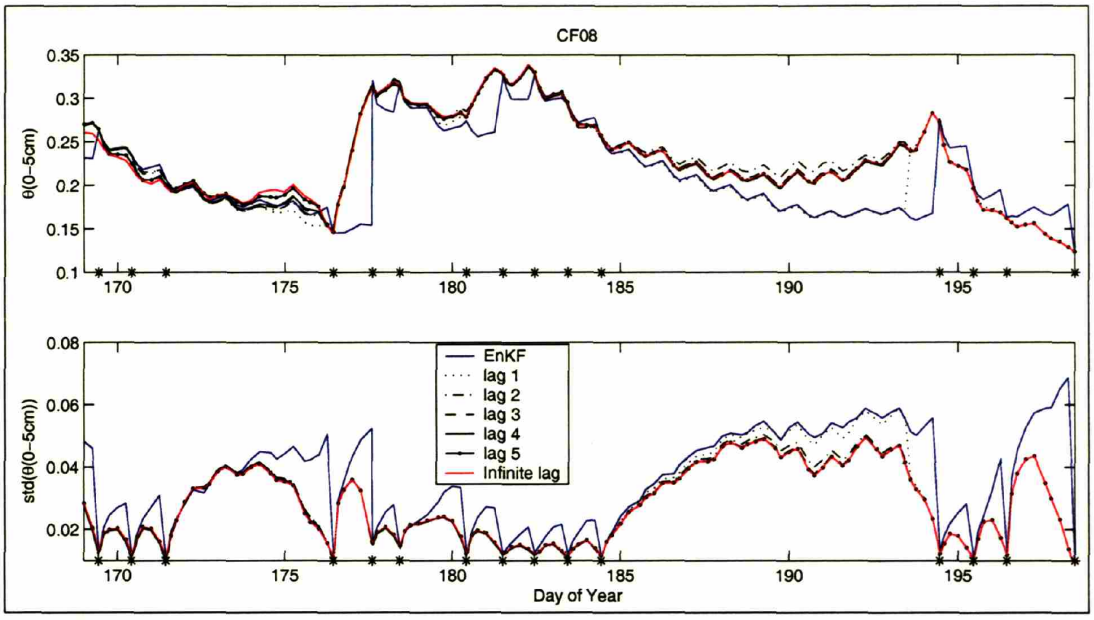

Figure 4-9: Estimated surface volumetric soil moisture (top) and ensemble standard deviation (bottom) for a fixed lag EnKS is compared to the EnKF and the EnKS using all observations, as the fixed lag is increased from one to five.

surface soil moisture and ensemble standard deviation at Central Facility CF08 using the EnKF, the EnKS with infinite lag, and EnKS with fixed lag. There is little difference between the EnKS with infinite lag and the EnKS with lag 1 (i.e. updating the states back to the last observation), both in terms of the estimate itself and the standard deviation, suggesting that the decorrelation time in surface soil moisture is quite short. Updating with the next observation produces most of the change in mean and variance. This is intuitive as soil moisture is precipitation-driven and dissipative between storm events. Including observations further into the future results in a negligible improvement.

At depth (Figure 4-10) the decorrelation time is longer than at the surface. After updating using the next five observations (lag 5), there is still a significant difference in the ensemble mean compared to the "infinite lag" case. There also remains a substantial reduction in standard deviation to be made through the inclusion of additional observations. So while a lag of 1 is sufficient for the surface soil moisture, a longer lag is needed at depth. Using the EnKS as a fixed lag smoother ([30], [8])

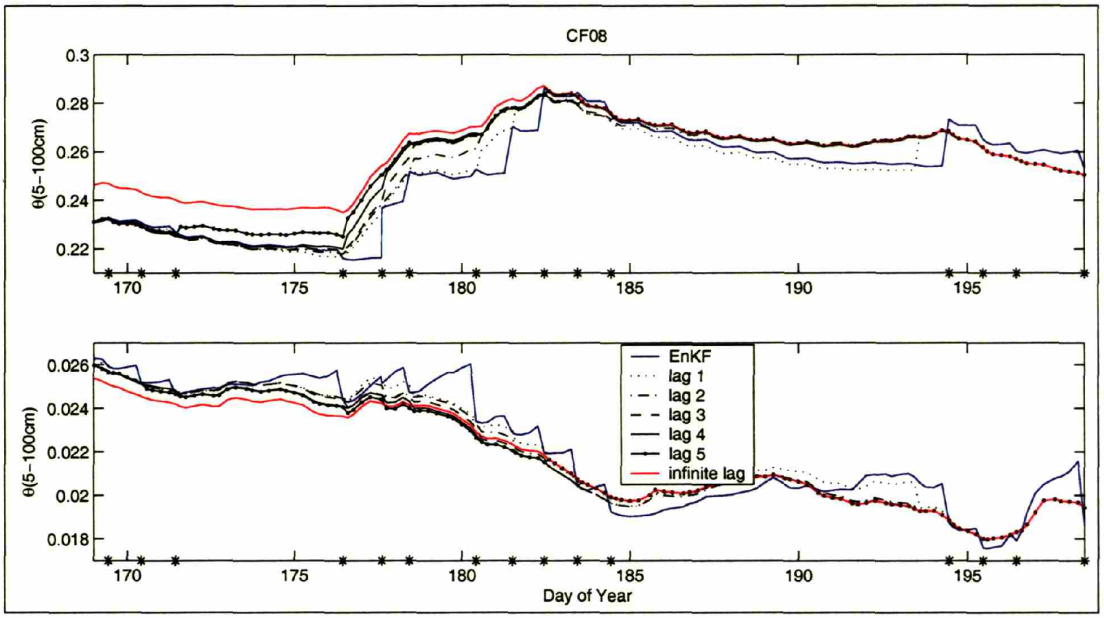

Figure 4-10: Estimated root zone volumetric soil moisture (top) and ensemble standard deviation (bottom) for a fixed lag EnKS is compared to the EnKF and the EnKS using all observations, as the fixed lag is increased from one to five.

Table 4.2: CPU requirement of algorithms relative to EnKF

Algorithm	EnOL	**EnKF**	EnKS (lag=1)	EnKS (lag=2)	EnKS (lag=3)	EnKS (lag=4)	EnKS (lag=5)	EnKS (all)
Relative Effort	0.977	**1.0**	1.208	1.471	1.655	1.874	2.068	2.832

considerably reduces the computational expense. In Table 4.2, the CPU time required by each algorithm is shown relative to that of the EnKF. Increasing the lag incurs greater computational expense, so there is a trade-off to be made between desired accuracy and cost. Nonetheless, this experiment shows that it is practicable to use a fixed lag EnKS alongside an EnKF in a real-time operational data assimilation system.

4.6 Conclusions and Discussion

This research demonstrates the feasibility of implementing the ensemble Kalman smoother (EnKS) in a soil moisture estimation problem. The EnKS is an extension of the EnKF in which observations are used as they become available to update the ensemble at prior estimation times in addition to the current forecast ensemble. The EnKS is computationally less expensive than the EnMB smoother used in Chapter 3. Its main advantage over the Ensemble Smoother [83] is that it uses the EnKF as its first guess, ensuring an estimate at least as good as the EnKF.

The EnKS is an inexpensive smoother as the most costly computational components are already performed in the EnKF. The remaining cost is that of retrieving the prior ensembles to be updated and their update through a single post-multiplication. The computational burden can be controlled by limiting the number of times at which an estimate is required. A key design parameter, therefore, is the frequency at which an estimate is needed for the application. This need not be the same as the model time step. In this application, an estimate four times daily is considered appropriate to capture the diurnal variation in the states, while the model used a half-hourly time step.

Estimated surface and root zone soil moisture were validated using gravimetric measurements at three ground truth sites, and flux tower observations of latent heat flux at Central Facility. The EnKS compared favorably to the EnKF, capturing the temporal variability in soil moisture due to storm events, while producing a smoother transition between observations. The backward propagation of information from subsequent observations also considerably reduced the standard deviation across the ensemble, indicating increased confidence in the smoothed estimate. The latent heat flux estimates show that using the EnKS to assimilate observations pertaining to surface soil moisture also impacted the terms of the energy balance. In periods with daily observations, and even occasionally missing data, the EnKS yielded an estimate closer to the observed latent heat flux. During a 10-day period without observations followed by precipitation, the EnKS spuriously raised the soil moisture at the sur-

face. This influenced the surface energy balance and resulted in a poor latent heat flux estimate, even after observations became available, demonstrating the difficulty in using a smoother if observations are not available with sufficient frequency.

Finally, the EnKS using all observations to update all states was compared to an implementation of the EnKS as a fixed lag smoother. It was found that the lag required to produce the best estimate at the surface was considerably shorter than that required at depth where there is greater memory. This experiment was 30 days long with just 16 observations. A lengthier experiment could investigate the lag required in a fixed lag EnKS for operational implementation. It would be interesting to investigate how the required lag might vary as a function of season, climatology or topography.

Chapter 5

Assimilation of multi-resolution L-band observations over the Arkansas-Red River basin using the EnKS

5.1 Introduction

The Hydros satellite mission will provide global L-band brightness temperature and microwave backscatter observations at 40km and 3km resolution respectively. A synthetic experiment is carried out over the Arkansas-Red river basin, in which "true" soil moisture is obtained from the TOPLATS (TOPMODEL-based land-atmosphere transfer scheme) model at 1km and used to generate synthetic Hydros observations. The ensemble Kalman smoother (EnKS) is used to merge these multi-resolution observations with modeled soil moisture from the Noah Land Surface Model to estimate surface and subsurface soil moisture at 6km resolution. This ensures that the physical processes which produced the "true" soil moisture are by definition different to those modeled in the EnKS framework. Consequently, this experiment simulates how data assimilation performs in real applications when the model is not a perfect represen-

tation of reality. The EnKS is an extension of the ensemble Kalman filter (EnKF) in which observations are used to update states at previous times. In Chapters 3 and 4, it was demonstrated that the EnKS provides a computationally inexpensive means to improve on the results from the EnKF, and that the limited memory in soil moisture can be exploited by employing it as a fixed lag smoother. In this chapter, it is shown that the EnKS can be used in a large problem, with a spatially distributed state vector, and spatially-distributed multi-resolution observations. The data assimilation framework built around the EnKS is used to study the synergy between passive and active observations which have different resolutions and measurement error distributions. The extent to which the design parameters of the EnKS vary depending on the combination of observations assimilated is investigated.

In Chapter 3 it was argued that soil moisture estimation is a reanalysis-type problem, and that smoothing is more appropriate than filtering. An ensemble moving batch smoother (EnMB) was used to merge synthetic L-band microwave brightness temperatures with modeled soil moisture from the Noah Land Surface Model to produce an improved estimate compared to the ensemble Kalman filter. The EnMB is a special case of the EnKF in which the state vector was distributed in time. In Chapter 4 a more computationally efficient approach was sought. The EnKS is an extension of the EnKF in which observations are used as they become available to update the ensemble at prior estimation times in addition to the current forecast ensemble. It was introduced by Evensen [26] who compared its performance to the ensemble Kalman filter (EnKF) and ensemble smoother (ES) of Van Leeuwen and Evensen [83]. The EnKS yielded identical results to the ES for linear dynamics, but produced a considerably improved estimate compared to the ES with a non-linear Lorenz equation because it started with the EnKF estimate as its first guess. The EnKS has been used in a fish stock assessment problem [34], and an operational ocean forecasting problem [2].

In Chapter 4 the EnKS was used to merge real brightness temperature data from the airborne ESTAR instrument with soil moisture modeled using the Noah Land Surface Model, demonstrating the feasibility of implementing the EnKS in a soil

moisture estimation problem. Estimated soil moisture was validated against gravimetric measurements, and the inferred latent heat flux was compared to tower observations. The EnKS captured the temporal variability in soil moisture due to storm events, producing a smoother transition between observations while the EnKF was drawn toward the ensemble open loop between observations. It was also shown that the backward propagation of information from subsequent observations reduced the ensemble spread, indicating increased confidence in the smoothed estimate. It was demonstrated that the most costly computational elements of the EnKS are already performed in the EnKF, making it a comparatively inexpensive smoother. Furthermore, the EnKS could be implemented as a fixed lag smoother. It was found that the lag required to produce the best estimate at the surface was considerably shorter than that required at depth where there is greater memory.

The objectives of this experiment are as follows:

1. Demonstrate that the ensemble Kalman smoother can be used in a large experiment.

2. Demonstrate that the ensemble Kalman smoother can be used to update a spatially distributed state vector with multi-resolution observations.

3. Use the EnKS to study the relative merits of radiometer (passive) and radar (active) observations in soil moisture estimation using data assimilation.

In Section 5.2, the Hydros OSSE is discussed in which the TOPLATS model is used to generate 1km soil moisture "truth" over the 575,000km^2 Arkansas-Red River basin. In this experiment, synthetic L-band brightness temperatures and microwave backscatter observations are generated using the Hydros Microwave Emission and Backscatter Model to simulate the observations which will be available from the future Hydros satellite mission.

Section 5.3 focuses on the details of the data assimilation component of this experiment. The EnKS is briefly reviewed, and the assimilation of multi-resolution observations is discussed. Results are presented in Sections 5.4 to 5.7. Section 5.4

looks at the estimation of surface and subsurface soil moisture evolution between two observation times as a storm passes across the basin. In Section 5.5, results are averaged in space and time to compare overall performance for combinations of observations using the EnKF and EnKS. Section 5.6 is concerned with the choice of lag when the EnKS is implemented as a fixed lag smoother. Finally, in Section 5.7 we investigate how the known dependence of measurement error on cross-swath position impacts the results from data assimilation.

5.2 Hydros OSSE

NASA's Hydrosphere State (Hydros) Satellite Mission is a pathfinder mission with the objective of providing global estimates of soil moisture at 10km resolution with a revisit time of three days [21]. Hydros will measure L-band microwave emission and backscatter from the Earth's surface using a combined passive and active sensor. A single feedhorn is shared by a 1.41 GHz passive channel and two active channels (1.26 H GHz and 1.29 V GHz). Hydros will be in a low earth orbit at an altitude of 670km, ensuring whole-earth coverage with a revisit time of 2-3 days. This antenna diameter results in a radiometer footprint of 40km (root ellipsoid area), and a radar two-way 3-dB real aperture footprint of 30km. Range and Doppler discrimination is used to subdivide the radar footprint to obtain 3km data. While radar resolution will be better than 3km over 70% of the swath, squint angle effects prevent high resolution observations at the center of the swath. In this chapter, a data assimilation framework is developed to merge synthetic passive and active Hydros observations with results from a land surface model to estimate surface and root zone soil moisture.

Crow et al. [14] presented an Observing System Simulation Experiment (OSSE) in which synthetic radiometer observations were generated over the Arkansas-Red river basin to examine the accuracy of soil moisture retrieval products. Observations were generated from May 26-June 28 1994, and were used to quantify the influence of land surface heterogeneity, instrument error and retrieval parameter uncertainty on retrieved soil moisture products. Our objective is to use the ensemble Kalman

smoother with combined active and passive observations. To this end, the modeled soil moisture from the TOPLATS model has been used to generate synthetic radar observations as well as radiometer observations. In this experiment, observations were generated for the full four-month period for which forcing data were available.

5.2.1 Arkansas Red River Basin

The Hydros OSSE site contains the Arkansas-Red river basin which stretches from (39.1^{o}N, 105.5^{o}W) in the northwest corner to (32.6^{o}N,92.7^{o}W) in the southeast, and has an area of approximately 575,000 km^2.

Figure 5-1 shows the distribution of soil texture, land cover across the Arkansas-Red River basin as well as the total cumulative precipitation during the experiment (April 1st to July 31st, 1994). As discussed by Crow et al. [14], 1km soil texture data were obtained from merged State Soil Geographic (Penn State University) products (http://www.essc.psu.edu) and land cover data are from a U. S. G. S. land cover database (http://nationalatlas.gov/landcvm.html).

Sand content varies from 10% to as high as 85% across the basin. While there is less range in the clay content, the occurence of relatively high clay deposits in otherwise sandy areas results in strong local variability in soil moisture.

The land cover classes are named in Table 5.1, which also includes the percentage of the basin covered by each type. The basin is dominated by evergreen shrub interspersed with short grass and crop in the west. In the east, there is mostly crop and some woodland or grass. Land cover is not as important as soil texture, but the fine scale variability and its effect on vegetation water content is significant in updating a spatially distributed state vector with multi-scale observations.

5.2.2 True soil moisture generation using TOPLATS model

TOPLATS (TOPMODEL-based land-atmosphere transfer scheme) is a spatially (explicitly) distributed water and energy balance model. Like most SVAT models, vertical heat and moisture diffusion in the unsaturated zone are modeled for each com-

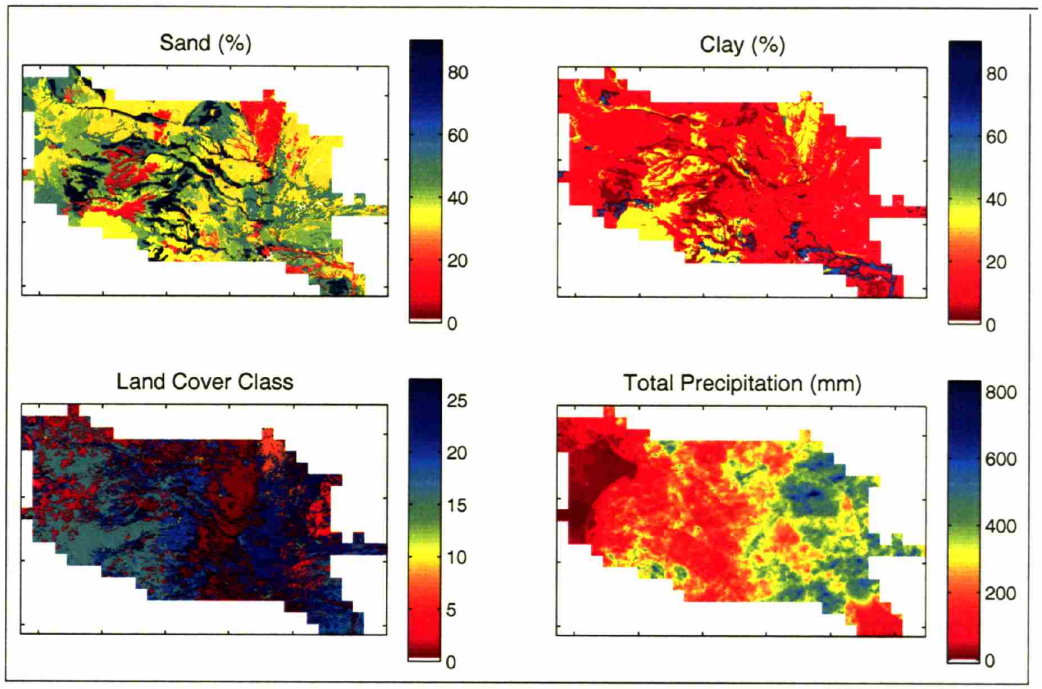

Figure 5-1: Sand content, clay content, land cover classification and total precipitation (April 1st to July 31st, 1994) over the Arkansas-Red River basin.

