Hillslope-scale Soil Moisture Estimation with a Physically-based Ecohydrology Model and L-band Microwave Remote Sensing Observations from Space

by

Alejandro Nicolas Flores
M.S. Civil Engineering, Colorado State University (2003)
B.S. Civil Engineering, Colorado State University (2001)

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 2009

Author .....
Department of Civil and Environmental Engineering
~~October~~ 24, 2008

Certified by .....
Rafael Luis Bras
Professor of Civil and Environmental Engineering
Dean of the Henry Samueli School of Engineering
University of California, Irvine
Thesis Supervisor

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

Accepted by .....
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

Hillslope-scale Soil Moisture Estimation with a Physically-based Ecohydrology Model and L-band Microwave Remote Sensing Observations from Space

by

Alejandro Nicolas Flores

Submitted to the Department of Civil and Environmental Engineering
on October 24, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in the Field of Hydrology

Abstract

Soil moisture is a critical hydrosphere state variable that links the global water, energy, and carbon cycles. Knowledge of soil moisture at scales of individual hillslopes (10's to 100's of meters) is critical to advancing applications such as landslide prediction, rainfall-runoff modeling, and wildland fire fuel load assessment. This thesis develops a data assimilation framework that employs the ensemble Kalman Filter (EnKF) to estimate the spatial distribution of soil moisture at hillslope scales by combining uncertain model estimates with noisy active and passive L-band microwave observations. Uncertainty in the modeled soil moisture state is estimated through Monte Carlo simulations with an existing spatially distributed ecohydrology model. Application of the EnKF to estimate hillslope-scale soil moisture in a watershed critically depends on: (1) identification of factors contributing to uncertainty in soil moisture, (2) adequate representation of the sources of uncertainty in soil moisture, and (3) formulation of an observing system to estimate the geophysically observable quantities based on the modeled soil moisture.

Uncertainty in the modeled soil moisture distribution arises principally from uncertainty in the hydrometeorological forcings and imperfect knowledge of the soil parameters required as input to the model. Three stochastic models are used in combination to simulate uncertain hourly hydrometeorological forcings for the model. Soil parameter sets are generated using a stochastic approach that samples low probability but potentially high consequence parameter values and preserves correlation among the parameters. The observing system recognizes the role of the model in organizing the factors effecting emission and reflection of L-band microwave energy and emphasizes the role of topography in determining the satellite viewing geometry at hillslope scales.

Experiments in which true soil moisture conditions were simulated by the model

and used to produce synthetic observations at spatial scales significantly coarser than the model resolution reveal that sequential assimilation of observations improves the hillslope-scale near-surface moisture estimate. Results suggest that the data assimilation framework is an effective means of disaggregating coarse-scale observations according to the model physics represented by the ecohydrology model. The thesis concludes with a discussion of contributions, implications, and future directions of this work.

Thesis Supervisor: Rafael Luis Bras
Title: Professor of Civil and Environmental Engineering
Dean of the Henry Samueli School of Engineering
University of California, Irvine

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

Acknowledgments

Many who know me know that I'm an avid fan of cycling and a cyclist myself. For me, the Ph.D. process has felt like I imagine a long stage race would; one of the Grand Tours[1]. It has had many stages along the way: long but flat stages where nothing happens, long but flat stages where the peloton seems in a hurry, undulating days with a maddening sequence of unclassified hills, individual time trials[2], and of course those big mountain stages of gargantuan climbs where the race is typically won or lost. In the Tour de France, the contenders all have designs on finishing in Paris in the *maillot jaune* three weeks later when the race starts. Yet at the end, those fortunate not to have crashed out, missed a time cut, or abandoned the race are happy just to be arriving with the peloton in Paris. Me, I'm just happy to be on the cobble-stoned Champs-Élysées in Paris.

Anyone that follows races like the Tour de France, knows that a rider requires the support of an army of people. This is my chance to thank those who supported my ride. And whether it's the Tour de France or NASCAR's Daytona 500, the first thing you always do during the post-race interview is thank the sponsors: thanks to the Army Research Office and Dr. Russ Harmon in particular for supporting this research and maintaining an interest in my academic training. I'd also like to thank Norman Leventhal and the Picower family who fund the fellowship in Norman's name for their support during my first year, and the Martin Family for their support of the Society of Fellows for Sustainability and providing one semester of support. Next I must profess my profound gratitude to my academic advisor Dean Rafael Bras, a man whom I respect deeply. Being able to contribute to and learn from him and his team has been an incredible experience. One of the reasons I decided to come to MIT to work with Rafael is that I knew that I would learn so much more than just hydrology. And while hydrology is at the core of what I've learned and accomplished, no one can escape Rafael's group without a more basic understanding of what it means to be a scientist, an educator, and a person balancing an immense passion for both of these with the other important things in life. That's why being a Bras student is transformational, and why there is such a bond between so many of his current and former students. I'd like to also thank my thesis committee members Professors Dara Entekhabi and Dennis McLaughlin of the Parsons Lab, Professor Enrique Vivoni of the New Mexico Institute of Mining and Technology (a Bras group alum), and (formerly) Dr. Jim Hansen of the Naval Postgraduate Research Lab. They were all pivotal in supervising the development and execution of the thesis work. In particular, I would like to thank Dara and Jim for their patience in helping me understand the in's and out's of data assimilation - a topic with which I was completely unfamiliar when I arrived at MIT.

I'd also like to acknowledge the members of the Bras group with whom I've

[1]There are three Grand Tours annually: the Giro D'Italia in May, the Tour de France in July, and the Vuelta a España in August

[2]Ahh, the general exams

had the privilege of working while at MIT. This includes current group members Jingfeng Wang, Homero Flores Cervantes, Ryan Knox, Gautam Bisht, Gajan Sivandran, Tony Parolari, Ujjwal Narayan, and Chiara Lepore. Former members to whom I owe much personal and professional gratitude include Valeriy Ivanov, Frederic Chagnon, Erkan Istanbulluoglu, Jean Fitzmaurice, Daniel Collins, Ola Hall, Giacomo Falorni, Hanan Karam, Fernando Nardi, Enrico Bono, Valerio Noto, and Miguel Campos. I'd like to thank Enrique Vivoni and Nicole Gasparini on a personal level for spending time with Emily and me during our initial visit to MIT in September 2002. Rafael's group also would cease functioning without the continuous support of Gayle Sherman, and her esteemed predecessor Elaine Pagliarulo.

Being a member of the Parsons Lab family during the past five years has allowed me to forge friendships that will last a lifetime and learn much about the world. I appreciate greatly my two closest friends in the lab Amy Mueller and Desirée Plata: thank you both so much for your friendship[3]. My infinite gratitude to members of my cohort on whom I have relied: Jon Winter, Eliza Peterson, Crystal Ng, Rebecca Neumann, Piyatida Hoisungwan, Arne Bomblies, and Behnam Jafarpour. I'd also like to thank other Parsonites, if not for helping my research then simply for their willingness to converse: Adel Ahanin, Susan Dunne, Sara Friedman, Sarah Jane White, Anne Lightbody, Charuleka Varadharajan, Sarah Bagby, (now Professor) Janelle Thompson, Paul Dombrowski, Marc Marcella, Matt Orosz, Mack Durham, Dana Hunt, Maureen Coleman, Jacob Waldbauer, Khandaker Ashfaque, Moji Bateni, Adam Rivers, Katya Frois-Moniz, and many, many others. Professors Janelle and Ruben Juanes offered great advice during my job hunt. I owe a special debt to Sheila Frankel and Harry Hemond who agreed to let me serve as Teaching Assistant for TREX. It was an important experience for me, and I'm truly grateful for the opportunity. I'd also like to thank Vicky Murphy for all of her help getting me out of financial predicaments that I knowingly walked into. I'm also grateful to Jim Long for routinely helping to make simple the many tasks that would have taken me hours on my own. Thanks also to the many people in Building 1 that have made life at MIT just a bit easier, and helped me when I was organizing the Hydrology Seminar: Kris Kipp, Cynthia Stewart, Donna Hudson, Debbie Levey, and Department Head Patrick Jaillet. I'd also like to thank Alison Hynd for twice giving me the opportunity to be a judge during the IDEAS competition.

I must also thank my academic advisors at Colorado State who were so instrumental in my academic training, and supportive of my decision to come to MIT. I believe my outlook on many things is still heavily influenced by Professor Brian Bledsoe, who together with Rafael have been most influential people in my education. I also thank Professors Ellen Wohl and Jorge Ramírez (another Bras group alum) for their continued support and encouragement. From those CSU days, thanks to my friends and fellow inaugural members of the Bledsoe group: David Pizzi, Russ Anderson, and Steve Sanborn. I have to thank a few other friends for their support both in Colorado and at MIT: Tracy Emanuel, Evan and Alison Schneider, David

[3]and being willing to slack off to go grab coffee at the "real" Starbuck's on a regular basis

Raff, and Joanna Doxey.

Thanks to my many extended family members in California for their support, particularly the Floreses and Tadiars, as well as my family in Texas. I am also grateful to my wife's family who has been so supportive of my work. I am and always will be so grateful to my dad Romeo Flores and mom Marcia Brislane who always supported me in their own ways, and always encouraged me to work hard. My dog Keana provides more support than she knows. Her lack of care about publications and failed experiments and simple needs have been much needed therapy for this MIT Ph.D. student.