Table 5.1: Dominant land cover classes and percentage of basin covered

Class	Description	%
16	Evergreen Shrub	22
1	Crop/mixed farming	18
22	Crop/mixed woodland	14
20	Short grass/crop	10
2	Short grass	9
26	Short grass/mixed woodland	9
5	Deciduous broadleaf tree	4
21	Tall grass/crop	4
18	Mixed woodland	3
7	Tall Grass	2
-	Others	5

putational grid cell. However, it also allows for the lateral redistribution of moisture at depth as a function of soil transmissivity and local topography. The details of the TOPLATS model are discussed by Peters-Lidard et al. [69]. To generate the "true" soil moisture for this experiment, four layers are modeled in the water balance; 0-5cm, 5-20cm, 20-50cm and a layer from 50cm to the top of the water table, the depth of which is a state variable.

The forcing data and land surface parameters used to generate the synthetic truth are identical to those used by Crow et al. [14] and Crow and Wood [15]. The precipitation data used to force the model are from 4-km Next Generation Weather Radar (NEXRAD) precipitation data [41]. Figure 5-1 maps the total precipitation in each grid cell of the 1km modeled domain. There is a clear West-East gradient in total precipitation, which will obviously have a significant effect on soil moisture. Shortwave radiation is derived from GOES [17], and the remaining hourly hydrometeorological data are obtained from 72 National Climate Data Center (NCDC) stations in the area. Data for this experiment were provided by Wade Crow (ARS-USDA).

5.2.3 Synthetic Hydros observations

Modeled volumetric soil moisture (0-5cm), surface skin temperature and soil temperature at 5cm from the TOPLATS model are used to generate brightness temperature and microwave backscatter "measurements" at 1km using the Hydros Microwave and Emission Backscatter Model (*Chan, S., E. G. Njoku (2005) - Personal Communication*). These 1km "measurements" are aggregated to the resolution of Hydros observations (36km for radiometer brightness temperature data, 3km for radar backscatter data). This ensures that the synthetic Hydros observations include sub-pixel heterogeneity, and their use in the EnKS framework tests the retrieval/estimation approach in the presence of errors arising from such heterogeneity. Brightness temperature with polarization p is given by:

$$T_{Bp} = T_s e_p exp(-\tau_c) + T_c(1-\omega)[1-exp(-\tau_c)][1+r_p exp(-\tau_c)] \qquad (5.1)$$

where T_s and T_c are the soil and skin/canopy temperatures (K), τ_c is the vegetation opacity along the slant path at look angle θ, ω is the single-scattering albedo, and r_p is the soil surface reflectivity which is obtained from the soil dielectric constant (a function of soil moisture) using the Fresnel equations and a modification to account for surface roughness. Both vertical (V) and horizontal (H) polarization observations are generated here and assimilated in the EnKS. Where the land cover data indicate the presence of water, the 1km brightness temperatures are given by:

$$T_{Bp} = T_c \times (1 - r_p) \tag{5.2}$$

The 1km brightness temperatures are aggregated up to 36km using simple linear averaging with equal weight given to each 1km pixel. Aggregated observations at 36km resolution are perturbed with additive observation error $\sim N(0, 1.5K)$.

Total co-polarized (pp = HH or VV) radar backscatter from the surface, σ^t_{pp} is the sum of three contributions:

$$\sigma^t_{pp} = \sigma^s_{pp} exp(-2\tau_c) + \sigma^{vol}_{pp} + \sigma^{int}_{pp} \tag{5.3}$$

The dominant contribution in bare soil or low-vegetation is the soil surface backscatter, a function of soil moisture and RMS surface roughness. In the presence of vegetation, this signal is subject to two-way attenuation through the vegetation layer along the slant path. The other backscatter contributions are from the vegetation volume (σ^{vol}_{pp}) and the interaction between the vegetation and soil surface (σ^{int}_{pp}). Backscatter is influenced by the geometry and orientation of vegetation components, as well as ground slope. The greatest obstacle in retrieval of soil moisture from backscatter data is vegetation. Accurate retrievals are limited to regions where vegetation water content is less than 0.5 kg m^{-2} [80]. Where inland water occurs, σ_{hh} and σ_{vv} are assigned values of -27dB and -23dB respectively. 3km microwave backscatter in hh and vv polarizations are obtained by simple linear averaging of the 1km observations.

Multiplicative error is applied to the 3km backscatter observations as follows:

$$\sigma_{perturbed} = \sigma_{nom}(1 + \omega K_p) \tag{5.4}$$

where ω are from the distribution $N(0, 1)$. K_p is a function of the signal-to-noise ratio and cross-swath position. The average value across the swath is 0.15, and unless otherwise stated that is the value used here.

We use $10 \log_{10} \sigma_{perturbed}$ as our observations, so the observation equation may be written in terms of an additive error which is not state-dependent:

$$10 \log_{10} \sigma_{perturbed} = 10 \log_{10} \sigma_{nom} + 10 \log_{10}(1 + \omega K_p) \tag{5.5}$$

5.3 Assimilation of radiometer and radar observations to estimate soil moisture

If any land surface model could model reality perfectly, it would obviate the need for observations and data assimilation. In practice, land surface models are developed with particular applications in mind and are very much output-driven. For example, in a distributed catchment model the primary objective is to model run-off, approximating or parameterizing land-atmosphere interactions to obtain reasonable available water and energy. A land surface model developed to provide forcing to an atmospheric model needs to capture land-atmosphere interactions correctly, so run-off and lateral flow are included to provide reasonable boundary conditions. One of the objectives of this chapter is to see how data assimilation works when the forward model does not model the same processes that produced the truth. While the TOPLATS model is used to generate "true" soil moisture, the forward model in the data assimilation framework is the Noah Land Surface Model [19],[5]. This is a different set-up to most synthetic experiments in data assimilation (so-called twin experiments), in which the truth consists of a single realization from the forward model. This means that the physical processes which produced the true soil moisture are by definition

different to those modeled in the EnKS framework. The design of our experiment is closer to how data assimilation performs in real applications when the model is not a perfect representation of reality.

5.3.1 Forward Model (Noah Land Surface Model)

The Noah Land Surface Model is a 1-D soil-vegetation-atmosphere transfer (SVAT) model that computes the terms of the water and energy balance for a single unified ground/vegetation surface. A land cover and soil class is assigned to each cell, which ordinarily determines the vegetation and soil parameters respectively. In this application, the porosity, wilting point, saturated hydraulic conductivity, surface roughness, leaf area index and minimum stomatal resistance are different for each ensemble member. Uncertainty in the porosity and wilting point impact canopy resistance and evapotranspiration and together with uncertain saturated hydraulic conductivity generate variability in the rate at which water can move within the soil column. Uncertainty in the leaf area index (LAI) and roughness length (z_0) affect the aerodynamic and stomatal resistance which in turn controls the rate of evapotranspiration and the turbulent fluxes in general.

Uncertainty in air temperature, shortwave radiation, relative humidity and wind influence evapotranspiration and thus the components of the energy balance at the surface. The error distributions applied to each of the model parameters and hydrometeorological forcing variables are identical to those used in Chapter 4.

Precipitation determines the amount of water introduced to the system, a key forcing variable in both the water and energy balances. Uncertainty in precipitation is the dominant source of uncertainty in the soil moisture estimate. Uncertainty in the timing and amount of precipitation is generated here using the same approach that was used in Chapter 4.

Our synthetic experiment is designed to simulate a data assimilation framework using remote-sensing observations from the Hydros satellite platform. With a view to developing a framework that could be implemented beyond the continental U.S. the Global Precipitation Climatology Project (GPCP) merged analysis of pentad precip-

itation was used to generate precipitation data which have both temporal and spatial uncertainty. These data consist of five-day precipitation totals available globally on a $2.5^o \times 2.5^o$ grid [92].

The data are disaggregated based on the approach used by Margulis et al. [61]. For each pentad in which precipitation occurred, it was assumed that the total precipitation occurred in a single event with uniform intensity. The duration and intensity are obtained from the Rectangular Pulses Model [76], using the parameters for Oklahoma City [37]. The timing of the start of the storm is uniformly distributed in the five day interval. The total precipitation from the storm is then conditioned on the pentad total assuming that the observed precipitation is lognormally distributed with a mean of the nominal value and a coefficient of variation of 0.6. This produces a time series of hourly precipitation forcing for each ensemble member for each 2.5^o grid cell. To introduce spatial variability, the data were disaggregated from the 2.5^o grid to the model resolution (6km) using a 6-level random cascade model [33] to produce a unique spatial structure for each ensemble member. It is assumed that the spatial structure of the storm is constant for the storm duration. The soil column is modeled as four layers; 0-5cm, 5-20cm, 20-50cm and 50-100cm. The first three are prescribed to match those in TOPLATS. In TOPLATS the fourth model layer is a variable depth to the water table, so our comparisons to the "truth" are limited to the top three soil layers.

5.3.2 Ensemble Kalman Smoother

An ensemble of states (soil moisture in four layers) is propagated forward using the full non-linear Noah land surface model A. The state y(t), at time t, is a function of the state at a previous time $y(\tau)$, the model parameters α, forcing $u(t)$ and model error $\omega(t)$:

$$y(t) = A[y(\tau), \alpha, u(\tau), t, \tau, \omega(t)] \tag{5.6}$$

Every three days at 6a.m. local time, a vector of observations $z(t)$ becomes available. These observations are related to the states through the measurement model:

$$z(t) = H\mathcal{M}[y(t)] + \epsilon(t) \tag{5.7}$$

Here, $\mathcal{M}$ is the Hydros Microwave Emission and Backscatter Model (MEBM), which takes the surface soil moisture and predicts the brightness temperature and microwave backscatter which would be observed (at 6km) if this were the true soil moisture. H is the transform matrix, which is discussed in Section 5.3.3, is required to relate the states to multi-resolution observations. As mentioned in Section 5.2.3, the brightness temperatures and microwave backscatter observations are uncertain, and their respective errors are included in the vector $\epsilon(t)$. At observation times, the ensemble Kalman filter (EnKF) updates the state vector using the following:

$$Y^a(t) = Y(t) + Y'(t)M'^T(t)(M'(t)M'^T(t) + \gamma\gamma^T)^{-1}(Z(t) - M(t)) \tag{5.8}$$

where $(^a)$ indicates the "analysis" or updated state. $Y(t)$, $Z(t)$, $M(t)$ and γ are matrices, each column of which are single realization of the state vector $y(t)$, observation vector $z(t)$, predicted measurements $H\mathcal{M}[y(t)]$and observation error $\epsilon(t)$ respectively. Primed quantities indicate that the mean across the ensemble has been removed from each column.

The ensemble Kalman smoother uses the observation at time t to update the state at previous times t' too:

$$Y^a(t') = Y(t') + Y'(t')M'(t)^T(M'(t)M'(t)^T + \gamma\gamma^T)^{-1}(Z(t) - M(t)) \tag{5.9}$$

As demonstrated in Chapter 4, this may be implemented as a fixed lag smoother to exploit the limited memory in soil moisture thereby reducing the computational expense. Experiments in Section 5.6 examine the effect of increasing the lag when multi-resolution observations are used with a spatially distributed state vector.

5.3.3 Assimilating Multi-resolution observations

This problem concerns quantities on three spatial scales; the coarse (36km) resolution of the microwave brightness temperature observations, the medium (6km) scale of the state vector and the fine (3km) resolution of the microwave backscatter observations. We need to define a transform matrix H to relate the predicted measurements on the estimation scale (6km) to the observations of brightness temperature and microwave backscatter at their respective resolutions.

On the coarse scale, there are two observations $T_{BH,obs}$ and $T_{BV,obs}$. At the finer scale there are 288 observations $(\sigma_{hh,obs}(1) \ldots \sigma_{hh,obs}(144))$ and $(\sigma_{vv,obs}(1) \ldots \sigma_{vv,obs}(144))$, where they are ordered column-wise within the radiometer pixel. At the medium scale, there are 36 modeled soil moistures, also ordered column-wise within the radiometer pixel. For each model grid cell (and each ensemble member), the Hydros Microwave Emission and Backscatter Model is used to obtain the microwave brightness temperatures and backscatters associated with the modeled states $\mathcal{M}[y(t)]$.

Radiometer

The coarse scale predicted measurements of of brightness temperature are the simple linear average of the medium scale predicted measurements, so that for a single ensemble member the measurement model would be:

$$\begin{bmatrix} T_{BH,obs} \\ T_{BV,obs} \end{bmatrix} = \left[\begin{array}{ccc|ccc} \frac{1}{36} & \cdots & \frac{1}{36} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{36} & \cdots & \frac{1}{36} \end{array}\right] \begin{bmatrix} T_{BH,1} \\ \vdots \\ T_{BH,36} \\ T_{BV,1} \\ \vdots \\ T_{BV,36} \end{bmatrix} + \epsilon$$

where ϵ contains the measurement error associated with the two brightness temperature observations.

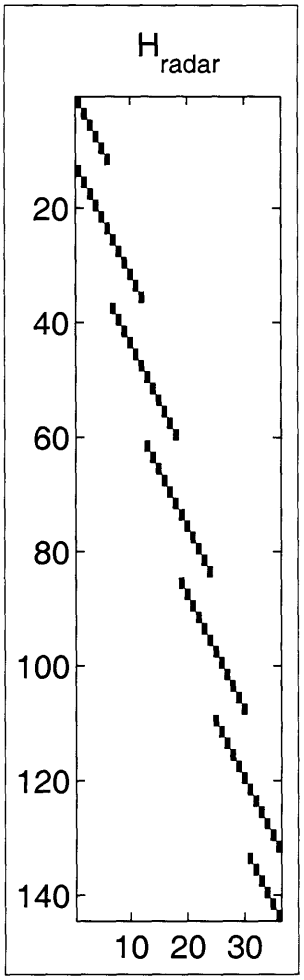

Figure 5-2: The transform matrix H_{radar} used to transform the predicted measurements of microwave backscatter to the scale of the radar observations (3km). The matrix is full of zeros, with ones where the cells are colored black.

Radar

Each 6km predicted microwave backscatter measurement is transformed into four 3km measurements using the transform matrix H_{radar} illustrated in Figure 5-2.

Combinations of Observations

As the observations are all assumed to be independent, a single block-diagonal H matrix can be constructed for any combination of observations (active alone/ passive alone/ active + passive).

Inland Water

Radar observations are omitted if the model grid cell is classified as water and therefore not included in the estimation, or if the radar pixel is partly or entirely covered by inland water. If the model grid cell is classed as inland water, the Hydros MEBM is used to generate the brightness temperature from the water to ensure that it is represented in the aggregated brightness temperature predicted measurement.

5.3.4 Evaluation of ensemble estimate using RMSE

A "good" estimate is one that results in a low root-mean-square-error (RMSE), which is defined here as the root mean squared difference between the ensemble members and the known truth. It is a measure of how the ensemble is spread around the truth. It is a useful metric as it includes both deviation between the mean and the truth and the ensemble spread in a single value. The RMSE is low if all ensemble members are close to the truth, therefore producing an ensemble mean which is close to the truth. If the ensemble spread is very large, but the mean happens to coincide with the truth, the RMSE is high because the ensemble mean happens to be right, but is highly uncertain. Similarly, if all ensemble members are clustered at a single value which is substantially different from the truth, the RMSE is high.

5.4 Results: Individual estimation times

Figure 5-3 shows the evolution of surface volumetric soil moisture as a precipitation event passes across the north of the domain. The top row contains the true soil moisture at four sequential estimation times from 2pm on Day 183 to 6am on Day 184. Radiometer (36km) and radar (3km) observations were available over the whole basin at 6am on Day 184.

The true soil moisture time series indicates that a precipitation event has moved across the northern boundary of the basin. At 6pm (Day 183), there are two small extremely wet areas in the north west of the domain. By 12am (Day 184), these areas

have dried somewhat, but the storm has clearly travelled east. By 6am, the storm has crossed to the north east of the domain. During this period, soil moisture across the remainder of the basin is largely unchanged.

The second and third rows show the estimated surface soil moisture at these times from the EnKF and EnKS respectively where only passive observations are assimilated. As observations are available only every three days, the EnKF estimate at non-update times is obtained by propagating an ensemble of realizations forward using the land surface model. Clearly, between observations the estimate has drifted away from the truth. The western half is not as dry as the truth, and the eastern half is too dry. When observations become available, the estimate is drawn back towards the truth. The estimate at the update time appears patchy because the passive observations are at coarser resolution than the estimation grid cells. The effect of each radiometer observation is to update the estimation cells therein to an average condition. As a result, using passive observations alone cannot capture the finer scale variability apparent in the true soil moisture. This patchiness is also apparent in the EnKS. Smoothing provides an improved estimate across most of the domain. Consistent with previous results, the smoother assumes a smooth transition between observations, and so performs best where there has been a gradual drying of the soil column. In the area which experienced the storm, the information on the moist state at the observation time is propagated back in time. The smoother incorrectly assumes that the state has evolved smoothly to this moist state since the previous observation and yields a spuriously moist soil condition for the previous three time steps.

In the last two rows both passive and active observations are assimilated. The fourth and fifth rows are results from the EnKF and EnKS (lag 2) respectively. As in the passive only case, there is little difference between the EnKF and EnKS results at the update time. Both capture the finer resolution variability in soil moisture better than the estimate using passive observations alone. The patchiness apparent in the radiometer results has been removed. The finer resolution observations capture the extremely wet conditions in the north and the very dry conditions in the west and center of the domain. As in the passive-only case, the EnKS generally yields an

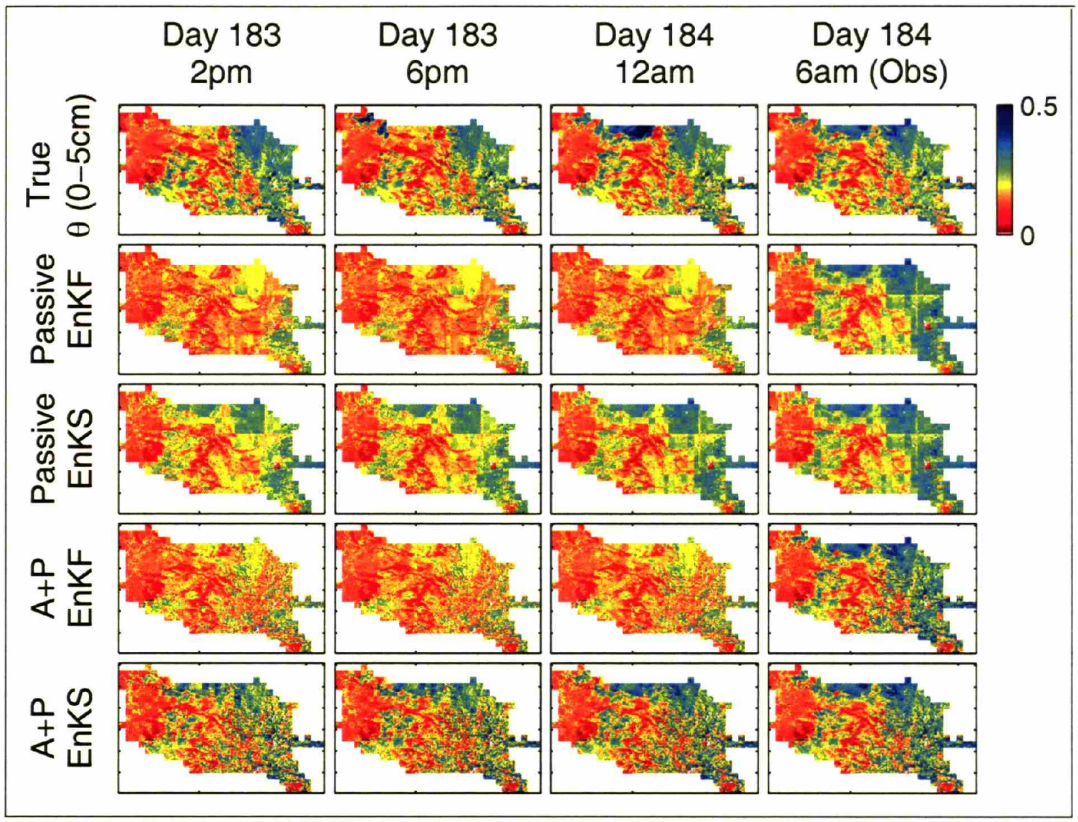

Figure 5-3: Surface volumetric soil moisture is estimated for four consecutive estimation times while a storm event crosses the north of the domain. Observations were available at Day 184.25. Estimates shown were obtained using the EnKF and EnKS with either passive observations alone (rows 2 and 3), or combined active and passive observations (rows 4 and 5).

improved result compared to the EnKF, particularly where there has been a smooth drydown between observations.

Figure 5-4 illustrates the capacity of the EnKS to estimate the soil moisture at a single observation time (Day 184.25), at various depths in the soil column through the assimilation of different combination of observations. At an observation time, there is minimal difference between the filter and the smoothed estimate, so only the EnKS solution is shown here. Assimilation of passive only, active only, and combined active and passive observations are compared to the known "true" soil moisture profile in the left-most column. This demonstrates that the soil moisture in the top two layers is well captured, but observations have little effect at depth. This is because the model

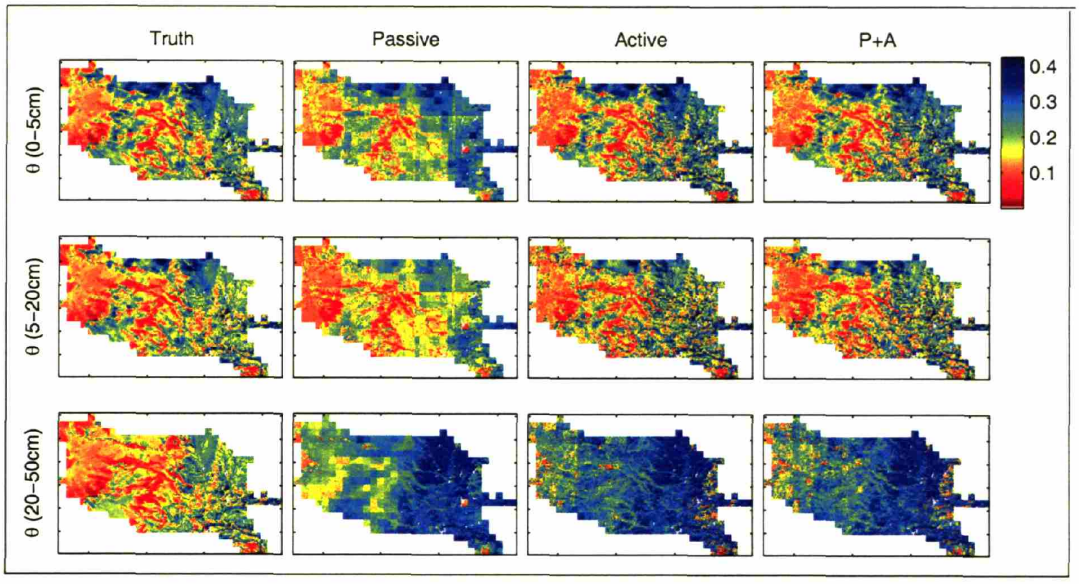

Figure 5-4: Volumetric soil moisture in each of the top three soil layers at Day 184.25 is estimated using the EnKS with lag 2, assimilating passive observations only (column 2), active observations only (column 3) and combined passive and active observations (column 4). Estimates are compared to the synthetic truth (column 1).

physics in the "true" model and the assimilation model are very different at depth. In order to estimate the non-observed states, there must be some correlation with the observations and the modeled states. If the model does not adequately reflect reality, it is difficult to extract any information from the observations through the Kalman update. It is clear that there is little discernible difference between using active alone or active with passive, particularly at the surface.

5.5 Summary Results

In the previous section, we have studied the performance of the EnKF and EnKS with passive/ active /passive+active observations estimating soil moisture at consecutive estimation times. We have seen the limitations of data assimilation using even high resolution data if the model does not adequately represent the evolution of the true state. In this section, rather than looking at just four snapshots of the soil moisture state, results are obtained for the whole four-month long experiment over the whole

domain. The objective is to make a more general statement regarding the estimation of soil moisture using the EnKS given the availability of multi-resolution L-band observations.

5.5.1 Temporally averaged results

In Figure 5-5 the image at the center maps true surface (0-5cm) volumetric soil moisture averaged across all 481 estimation times of the four month long experiment. Clearly, soil moisture is a strong function of soil texture. The extremely dry areas correspond to areas of high sand content, and conversely soil in the wettest areas has highest clay content. The top row shows $\theta(0-5cm)$ as estimated using the EnKF assimilating passive observations only (left), or combined passive and active observations (right). The third row shows $\theta(0-5cm)$ estimated using the EnKS for the same combination of observations. There is little difference overall between active alone and active with passive, so the active alone case is not shown here.

Using passive observations only, EnKF and EnKS estimates are patchy due to the coarse resolution of the observations. Inclusion of active observations produces an estimate in which finer scale features are captured. Both the EnKF and EnKS produce an estimate which reflects the soil texture map. The EnKS prevents the estimate from drifting away from the truth between observation times, and so produces a better estimate between observation times. This is particularly noticeable in the eastern half of the domain in areas of elevated soil moisture and in the extremely dry sandy areas in the west.

The histograms in Figure 5-6 illustrate that inclusion of radar observations reduces the largest errors considerably. While the passive (36km) observations update each of the estimation cells towards an average value across the 36km cell, the active (3km) observations capture fine scale spatial distribution of soil moisture. This is true in both the EnKF and EnKS results. The results from the EnKS are better than the EnKF because it prevents the estimate from drifting between observations. Recall that observations are only available at every 12th estimation time (i.e. every 3 days).

Figure 5-7 shows the RMSE averaged over all estimation times for each 6km esti-

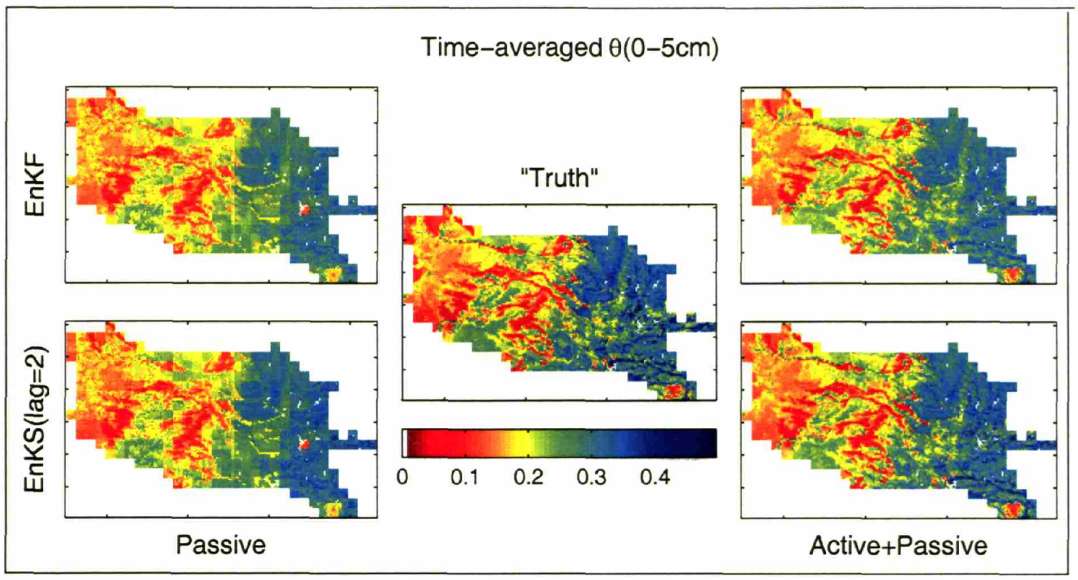

Figure 5-5: Estimated surface volumetric soil moisture from the EnKF (top row) and EnKS (bottom row) is averaged across all estimation times and compared to the truth (center). Assimilation of passive observations alone is shown on the left, active observations alone in the center), and combined active and passive observations are shown on the right.

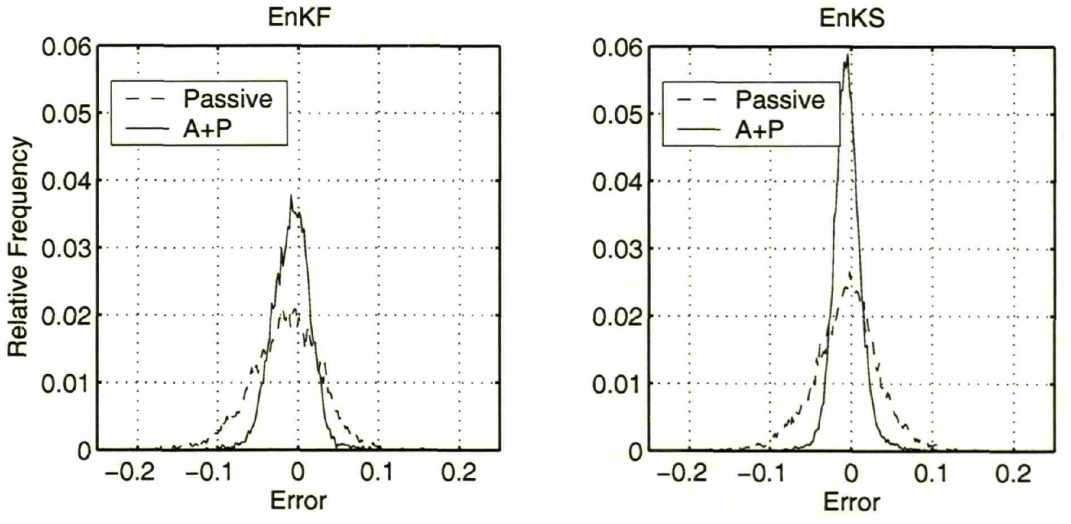

Figure 5-6: Relative frequency distribution of "error" defined as the difference between time-averaged estimated $\theta(0-5cm)$ and the truth in each of the 13163 estimation pixels. 200 bins were used.

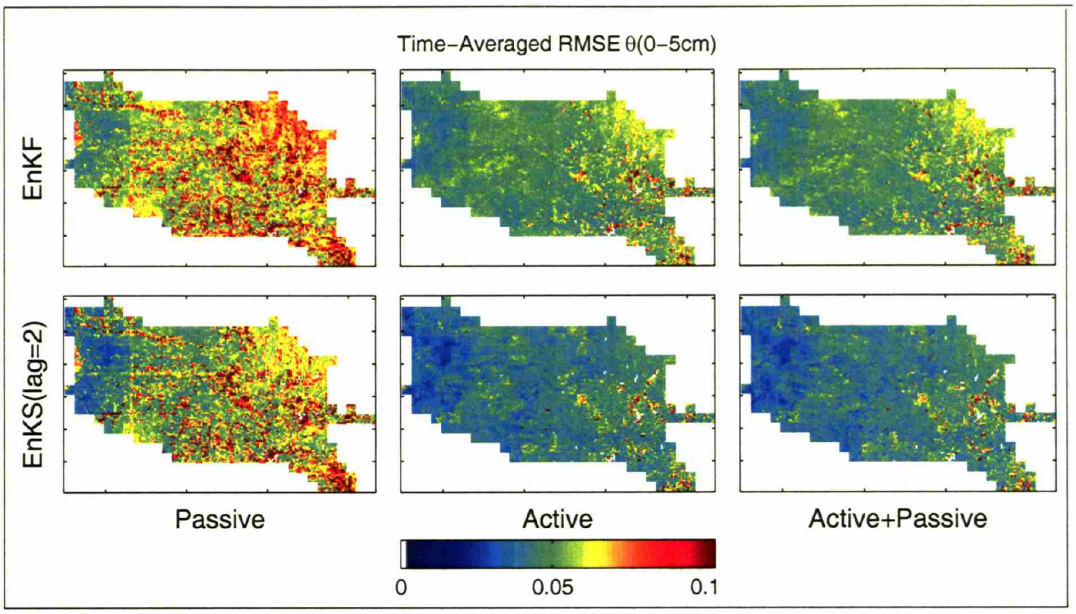

Figure 5-7: Time-averaged RMSE in estimated volumetric soil moisture estimated using the EnKF (top) and EnKS (bottom), assimilating passive observations only (left), active observations only (center) and combined active and passive observations (right).

mation pixel using the EnKF and EnKS after assimilation of different combinations of observations. Errors are greatest when only coarse resolution passive observations are assimilated. There is a huge reduction in RMSE across the domain when 3km radar observations are assimilated either alone, or combined with passive observations. In all cases, the errors are smallest in the western half of the domain where there is least spatial variability. The RMSE is almost uniform across the domain when radar corrects for spatial variability. The highest errors when radar observations are included are in areas containing some water (in the east).

For any choice of observations, use of the EnKS results in a general reduction in RMSE. However, where only passive observations are assimilated, the EnKS cannot correct for the lack of information on spatial variability. As a result, RMSE remains high, particularly in the eastern half of the domain.

5.5.2 Spatially averaged results

Figure 5-8 shows the average volumetric soil moisture in each of the top three model soil layers, calculated across the whole domain for each estimation time during the experiment. Results shown were obtained using the EnKS with lag 2. Figure 5-9 shows the spatially-averaged RMSE for the same quantities.

For the first 10 days, true soil moisture in each of the three layers is high, so there is little downward movement of water in the soil column. In the EnKS estimates (for any combination of observations), the spurious dry condition at depth results in excessive infiltration between observations during this period. After 10 days the soil moisture at depth is closer to the true soil moisture, improving the estimated surface soil moisture in all three cases. It is clear from Figures 5-8 and 5-9 that when averaged over the whole domain, all three combinations of observations can capture surface soil moisture quite well. While there appears to be little difference between the estimates, recall that these are values averaged across the whole domain. Although the radiometer data can capture the mean at 36km quite well, they cannot reproduce the true spatial variability as discussed in Section 5.5.1.

In the second soil layer, using only passive observations seems to produce the estimate closest to the truth, particularly between days (105-160). In this period, the RMSE in $\theta(5-20cm)$ does not exhibit any rise/fall due to observations, in any of the three cases. Without observations, the states evolve as they would in an ensemble open loop, with the passive only case closer to the truth because it is closer on Day 105. In Section 5.6, it will be shown that increasing the lag in the combined active/passive case yields a considerable improvement in the estimate in this interval.

It is clear from both Figures 5-8 and 5-9 that the observations have no impact on the estimated soil moisture in the third soil layer (20-50cm) beyond the first 10 days. While the true soil moisture at depth decreases with time, the estimated soil moisture is essentially unchanged. The RMSE grows unabated from Day 112 to the end of the experiment. This suggests that if the forward model does not adequately model the unobserved states, allowing the true correlations between the observed and

unobserved states to evolve correctly, a data assimilation approach will be unable to extract information from the observations to update the unobserved states.

Figure 5-9 shows the average RMSE in the estimated volumetric soil moisture for each of the top three model layers estimated using the EnKS (lag 2) to assimilate combinations of active and passive observations. As both the passive and active observations are functions of soil moisture in the top 0-5cm, the impact of observations is most apparent in the surface layer. In all three experiments, the RMSE rises after each observation but the impact of the smoother causes the RMSE to fall again before the next observation time as information is propagated backwards. In general, the RMSE in the passive only case is highest, and the RMSE in the active alone and active with passive case are very similar and less than passive observations alone. The finer resolution of the active observations allow the EnKS to capture the spatial variability at finer scales.

5.6 Impact of smoother lag

Figure 5-10 shows the impact of increasing lag on the estimated volumetric soil moisture estimate in each of the top two soil layers. A lag 0 corresponds to the EnKF case as observations are used to update the current state only, and are not used to update the state at previous estimation times. Results are shown for active observations alone, passive observations alone and combined active/passive observations.

In both soil layers, it is noteworthy that the RMSE when only passive measurements are assimilated is always higher than active alone or combined active/passive regardless of any increase in lag. For any choice of lag, the results for active alone and combined active/passive are very close.

In each combination of observations, the greatest improvement in RMSE in both soil layers is achieved when the EnKS is used rather than the EnKF (lag 0). In the radiometer case, increasing the lag improves the estimate in both layers, albeit marginally. In the surface layer, the active alone and active with passive cases are very similar. Increasing the lag from 1 to 2 results in an additional decrease in RMSE,

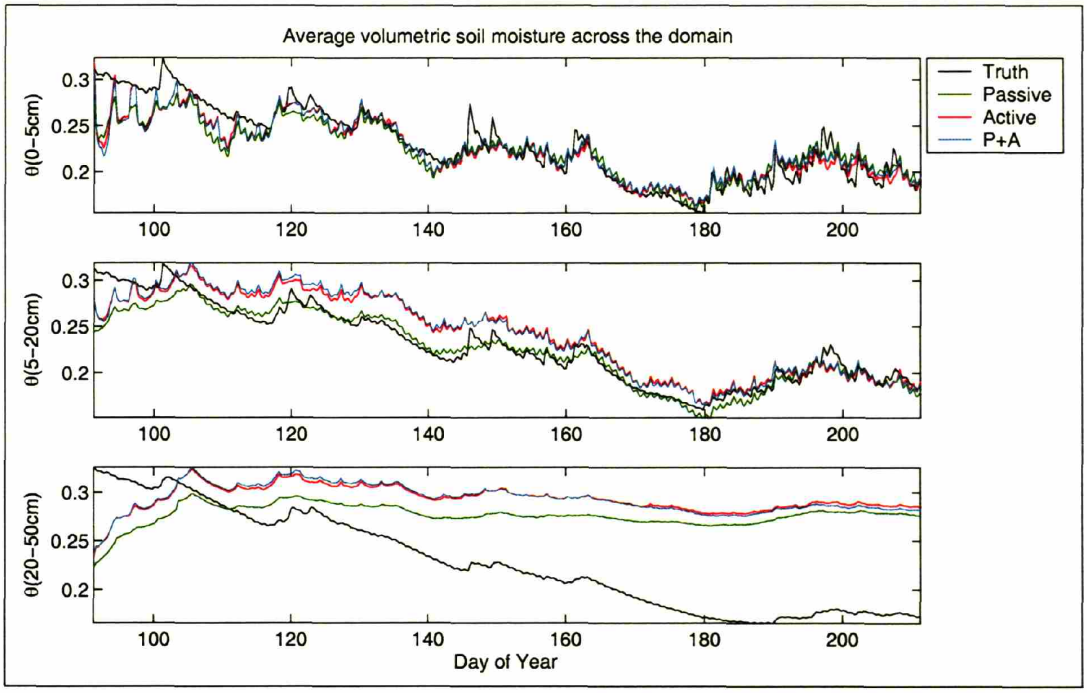

Figure 5-8: Estimated volumetric soil moisture in each of the four modeled soil layers is averaged across the whole Arkansas-Red River basin and compared to the synthetic truth. Solid and dashed lines indicate quantities estimated using the EnKF and EnKS respectively. Green, red and cyan indicate assimilation of passive observations alone, active observations alone and combined active and passive respectively.