Most importantly I'd like to thank my wife, partner, and best friend Emily. Without her support and love, none of this would be possible.

Lyrics to "Throwing Stones"
Performed by The Grateful Dead
Words by John Perry Barlow
Music by Bob Weir

Picture a bright blue ball, just spinning, spinnin free,
Dizzy with eternity.
Paint it with a skin of sky,
Brush in some clouds and sea,
Call it home for you and me.

A peaceful place or so it looks from space,
A closer look reveals the human race. Full of hope, full of grace
Is the human face,
But afraid we may lay our home to waste.

There's a fear down here we can't forget.
Hasn't got a name just yet.
Always awake, always around,
Singing ashes, ashes, all fall down.
Ashes, ashes, all fall down.

Now watch as the ball revolves
And the nighttime falls.
Again the hunt begins,
Again the bloodwind calls.
By and by, the morning sun will rise,
But the darkness never goes
From some men's eyes.

It strolls the sidewalks
and it rolls the streets,
Staking turf, dividing up meat.
Nightmare spook, piece of heat,
It's you and me.
You and me.

Click flash blade in ghetto night,
Rudies looking for a fight.
Rat cat alley, roll them bones.
Need that cash to feed that jones.
And the politicians throwin' stones,
Singing ashes, ashes, all fall down.
Ashes, ashes, all fall down.

Commissars and pin-stripe bosses
Roll the dice.
Any way they fall,
Guess who gets to pay the price.
Money green or proletarian gray,
Selling guns 'stead of food today.

So the kids they dance
And shake their bones,
And the politicians throwin' stones,
Singing ashes, ashes, all fall down.
Ashes, ashes, all fall down.

Heartless powers try to tell us
What to think.
If the spirit's sleeping,
Then the flesh is ink
History's page will thus be carved in stone.
And we are here, and we are on our own
On our own.
On our own.
On our own.

Instrumental

If the game is lost,
Then we're all the same.
No one left to place or take the blame.
We can leave this place and empty stone
Or that shinin' ball we used to call our home.

So the kids they dance
And shake their bones,
And the politicians throwin' stones,
Singing ashes, ashes, all fall down.
Ashes, ashes, all fall down.

Shipping powders back and forth
Singing black goes south and white comes north.
In a whole world full of petty wars
Singing I got mine and you got yours.
And the current fashion sets the pace,
Lose your step, fall out of grace.
And the radical, he rant and rage,
Singing someone's got to turn the page.
And the rich man in his summer home,
Singing just leave well enough alone.
But his pants are down, his cover's blown...

And the politicians throwin' stones,
So the kids they dance
And shake their bones,
And it's all too clear we're on our own.
Singing ashes, ashes, all fall down.
Ashes, ashes, all fall down.

Picture a bright blue ball,
Just spinnin', spinnin, free.
Dizzy with the possibilities.
Ashes, ashes, all fall down.
Ashes, ashes, all fall down.
Ashes, ashes, all fall down.
Ashes, ashes, all fall down.
Ashes, ashes, all fall down.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis addresses the estimation of soil moisture using both observations and models. Prior to delving into the methodologies used in this thesis, the technical aspects of implementing those methods, and presentation of results, it is critical to make the case for the importance of this work. This chapter introduces the importance of soil moisture in the global hydrologic cycle, including a survey of the magnitude of the stores of water and fluxes that make up the water cycle. After this survey, which is designed to highlight the role of soil moisture as a dynamic interface for water vapor exchange between the land and atmosphere, the importance of knowledge at hillslope scales (10's to 100's of meters) is stressed. The chapter concludes with an outline of the remainder of the thesis.

1.1 Soil moisture and the global hydrologic cycle

Water is the chemical compound most responsible for the presence of life on planet Earth. For example, water is central to plant photosynthesis and the citric acid cycle, two chemical reactions that together are responsible for the conversion of light into carbohydrate and then from carbohydrate to metabolic energy. Water is also a critical component in moderating Earth's weather and climate. Atmospheric water vapor is one of the greenhouse gases (along with carbon dioxide, methane,

hydro-fluorocarbons and a host of other atmospheric gases) that acts to trap infrared radiation emitted at the land surface in the atmosphere. This "Greenhouse Effect," first quantitatively studied by Arrhenius 1896, leads to the ambient temperatures that are higher than those predicted by a simple equilibrium black-body radiative transfer model. Because the triple point of water is 273.16 K, the moderate temperatures associated with the "Greenhouse Effect" on Earth imply that water can exist simultaneously in all three phases in the Earth system. Thus, the chemical composition of the Earth's atmosphere, in the context of the energy output of the Sun and the distance between the Earth and the Sun are the factors that lead to a planet in which spatiotemporal thermodynamic gradients can create a dynamic water cycle [1].

Ultimately, incoming solar radiation serves as the fuel that drives the engine of the global hydrologic cycle (figure 1-1). Solar radiation is supplied at the top of the atmosphere at a rate of approximately 1350 W/m^2 [Eagleson, 1970]. The Earth's rotation about its axis and curvature lead to spatiotemporally varying solar input, leading to a diurnal temperature cycle and latitudinal variation in long-term averaged temperatures. In the atmosphere, differential solar heating induces poleward transport of heat energy. Because water has a high latent heat of vaporization and a high specific heat capacity, it plays a critical role in the redistribution of energy in the atmosphere. Evaporation from ocean and land surfaces dissipate input solar radiation, while condensation and precipitation release heat energy into the atmosphere (figure 1-1). Associated with this energy transport is the mass transport of water. From the perspective of the Earth's surface, evaporation removes water from the ocean and land surface, while precipitation injects water into the ocean and land systems. At the land-atmosphere interface, incoming solar energy is converted into sensible, latent, and ground heat. The magnitude of the latent heat flux is limited by: (1) the availability of liquid water, (2) the availability of energy to vaporize liquid water, and (3) the existence of a transport mechanism to remove moist air

[1] Astronomers have defined the so-called Circumstellar Habitable Zone as a region in the space surrounding stars where, based on the luminosity of the star, orbiting exoplanets could exhibit surface temperatures that would notionally sustain liquid water.

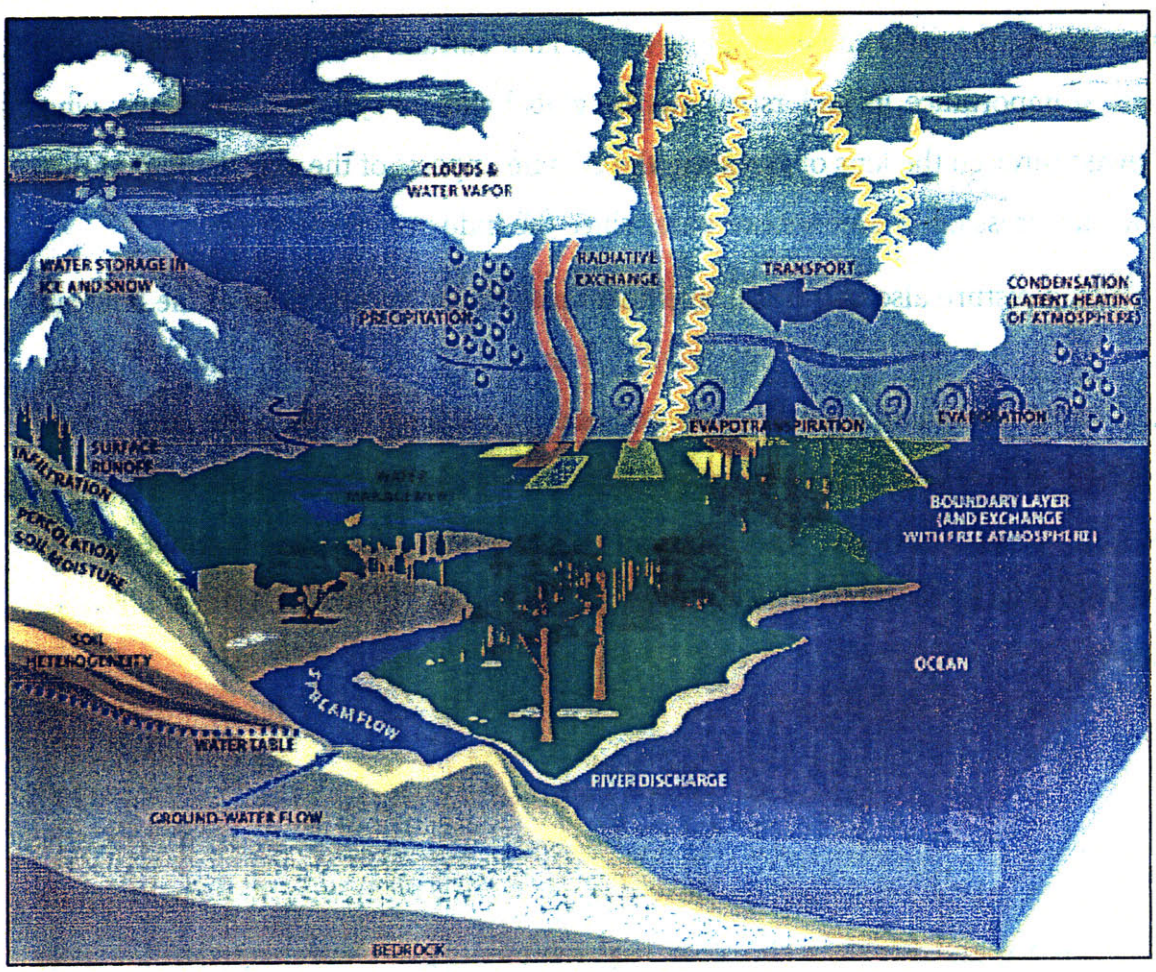


Figure 1-1: A diagrammatic representation of the hydrologic cycle, including both fluxes of mass and energy. Taken from a Climate Change Science and Subcommittee on Global Change Research report 2003.

from the near surface in order to sustain a vapor pressure deficit required to evaporate moisture into the air. Latent heat flux, in turn, serves as a forcing at the bottom boundary of the atmospheric column that, depending on atmospheric conditions, can initiate a suite of processes that ultimately lead to rainfall.