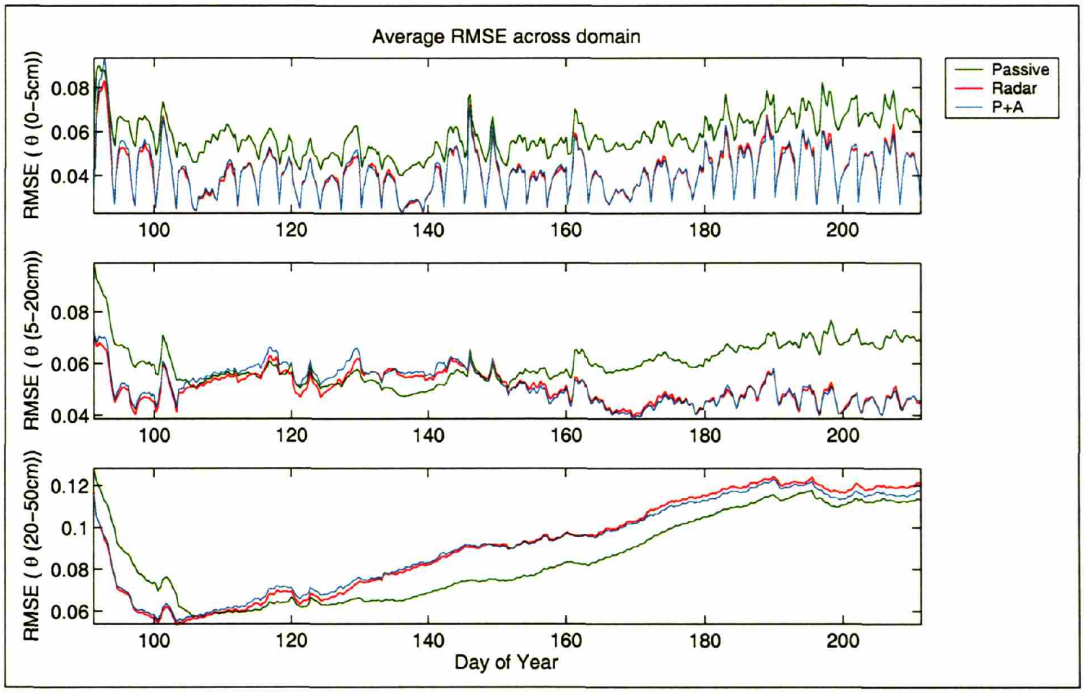

Figure 5-9: At each estimation time, RMSE is averaged across the whole Arkansas-Red River basin. Solid and dashed lines indicate use of the EnKF and EnKS respectively. Green, red and cyan indicate assimilation of passive observations alone, active observations alone and combined active and passive observations.

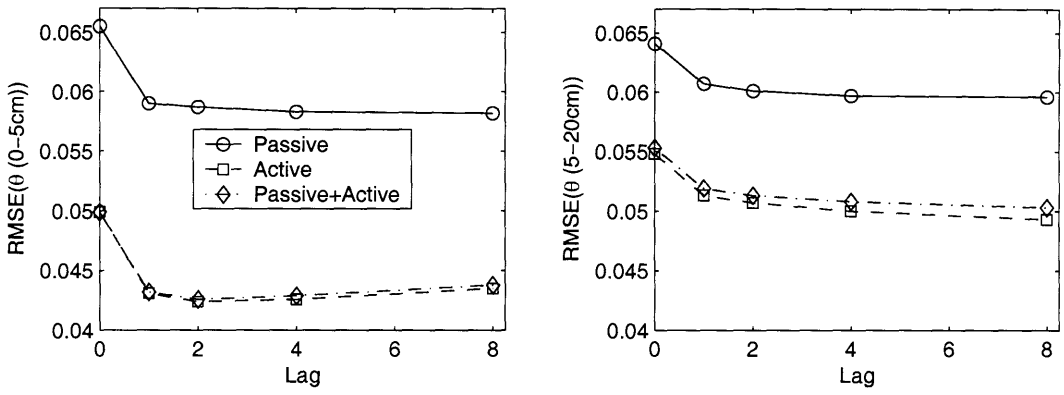

Figure 5-10: Averaged RMSE across all estimation times and all 6km estimation pixels is plotted as a function of increasing lag for volumetric soil moisture at the surface (left) and in the second soil layer (right). A lag of 0 implies that the EnKF was used without using the EnKS.

but further increases make the estimate marginally worse. With increasing lag, the difference between the active alone and the active with passive case increases. In the layer (5-20cm), increasing the lag continues to reduce the RMSE and increases the difference between using active alone or active with passive. It is also noteworthy that the RMSE in the two soil layers are so close in the radiometer only case, but very different for the active alone and the combined active/passive case.

Figure 5-11 maps the difference in RMSE (averaged over all estimation times) between the EnKS estimate using lag 8, and lag 1. A negative value indicates that increasing the lag to 8 improves the estimate. Including additional observations in the passive only case resulted in a modest improvement in $\theta(0-5cm)$ and $\theta(5-20cm)$. Here, increased lag causes no change in western half of the domain, some improvement in the east at the surface, and an improvement across the domain at depth. RMSE increased in some cells, particularly in the east where there is more precipitation do disrupt drydown and more spatial variability than in the western half of the basin.

The greatest improvement occurs in $\theta(5-20cm)$ when active observations are used, either alone or with passive observations. The areas which improve correspond to those with drier conditions, i.e. the center and west of the basin. This is consistent with results in Sections 5.4 and 5.5, as well as Chapters 3 and 4 which have shown

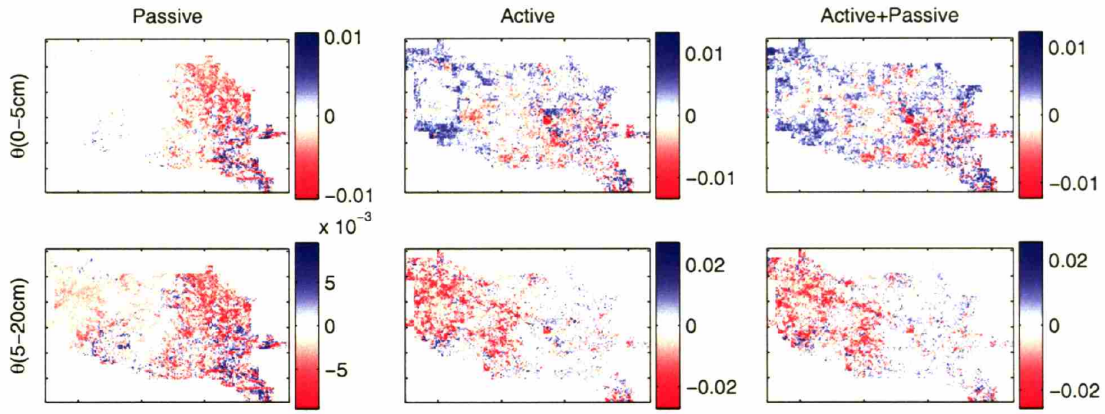

Figure 5-11: Difference in RMSE (averaged over all estimation times) between the EnKS estimate using lag 8, and lag 1. A negative value indicates that increasing the lag to 8 improves the estimate.

that the EnKS and indeed the EnMB perform best when the soil is drying down.

Figure 5-12 shows the impact of increasing the lag of the EnKS on the RMSE of the estimate of $\theta(0-5cm)$ and $\theta(5-20cm)$ as a function of time. In the surface soil moisture, there is generally little improvement made by increasing the lag beyond 1. The only time when this results in an improvement is at the very beginning of the experiment. Recall that the surface layer was drying down too quickly due to excessively dry conditions at depth. In the layer 5-20cm, increasing the lag in the passive only case has little effect, except in the first 15 days or so, where the estimate is improving from the spurious initial conditions. In both the active and active with passive case, increasing the lag clearly reduces the RMSE particularly between days 130 and 180. Recall that during this period, the radiometer-based results were closer to the truth than the radar-based results. Propagating observations backwards in time during this interval yields an improved result.

5.7 Impact of Observation Error (K_p)

The radar two-way 3-dB real aperture footprint is 30km, so range and Doppler discrimination is used to subdivide the footprint producing data at 3km. While radar

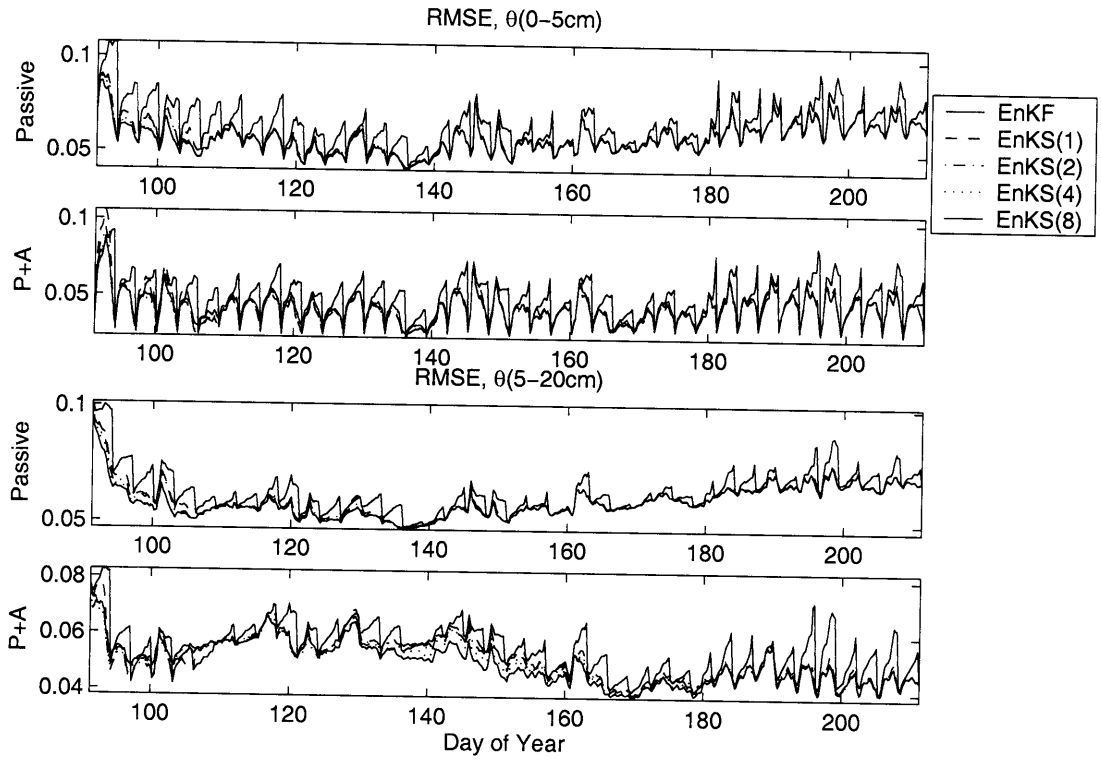

Figure 5-12: RMSE in surface (top three panels) and second layer volumetric soil moisture (lower three panels), is averaged across the whole Arkansas-Red River basin, to show how the impact of increasing lag affects RMSE as a function of time.

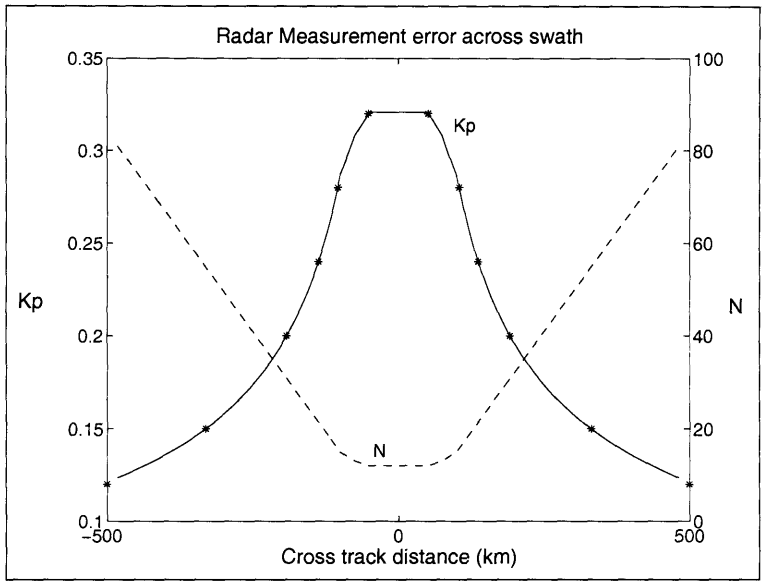

Figure 5-13: Number of looks per cell (dashed) and resultant K_p (solid) as a function of cross-swath position. Asterisks indicate the values of K_p for which results are shown in Figure 5-14

resolution is better than 3km over 70% of the swath, squint angle effects prevent high resolution observations at the center of the swath.

Cells at the edge of the swath are "seen" more frequently than cells at the swath center as illustrated in Figure 5-13. As K_p is a function of the number of looks per cell:

$$K_p = \frac{1}{\sqrt{N}}\left(1+\frac{2}{SNR}+\frac{1}{SNR^2}\right)^{\frac{1}{2}} \tag{5.10}$$

K_p decreases with distance from the center of the swath (the signal-to-noise ratio, SNR is $+9[dB]$). Consequently, the multiplicative measurement error on the backscatter observations is a function of position with the the swath. In this section, we want to examine how the known dependence of measurement error on position in the swath impacts the results from data assimilation. From Section 5.5, it is clear that the RMSE varies in space and is highly dependent on incident precipitation, and soil texture in particular. To isolate the effect of K_p alone, everything else being equal, the experiment was repeated assuming a single value of K_p across the whole domain

for the four-month experiment duration. This ensures that statistics are calculated across varying soil moisture conditions and land surface parameters. For each of the six values of K_p represented by an asterisk in Figure 5-13 the EnKS with lag 2 was implemented using active observations alone, and combined active/passive observations. Each data point in Figure 5-14 represents the average of the RMSE calculated across all 13163 estimation pixels and 481 estimation times.

RMSE in $\theta(0-5cm)$ increases steadily as K_p is increased for both choices of observations. For low K_p, including passive observations makes estimation marginally worse than using active observations alone. Beyond about K_p=0.2, the radar observations are sufficiently poor that inclusion of radiometer data improves the estimate. This indicates that including passive observations is most significant towards the center of the swath where K_p is high.

For $\theta(5-20cm)$, it is also true that beyond K_p=0.2, the inclusion of radiometer results improves the estimate. However as $\theta(5-20cm)$ is not observed, it is not as sensitive to the observations as surface soil moisture which is observed. In fact, not only is there little net difference between using $K_p = 0.12$ or $K_p = 0.32$, but RMSE is seen to initially decrease with increasing K_p.

5.8 Conclusions And Discussion

The EnKS is an extension of the EnKF in which observations are used as they become available to update the ensemble at prior estimation times in addition to the current forecast ensemble. Because it takes the EnKF as its first guess, it ensures an estimate at least as good as the EnKF. In this chapter, the EnKS has been implemented in a synthetic experiment over the 575,000 km^2 Arkansas-Red River basin to estimate surface and root zone soil moisture through the assimilation of multi-resolution observations.

Either passive (36km) obervations alone, active (3km) observations alone or combined active and passive observations were used to estimate the soil moisture state at 6km resolution. In terms of mean estimated soil moisture and RMSE, the EnKS

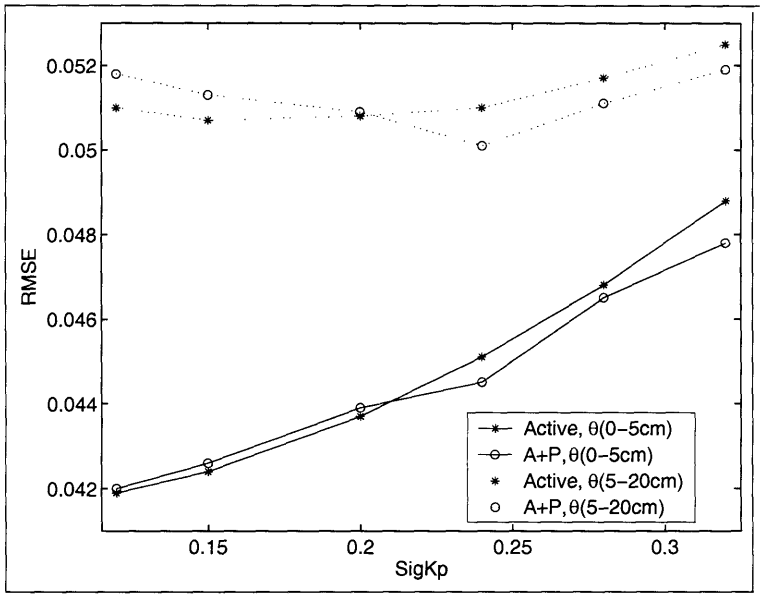

Figure 5-14: RMSE, averaged across all estimation pixels and times is shown as a function of K_p used in the assimilation. In a given run, K_p is assumed constant across the whole domain, to increase the sample size. Results are shown for RMSE in surface soil moisture (solid), and the second soil layer (dotted).

generally produced a better estimate of soil moisture than the EnKF. Assimilating passive observations could update the mean state over a 36km pixel well, but lack of information on spatial variability resulted in patchiness. In all experiments except Section 5.7, K_p was fixed at 0.15. At this value, there was little difference between using active observations alone or in conjuction with passive observations. Including finer scale observations produces an estimate which reflects the true spatial varibility in soil moisture which is primarily a result of fine scale variability in soil texture.

The EnKS was implemented as a fixed lag smoother in this paper, as previous work had demonstrated that there is limited memory in soil moisture. A set of experiments were conducted in which the lag in the EnKS was varied from 1 to 8. In general, it was shown that the most significant improvement over the EnKF occurred when the EnKS was used at all; using a lag=1 prevented the ensemble from drifting towards the open loop estimate as occurs in the EnKF. Further increase in the prescribed lag resulted in a comparatively small improvement. The results varied depending on which combination of observations were assimilated. Consistent with previous work,

increasing the lag was more significant at depth where the memory is longer than at the surface. The results also varied in space. The greatest improvement in increasing the lag from 1 to 8 occurred when either active observations alone or combined active and passive observations were assimilated. In this case, the improvement was most significant at depth in the drier western half of the domain.

Finally, the data assimilation framework developed around the EnKS was used to examine how the impact of spatially variable observation error on the microwave backscatter observations would influence the choice of observations assimilated. The measurement error on the microwave backscatter observations is a function of position within the swath. Errors decrease with distance from the center of the swath. It was shown that including passive observations achieved little difference compared to using active observations alone at the edge of the swath where K_p is less than about 0.2, its inclusion results in an improved estimate at the center of the swath. Consequently, the estimation of soil moisture from Hydros observations should include both active and passive observations.

Chapter 6

Soil Moisture Estimation using the EnKF and EnKS with biased observations.

6.1 Introduction

A synthetic experiment based on data from the Southern Great Plains Experiment 1997 is used to study the impact of observation bias on the soil moisture estimation problem using the ensemble Kalman filter and smoother. Synthetic "true" soil moisture at El Reno (Oklahoma) was obtained by running a single realization of the Noah land surface model, forced with randomized meteorological forcing data, and randomized soil and vegetation parameters as in Chapter 4. The true precipitation recorded at El Reno during SGP97 was used to generate the synthetic "true" soil moisture. The ESTAR radiative transfer model discussed in Appendix B was used to generate so-called perfect radiobrightness temperature observations. These observations were perturbed by measurement error $\sim N(0, 3K)$, to produce unbiased though noisy observations. The biased observations of this experiment were obtained by adding a bias to each of these observations. The goal of this chapter is to estimate the synthetic "true" soil moisture from these biased observations using the EnKF and EnKS by

estimating the bias inline. This is achieved through its inclusion in an augmeneted state vector as discussed in Section 6.2.

Section 6.3 discusses the impact of initial conditions, particularly assumed initial error covariance on the final estimate. In Section 6.4, results are presented which demonstrate the feasibility of using this simple approach to estimate the observation bias. The issue of bias consisting of a combination of model bias and observation bias is introduced. The impact of bias estimation on the soil moisture estimates from the EnKF and EnKS is discussed in Section 6.5. Finally, Section 6.6 addresses the impact of implementing the EnKS as a fixed lag smoother on the bias estimate and the resultant soil moisture estimates.

6.2 Bias Estimation

The state vector is augmented to include the bias along with the soil moisture states of interest, to be updated in the EnKF and EnKS. Between update steps, the soil moisture is modeled using the Noah Land Surface Model denoted A in (6.1); the soil moisture state y(t), at time t, is a function of the state at a previous time $y(\tau)$, the model parameters α, forcing $u(t)$ and model error $\omega(t)$. Model error is introduced in the same manner as in Chapter 4. Between observations, it is assumed that the bias is constant, so that:

$$\begin{bmatrix} y(t) \\ b(t) \end{bmatrix} = \begin{bmatrix} A[y(\tau), \alpha, u(\tau), t, \tau, \omega(t)] \\ b(\tau) \end{bmatrix} \tag{6.1}$$

where $t > \tau > 0$. The measurement model, adapted to include the bias is as follows:

$$Z(t) \;=\; \mathcal{M}[y(t)] + b(t) + \varepsilon(t) \tag{6.2}$$

where $\mathcal{M}$ is a non-linear operator (in this case the ESTAR radiative transfer model described in Appendix B) which relates the states $y(t)$ to the brightness temperatures in $Z(t)$. The observations are assumed to have an additive error, the sum of an

unknown bias $b(t)$ and a mean-zero gaussian noise with standard deviation of 3K. Observations are available daily, and estimates are obtained four times daily. At each update time, the observation is used to update the current state estimate in the EnKF and to update the ensemble at previous times using the EnKS. The EnKF and EnKS are discussed in Chapter 4.

6.3 Results: Dependence on assumed initial distribution

Figure 6-1 shows that the final mean estimated bias and its standard deviation depend on the initial error covariance of the bias. The mean initial value is less significant. If the standard deviation of the initial bias guess is too low, the bias estimate is deemed perfect by the EnKF and is not adjusted at the update steps. Increasing the standard deviation implies uncertainty in the bias estimate, so that as observations are assimilated the bias gets updated, ultimately converging towards the true observation bias. As the bias estimate is improved through assimilation of observations, the standard deviation in the bias estimate decreases as it becomes more certain. In Figure 6-1, the relative reduction in standard deviation between the initial standard deviation and that at the final time step is greatest where the initial standard deviation was high.

6.4 Results: Observation or Model Bias?

If the final estimate has not yet converged on the true bias, it could require assimilation of additional observations. Nonetheless, at the final time step, the bias estimate should be somewhere between the initial guess and the imposed observation bias of 7K. This is clearly not the case in Figure 6-1. For example, where the mean initial bias is greater than 7K, and the standard deviation is high, final bias estimates are between 6 and 7K. This suggests that the bias $b(t)$ that we have estimated may be a combination of model bias as well as observation bias.

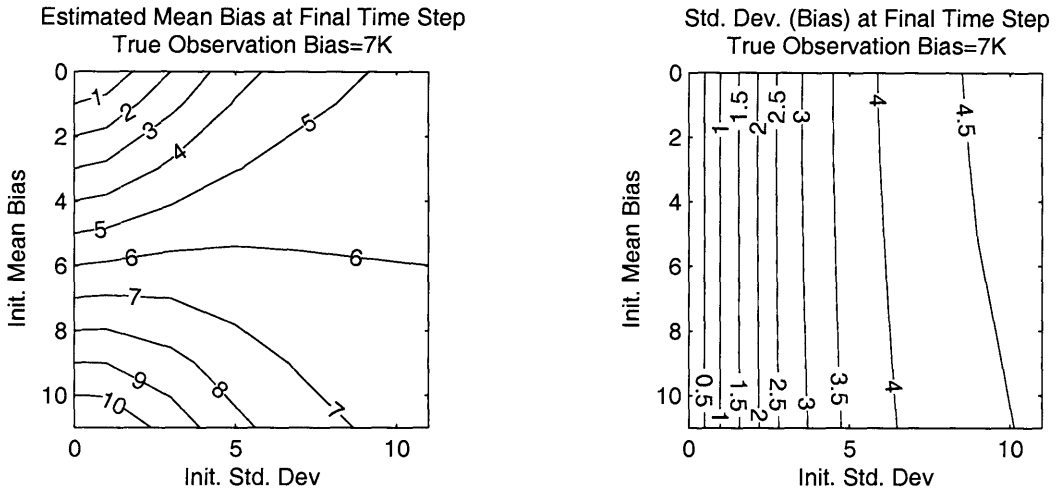

Figure 6-1: The mean (left) and standard deviation (right) in the estimated bias when the true observation bias is constant and equal to 7K, as a function of the mean and standard deviation of the initial distribution of bias.

Figure 6-2 shows the time series of estimated bias and its standard deviation from the EnKF and EnKS where the true observation bias has a constant value of 7K where the initial model bias estimate was $\sim N(0.0, 9.0K)$. From the initial guess, the estimate quickly rises to about 7K, but rather than remain at this value, fluctuates for 40 estimation time steps before falling to below 5K. Our assertion that there is model bias present is supported by this temporal fluctuation. Due to the constant bias model used between observations, the EnKS will simply update all previous estimates to the best estimate given the latest observation. This is why the mean bias estimate and its standard deviation from the EnKS are constant for the study duration, and identical by definition to the final values from the EnKF.

As this is a synthetic experiment, the unbiased observations are available. Implementing the same bias estimation approach using unbiased observations allows us to estimate the model bias. The resultant estimated model bias is shown in Figure 6-3, and demonstrates that the model bias varies in time. There is an initial model bias arising from the spurious initial soil moisture conditions. As the EnKF corrects this through assimilation of observations, the estimated bias decreases. However, as it does not simply converge to a zero value, one can assume that some model

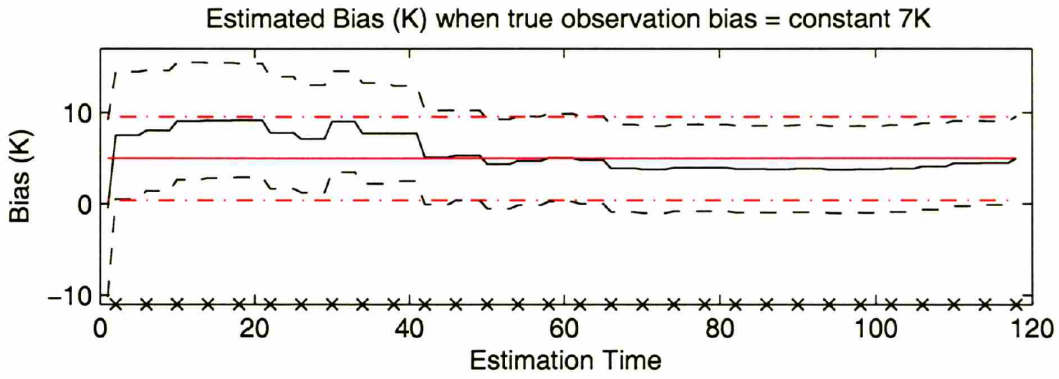

Figure 6-2: Time series of estimated bias from the EnKF (black) and EnKS (red) when the true observation bias is constant and equal to 7K. Times at which observations were available are indicated with an 'x'. Solid lines indicate the mean estimated bias. Dashed and dash-dot lines indicate $\pm$ one standard deviation in the EnKF and EnKS estimates respectively. Initial bias estimate $\sim N(0.0, 9.0K)$.

bias persists after the impact of the initial conditions has dissipated. In Figure 6-4, the estimated model bias (as in Figure 6-3) and the estimated (combined) model and observation bias (as in Figure 6-2) are shown. The difference between these two quantities is the estimated observation bias (shown in bold). The solution rises sharply from the initial zero-mean guess at the first update and then converges slowly to the correct value of 7K. In applications with real observations model bias is likely to be unknown, so its removal from the estimated bias may not be possible. However, its removal here allows us to demonstrate that augmenting the state vector yields a reasonable estimate of the bias which, if the model is unbiased, is the observation bias.

6.5 Impact of bias estimation on soil moisture estimation

From the previous section it is clear that estimated bias in this problem is a combination of model and observation bias. In this section, we examine the impact that estimation of this bias has on soil moisture estimation using both the EnKF and

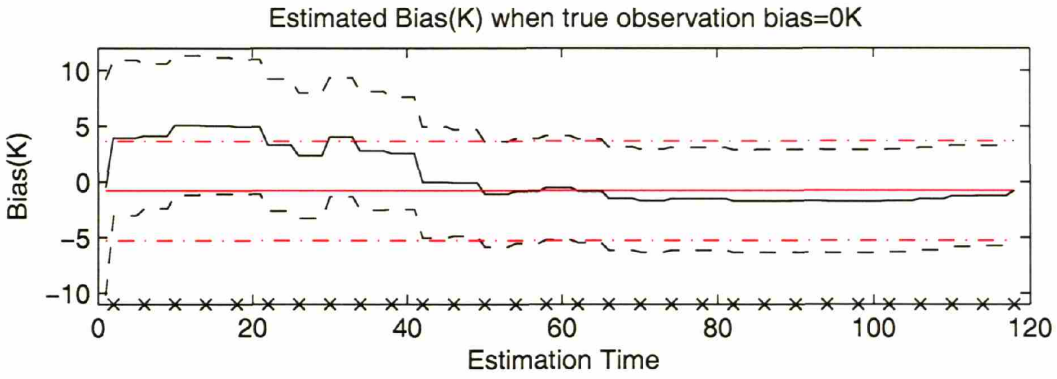

Figure 6-3: Time series of estimated bias from the EnKF (black) and EnKS (red) when the true observation bias is constant and equal to 0K. Times at which observations were available are indicated with an 'x'. Solid lines indicate the mean estimated bias. Dashed and dash-dot lines indicate $\pm$ one standard deviation in the EnKF and EnKS estimates respectively. Initial bias estimate $\sim N(0.0, 9.0K)$. As the observations are unbiased in this case, the estimated bias corresponds to the model bias.

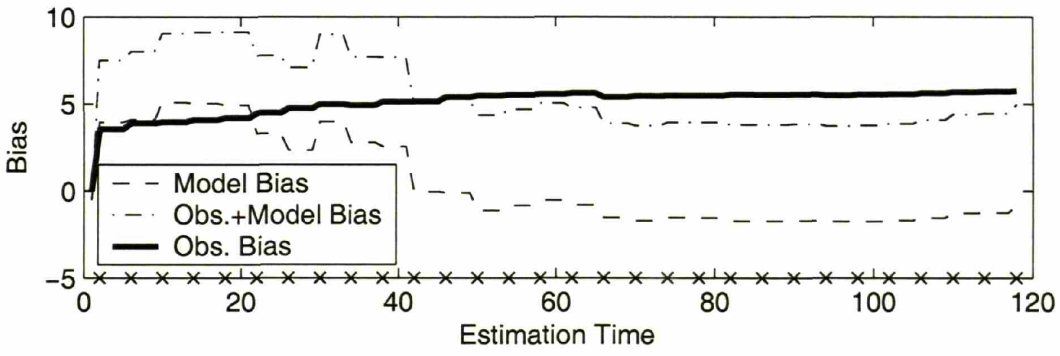

Figure 6-4: Time series of mean estimated bias from the EnKF. Times at which observations were available are indicated with an 'x'.The EnKS estimate at each time step is equal to the final value from the EnKF. Subtracting the known model bias from the estimated bias yields the estimated observation bias (bold).

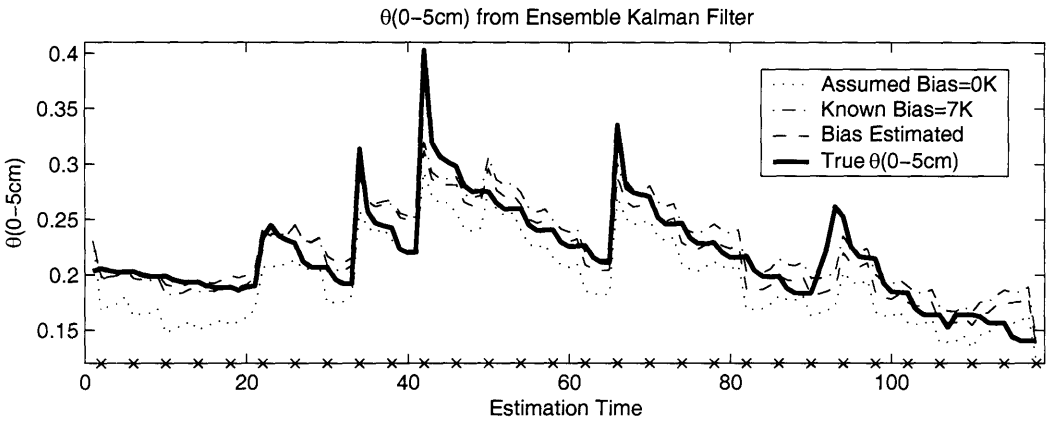

Figure 6-5: Volumetric soil moisture $\theta(0-5cm)$ estimated using the EnKF with biased observations is compared to the synthetic "true" value (bold). Times at which observations were available are indicated with an 'x'. In all cases, observations were perturbed by a constant bias of 7K. In the "No Bias Correction" case, and the "Known Bias=7K" case $b(t)$ is not estimated but is set equal to 0K and 7K respectively in the measurement model. In the "Bias Estimated" case, the initial bias is $\sim N(0.0, 9.0K)$, and $b(t)$ is estimated through its inclusion in the state vector.

EnKS.

Figure 6-5 shows the improvement in the soil moisture estimate obtained using the EnKF when the bias is estimated along with the soil moisture state. The true observation bias in this experiment is constant and equal to 7K on all observations. The worst-case-scenario is when the observations are incorrectly assumed to be unbiased, producing a persistent dry bias in the soil moisture estimate. The best-case-scenario is when $b(t)$ is known to be 7K and simply inserted into the measurement model. This produces an estimate which tracks the true evolution of soil moisture very well. While it deviates from the truth between observations (e.g. times 75-80 and 113-118), it is drawn back to the truth by the EnKF when the next observation becomes available. When the bias is estimated (shown as the dashed line), the estimated soil moisture is always somewhere between these two scenarios, and is generally closer to the estimate obtained when the observation bias is known and accounted for. This demonstrates that using the EnKF to estimate the observation bias reduces the dry bias in the soil moisture estimate associated with the biased observations.

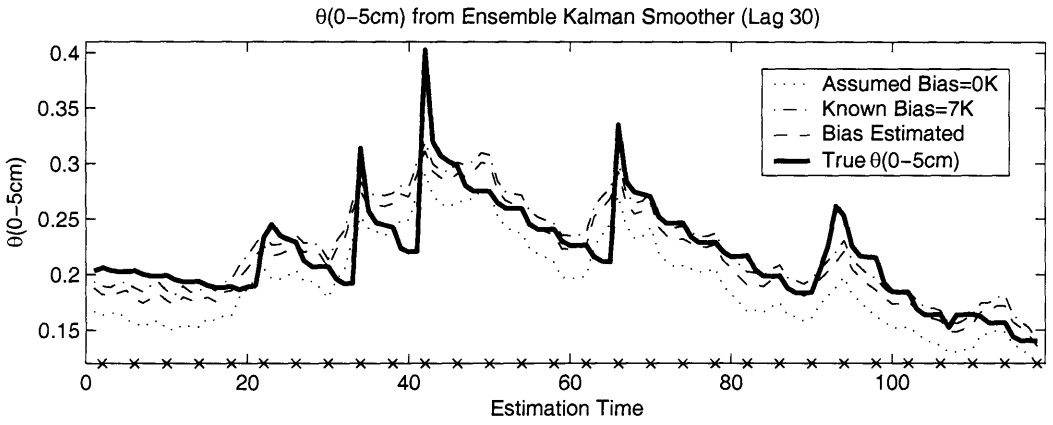

Figure 6-6: Volumetric soil moisture $\theta(0-5cm)$ estimated using the EnKS with biased observations is compared to the synthetic "true" value (bold). Times at which observations were available are indicated with an 'x'. Lag=30 means that as an observation is used to update the state, it is also used to update the state at all previous times. In all cases, observations were perturbed by a constant bias of 7K. In the "No Bias Correction" case, and the "Known Bias=7K" case $b(t)$ is not estimated but is set equal to 0K and 7K respectively in the measurement model. In the "Bias Estimated" case, the initial bias is $\sim N(0.0, 9.0K)$, and $b(t)$ is estimated through its inclusion in the state vector.

In Figure 6-6 the true soil moisture is compared to the estimated soil moisture using the EnKS when the bias is assumed , known and estimated. Results are generally consistent with those from Chapters 4 and 5. The EnKS performs best when the soil column is drying down and is least effective when precipitation occurs between observations. As in the EnKF results from Figure 6-5, the soil moisture estimate when the bias is included in the state vector and estimated is closer to the soil moisture estimated when bias is known and accounted for than the case where observations are assumed to be unbiased. However, between estimation times 0 and 40, the EnKF is more successful in correcting the dry bias associated with the 7K observation bias. Recall from Figure 6-2 that due to the constant bias model, the EnKS assumes that the bias estimate at the final time step is the best estimate for all estimation times in the experiment. While the EnKF detects the considerable model bias in this interval, bias is grossly underestimated by the EnKS. As a result, the EnKF actually yields a better estimate of soil moisture in the first 10 days of the experiment.

6.6 Results: Impact on bias estimation if the EnKS is implemented as a fixed-lag smoother

The EnKS uses observations as they become available to update the current state but also uses the correlation between the model forecast at the current time and the state at previous times to update the ensemble at previous times. In a fixed lag smoother, observations are used to update past estimates within a fixed time window.

In Chapters 4 and 5, it was demonstrated that the EnKS could be implemented as a fixed-lag smoother in the soil moisture estimation problem. While even a lag of 1 (i.e. updating the states back to the last observation) resulted in an improved memory at the surface, a longer lag was required at depth where memory is greater. In previous sections of this chapter, results from the EnKS have been obtained by using observations when they become available to update all previous states in addition to the current state. In this section, the impact of using a shorter lag on the bias estimate and resultant soil moisture estimate are studied.