Because a supply of liquid water in the form of near-surface soil moisture or water stored in plants is needed for evaporation, and because latent heat flux can trigger atmospheric convection, the precipitation and latent heat flux processes are coupled over land masses. The strength of this coupling varies varies, although modeling studies suggest that coupling is weakest nearer the ocean and equator and highest in interior semiarid regions [Koster *et al.*, 2004]. Understanding and

quantifying this non-linear coupling of the land surface with the atmosphere is of critical importance in understanding the global climate system, particularly when viewed through the lens of assessing the climate impacts of the accelerating anthropogenic emission of greenhouse gases since the Industrial Revolution.

Soil moisture also imposes a significant mass-balance constraint at the land surface because it controls the partitioning of incoming precipitation into infiltration and runoff. The nature of the mass and energy coupling between the land surface and the atmosphere is best illustrated by briefly reviewing the simplified equations of state for soil moisture. The change in volumetric soil moisture mixture fraction (θ) with time is given by Richards equation [Celia *et al.*, 1990; van Dam and Feddes, 2000],

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[D(\theta)\frac{\partial \theta}{\partial z} + K(\theta)\right] - S(z), \tag{1.1}$$

where t is time, z is depth (positive down), D is the diffusivity, K is the hydraulic conductivity, and $S(z)$ is a sink term that represents plant water uptake. Equation 1.1 specifically highlights the vertical direction (z) because moisture gradients are generally the sharpest, and moisture fluxes therefore the greatest, in this direction. Since both D and K depend on the moisture status at time t, equation (1.1) is a nonlinear partial differential equation for which no generic analytical solution exists. In general, D and K depend not only on the current moisture, but also on the time-history of wetting and drying. Furthermore, because θ is bound to lower and upper limits, there exist fundamental constraints on how quickly (or slowly) water can infiltrate into partially-saturated soils. In numerical solutions of equation (1.1) this hysteresis is often neglected and nonlinear functional relationships between D and K and θ assumed (e.g., Brooks and Corey [1964]). Adding to the complexity of the problem, D and K are anisotropic and vary at significantly small spatial scales [Vogel and Roth, 2003]. Given initial and boundary conditions, equation (1.1) can be solved numerically, but the nonlinear nature of the constitutive relationships between D, K and θ requires significant attention to the discretization of the area

under study and the numerical solution scheme (Celia *et al.* [1990]; van Dam and Feddes [2000]). Hydrologic and applied mathematics literature burgeon with studies dedicated to schemes solving equation (1.1) and its more simplified forms under a spectrum of different conditions and assumptions. In modeling the spatial distribution of soil moisture, most models impose the following flux boundary conditions at the surface:

1. Infiltration flux, f is equal to precipitation, P, for $0 < P \leq f_{limit}$.
2. f is equal to $K(\theta)$ for $P > f_{limit}$. The difference $P - f_{limit}$ becomes the instantaneous runoff rate R.
3. If $P = 0$ and sufficient available energy and transport mechanisms are present, the bare soil evaporation (E) is the minimum of f_{limit} and the potential evaporation E_p that depends on the available energy and the vapor pressure deficit.

The maximum infiltration rate, f_{limit} is equal to the product of the soil moisture-dependent hydraulic conductivity ($K(\theta)$) and the vertical moisture gradient at the land-atmosphere interface. Specifically,

$$f_{limit} = K(\theta)\frac{\partial \theta}{\partial z}\bigg|_{z=0}. \tag{1.2}$$

Note the latent heat flux associated with the mass transfer of E from the soil to the atmosphere is simply λE, where λ is the latent heat of vaporization of water.

The term $S(z)$ in equation (1.1) suggests an interaction between soil moisture and vegetation over a range of depth within the soil column, the dynamics of which can be controlled by biophysical and biochemical properties of the plant itself. Liquid water enters plant root tissue through osmosis, and then is transported in the xylem tissue to the leaves where it is required for the carboxylation component of the Calvin Cycle [Taiz and Zeiger, 2002]. The photosynthetic process, in which atmospheric carbon dioxide diffuses into stomatal tissues while water vapor diffuses out, is at the core of soil moisture control by plants, because the water vapor in

plant stomata is ultimately derived from soil water. The photosynthesis process also underscores the coupling between terrestrial water, energy, and carbon cycles. Instantaneous plant water demand in plants is set by the the rate at which photosynthesis can proceed and depends on a number of variables that include energy availability, plant stress, and phenology [Kramer, 1983]. However, the supply of soil moisture ultimately serves as a constraint on the ability of plants to meet those demands for transpiration. In hydrologic and climate sciences, the dynamics of vegetation are being treated in increasingly complex ways that capture the growth, stress, and senescence of vegetation, therefore incorporating feedbacks between vegetation and soil moisture dynamics.

The above discussion is an introductory depiction of the nonlinear, multidirectional, and conditional nature of soil moisture dynamics and underscores some of the complications involved in modeling soil moisture. From this discussion we can conclude that soil moisture is: (1) a critical hydrosphere state variable because it couples global water, energy, and carbon cycles, and (2) difficult to model owing to the nonlinear nature of the equations governing soil moisture dynamics and the intrinsic coupling of soil moisture to atmospheric, biochemical, and biophysical processes. To motivate understanding about the potential magnitude of soil moisture dynamics in the global water cycle, what follows is an overview of previous work to quantify the stores and fluxes that comprise the global water cycle.

1.2 Magnitude of the global hydrologic cycle

This section provides a brief overview of the magnitude of both the storage terms and fluxes of the global hydrologic cycle. This is designed to motivate the discussion of soil moisture in the context of the terrestrial hydrologic cycle by providing some quantitative estimates of the mass and energy fluxes that occur across the soil-atmosphere interface. A significant body of scientific literature surveys the terms of the global water cycle, both for purposes of constraining climate and eval-

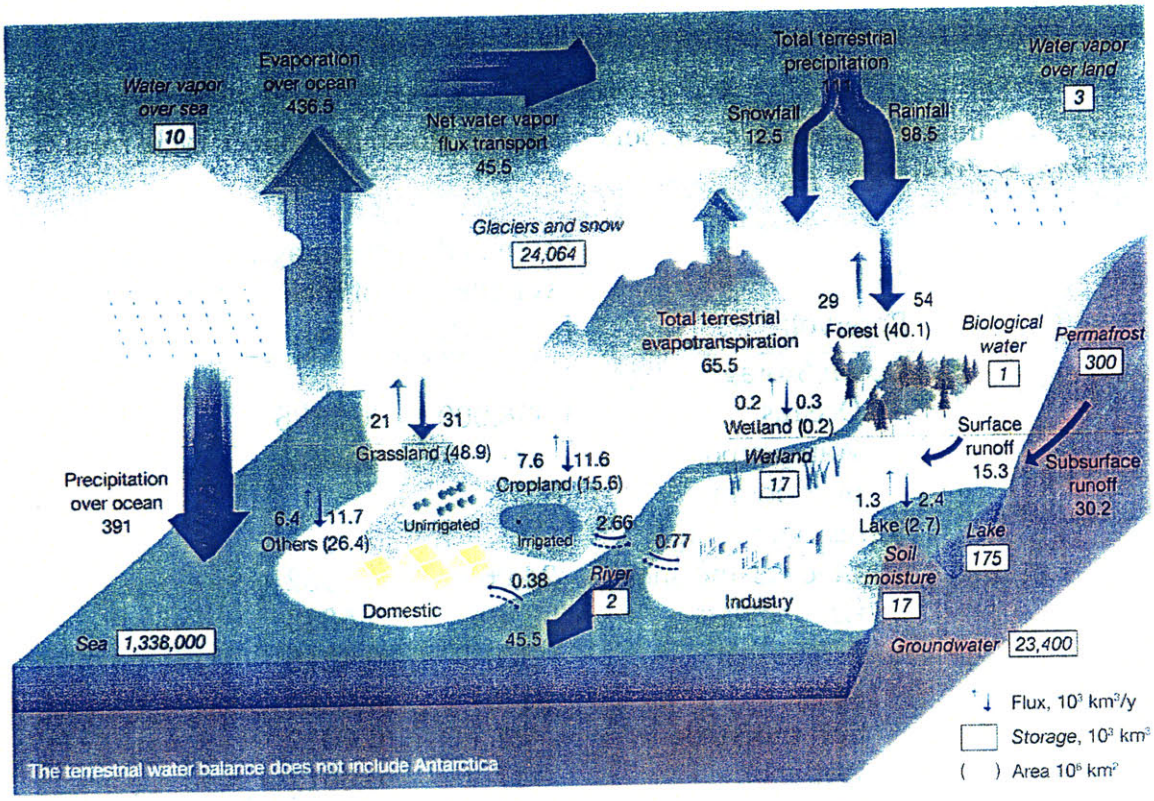


Figure 1-2: A diagrammatic representation of the hydrologic cycle together with estimates of storages and fluxes from Oki and Kanae [2006]. Storage terms are given in units of 10^3 km^3, fluxes in 10^3 km^3/y, and areas in 10^6 km^2.

uating global water resources (e.g., see Korzun [1978]; Shiklomanov [1997]; Chen *et al.* [1998]; Douville [1998]; Vorosmarty *et al.* [2000]; Oki [2005]; Dirmeyer *et al.* [2006]; Oki and Kanae [2006]). In a recent review paper that addresses potential future challenges to global water supply, Oki and Kanae [2006] synthesized the works of Korzun [1978]; Shiklomanov [1997]; Oki [2005]; Dirmeyer *et al.* [2006] and provide a useful graphic illustration of the stores and fluxes of the global hydrologic that is shown in Figure 1-2. Their survey provides adequate detail for the purposes of this discussion, and are therefore referred to throughout this section.