Figures 6-7 and 6-8 show the variation in mean estimated bias and the associated standard deviation during the experiment, when the lag is varied from 1 to 30. Because the forward model for the model bias simply assumes that $b(t) = b(t-1)$, the EnKS will set all estimates in the smoother window to the latest estimate. So, for lag 1 all values back as far as the last observation are set to the updated bias upon assimilation of the current observation. If the lag is 30 (i.e. each observation updates the ensemble at all previous estimation times), the ensemble of bias estimates at all estimation times in the experiment are set to the values obtained at the final update. This value is by definition identical to the EnKF estimate at the final update. This causes the banded appearance of Figures 6-7 and 6-8. Recall that observations are available daily or every 4 estimation time steps.

As shown earlier, there is a bias associated with forward model which varies in time, particularly at the beginning of the experiment. As a result, using the constant bias model with an EnKS of infinite lag proves detrimental as it fails to account for the time-varying model bias component of the bias estimate. Figure 6-9 shows the

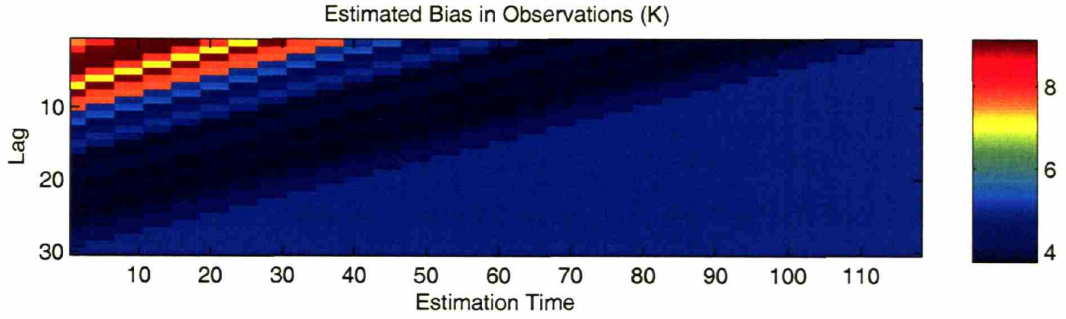

Figure 6-7: Mean estimated bias, from the EnKS as a fixed lag smoother, for lag varied from 1 to 30.

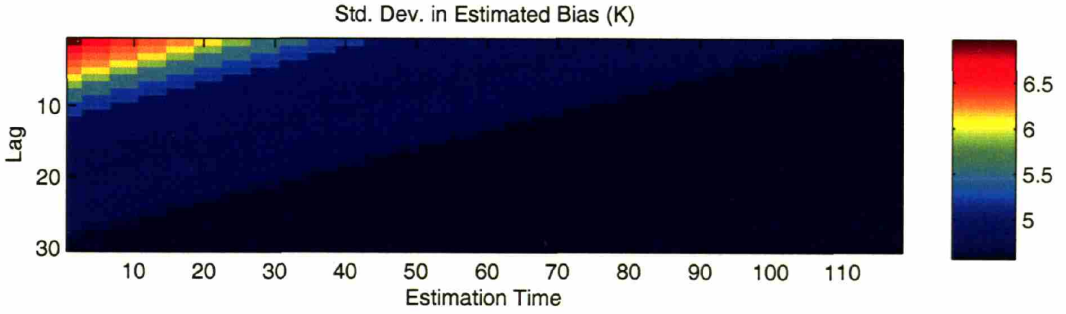

Figure 6-8: Standard deviation in the estimated bias, obtained using the EnKS as a fixed lag smoother, for lag varied from 1 to 30.

estimated soil moisture for the experiment duration calculated using the EnKS using a lag varying from 1 to the maximum value of 30. Results are shown for three cases. Results in the bottom panel, where the observation bias is known are taken to be the best results attainable from the EnKS. Results in the top panel are the worst of the cases studied, because the observation bias is assumed to be zero and is not estimated at all. A persistant dry bias is apparent at all times regardless of lag. In the center, results are shown where the observation bias is included in the state vector and estimated inline. Clearly, bias estimation reduces the dry bias. The impact of lag is most apparent at the first 30 or so estimation times. Where the lag is large enough to set the bias value to the estimated value at time ~ 40, the bias is too low, and the dry bias is more pronounced than the value obtained from the EnKF and the EnKS with lower lag.

6.7 Final Remarks

In this chapter, a synthetic experiment was used to study the soil moisture estimation problem when observations are subject to bias. True soil moisture for a 30-day period was generated using the true recorded precipitation at El Reno (Oklahoma) during the Southern Great Plains to force a single realization of the model. This true soil moisture was used to generate synthetic unbiased and biased observations. In the biased case, all observations were subject to a constant bias of 7K.

In this experiment, the bias is included in the state vector and estimated in the ensemble Kalman filter and smoother. Results demonstrate that standard deviation in the initial estimate must be large enough to allow the filter to update the bias estimate. It was demonstrated that the estimated bias is actually a combination of model bias and observation bias. In this experiment, as observations were synthetic, the unbiased observations could be used to determine the model bias. It was argued that if the model is unbiased, or the model bias is known and can be subtracted, the observation bias can be determined.

This simple experiment demonstrates that estimation of the bias through its inclu-

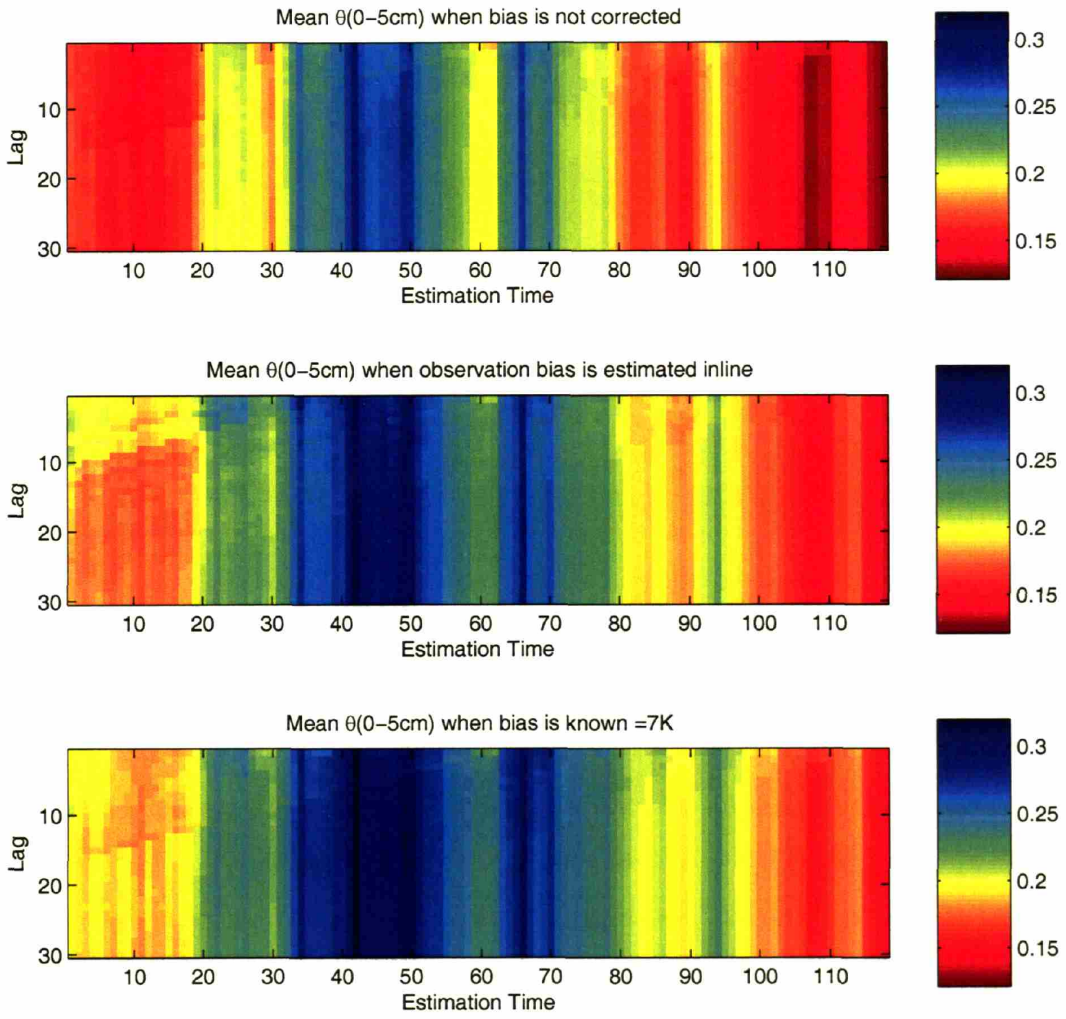

Figure 6-9: Estimated volumetric soil moisture (0-5cm) from the EnKS with varying lag, when observation bias is incorrectly assumed to be zero (top), estimated (center) or set equal to its true value of 7K.

sion in the state vector is feasible and potentially valuable in soil moisture estimation. Additional studies might address the issue of non-constant bias, for example drifting bias or state-dependent bias.

Chapter 7

Original Contributions and Future Directions

7.1 Conclusions and Original Contributions

The role of soil moisture in the climate system and its influence on land-atmosphere interactions at different spatial and temporal scales were discussed in Chapter 1. Soil moisture observation using in-situ measurements and remote-sensing technology were discussed. It was argued that study of hydrometeorological and hydroclimatological problems requires global measurements at resolutions on the order of 10-40km at least once every 2-3 days, and therefore remote-sensing of soil moisture is a more feasible solution compared to reliance on in-situ observations. The advantages of using L-band microwave observations were outlined, with particular emphasis on observations from NASA's future Hydros mission. Advances in hydrological data assimilation were reviewed. Variational methods were deemed unattractive as they require the development of an adjoint model. Methods derived from the classic linear Kalman filter are considered instead. As land surface models are non-linear and contain thresholds, the classic Kalman filter is unsuitable. To use the extended Kalman filter, the model must be approximated to propagate the mean and covariance forward between updates. There are also potential problems with instability in non-linear systems. The key advantages of ensemble-based techniques is that the allow the use

of any non-linear model. There is no need to develop an adjoint, and they allow great flexibility in the specification of model error. Finally, it was argued that soil moisture estimation is a reanalysis-type problem rather than a control-type problem and consequently that a smoothing approach is more suitable than filtering. So, the goal of this thesis was to investigate ensemble smoothing techniques, and use such a technique to develop a data assimilation framework in which multi-resolution observations such as those from Hydros could be merged with modeled soil moisture from a conventional land surface model to estimate soil moisture and surface energy fluxes.

A simple linear model was used in Chapter 2 to illustrate the advantages of smoothing rather than filtering. Two ensemble smoothing algorithms were introduced in which the state vector was distributed in time and updated in a batch using all observations within the same time window. In the ensemble single batch (EnSB) smoother, the states at all estimation times in the experiment are updated in a single batch using all observations during the experiment. In the ensemble moving batch smoother (EnMB) the window spans a few observations. For the linear problem, solutions from these two smoothers were compared to the optimal filter (classic Kalman filter) and the Rauch-Tung-Striebel form of the optimal smoother. It was found that for the simple linear gaussian problem, both algorithms approximated the solution from the optimal smoother. This was shown in terms of the state estimate itself, reduced ensemble spread and overall RMSE. The length of the smoother window used in the ensemble moving batch smoother was allowed to vary, and it was found that provided the window was sufficiently long to capture the memory in the system, the EnMB could estimate the state as well as the EnSB at a fraction of the computational expense. By estimating fewer states at a time, it also required fewer ensemble members than the EnSB to converge resulting in a further reduction in computational cost.

In Chapter 3, the ensemble moving batch smoother is used to estimate soil moisture in a synthetic experiment based on the Southern Great Plains Experiment 1997. The inclusion of additional observations compared to the EnKF resulted in an im-

proved estimate of soil moisture at the surface and at depth. The estimated soil moisture was closer to the synthetic "true" soil moisture, and the ensemble spread using the EnMB was reduced relative to the EnKF indicating increased confidence in the estimate. The experiment was repeated using two different precipitation forcing datasets generated using different assumptions about precipitation uncertainty. In one case, it was assumed that the timing of precipitation was known perfectly, but the amount was uncertain. In the second case, it was assumed that both the timing and amount of precipitation was uncertain. The markedly different soil moisture estimates obtained in the two experiments underline the importance of accurately representing sources of model error in a data assimilation framework. Finally, in this chapter, it was shown that implementing a smoothing algorithm in hydrologic data assimilation is complicated by the occurrence of precipitation. A smoother will, by definition, produce an estimate which transitions smoothly between observations. Intuitively, this is ideally suited to estimating soil moisture when the soil is drying. A time series of soil moisture can be broken into a series of drydown curves, separated by the occurrence of precipitation. This can be done in-line as the occurrence of precipitation causes a peak in soil moisture and therefore a sharp decrease in brightness temperature. A hybrid smoother/filter approach was presented in which the smoother window length is dynamic rather than a single prescribed value, defined so that the soil moisture for the whole dry-down is determined in one batch, using all observations during the drydown. This method improves the estimate by preventing the backward propagation of information from precipitation events after an observation at the end of a smoothing window although the benefit in doing so depends on the relative timing of precipitation and observations.

While results from Chapter 3 demonstrated that ensemble smoothing techniques could be used in soil moisture estimation, the computational cost of a spatially-distributed state vector which must be augmented to include states distributed in time would render the EnMB approach infeasible for spatially-distributed problems. In Chapter 4, the ensemble Kalman smoother (EnKS) was studied as a computationally less expensive alternative. It was shown that the EnKS is an extension of

the EnKF in which observations are used as they become available to update the ensemble at prior estimation times in addition to the current forecast ensemble. As its starting point is the EnKF, the estimate is always at least as good as that obtained using the EnKF. It was demonstrated that the most costly calculations are performed already in the EnKF, rendering it an inexpensive ensemble smoother. Validation of estimated surface and subsurface soil moisture against gravimetric measurements at three ground truth sites, and flux tower observations of latent heat flux at Central Facility demonstrated that the EnKS resulted in an improved estimate compared to the EnKF. Finally, it was shown that the limited memory in soil moisture can be exploited by implementing the EnKS as a fixed lag smoother. The required lag was shown to increase with depth, consistent with there being greater memory at depth.

In Chapter 5 the EnKS was used in a synthetic experiment over the 575,000 km^2 Arkansas-Red River basin to estimate surface and root zone soil moisture through the assimilation of multi-resolution observations. Passive (36km) observations alone, active (3km) observations alone or combined active and passive observations were used to estimate the soil moisture state at 6km resolution. In terms of mean estimated soil moisture and RMSE, the EnKS generally produced a better estimate of soil moisture than the EnKF. Assimilating passive observations alone estimated the mean state over a 36km pixel well, but failed to capture finer scale spatial variability. Including finer scale observations either alone or in conjunction with passive observations produced an estimate reflecting the true spatial variability in soil moisture which is primarily a result of fine scale variability in soil texture. In general, it was shown that the most significant improvement over the EnKF occurred when the EnKS was used at all (even just between consecutive observations). Further increase in the prescribed lag resulted in a comparatively small improvement, and was more significant with increasing depth. Results were shown to vary depending on the combination of observations, with greater improvement upon assimilation of radar observations. Finally, the EnKS was used to examine how the impact of spatially variable observation error on the microwave backscatter observations would influence the choice of observations assimilated. It was shown that including passive observations made

little difference compared to using active observations alone at the edge of the swath where radar observation error is lowest, however their inclusion results in an improved estimate at the center of the swath where radar measurements are highly uncertain.

7.2 Future Research Directions

The term "data assimilation" refers to techniques which merge noisy observations with uncertain model estimates, weighing their respective uncertainties to produce an estimate which is better than either observations or modeling alone. The results from a data assimilation framework can be improved through improvements in any component of the data assimilation framework; obtaining better or more frequent observations, using a model which is more capable of representing reality or using a data assimilation technique that makes fewer or more valid assumptions or is more computationally efficient. In this section, potential improvements in each of these areas are discussed.

7.2.1 Model

The forward model in the data assimilation framework is used to propagate the states forward between update times. Achieving the correct prior or background covariance at an update step depends on the model's ability to correctly propagate the initial state forward. This in turn depends on the model physics being an adequate representation of reality, and a realistic characterization of model error.

Model Physics

An advantage of ensemble techniques is that the model may be considered as an offline subroutine. As both the EnKF and EnKS require only forward model runs, there is no need to re-calculate the adjoint every time a change is made to the model. In fact, the burden in replacing the forward model completely is in tailoring the interface between the forward model, measurement model and data assimilation algorithm.

In Chapter 5 it was shown that neither the EnKF nor the EnKS could estimate subsurface soil moisture particularly well when the true soil moisture was generated using the spatially distributed TOPLATS model and the 1-D Noah land surface model was used as the forward model. This suggests that if non-observed states are to be estimated, the model must be such that the correlation between the observed and non-observed states develops correctly. Our choice of the 1-D Noah land surface model worked well in our two experiments because of the low relief landscape. If this data assimilation framework is to be implemented over varied terrain, the 1-D model should be replaced by a model capable of modeling the lateral redistribution of moisture. Moving to a distributed model is non-trivial. In a 1-D model, estimation cells are independent so it is easy to subdivide estimation over the domain into smaller problems to overcome memory or storage limitations. In distributed models such as tRIBS [42] or TOPLATS [69], the whole watershed must be modeled. Furthermore, the state vector must be defined so they are uniquely related to the states required to re-initialize the model.

Model Error - Parameters

In this thesis, model error was included by assuming key parameters were uncertain and characterize by an assumed probability distribution. Clearly, the estimate would be more meaningful if the uncertainty in these parameters could be quantified more objectively through observations and/or calibration.

Model Error - Precipitation Forcing

Soil moisture is more sensitive to incident precipitation than any other forcing variable. As demonstrated in Chapter 3, results from the EnOL, EnKF and the EnMB depend heavily on the assumed precipitation forcing. Generation of appropriate precipitation forcing ensembles is the key to estimating soil moisture using ensemble data assimilation. Current precipitation forcing includes uncertainty in the timing and amount of precipitation. However, downscaling precipitation using a cascade does not produce a field which is physically realistic in time and space. A more

physically-based downscaling scheme, conditioned on available NEXRAD or GPCP data could more adequately represent the true error in precipitation. This would yield an improved estimate by providing more realistic forcing, and by representing the primary source of spatial variability.

7.2.2 Observations

Some improvements in the "observations" component are beyond our control e.g. revisit time, orbit and raw data resolution. However, some improvements may be made to the measurement model. Just like the forward model, the measurement model in an ensemble-based data assimilation framework may be considered as an off-line subroutine. It may be updated or replaced as the retrieval algorithm develops to include more diverse conditions e.g. denser vegetation, inland water or sub-pixel heterogeneity.

The measurement model component of the data assimilation framework could also be improved by including the assimilation of remotely-sensed land surface temperature. Currently, the soil temperature comes from the land surface model. While it is affected by the soil moisture estimation through the coupling of the water and energy balances, constraining it with observations can only improve the performance of the measurement model and the state estimation.

It is currently assumed that all observation error is independent in space and time. In the SGP97-based experiments the measurement error on the synthetic or real ESTAR observations is $\sim N(0, 3K)$. The synthetic Hydros radiometer observations in the Arkansas-Red River experiment are assumed to have additive noise $\sim N(0, 1.5K)$, and the synthetic backscatter observations have multiplicative error, a function of $K_p = 0.15$. We have also assumed that observations are available for the whole domain at the same time. Future work on this data assimilation framework should include the use of an orbit simulator to simulate the real availability of observations across the domain. Doing so would allow for a more accurate characterization of the likely model error associated with Hydros observations.

7.2.3 Data Assimilation Techniques

An undeniable disadvantage of smoothing techniques is the increased computational expense compared to filtering. In Chapter 4, it was argued that the EnKS is a relatively inexpensive smoother as the most costly computational components are already performed in the EnKF. The increased cost above filtering is that of retrieving the prior ensembles to be updated and their update through a single post-multiplication. It was shown that the EnKS could be implemented as a fixed lag smoother resulting in a reduction in computational cost without any loss in accuracy. The lag required is determined by the memory in soil moisture, and was shown to be greater at depth. This issue merits further investigation, particularly if the EnKS is to be implemented in a larger scale problem where computational cost will likely determine the choice of data assimilation algorithm.

7.2.4 Relevance to other problems in hydrology

The conclusions of this research may be applicable in other areas of hydrology. For example, a time series of streamflow shares the key features of a time series of soil moisture. Streamflow consists of several contributions which represent the response of the basin to precipitation on different timescales. Overland flow is the fast-response component. Subsurface stormflow is slower as water has to infiltrate the soil and travel through the unsaturated zone to the stream. Baseflow has the slowest response, as it represents the slowest lagged response as it is the groundwater response to precipitation. Smoothing may prove even more useful in streamflow estimation than it is in soil moisture estimation as the varied response times of streamflow components result in a smoother response to precipitation forcing.

There are many areas in which variational smoothing has been used successfully, e.g. groundwater flow, chemical fate and transport and surface flux estimation. Provided the scale of the problem is such that ensemble model runs are computationally feasible, the techniques used in this thesis could be suitable for use in these areas.

Appendix A

Model Error & Noah Land Surface Model

In the Noah land surface model [5], the terms of the water and energy budgets for a 1-D soil column are computed for a single unified ground/vegetation surface (Figure A-1). The volumetric soil moisture profile is modeled using the bulk form of the Richards Equation. Functions for hydraulic conductivity $K(\theta)$ and matric head $\Psi(\theta)$ are taken from [10]:

$$K(\theta) = K_s \left(\frac{\theta}{\theta_s} \right)^{2b+3} \tag{A.1}$$

$$\Psi(\theta) = \Psi_s \left(\frac{\theta}{\theta_s} \right)^{-b} \tag{A.2}$$

The saturated hydraulic conductivity (K_s), porosity (θ_s), saturated soil potential (Ψ_s) and the parameter b are determined by the user-defined soil class of the grid cell.

Total evaporation from the ground/canopy surface includes three contributions:

1. Direct evaporation from the bare soil (E_{dir}) is at the potential rate (E_p) if energy is the limiting factor for latent heat flux. Otherwise it is limited by the rate at which water can be transferred to the surface through the soil, a function of the parameters of Richard's equation. E_p assumes unlimited moisture availability and zero stomatal resistance, and is constrained by a surface energy balance to

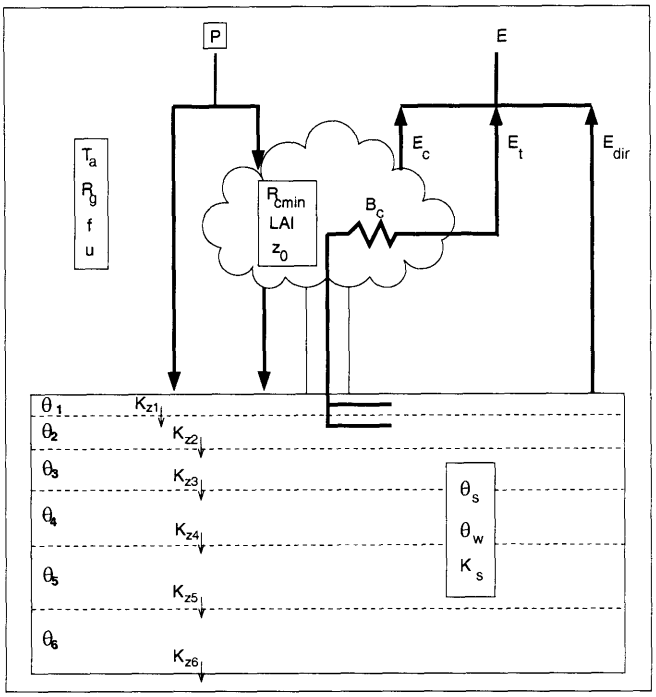

Figure A-1: Soil hydrology and evaporation components of the Noah land surface model.

produce diurnal variation.

2. Evaporation of intercepted water from the canopy (E_c) is controlled by the availability of intercepted water (W_c) compared to the maximum capacity (S):

$$E_c = \sigma_f E_p \left(\frac{W_c}{S}\right)^n \tag{A.3}$$

where $n = 0.5mm$.

3. Evapotranspiration from the root zone through the roots and canopy, E_t is controlled by a resistance term B_c:

$$E_t = \sigma_f E_p B_c \left[1 - \left(\frac{W_c}{S}\right)^n\right] \tag{A.4}$$

$$B_c = \frac{1 + \frac{\Delta}{R_r}}{1 + R_c C_h + \frac{\Delta}{R_r}} \tag{A.5}$$

Δ is a function of the slope of the saturation specific humidity curve. C_h is is

the surface exchange coefficient for heat and moisture, and is a function of wind speed and roughness length (z_0). The aerodynamic resistance, R_r, is a function of air temperature, surface pressure and C_h. R_c is the canopy resistance:

$$R_c = \frac{R_{cmin}}{LAI \times F_1 \times F_2 \times F_3 \times F_4} \tag{A.6}$$

where R_{cmin} is the minimum stomatal resistance (a function of vegetation type), and F1 to F4 are stress functions representing the effects of solar radiation, vapor pressure deficit, air temperature and soil moisture availability.

Uncertainty in precipitation is the dominant source of ensemble spread in the soil moisture estimate as it determines the amount of water introduced to the system. It is a key forcing variable in both the soil moisture diffusion equation and each of the sources of evaporation. Uncertainty in air temperature, shortwave radiation, relative humidity and wind influence evapotranspiration through the stress functions in B_c and the components of the energy balance at the surface. Randomness in the soil parameters K_s and θ_s causes uncertainty in the $K(\theta)$ and $\Psi(\theta)$ functions, resulting in variability in the rate at which water can move within the soil column. Through the soil moisture stress function, uncertainty in θ_s and wilting point θ_w impact canopy resistance and evapotranspiration. Noise in the vegetation parameters, LAI and z_0, affects the aerodynamic and stomatal resistance which controls the rate of evapotranspiration and the turbulent fluxes in general.

Appendix B

ESTAR Radiative Transfer Model

Microwave remote-sensing would provide a way of obtaining global observations of soil moisture. However, to obtain data at a reasonable distribution (O 10km), any instrument intended for low earth orbit would require a very large antenna (O 10m). Clearly there are engineering difficulties associated with placing such a large antenna in orbit. ESTAR used an innovative interferometric technique to obviate the need for a large antenna.

ESTAR was developed by the Goddard Space Flight Center and the University of Amherst to demonstrate the potential of aperture synthesis in microwave remote-sensing [56]. ESTAR is implemented in a hybrid configuration during SGP97. Real antennas (stick antennas oriented with their long axis in the direction of motion) are used to obtain resolution along-track. Resolution across track is achieved using aperture synthesis (i.e. measuring the correlation of the voltage from pairs of sticks [57]. ESTAR proved invaluable in demonstrating the value of L-band passive observations in soil moisture observation and estimation. It was used in Washita'92 (e.g. [47]), SGP97 (e.g. [16], [64]), SGP99 (e.g. [55],[35],[29]). In this thesis, real and synthetic ESTAR observations are used to examine the feasibility of using ensemble smoothing techniques in soil moisture data assimilation.

The dielectric constant of pure water is obtained using the Debye Equation ([81])

$$\epsilon_w = \epsilon_{w\infty} + \frac{\epsilon_{w0} - \epsilon_{w\infty}}{1 + j2\pi f \tau_w} \tag{B.1}$$

where

- ϵ_{w0} is the static dielectric constant of pure water, a function of temperature in degrees Celsius [52]:

$$\epsilon_{w0}(T) = 88.045 - 0.4147T + 6.295 \times 10^{-4}T^2 + 1.075 \times 10^{-5}T^3 \quad \text{(B.2)}$$

- $\epsilon_{w\infty}$ is the optical limit of ϵ_w, found to have a value of 4.9 (dimensionless) [79].

- τ_w is the relaxation time of pure water (in seconds) given by [79]:

$$\begin{aligned}\tau_w(T) = \frac{1}{2\pi}(&1.1109 \times 10^{-10} - 3.824 \times 10^{-12}T \\ &+ 6.938 \times 10^{14}T^2 - 5.096 \times 10^{16}T^3)\end{aligned}$$

  where temperature (T) is in degrees Celsius.

- f is the frequency in Hertz. For ESTAR, f=1.4GHz.

From B.2 real and imaginary parts of the dielectric constant for pure water may be written:

$$\epsilon'_w = \epsilon_{w\infty} + \frac{\epsilon_{w0} - \epsilon_{w^\infty}}{1 + (2\pi f\tau_w)^2} \quad \text{(B.3)}$$

$$\epsilon''_w = \frac{2\pi f\tau_w(\epsilon_{w0} - \epsilon_{w^\infty})}{1 + (2\pi f\tau_w)^2} \quad \text{(B.4)}$$

In observing soil moisture, we are observing a soil composed of solids, air and water. To obtain the dielectric constant of a soil as a function of soil moisture, the empirically-based mixing model of [90] is used. In this model, a transition point in soil moisture is identified as a function of soil texture. At soil moisture values less than this transition point, the observed dielectric constant of the soil increased slowly as a function of increasing soil moisture. Above this transition point, the increase is more rapid.

The value of this transition point is obtained from:

$$W_t = 0.49WP + 0.165 \tag{B.5}$$

where WP is the wilting point of the soil in precent of dry weight. This is given in terms of volumetric water content (cm^3/cm^3) by:

$$WP = 0.06774 - 0.00064 \times S + 0.00478 \times C \tag{B.6}$$

S and C are the sand and clay content of the soil, expressed as percentage of dry weight of soil.

The complex dielectric constant ϵ of a soil-water mix is given by:

$$\epsilon = W_c\epsilon_x + (n - W_c)\epsilon_a + (1 - n)\epsilon_r \ , \ W_c \leq W_t \tag{B.7}$$

where

$$\epsilon_x = \epsilon_i + (\epsilon_w - \epsilon_i)\frac{W_c}{W_t}\gamma \tag{B.8}$$

and

$$\epsilon = W_t\epsilon_x + (W_c - W_t)\epsilon_w + (n - W_c)\epsilon_r \ , \ W_c > W_t \tag{B.9}$$

where

$$\epsilon_x = \epsilon_i + (\epsilon_w - \epsilon_i)\gamma \tag{B.10}$$

In these expressions, n is the porosity of the soil, ϵ_a,ϵ_w, ϵ_r and ϵ_i are the dielectric constants of air (1.0+0.0j), water, rock (5.5+0.2j) and ice (3.2+0.1j) respectively.

ϵ_x represents the dielectric constant of the initially absorbed water, and γ is a

parameter found empirically to be a function of wilting point:

$$\gamma = -0.57WP + 0.481 \tag{B.11}$$

Having found the dielectric constant of the wet soil, the smooth microwave emissivity (horizontal polarization) is found using the Fresnel Equation:

$$\varepsilon = 1 - \left| \frac{cos\theta - \sqrt{\epsilon - \sin^2\theta}}{\cos\theta + \sqrt{\epsilon - \sin^2\theta}} \right|^2 \tag{B.12}$$

where θ is the look angle (0 for ESTAR). The Fresnel equations assume that the air-soil interface is perfectly smooth within the area viewed by the radiometer. When the surface is rough, the area of the air-soil interface is greater, so there is a larger area to transmit upwelling energy. This results in a higher emissivity over a rough surface than that predicted by the Fresnel equations. [7] provides a simple expression to correct for roughness:

$$\varepsilon_s = 1 + (\varepsilon - 1)\exp(h) \tag{B.13}$$

where h is an empirically determined roughness parameter proportional to the rmse height variations of the surface.

The total microwave brightness temperature including contributions from bare soil and vegetation canopy is given by:

$$T_b = (1 - V_f)\varepsilon_s T_{surf} \tag{B.14}$$

$$+ V_f[\varepsilon_s T_{surf}\exp(-\tau) \tag{B.15}$$

$$+T_c(1-\omega)(1-\exp(-\tau))(1+(1-\varepsilon_s)\exp(-\tau)] \tag{B.16}$$

T_{surf} is the soil temperature, T_c is the canopy temperature, which can be assumed to be the same as the skin temperature of the soil. The vegetation optical depth (τ) is

calculated from [46]. For low frequencies (1-5GHz):

$$\tau = bW \tag{B.17}$$

where W is the vegetation water content and b is a parameter dependent on plant characteristics. All parameters required in Chapters 3 and 4 were obtained from the Southern Great Plains experiment website at

http://disc.gsfc.nasa.gov/fieldexp/SGP97/

Appendix C

Hydros Microwave Emission and Backscatter Model (MEBM)

NASA's Hydrosphere State (Hydros) Satellite Mission is a pathfinder mission with the objective of providing global measurements of soil moisture at 10km resolution with a revisit time of three days [21]. The information in this appendix was obtained from *S. Chan and E. G. Njoku (personal communication).* It is reproduced here to provide additional detail on the observations simulated and used in Chapter 5.

In the following discussion, the look angle θ is 40^o. The frequency for the radiometer observations is 1.41GHz, and 1.26GHz for the radar.

C.1 Emission Model

Neglecting atmospheric effects, the brightness temperature observed from the land surface is given by:

$$\begin{aligned} T_{Bp} &= T_s e_p \exp\left(\frac{-\tau_o}{\cos\theta}\right) \\ &\quad + T_c(1-\omega)\left[1-\exp\left(\frac{-\tau_0}{\cos\theta}\right)\right]\left[1+r_p\exp\left(\frac{-\tau_0}{\cos\theta}\right)\right] \end{aligned} \tag{C.1}$$

where:

- p indicates the polarization. Both horizontal (h) and vertical (v) polarizations

are obtained.

- The soil microwave effective temperature, T_s, is an average of the surface skin temperature T_0 and the soil temperature at a depth of 5cm:

$$T_s = \frac{1}{2}(T_0 + T_5) \tag{C.2}$$

- e_p is the soil emissivity ($= 1 - r_p$). The soil reflectivity, r_p is obtained from the smooth surface reflectivity r_{sp} which is modified to account for surface roughness using the parameter h:

$$r_p = r_{sp} \exp(-h) \tag{C.3}$$

  The parameter h is related empirically to the RMS surface height ($h = 10 \times s(cm)$). The smooth surface reflectivity is obtained from the soil dielectric constant ϵ using the Fresnel equations, as in the ESTAR algorithm in Appendix B.

- τ_0 is the nadir vegetation opacity, a function of the vegetation opacity coefficient b and the total vegetation water content W kg m^{-2}.

$$\tau_0 = b_0 W \tag{C.4}$$

  b is a function of vegetation type. W is obtained from $W = \frac{W_c}{1-f_T}$, where f_T is the woody component fraction. W_c, the canopy vegetation water content is obtained from 1km NDVI data using the following relationship from [45]:

$$W_c = 1.9134(NDVI)^2 - 0.3215(NDVI) \tag{C.5}$$

Over **water surfaces**, the brightness temperature is given by:

$$T_{Bp} = T_0(1 - r_p) \tag{C.6}$$

where T_0 is the skin temperature and r_p is the smooth surface reflectivity for fresh water from [52].

C.2 Backscatter Model

In Chapter 5, only co-polarized observations were used, i.e. σ_{hh} and σvv. Total co-polarized (pp = HH or VV) radar backscatter from the surface, σ^t_{pp} is the sum of three contributions:

$$\sigma^t = \sigma^s \exp\left(\frac{-2\tau_0}{\cos\theta}\right) + \sigma^v + \sigma^{sv} \tag{C.7}$$

where:

- σ^s is the scattering cross-section of the soil surface. Where there is vegetation, this signal is subject to two-way attenuation through the vegetation layer along the slant path. The dominant contribution in bare soil or low-vegetation is the soil surface backscatter, a function of soil moisture and RMS surface roughness. The co-polarized backscatter is obtained from:

$$\sigma^o_{hh} = 10^{-2.75}\left(\frac{\cos^{1.5}\theta}{\sin^5\theta}\right) 10^{0.028\epsilon' \tan\theta} (ks\sin\theta)^{1.4}\,\lambda^{0.7} \tag{C.8}$$

$$\sigma^o_{vv} = 10^{-2.35}\left(\frac{\cos^3\theta}{\sin^3\theta}\right) 10^{0.046\epsilon' \tan\theta} (ks\sin\theta)^{1.1}\,\lambda^{0.7} \tag{C.9}$$

  where k is the wavenumber ($= \frac{2\pi}{\lambda} cm^{-1}$), and s is the surface RMS height in centimeters. ϵ' is the real part of the dielectric constant.

- σ^v is the scattering cross-section of the vegetation volume:

$$\begin{aligned}\sigma^v_{pp} &= 0.74\omega\cos\theta[1 + 0.54\omega\tau_o - 0.24(\omega\tau_o)^2] \\ &\quad \times[1 - \exp(-2.12\tau_o \sec\theta)]\end{aligned} \tag{C.10}$$

Table C.1: Dominant land cover classes, percentage of basin covered, and vegetation and roughness parameters.

Class	Description	% Area	h	ω	b_o	f_T
16	Evergreen Shrub	22	0.10	0.05	0.11	0.0
1	Crop/mixed farming	18	0.15	0.05	0.13	0.2
22	Crop/mixed woodland	14	0.12	0.08	0.12	0.5
20	Short grass/crop	10	0.12	0.05	0.13	0.1
2	Short grass	9	0.10	0.05	0.10	0.0
26	Short grass/mixed woodland	9	0.10	0.08	0.11	0.4
5	Deciduous broadleaf tree	4	1.0	0.12	0.12	0.8
21	Tall grass/crop	4	0.12	0.05	0.13	0.1
18	Mixed woodland	3	0.10	0.12	0.11	0.8
7	Tall Grass	2	0.10	0.05	0.10	0.0
-	Others	5	-	-	-	-

- σ^{sv} is the scattering interaction between soil and vegetation:

$$\begin{aligned}\sigma_{pp}^{sv} &= 1.9\omega\cos\theta[1+0.9\omega\tau_o+0.4(\omega\tau_o)^2][1-exp(-1.93\tau_o\sec\theta)] \\ &\quad\times\exp(-1.37\tau_o^{1.12}\sec\theta)\exp(-0.84(ks)^2\cos\theta)r_{sp} \qquad \text{(C.11)}\end{aligned}$$

Over **water surfaces**, backscatter is given experimental values of $\sigma_{vv} = -23dB$ and $\sigma_{hh} = -27dB$. These are based on observations taken during SGP99 with the PALS instrument over Lake Ellsworth in Oklahoma.