The stores are summarized in Table 1.1. Not surprisingly, the world's oceans account for the vast majority of the water on Earth, comprising more than 96% of the total. Of the remaining water on Earth, more than 3% is resides in groundwater aquifers, glaciers and snowpack. Soil moisture, by contrast accounts for about one one-thousandth of one percent of global water, containing only about 30% more

water than exists in the atmosphere globally at any time (Table 1.1).

Table 1.1: Summary of storage components in the global hydrologic cycle from Oki and Kanae [2006]

Reservoir	Estimated volume [10^3 km^3]	Percentage of total [%]
Ocean Storage		
Oceans	1,338,000	96.5376
Water vapor over ocean	10	0.0007
Terrestrial Storage		
Glaciers and snow	24,064	1.7362
Groundwater	23,400	1.6883
Permafrost	300	0.0216
Lakes	175	0.0126
Soil moisture	**17**	**0.0012**
Wetlands	17	0.0012
Water vapor over land	3	0.0002
Rivers	2	0.0001
Biological water	1	0.0001
Approximate total	1,385,989	100.0000

To understand the magnitude of the fluxes coming into the terrestrial hydrologic cycle, the breakdown of global precipitation is shown if Table 1.2. Note that the precipitation fluxes are expressed both in terms of mass flux into each store, and as equivalent depth of mean annual precipitation by normalizing the mass flux by the appropriate area. Annually, some 502×10^3 km^3 of rainfall (roughly 1 m in equivalent depth) falls on the surface of the Earth. About one-fifth of the total precipitation (111.0×10^3 km^3/y or 829 mm/y) falls over land masses, approximately 10% of it as snow. Given the annual terrestrial precipitation rate and given that the amount of water vapor stored in the atmosphere over land is approximately 3000 km^3, the average residence time for water vapor in the atmosphere over land is approximately 10 days (the global average residence time is roughly the same, although both atmospheric water vapor storage over the ocean and ocean precipitation are much larger than the corresponding terrestrial terms). This average

terrestrial atmospheric water vapor residence time underscores the temporally dynamic nature of the precipitation process as a forcing to the soil moisture process at the land-atmosphere interface. The presence of semiarid and arid lands, as well as tropical and temperate rainforests underscores the significant spatial heterogeneity in annual rainfall patterns, despite a relatively short residence time for water vapor in the atmosphere.

Table 1.2: Summary of precipitation fluxes comprising the global hydrologic cycle from Oki and Kanae [2006]

Precipitation flux	Area [10^6 km^2]	Annual rate [10^3 km^3/y]	Annual rate [mm/y]
Forest precipitation	40.1	54.0	1347
Grassland precipitation	48.9	31.0	634
Cropland precipitation	15.6	11.6	744
Lake precipitation	2.7	2.4	889
Wetland precipitation	0.2	0.3	1500
Other terrestrial precipitation	26.4	11.7	443
Total terrestrial precipitation	133.9	111.0	829
Snow		12.5	93
Rain		98.5	736
Ocean precipitation	376.2	391.0	1039
Global precipitation	510.1	502.0	984

In contrast with terrestrial precipitation, which injects mass to the terrestrial hydrologic cycle, Table 1.3 provides a breakdown of the global annual evapotranspiration process. In a similar fashion to Table 1.2, evapotranspiration fluxes are expressed both in terms of mass flux out of each store, and as equivalent depth of mean annual evapotranspiration by normalizing each mass flux by the appropriate area. In terms of mass flux, evapotranspiration from forests constitutes the largest single source of terrestrial evapotranspiration. This is likely no surprise, owing to the biomass productivity associated with forests in tropical regions of the globe. Although the contribution of wetlands to global evapotranspiration is small (it is roughly one-sixth the annual amount of lake evaporation), wetlands exhibit

the highest annual rate of evapotranspiration, which underscores the productivity of wetlands. If it is assumed that the size of soil moisture storage does not change substantially over time, and that evapotranspiration in forests, grasslands, and croplands originates from the soil moisture and biological water stores[2], then the average residence time for moisture in the soil is on the order of 25 days. It is expected, however, that there is significant variation about this average residence time throughout the terrestrial ecosystem, owing to variation ecological climatology throughout the world. Nevertheless, given the relatively short residence times of moisture in the atmosphere over land and in soil and biological moisture stores, it is readily apparent that the soil moisture process is both temporally dynamic and spatially variable.

The final discussion in this section seeks to place the hydrologic cycle in the context of the energy required to sustain it. As mentioned previously it is the Sun that drives the water cycle on planet Earth, and the amount of solar energy absorbed by Earth surfaces annually is 2700 ZJ [Smil, 2003][3]. To provide a benchmark of the energy required to vaporize water to sustain global evapotranspiration, each annual volume flux is converted to an annual energy flux by multiplying the mass flux by the density of water (taken as 1000 kg/m^3) and its latent heat of vaporization (taken as 2260 J/kg). Understandably, these energy flux numbers are astronomically high when expressed using Joules per year. Therefore, the energy required to sustain each evapotranspiration flux in Table 1.3 is expressed as the number of 100 megaton TNT Tsar Bombas[4] that would need to be detonated **per day** to achieve the same energy yield. As seen, the energy required for ocean evaporation is greater than any single terrestrial evapotranspiration source by an order of magnitude. What is clear in Table 1.3 is that the amount of energy required to drive the

[2]These assumptions are admittedly suspect. For instance: (1) tropical forest wet canopy evaporation is likely non-negligible, (2) many tree species have shown an ability to utilize water from saturated aquifers, and (3) a substantial amount of cropland evapotranspiration likely arises from evaporation of irrigated water.

[3]1 ZJ = 1×10^{21} J

[4]The largest thermonuclear weapon ever detonated and the most powerful single device created by humans, the Soviet "Tsar Bomba" had a nominal yield of 100 megatons TNT. Its actual yield was closer to 50 megatons TNT.

global hydrologic cycle, even on a daily basis, is thousands of times greater than the most energetic device every produced by humans. It is interesting to note, however, that the Intergovernmental Panel on Climate Change's (IPCC) Fourth Assessment Report (AR4) estimated the anthropogenic component of radiative forcing to be approximately 1.6 W/m^2. When integrated over the surface of the Earth and converted this corresponds to approximately an additional 169 Tsar Bomba's worth of energy per day.

Table 1.3: Summary of evapotranspiration fluxes comprising the global hydrologic cycle from Oki and Kanae [2006]

Evapotranspiration flux	Annual rate [10^3 km^3/y]	Annual rate [mm/y]	Number of Tsar Bombas per day
Forest evapotranspiration	29.0	723	429
Grassland evapotranspiration	21.0	429	311
Cropland evapotranspiration	7.6	487	112
Lake evaporation	1.3	481	19
Wetland evapotranspiration	0.2	1000	3
Other terrestrial precipitation	6.4	242	95
Total terrestrial evapotranspiration	65.5	489	969
Ocean evaporation	436.5	1160	6460
Global evapotranspiration	502.0	984	7429

Although it is quite clear that human activity exerts a positive radiative forcing, it is considerably less clear how that forcing translates to changes in the hydrologic cycle. In preparation of the IPCC's AR4, Meehl *et al.* [2007] studied the effects of anthropogenic activity on hydroclimatology through the use of a suite of climate models. Figure 1.3 shows the global distribution of changes in the mean annual precipitation, soil moisture, runoff, and evaporation for the period 2080-2099 relative to 1980-1999 for the SRES A1B scenario. Stippled areas in these plots indicate locations around the globe where at least 80% of the models agree in sign. Although the models fairly consistently model the directionality of impacts on global precipitation and evaporation over ocean areas (figure 1.3(a) and 1.3(d)), over

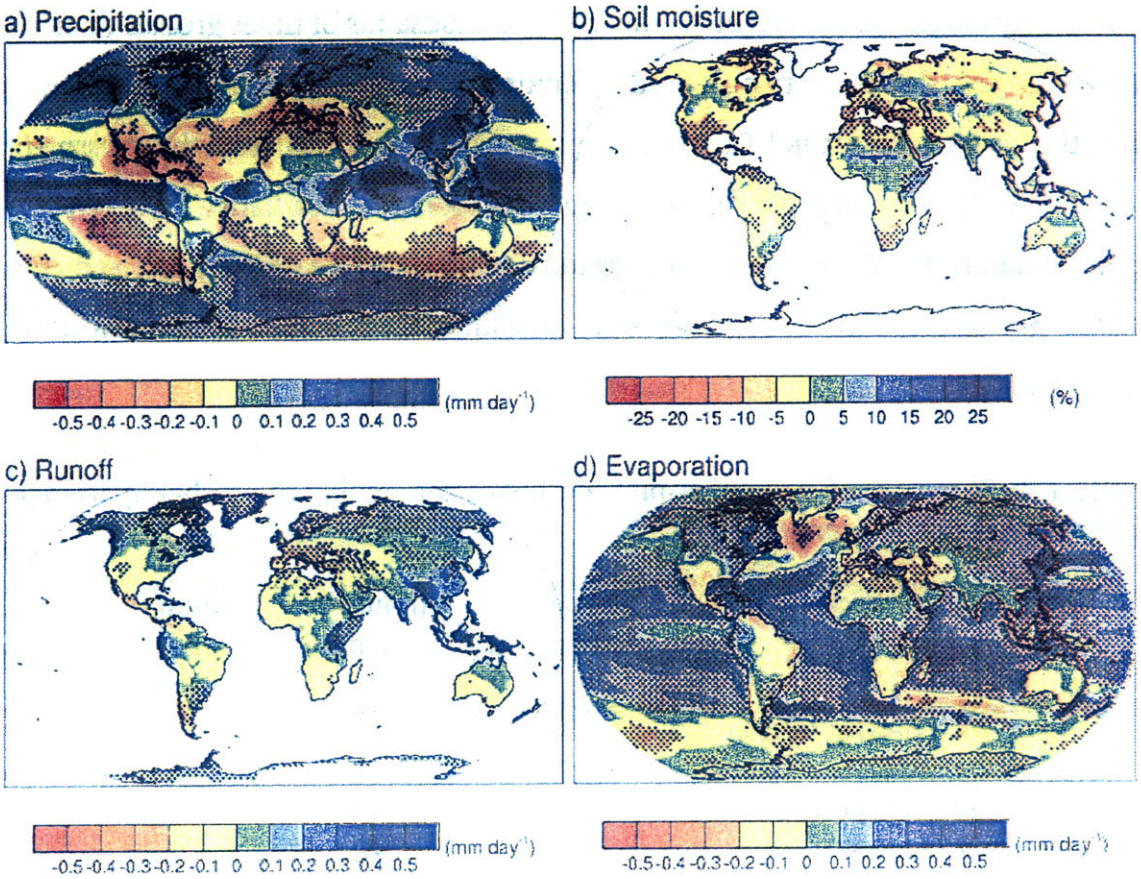


Figure 1-3: Figure and caption taken from Meehl *et al.* [2007]. Multi-model mean changes in (a) precipitation [mm/day], (b) soil moisture content [%], (c) runoff [mm/day], and (d) evaporation [mm/day]. To indicate consistency in the sign of the change, regionds are stippled where at least 80% of the models agree on the sign of the mean change. Changes are annual means for the SRES A1B scenario for the period 2080 to 2099 relative to 1980 to 1999. Soil moisture and runoff changes are shown at land points with valid data from at least 10 models.

land there is considerable disagreement between model projections. In particular, soil moisture is predicted to modestly decrease over most of the globe, however there is little consensus in the model predictions about the local direction of change (figure 1.3(b)). Furthermore, many of the locations where the models do agree about the direction of soil moisture change are relatively dry locations (e.g., the Northern Sahara, the Tibetan Plateau, the High Andes, Western Australia, the Kalahari Desert, and much of the Middle East). The implications of these results, which are presented in the AR4, together with the overview of the global hydrologic cycle presented above, highlights the need for improved predictability of soil moisture.

1.3 Practical necessity of soil moisture knowledge

The above discussion underscores the idea that soil moisture is a temporally dynamic and spatially heterogeneous hydrosphere state variable, and because soil moisture affects the partitioning of mass and energy exchange between the land surface and atmosphere there applications for which knowledge of soil moisture is critical. As a way of outlining the motivation for advancing the ability to predict soil moisture at the spatial resolutions addressed in this thesis, this section briefly discusses five of these applications that would potentially benefit from soil moisture estimates at the targeted spatial scales. Specifically, this section identifies and outlines how soil moisture could benefit each of the following applications: (1) prediction of terrain mobility and maneuverability in military applications, (2) irrigation scheduling, (3) rainfall-induced landslide susceptibility forecasting, (4) flood forecasting, and (5) wildland fire fuel load estimation. In discussing each of these potential applications, specific attention is paid to the scale at which knowledge of soil moisture is required. Prior to discussing each individual application it is important to highlight the similarities they all share, aside from a requirement of knowing soil moisture in high spatial resolution. The common threads that these applications share impact the viewpoint of the thesis work in ways that will become apparent later on. First, they are all ***real-time*** applications. That is, they are all applications in which some estimate of the spatial distribution of soil moisture is required now through some time in the near future. Second, these applications all implicitly involve identification of some kind of ***risk*** of occurrence of a potentially negative consequence (i.e., landslides, floods, and fire), which a knowledge of soil moisture may help mitigate against. Evoking the concept of risk in these applications necessitates a treatment of uncertainty in estimates of the spatial distribution of soil moisture, which will become a central part of the thesis work. Finally, these applications all involve discrete events that occur in the Earth system as a consequence of forcings and boundary conditions imposed by climatology, ecology, and geology. Therefore, changes in the frequency, magnitude, and any relationship be-

tween frequency and magnitude (the regime) of these discrete events is a potential local expression of global change.

The moisture and energy states and fluxes significantly impact U.S. Army operations. Soil moisture directly impacts the ability to move troops and materiel over land, as soil wetness impacts the dynamic load-bearing capabilities of natural soils [Hillel, 1998]. Evaluation of soil strength for Army applications is traditionally done through a combination of in situ measurements involving, for instance, cone penetrometers, shear vanes, and other handheld equipment. The trafficability of the soil is then assessed by comparing the measurements against vehicle-specific thresholds meant to convey the ease or difficulty of passibility [Shoop, 1993]. Not only is realtime trafficability assessment of importance to planning and execution of Army operations, but it is also critical to the sustainable management of military lands, as training exercises conducted under conditions that can significantly degrade soils may have long-lasting and undesirable consequences. Therefore, the ability to remotely assess trafficability conditions in realtime over large areas in adequate spatial detail (e.g., 10s to 100s of m) with some combination of models and observational data is of significant importance to Army operations. Remote trafficability assessment is a complex logistical issue because it involves the interplay of several important factors: (1) the spatial heterogeneity of soils and their intrinsic and moisture-dependent load-bearing capacities, (2) the spatial variability of soil moisture and its effect on the load-bearing capabilities of soil, (3) the timing of and likely weather conditions during the intended operation as it pertains to the local trafficability conditions *as the materiel arrives*, (4) the composition of the tactical formation and the corresponding vehicle-specific sensitivities to soil load-bearing capacities, and (5) the size and arrangement of the tactical formation and the degree to which the operation will itself impact local trafficability. It must be underscored that other secondary and tertiary sensitivities to moisture and energy conditions exist. For instance, optical-electrical night vision equipment is sensitive to the physical soil and canopy surface, and sky temperatures, and therefore impacted by soil moisture. Additionally, excessive helicopter rotorwash, which is a

function of soil texture and wetness, can compromise a concealed presence during the daytime, and endanger personnel during night vision-assisted landings. The important contribution of this thesis work to this application lies in the critically-important prediction of soil moisture in sufficiently high detail for Army operations using data and models.

In many cultivated lands, precipitation must be augmented with irrigation from either groundwater aquifers or surface waters to realize sufficient crop yields to achieve some desired utility (e.g., profit maximization). In market-based agricultural settings, the realized profit of a crop at harvest is determined by the market or negotiated price of the crop, the amount of the crop produced (the yield), and the integrated costs incurred to produce a particular harvest. The incurred costs range from taxes paid on the land to operating costs for planting and harvesting machinery to the unit costs of water. The amount, timing, and duration of irrigated water that must be applied (the irrigation schedule) is a function of: (1) the type of crops being grown and their developmental stage as it pertains to crop water demand, stress, and water use efficiency, (2) the current soil wetness conditions (or from the agronomist's perspective, the soil moisture deficit), (3) the expected near-term weather conditions, (4) the structure and terms of the water allotment (i.e., the water rights), and (5) any external demands related to, for example, leaching requirements to prevent soil salinization [Rhenals and Bras, 1981]. Many irrigation technologies, such as center pivot sprinkler and drip irrigation, allow some degree of control over the plot-scale distribution of applied irrigation water. In areas where the price of water constitutes a significant portion of growing costs or where the allotment of water is fixed, scheduling irrigation in a manner that minimizes water use while negligibly affecting yields can improve profit margins. Irrigation scheduling can potentially be substantially improved through an enhanced ability to characterize and predict the spatial distribution of soil moisture, as has been previously found [Bras and Cordova, 1981; Aboitiz *et al.*, 1986]. The integrated demand of irrigation water for a growing season within an irrigation district, along with the structure of the water rights within the district, are of critical importance

to agencies tasked with providing adequate water supply (such as the Army Corps of Engineers and the Bureau of Reclamation in the U.S.). Management of reservoirs, vis-á-vis the timing, duration, and rate of release, is a complex optimization and control problem that involves not only demands from agriculture, but also potentially power generation, recreation, and increasingly in-stream channel maintenance. Although seasonal scale planning of reservoir release schedules is most sensitive to the current storage, climatology, and forecast climatological anomalies (e.g., the El Niño Southern Oscillation), realtime *adaptive* management of reservoirs could be improved if farmers and growers had access to accurate knowledge of soil moisture in sufficiently high detail to allow for more optimal irrigation scheduling Georgiou *et al.* [2006]. Hence, this thesis makes a potentially important contribution to agricultural applications such as irrigation scheduling, and indirectly to realtime adaptive management of surface water impoundments.