C.3 Typical Parameter Values

Table C.1 lists the dominant land cover types in the Arkansas-Red River basin and gives typical values for the key parameters required in the Hydros MEBM.

Bibliography

[1] G. Boni, F. Castelli, and D. Entekhabi. Sampling strategies and assimilation of ground temperature for the estimation of surface energy balance components. *IEEE Transactions in Geoscience and Remote Sensing*, 39:165–172, 2001.

[2] K. Brusdal, J.M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P. J. Van Leeuwen, E. Dombrowsky, and J. Verron. A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *Journal of Marine Systems*, 40-41:253–289, 2003. doi:10.1016/S0924-7963(03)00021-6.

[3] G. Burgers, P.J. van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6):1719–1724, 1998. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

[4] F. Castelli, D. Entekhabi, and E. Caporali. Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance. *Water Resources Research*, 35(10):3115–3125, 1999.

[5] F. Chen, K Mitchell, J. Schaake, Y. Xue, H-L. Pan, V. Koren, Q.Y. Duan, M. Ek, and A. Betts. Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research - Atmospheres*, 101(D3):7251–7268, March 20 1996. doi:10.1029/95JD02165.

[6] F. Chen, T.T. Warner, and K. Manning. Sensitivity of orographic moist convection to landscape variability: A study of the buffalo creek, colorado, flash flood case of 1996. *Journal of the Atmospheric Sciences*, 58(21):3204–3223, 2001.

[7] B. J. Choudhury, T.J. Schmugge, A. Chang, and R. W. Newton. Effect of Surface Roughness on the Microwave Emission From Soils. *Journal of Geophysical Research - Oceans and Atmospheres*, 84(C9):5699–5706, 1979.

[8] S.E. Cohn, N.S. Sivakumaran, and R. Todling. A fixed-lag Kalman smoother for retrospective data assimilation. *Monthly Weather Review*, 122:2838–2867, 1994. doi:10.1175/1520-0493(1994)122<2838:AFLKSF>2.0.CO;2.

[9] D.R. Cook. Energy Balance Bowen Ratio (EBBR) handbook. Technical Report ARM TR-037, Atmospheric Radiation Measurement (ARM) Program, 2005.

[10] B. J. Cosby, G. M. Hornberger, R. B. Clapp, and T. R. Ginn. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, 20:682–690, 1984.

[11] P. Courtier, J. Derber, R. Errico, J.-F. Louis, and T. Vukicevic. Important Literature on the use of Adjoint, Variational Methods and the Kalman Filter in Meteorology. *Tellus A - Meteorology and Oceanography*, 45A:342–357, 1993.

[12] W. L. Crosson, C.A. Laymon, R. Inguva, and M.P. Schamschula. Assimilating remote sensing data in a surface flux-soil moisture model. *Hydrological Processes*, 16(8):1645–1662, 2002.

[13] W. T. Crow. Correcting land surface model predictions for the impact of temporally sparse rainfall rate measurements using an ensemble kalman filter and surface brightness temperatures. *Journal of Hydrometeorology*, 4(5):960–973, 2003.

[14] W. T. Crow, S. T. K. Chan, D. Entekhabi, P. R. Houser, A. Y. Hsu, T. J. Jackson, E. G. Njoku P. E. O'Neill, J. Shi, and X. Xhan. An Observing System Simulation Experiment for Hydros Radiometer-Only Soil Moisture Products. *IEEE Trans. Geosci. Remote Sens.*, 43(6):1289–1303, 2005.

[15] W. T. Crow and E. F. Wood. The value of coarse-scale soil moisture observations for regional surface energy balance modeling. *Journal of Hydrometeorology*, 3(4):467–482, 2002.

[16] W. T. Crow and E. F. Wood. The assimilation of remotely-sensed soil brightness temperature imagery into a land surface model using ensemble kalman filtering: a case study based on estar measurements during sgp97. *Advances in Water Resources*, 26(2):137–149, 2003. doi:10.1016/S0309-1708(02)00088-X.

[17] G. R. Diak and C. Gauthier. Improvements to a simple physical model for estimating insolation from GOES data. *Journal of Climate and Applied Meteorology*, 22(3):505–508, 1983.

[18] M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. El-Reyes. Microwave dielectric behavior of wet soil - Part II: Dielectric mixing models. *IEEE Trans. Geosci. Rem. Sens.*, GE-23:35–46, 1985.

[19] M. B. Ek, K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley. Implementation of Noah land surface model advances in the national centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, 108(D22):8851, 2003. doi:10.1029/2002JD003296.

[20] D. Entekhabi, H. Nakamura, and E.G. Njoku. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Transactions in Geoscience and Remote Sensing*, 32:438–448, 1994. doi:10.1109/36.295058.

[21] D. Entekhabi, E. Njoku, P. Houser, M. Spencer, T. Doiron, J. Smith, R. Girard, S. Belair, W. Crow, T. Jackson, Y. Kerr, J. Kimball, R. Koster, K. McDonald, P. O'Neill, T. Pultz, S. Running, J.C. Shi, E. Wood, and J. van Zyl. The Hydrosphere State (hydros) mission concept: An earth system pathfinder for global mapping of soil moisture and land freeze/thaw. *IEEE Transactions in Geoscience and Remote Sensing*, 2004. In Press.

[22] D. Entekhabi, I. Rodriguez-Iturbe, and F. Castelli. Mutual interaction of soil moisture state and atmospheric processes. *Journal of Hydrology*, 184(1-2):3–17, 1996.

[23] D. et al. Entekhabi. An agenda for land surface hydrology research and a call for the second international hydrological decade. *Bulletin of the American Meteorological Society*, 80(9):1–16, 1999.

[24] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research - Oceans*, 99(C5):10143–10163, 1994. doi:10.1029/94JC00572.

[25] G. Evensen. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003. doi:10.1007/S10236-003-0036-9.

[26] G. Evensen and P.J. Van Leeuwen. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128(6):1852–1867, 2000. doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

[27] G. Evensen and P.J. VanLeeuwen. Assimilation of Geosat Altimeter Data for the Agulhas Current Using the Ensemble Kalman Filter with a Quasigeostrophic Model. *Monthly Weather Review*, 124(1):85–96, January 1996.

[28] J.F. Galantowicz, D.Entekhabi, and E.G. Njoku. Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. *IEEE Trans. Geoscience and Remote-Sensing*, 37(4):1860–1870, 1999. doi:10.1109/36.774699.

[29] H. Gao, E. F. Wood, M. Drusch, W. Crow, and T. J. Jackson. Using a microwave emission model to estimate soil moisture from ESTAR observations during SGP99. *Journal of Hydrometeorology*, 5(1):49–63, 2004.

[30] A. Gelb. Applied Optimal Estimation. *The MIT Press, Cambridge Massachusetts*, 1974.

[31] M. Ghil and P. Manalotte-Rizzoli. Data assimilation in meteorology and oceanography. *Advances in Geophysics*, 33:141–266, 1991.

[32] R. Giering and T. Kaminski. Recipes for adjoint code construction. *ACM Transactions on Mathematical Software*, 24(4):437–474, 1998.

[33] I. P. Gorenburg, D. McLaughlin, and D. Entekhabi. Scale-recursive assimilation of precipitation data. *Advances in Water Resources*, 24:941–953, 2001.

[34] R. Gronnevik and G. Evensen. Application of ensemble-based techniques in fish stock assessment. *Sarsia*, 86:517–526, 2001.

[35] A. Guha, J. M. Jacobs, T. J. Jackson, M. H. Cosh, E-C Hsu, and J. Judge. Soil moisture mapping using ESTAR under dry conditions from the southern great plains experiment (SGP99). *IEEE Trans. Geosci. Rem. Sens.*, 41(10):2392–2397, 2003.

[36] M. T. Hallikainen, F. T. Ulaby, and M. C. Dobson. Microwave dielectric behavior of wet soil - Part I: Empirical models and experimental observations. *IEEE Trans. Geosci. Rem. Sens.*, GE-23(1):25–34, 1985.

[37] K.L. Hawk and P.S. Eagleson. Climatology of station storm rainfall in the continental united states: parameters of the bartlett-lewis and poisson rectangular pulses models. Technical Report 336, Massachusetts Institute of Technology, 1992.

[38] A.W. Heemink, M. Verlaan, and A.J. Segers. Variance reduced ensemble Kalman filtering. *Monthly Weather Review*, 129(7):1718–1728, 2001.

[39] P.L. Houtekamer and H.L. Mitchell. Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, 126(3):796–811, 1998.

[40] P.L. Houtekamer and H.L. Mitchell. A Sequential Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, 129(1):123–137, 2001.

[41] M. D. Hudlow, J. A. Smith, M. L. Walton, and R. C. Shedd. *Hydrological Applications of Weather Radar*, chapter NEXRAD: New era in hydrometeorology, pages 602–612. I. Cluckie and C. Colliers, Eds.,Ellis Horwood, New York, New York, 1991.

[42] V. Y. Ivanov, E. R. Vivoni, R. L. Bras, and D. Entekhabi. Catchment hydrologic response with a fully distributed triangular irregular network model. *Water Resources Research*, 40:doi:10.1029/2004WR003218, 2004.

[43] T. J. Jackson, R. Bindlish, A. J. Gasiewski, B. Stankov, M. Klein, E. G. Njoku, D. Bosch, T. L. Coleman, C. A. Laymon, and P. Starks. Polarimetric scanning radiometer c- and x-band microwave observations during SMEX03. *IEEE Trans. Geosci. Rem. Sens.*, 43(11):2418–2430, 2005.

[44] T.J. Jackson. Soil moisture estimation using special satellite microwave/imager(SSM/I) satellite data over a grassland region. *Water Resources Research*, 33(6):1475–1484, 1997.

[45] T.J. Jackson, D.M. LeVine, A.Y. Hsu, A. Oldak, P.J. Starks, C.T. Swift, J.D. Isham, and M. Haken. Soil moisture mapping at regional scales using microwave radiometry: The Southern Great Plains Hydrology Experiment. *IEEE Trans. Geosci. Rem. Sens.*, 37:2136–2151, 1999. doi:10.1109/36.789610.

[46] T.J. Jackson and T.J. Schmugge. Vegetation effects on the microwave emission of soils. *Remote Sens. Environ.*, 36(3):203–212, 1991. doi:10.1016/0034-4257(91)90057-D.

[47] T.J. Jackson, D.M. Le Vine, C.T. Swift, T.J. Schmugge, and F.T. Schiebe. Large area mapping of soil moisture using the ESTAR passive microwave radiometer in Washita'92. *Remote Sensing of the Environment*, 54(1):27–37, 1995.

[48] A. H. Jazwinski. Stochastic processes and filtering theory. *Academic Press, New York, NY*, 1970.

[49] Thomas Kailath, A.H. Sayed, and B. Hassibi. Linear estimation. *Prentice Hall, Upper Saddle River, N.J.*, 2000.

[50] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (ASME)*, 1963.

[51] R.E. Kalman and R. Bucy. New results in linear filtering and prediction. *Journal of Basic Engineering (ASME)*, 1961.

[52] L. A. Klein and C. T. Swift. An improved model for the dielectric constant of sea water at microwave frequencies. *IEEE Trans. Antennas Propag.,*, AP-25:104–111, 1977.

[53] J. A. Kong. *Electromagnetic Wave Theory.* Wiley-Interscience,New York, 2nd edition, 1990.

[54] R. D. Koster, P. A. Dirmeyer, Z. Guo, G. Bonan, E. Chan, P. Cox, C. T. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C.-H. Lu, S. Malyshev, B. McAvaney, K. Mitchell, D. Mocko, T. Oki, K. Oleson, A. Pitman, Y. C. Sud, C. M. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada. Regions of Strong Coupling Between Soil Moisture and Precipitation. *Science*, 305:1138–1141, August 2004.

[55] D. M. Le Vine, T. J. Jackson, C. T. Swift, M. Haken, and S. W. Bidwell. ESTAR measurements during the southern great plains experiment (SGP99). *IEEE Trans. Geosci. Rem. Sens.*, 39(8):1680–1685, 2001.

[56] D. M. Le Vine, C. T. Swift, and M. Haken. Development of the synthetic aperture microwave radiometer, estar. *IEEE Trans. Geosci. Rem. Sens.*, 39(1):199–202, 2001.

[57] D.M. Le Vine, A.J. Griffis, C.T. Swift, and T.J. Jackson. ESTAR: A synthetic aperture microwave radiometer for remote sensing applications. *IEEE Proc.,*, 82:1787–1801, 1994.

[58] L. Ljung. Asymptotic behaviour of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 24(1):36–50, 1979.

[59] D. Lohmann, K.E. Mitchell, P.R. Houser, E.F. Wood, J.C. Schaake, A. Robock, B.A. Cosgrove, J. Sheffield, Q.Y. Duan, L.F. Luo, R.W. Higgins, R.T. Pinker, and J.D. Tarpley. Streamflow and water balance intercomparisons of four land surface models in the north americal land data assimilation system project. *Journal of Geophysical Research - Atmospheres*, 109(D7), April 2004. doi:10.1029/2003JD003517.

[60] A. C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.

[61] S. Margulis, D. Entekhabi, and D. McLaughlin. Spatio-temporal disaggregation of remotely-sensed precipitation for ensemble hydrologic modeling and data assimilation. *Journal of Hydrometeorology*, 2005. Submitted, 54pp.

[62] S. A. Margulis. *Variational sensitivity Analysis and data assimilation studies of the coupled land surface-atmospheric boundary layer system.* PhD thesis, Massachusetts Institute of Technology, 2002.

[63] S. A. Margulis and D. Entekhabi. Temporal disaggregation of satellite-derived monthly precipitation estimates and the resulting propagation of error in partitioning of water at the land surface. *Hydrology and Earth System Sciences*, 5(1):27–38, 2001.

[64] S. A. Margulis, D. McLaughlin, D. Entekhabi, and S. Dunne. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resources Research*, 38(12), 2002. doi:10.1029/2001WR001114.

[65] H. Nakamura, D. Entekhabi, and E. Njoku. Soil moisture profile retrieval by assimilation by multispectral remote sensing and surface observations. In *IEEE Geosciences and Remote-Sensing Annual Conference*, Pasadena, 1994.

[66] E. G. Njoku and D. Entekhabi. Passive microwave remote sensing of soil moisture. *Journal of Hydrology*, 184:101–129, 1996.

[67] J. Paegle, K.C. Mo, and J. Nogues-Paegle. Dependence of simulated precipitation on surface evaporation during 1993 United States summer floods. *Monthly Weather Review*, 124(3):345–361, 1996.

[68] Z. Pan, M. Segal, R. Turner, and E. Takle. Model simulation of impacts of transient surface wetness on summer rainfall in the United States Midwest during drought and flood years. *Monthly Weather Review*, 123:1575–1581, 1995.

[69] C. D. Peters-Lidard, M. S. Zion, and E. F. Wood. A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes. *Journal of Geophysical Research - Atmospheres*, 102(D4):4304–4324, 1997.

[70] H.E. Rauch, F.Tung, and C.T.Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450, 1965.

[71] R. Reichle. *Variational Assimilation of Remote Sensing Data for Land Surface Hydrologic Applications.* PhD thesis, Massachusetts Institute of Technology, 2000.

[72] R.H. Reichle, D. Entekhabi, and D.B. McLaughlin. Downscaling of radiobrightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. *Water Resour. Res.*, 37(9):2353–2364, 2001a. doi:10.1029/2001WR000475.

[73] R.H. Reichle, D. B. McLaughlin, and D. Entekhabi. Variational data assimilation of microwave radiobrightness observations for land surface hydrol-

ogy applications. *IEEE Trans. Geosci. Rem. Sens.*, 39(8):1708–1718, 2001. doi:10.1109/36.942549.

[74] R.H. Reichle, D.B. McLaughlin, and D. Entekhabi. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1):103–114, 2002. doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.

[75] A. Robock, K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai. The Global Soil Moisture Data Bank. *Bull. Amer. Soc. Meteor.*, 81(6):1281–1299, 2000.

[76] I. Rodriguez-Iturbe, V. K. Gupta, and E. Waymire. Scale considerations in the modeling of temporal rainfall. *Water Resources Research*, 20:1611–1619, 1984.

[77] A.J. Segers, A.W. Heemink, M. Verlaan, and M. VanLoon. A modified RRSQRT-filter for assimilating data in atmospheric chemistry models. *Environmental Modelling Software*, 15:663–671, 2000.

[78] B.L. Shaw, R.A. Pielke, and C.L. Ziegler. A three-dimensional numerical simulation of a great plains dryline. *Monthly Weather Review*, 125(7):1489–1506, 1997.

[79] A. Stogryn. Equations for calculating the dielectric constants of saline water. *IEEE Trans. Microwave Theory Techn.*, MIT-19:733–736, 1971.

[80] F. T. Ulaby, P. C. Dubois, and J. van Zyl. Radar mapping of surface soil moisture. *J. Hydrol.*, 184:57–84, 1996.

[81] F.T. Ulaby, R.K. Moore, and A.K. Fung. Microwave remote sensing, vol. I-III. *Artech House Inc., Norwood, MA.*, 1986.

[82] P. J. Van Leeuwen and G. Evensen. An ensemble smoother with error estimates. *Mon. Weather Rev.*, 129(4):709–724, 2001.

[83] P.J. van Leeuwen and G. Evensen. Data assimilation and inverse methods in terms of a probabalistic formulation. *Monthly Weather Review*, 124(12):2898–2913, 1996. doi:10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.

[84] M. Verlaan. *Efficient Kalman Filtering Algorithms for Hydrodynamic Models.* PhD thesis, Technische Universiteit Delft, 1998.

[85] M. Verlaan and A.W. Heemink. Nonlinearity in data assimilation applications: A practical method for analysis. *Monthly Weather Review*, 129(6):1578–1589, 2001.

[86] J. P. Walker and P. R. Houser. A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *Journal of Geophysical Research*, 106(D11):11,761–11,774, 2001.

[87] J. P. Walker, G. R. Willgoose, and J. D. Kalma. One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms. *Advances in Water Resources*, 24:631–650, 2001. doi:10.1016/S0309-1708(00)00043-9.

[88] J. R. Wang and B. J. Choudhury. *"Passive microwave radiation from soil: examples of emission models and observations", in Passive Microwave Remote Sensing of Land-Atmosphere Interactions.* VSP Publishing, Utrecht, 1995.

[89] J.R. Wang, P.E. O'Neill, T.J. Jackson, and E.T. Engman. Multifrequency measurements of the effects of soil moisture, soil texture, and surface roughness. *IEEE Trans. Geoscience and Remote Sensing*, 21:44–50, 1983.

[90] J.R. Wang and T.J. Schmugge. An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Trans. Geoscience and Remote-Sensing*, 18(4):288–295, 1980.

[91] C. Wunsch. The Ocean Circulation Inverse Problem. *Cambridge University Press, New York*, 1996.

[92] P. Xie, J. E. Janowiak, P. A. Arkin, R. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis. GPCP Pentad Precipitation Analyses: An Experimental Dataset Based on Gauge Observations and Satellite Estimates. *Journal of Climate*, 16(13):2197–2214, 2003. doi:10.1175/2769.1.

[93] Chen D. Y., J. F. Huang, and T. J. Jackson. Vegetation water content estimation for corn and soybeans using spectral indices derived from modis near- and short-wave infrared bands. *Remote Sensing of Environment*, 98(2-3):225–236, 2005.