Landslides pose a significant hazard to life and property throughout the U.S. and the world, and are ubiquitous in steep terrain. It has been previously estimated that casualties and economic losses due to landslides in the U.S. equate annually to approximately 25-50 deaths and $1-2B [Schuster and Fleming, 1986]. These figures are likely greater today due to development in landslide-prone areas, changing climate, and other anthropogenic disturbance. Effects of individual landslides can be staggering to individual communities as evidenced by several widely reported events in Guinsaugon, Philippines; La Conchita, California; Cordillera de la Costa, Venezuela; and Mameyes, Puerto Rico. These risks have motivated significant scientific effort to understand the interplay between geologic, hydrologic, and meteorologic processes associated with mass-wasting, and to develop methodologies to predict and mitigate effects of mass-wasting. Mass-wasting is commonly triggered by significant rainfall, seismic activity, volcanic eruption, and human disturbance. Dai *et al.* [2002] report that extreme rainfall is one of the most common triggering mechanisms. Hillslope hydrology ultimately modulates rainfall triggering of landslides, and is sensitive to: (1) the antecedent conditions, (2) the hydraulic properties of the soil-bedrock matrix, (3) the near-surface soil and vegetation conditions,

and (4) the spatiotemporal characteristics of rainfall. High-resolution digital elevation models (DEMs) are an enabling factor in the study of mass-wasting from a physically-based perspective. This is largely because DEMs can serve as the topographic boundary conditions for dynamic hydrology models that resolve the soil pore pressure distribution, which affects slope stability, at spatial scales of individual hillslopes and temporal scales coinciding with landslide occurrence. Studies coupling slope-stability and seepage analysis have identified the promise of DEMs for mass-wasting assessment [Okimura and Nakagawa, 1988; Montgomery and Dietrich, 1994; Casadei *et al.*, 2003]. In realtime forecasting of landslides, the initial soil moisture state and the associated pore pressure distribution is critical to being able to issue a reliable forecast about the risks of landsliding in an area given uncertain forecasts of precipitation. It is apparent that reliable prediction of individual landslide events requires knowledge of soil moisture at spatial scales of individual hillslopes and with lead times sufficiently long to allow implementation of mitigation activities.

Flood forecasting has long been a central application in hydrology, and it has long been recognized that river discharge depends on the nature of the rainfall falling on a watershed, and on the initial moisture conditions in the watershed. For instance, even the Soil Conservation Service curve number approach (1968) to flood runoff prediction allows for variation of curve number (and therefore runoff potential) on the basis of "antecedent moisture conditions." Channel networks serve as conduits for the runoff produced from hillslopes in a watershed, aggregating runoff in a downstream direction through lateral influxes of water from adjoining hillslopes and (more importantly) through interception of other tributary channels. The spatial density and connectivity of channel networks accounts for a significant degree of the variation in discharge regimes between watersheds of similar size, soil characteristics, and climate. However, variation in the timing and magnitude of discharge in a particular watershed are a function of variation in the spatiotemporal distribution of runoff production between those events. As outlined in the above survey of the Richards equation, runoff production at a point is a nonlinear func-

tion of soil moisture conditions, the local rainfall rate, and the properties of the soil. Hence, the distribution of runoff production in a watershed depends on the spatial organization of soil moisture, soil heterogeneity, and rainfall. This is a major limiting factor in application of the unit hydrograph methodology [Sherman, 1932], which assumes a linear time-invariant system, and a contributing factor in the increasing use of spatially distributed watershed models for flood forecasting [Wigmosta *et al.*, 1994; Downer *et al.*, 2002; Ivanov *et al.*, 2004b,a, 2008a,b]. Many contemporary physically-based continuous simulation models pose watershed rainfall-runoff simulation in three distinct phases: (1) simulating the spatial distribution of runoff production based on local soil moisture, soil character, and rainfall conditions, (2) routing produced runoff as overland flow to channels, and (3) routing runoff from distal channel reaches to the outlet. This approach links the discharge hydrograph at the outlet of the watershed with the upstream spatial distribution of soil moisture dynamics, allowing greater generality in model application. A consequence of this approach is that flood forecasting requires an accurate initialization of the spatial distribution of soil moisture. Thus, improved knowledge of the spatial distribution of soil moisture upon initialization of the models will enhance the ability to predict floods. Related to flood forecasting is the coupling between soil moisture and atmospheric conditions, vis-á-vis its impact on evapotranspiration. In a hindcast of a flash flood event in a topographically complex area of the Colorado Front Range, Chen *et al.* [2001] found that realistic initialization of soil moisture patterns significantly improved 24 hr lead time precipitation forecasts.

Wildland fires play a fundamental role in the disturbance regimes of many terrestrial ecosystems, yet they pose a significant threat to life in areas where these ecosystems intersect. Fires are also a significant source of atmospheric CO_2, water vapor, and dark particulate matter, which both tend to have a positive impact on radiative forcings. Furthermore, because wildfires are sensitive to temperature and precipitation regimes, there is substantial evidence that contemporary climate change has resulted in increases in the frequency and magnitude of wildland fires [Running, 2006]. Rough estimates on the amount of biomass consumed through

burning of biomass in the conterminous United States range from 77-189 Tg annually [4] [Leenhouts, 1998]. A significant amount of scientific work has gone into understanding the role of wildland fire in terrestrial ecosystems, including the development of fuel beds, fire growth and evolution, and impacts of fire as a disturbance mechanism in terrestrial ecosystems. Although mapping of fuel loads for wildfire is incredibly complex, fuel maps are nevertheless essential tools for risk management and mitigation in fire-prone ecosystems [Keane *et al.*, 2001]. Fuel maps are typically produced through some combination of input from field, remote sensing, and simulation data that attempts to characterize the spatial distribution of fuel and its associated moisture content, along with observations and forecasts of meteorological variables that affect fuel moisture (e.g., Keane *et al.* [2001]; Rollins *et al.* [2004]). These maps have relatively high spatial resolutions (e.g., 10's of meters), however the rate at which maps can be updated is sensitive to the revisit interval of the satellite data used as input to the models. In many cases, remote sensing data sources use spectral bands and are therefore associated with longer revisit intervals, limitations on canopy penetration, and the potential of corruption of the observation by clouds. Hence, ecosystem simulation models are becoming increasingly important in the production of fuel maps. These models fuse field and remote sensing data to estimate parameters that are critical to determining fire risk, including: net ecosystem productivity, ecosystem respiration, and leaf area index [Rollins *et al.*, 2004]. Because soil moisture is inextricably these and other variables required as input to fuel loading models, knowledge of soil moisture at spatial resolutions of 10's to 100's of meters would presumably enhance the ability to model fuel loads and moisture, potentially yielding improved understanding of fire risk.

[4]This is substantially lower than the preindustrial era because of bias in contemporary fire management toward suppression.

1.4 Problem definition, objective, and outline of thesis

The above discussion is designed to convey the importance of soil moisture in the global hydrologic cycle, as well as three practical applications that would benefit from knowledge of soil moisture at spatial resolutions of 10's to 100's. The overarching theme of this work is to develop a framework for estimating soil moisture at these spatial scales that combines the ability of spatially distributed ecohydrology models to represent the physical processes governing the dynamics of moisture and the capabilities of satellite platforms to provide repeated observations of variables related to near-surface soil moisture. The approach taken is a data assimilation framework that recognizes uncertainty in both the observations and model estimates, producing an estimate of the spatial distribution of soil moisture that weights each source of information according to its degree of certainty. While process ecohydrology models are capable of simulating soil moisture at the required spatial scales, they are subject to both structural and input uncertainty. By contrast, remotely sensed observations from space yield relatively accurate near-surface moisture information at spatial scales too coarse to capture hillslope-scale variation. Therefore the assimilation framework seeks to combine these sources of information in a way that leverages their respective strengths while compensating for their respective weaknesses. Chapter 2 provides a detailed review of the literature relating to observation of soil moisture through remote sensing, modeling soil moisture, and previous works that employ data assimilation to estimate soil moisture. The uniqueness and potential contributions of this work in the context of these previous works is also discussed. In chapter 3, the mathematical framework of the data assimilation methods used in this thesis are developed, and the necessary developments to constrain the observations to the model are outlined. The mathematical framework of an observing system that relates the modeled spatial distribution of soil moisture to observation of coarse scale observable quantities, particularly microwave brightness temperature and radar backscatter, is developed in chapter 4. Chapter 5 outlines a stochastic rainfall generator that can represent

uncertainty in the spatiotemporal distribution of rainfall. Development of a Latin Hypercube-based technique to represent uncertainty in the soil hydraulic and thermal properties required as input to the ecohydrology model is covered in chapter 6. Chapter 7 is describes a set of experiments that investigate the sensitivity of soil moisture to the various factors contributing uncertainty to soil moisture prediction. Chapter 8 outlines and presents the results of the data assimilation experiments. Finally, chapter 9 provides a discussion of results and the contribution of this work to hillslope-scale soil moisture estimation and concludes with an overview of potential future research directions.

CHAPTER 2

LITERATURE REVIEW

The overarching objective of this research is to construct a data assimilation system to estimate soil moisture at hillslope scales (10's to 100's of meters) that combines noisy remotely sensed observations with an uncertain estimate of soil moisture derived from a physically-based ecohydrology model. This literature review is designed to outline the state of the art of methods to constrain knowledge of the spatial distribution of soil moisture. Two particular techniques, remote sensing using microwave technology and spatially distributed process modeling, receive the most focus because they are central to the development of the data assimilation system. Specifically, a detailed review of techniques to observe soil moisture through remote sensing satellites is provided. The spatially distributed ecohydrology model used throughout this thesis, the TIN-based Real-time Integrated Basin Simulator and VEGetation Integrated Development (tRIBS-VEGGIE) model [Ivanov *et al.*, 2004b,a; Vivoni *et al.*, 2004; Ivanov *et al.*, 2008a,b, 2007] and its conceptual heritage are then overviewed. The remainder of the chapter is devoted to a discussion of previous hydrologic data assimilation work, largely based on a chronological review of notable historical work. This historical outline leads to contemporary identification of soil moisture data assimilation at hillslope scales as an area where research effort is needed. The review concludes by enumerating the developments required for construction of a system to estimate soil moisture at hillslope scales

through data assimilation, and linking the required developments enumerated with chapters in the thesis.

2.1 Estimation of soil moisture

As the introduction demonstrates, knowledge of soil moisture and its spatial distribution is of critical importance to understanding the hydrosphere because soil moisture links global water, energy, and vegetation dynamics. This section deals with methods to measure or otherwise estimate soil moisture. The methods by which soil moisture is typically estimated can be broadly categorized into three distinct classes: (1) *in situ* measurement through a host of analytical techniques such as gravimetric sampling or water content reflectometry, (2) active and passive remote sensing through satellites or aircraft, or (3) estimation by models. Within each of these categories there is substantial conceptual, analytical, and technical heterogeneity in techniques. For instance, *in situ* measurement can vary from instantaneous "grab sampling" in which samples are taken to a geotechnical lab for analysis, to continuous monitoring through devices such as time-domain reflectometers (TDRs) or tensiometers. Further, modeling of soil moisture can vary from geostatistical modeling to dynamic conceptual models to process-based continuous forecasting. An exhaustive review of techniques within each of these categories is beyond the scope of this thesis, and detailed discussion of these techniques, their assumptions, limitations, and accuracy could fill several volumes. It is useful to highlight some of the potential benefits and drawbacks of a few particular subsets of techniques within each category here:

1. *In situ* measurement:
    - **Advantages:** Inexpensive samles, entire profile accurately, continuous
    - **Drawbacks:** Labor intensive, point-scale observation only leading to considerable spatial uncertainty

2. Microwave ($\sim$ 1 GHz) remote sensing:

   - **Advantages:** Global, high revisit rate, sensitive to moisture
   - **Drawbacks:** Near surface observation only, coarse scale, requires inverse model

3. Physically-based models:

   - **Advantages:** Represent processes, integrates multiple-scale data, continuous in time, high spatial resolution
   - **Drawbacks:** Input uncertainty, model uncertainty, numerical complexity, computational burden

Presently, estimating the spatial distribution of soil moisture at scales of interest to this thesis is performed using either deterministic prediction using distributed basin hydrology models, or downsampling of coarse-scale observations according to empirical scaling relationships and/or geostatistical trends. Deterministic hydrologic forecasting neglects the uncertainty associated with the input data. Geostatistical interpolation tends to impose stationary patterns on the disaggregated moisture product (e.g., Kim and Barros [2002a,b]). Reviewing the advantages and drawbacks of each source of soil moisture information, it becomes apparent that microwave remote sensing of soil moisture and physically-based modeling provide fairly complementary sources of information. Therefore, fusion of these two important sources of soil moisture information leverages the benefits of each, while minimizing their drawbacks. The underlying hypothesis of this thesis work is that coarse-scale remotely sensed observations of geophysical variables related to soil moisture can be disaggregated to hillslope scales, conditioned by a physically-based, spatially distributed estimate of soil moisture produced via a modeling framework in which the sources of uncertainty are explicitly represented. The value added by these coarse-scale remote sensing observations is a decrease in uncertainty of soil moisture at hillslope scales. An overview of soil moisture remote sensing and an introduction to the process ecohydrology model used in this research are the topic of the remainder of this section.

2.1.1 Microwave remote sensing of soil moisture

The advent of spaceflight in the 1950's was a watershed moment in the observation of the Earth system. Spaceborne remote sensing provides tremendous flexibility in observing the Earth and its hydrologic cycle. For instance, satellites can be placed into orbits that allow for rapid observation of the entire planet on a regular basis (e.g., low earth orbit). Other orbits allow for continuous monitoring of a particular portion of the planet (i.e., geostationary orbit). There are two fundamental techniques for remote sensing: active and passive. Active remote sensing requires a powered transmitter to broadcast energy toward a target, and a (not necessarily co-located) receiver to measure the intensity of the scatter (often expressed as the ratio of power measured at the receiver relative to the power transmitted). When the transmitter and receiver are co-located in a remote sensing package, fractional energy returned is often quantified in terms of a dimensionless backscatter coefficient [Ulaby *et al.*, 1986; Rees, 2001]. Passive remote sensing, by contrast, measures only naturally emitted or reflected radiation from a target. The intensity of the emitted or reflected radiation being observed is expressed as a brightness temperature in units of kelvins. It is important to note that active and passive sensors are not mutually exclusive and can be used aboard the same vehicle. In the L-band microwave region of the electromagnetic spectrum (i.e., < 5 GHz), the dielectric constant of the near surface (i.e., ≤ 10 cm) soil-water-air matrix and therefore emission and backscatter are sensitive to moisture content [Njoku and Kong, 1977; Ulaby *et al.*, 1986]. Therefore, lower microwave remote sensing is a powerful tool for remote sensing of soil moisture [Njoku and Entekhabi, 1996; Crow *et al.*, 2001; Kerr *et al.*, 2001; Entekhabi *et al.*, 2004; Paloscia *et al.*, 2006].

Spaceborne low frequency microwave radiometry is emerging as an important tool for passive remote sensing of the hydrologic state of the Earth system. This is largely because radiometer sensitivity is typically much smaller than the soil moisture dependent dynamic range of radiobrightness variation, resulting in a large signal-to-noise ratio in the measurement of microwave radiation emission and a

theoretical accuracy in soil moisture observation of approximately 1 - 2% [Njoku and Entekhabi, 1996]. While land surface factors such as soil surface roughness [Choudhury *et al.*, 1979; Tsang and Newton, 1982; Mo *et al.*, 1987, 1982] and vegetation canopy attenuation [Jackson *et al.*, 1982; Ulaby *et al.*, 1983; Pampaloni and Paloscia, 1986; Jackson and Schmugge, 1991] render the theoretical accuracy unrealizable in practice, microwave radiometry is nevertheless a powerful tool to observe near-surface soil moisture over much of the planet [Njoku and Entekhabi, 1996].

For instance, L-band microwave radiometry technology is vital to both the European Space Agency's Soil Moisture and Ocean Salinity (SMOS) mission, scheduled to launch in 2008 [Wigneron *et al.*, 2000; Pellarin *et al.*, 2003], and the National Aeronautical and Space Administration's Soil Moisture Active-Passive (SMAP) mission, scheduled to launch between 2010-2013 (`http://smap.jpl.nasa.gov/`). A change in brightness temperature of 10 K in both horizontal and vertical polarizations is associated approximately with an 8% and change in volumetric water content at incidence angles ranging from 40° to 60° for a loam soil with sparse natural grass cover [Njoku and Kong, 1977; Mo *et al.*, 1982; Njoku and Entekhabi, 1996]. This range of incidence angles coincides with off-nadir look angles of existing and planned radiometers Wigneron *et al.* [2000]; Crow *et al.* [2001]; Pellarin *et al.* [2003]; Entekhabi *et al.* [2004]. The forthcoming SMOS satellite [Kerr *et al.*, 2001] will be the first satellite to observe soil moisture and ocean salinity using L-band microwave radiometery. It incorporates a multiangular viewing concept that observes the vertically and horizontally polarized brightness temperature of a particular location at viewing angles ranging from 0° (nadir) to 55° off-nadir in a 7 MHz band centered at 1.4 GHz. A consequence of this synthetic aperture approach is that the spatial resolution of the brightness temperature products varies with viewing angle. The expected ground resolution of released SMOS products, however, is approximately 40 km [Kerr *et al.*, 2003; Merlin *et al.*, 2005]. Since the sensor provides multiple observations over the same location nearly instantaneously, the relationship between brightness temperature at a location and satellite viewing an-

gle can, in principle, be used to better constrain soil moisture than a single-look observation at one viewing angle. NASA's SMAP mission is now in formulation, and inherits much of its conceptual heritage from the cancelled Hydrosphere State (HYDROS) satellite mission. The purpose of SMAP, like HYDROS before it, is to provide global observation of soil moisture at a revisit rate of 2-3 days. Similar to SMOS, satellite design also calls for a microwave radiometer to observe land surface microwave brightness temperature at a frequency of 1.41 GHz. However, in the case of SMAP the radiometer will be conically-scanning, with a single viewing angle of 40° off-nadir. The ground resolution of the SMAP radiometer products will also be 40 km. Microwave radiometer observation of soil moisture at an altitude that will allow frequent satellite revisit necessitates large antennae on both these missions [Long *et al.*, 2005; Njoku *et al.*, 2000]. The SMOS mission accomplishes a larger antenna via the multiangular viewing aperture synthesis technique mentioned above [Wigneron *et al.*, 2000], while the SMAP mission employs an expandable antenna [Entekhabi *et al.*, 2004].

Estimates of near-surface soil moisture are typically retrieved from brightness temperature observations through numerical inversion of radiative transfer models (RTMs), which relate geophysically observable quantities like brightness temperature to land surface states like soil moisture and temperature [Wigneron *et al.*, 2000; Pellarin *et al.*, 2003; Crow *et al.*, 2005]. These algorithms are often developed by inverting simulated brightness temperature observations to retrieve known inputs to the generally nonlinear RTM. The known inputs to the RTM are typically spatial maps of near-surface moisture and temperature, along with ancillary data related to surface vegetation conditions and soil roughness. Moisture and temperature states input to the RTM are frequently derived from physically-based land surface simulation models. As part of the HYDROS mission, three soil moisture retrieval algorithms that differ in ancillary data requirements were tested using simulated brightness temperature observations over the Arkansas-Red River Basin, US. Results indicate that volumetric soil moisture accuracy of 4% is obtainable except in woodland areas [O'Neill *et al.*, 2004]. Previous work requiring extensive

RTM use typically resolve the land surface at resolutions coarser than 1 km [Crow *et al.*, 2001, 2005; Pellarin *et al.*, 2006; Paloscia *et al.*, 2006; Holmes *et al.*, 2008]. It has been well established that the spatial distribution of inputs to existing radiative transfer models, specifically soil moisture, vary significantly over a range of spatial scales [Rodríguez-Iturbe *et al.*, 1995; Schmugge and Jackson, 1996; Haverkamp *et al.*, 1998]. Variation at the hillslope scale (10s of meters) is associated with spatial variability of topography, soils and vegetation [Western *et al.*, 1999]. In these studies, spatial variation in soil moisture and vegetation states at hillslope scales has largely been ignored. This is primarily because models capable of resolving moisture and vegetation states at these scales are of very high dimension when spatial domains of sufficient size to develop soil moisture retrieval algorithms are represented. Besides influencing hydrology and vegetation, local topography and relative sky position of the observing satellite dictate the incidence angle [Njoku and Kong, 1977; Mo *et al.*, 1982; Njoku and Entekhabi, 1996] and the polarization [Mätzler and Standley, 2000], to which observed brightness temperature is sensitive. The impact of topography on viewing geometry is increasingly gaining attention in the remote sensing literature [Mätzler and Standley, 2000; Kerr *et al.*, 2003; Mialon *et al.*, 2008; Sandells *et al.*, 2008]. In a modeling study using digital elevation models Kerr *et al.* [2003] find that modeled brightness temperatures in areas of variable topography can be several kelvins different than a corresponding flat surface. Further, Mialon *et al.* [2008] recently discussed these effects in the context of the SMOS mission, and developed a criterion to identify SMOS brightness temperature pixels in which topographic effects on incidence angle are likely to result in observation errors greater than the required 4 K accuracy [Kerr *et al.*, 2001]. The works of Sandells *et al.* [2008] and Mätzler and Standley [2000] are notable because they include the effects of topographic slope on the geometry of observation, they did not consider the covariation between topography and the land surface factors affecting emission of microwave energy. It is important to note, however, that despite limitations of developed retrieval algorithms in topographically variable areas, as will become apparent later these RTMs are invaluable in

soil moisture data assimilation because they mathematically relate the land surface state to the geophysically observed quantity.

The coarse spatial resolutions associated with passive L-band soil moisture observations, however, make them difficult to meaningfully use for hillslope-scale moisture estimation, at least in the absence of ancillary data that can be useful for disaggregation. Significant effort has gone into development of brightness temperature disaggregation schemes. For instance, Crow and Wood [2002] used a scaling-based downscaling algorithm to enhance the resolution of microwave brightness imagery to better estimate regional-scale surface energy fluxes. Using a multifractal interpolation technique with ancillary land surface data Kim and Barros [2002a] downscaled remotely sensed soil moisture from the Southern Great Plains 1997 Field Experiment to a resolution of 1 km. While these downscaling techniques are able to estimate soil moisture at resolutions finer than observation in a relatively computationally efficient manner, they represent the processes responsible for the redistribution of moisture in the subsurface in a statistical sense and the spatial scale of the downscaled products is nevertheless significantly coarser than hillslopes.

The uniqueness of the SMAP mission lies in the pairing of the radiometer with an active microwave real aperture radar measuring backscatter at 1.26 GHz. One of major advantages of active remote sensing of soil moisture is a considerable improvement in the resolution of radar products. For instance, while the SMAP radiometer product will have a resolution of 40 km, the active component will have a resolution of $\sim$3 km [Entekhabi *et al.*, 2004]. Reconstruction and resolution enhancement techniques will fuse the radiometer and radar data to yield a soil moisture product at 10 km resolution [Long *et al.*, 2005]. The inclusion of the radar component on the SMAP payload underscores a broader increase in interest in radar observation of soil moisture (e.g., Wang *et al.* [1986]; Engman [1991]; Evans *et al.* [1992]; Chen *et al.* [1995]; Altese *et al.* [1996]; Hoeben and Troch [2000]; Entekhabi *et al.* [2004]). Like passive microwave observation, radar backscatter is sensitive to surface roughness, as well as the dielectric constant and therefore

soil moisture [Dobson and Ulaby, 1986; Wang *et al.*, 1986; Engman, 1991; Evans *et al.*, 1992]. The first opportunity to investigate the usefulness of L-band synthetic aperture radar for geophysical measurement was the Shuttle Imaging Radar payload, which flew aboard NASA's space shuttle on mission STS-41G in October 1984 [Dobson and Ulaby, 1986]. The radar operated at a frequency of 1.28 GHz and at horizontally transmitted and received polarizations [Wang *et al.*, 1986]. Overflying several experimental ground-truthing stations throughout the world, the ground resolution of the radar products was as high as 12.5 m, and analysis of radar images collected near Fresno, California by Wang *et al.* [1986] showed that L-band radar is a potentially powerful tool for observation of soil moisture.

L-band radar measurement of soil moisture is not, however, without problems. Existing models of radar backscatter are significantly more complex than the RTMs used for passive microwave observation (e.g., see Eom and Boener [1986]; Fung *et al.* [1992]; Fung [1994]). Radar backscatter, as will be shown in a later chapter, is significantly more sensitive to the roughness at the surface than is passive brightness temperature [Fung *et al.*, 1992]. The presence of vegetation can significantly impact the ability to infer soil moisture from observations of microwave radar backscatter. Up to a Normalized Difference Vegetation Index (NDVI) of approximately 0.2, the microwave backscatter is largely due to interaction of transmitted energy with the surface [Dubois *et al.*, 1995]. However, as NDVI increases beyond 0.2, volume scattering by the vegetation canopy becomes progressively more important [Dubois *et al.*, 1995]. Vegetation impacts scattering because the vegetation canopy is semitansparent to the transmitted energy, resulting in backscatter from the vegetation volume itself and an interaction between surface scattering and the vegetation volume scattering [Njoku *et al.*, 2002]. Three approaches to treating vegetation are generally pursued. The first approach involves applying surface scattering models and inversion techniques (e.g., Fung [1994]) only to areas where vegetation cover is known to be sparse, such as semiarid to arid settings [Oh *et al.*, 1992; Altese *et al.*, 1996; Chen *et al.*, 1995; Shi *et al.*, 1997; Hoeben and Troch, 2000; Thoma *et al.*, 2006]. Another approach seeks to use more complex backscatter models for lay-

ered, heterogeneous media such as the Matrix Doubling Method [Eom and Boener, 1986]. The final approach is based on the assumption that changes in the volume of the vegetation canopy are slow relative to the revisit time of the satellite sensor, and therefore the temporal change detected in backscatter observations between successive observations is due largely to changes in surface soil moisture [Njoku *et al.*, 2002; Narayan *et al.*, 2006; Narayan and Lakshmi, 2008].

The preceding review of active and passive microwave soil moisture observation highlights the potential benefits, drawbacks, and limitations of soil moisture remote sensing. Spaceborne soil moisture observing satellites have the potential to globally measure geophysical variables related to soil moisture (brightness temperature and radar backscatter) with revisit intervals ranging from 2-3 days. Brightness temperature and backscatter are particularly sensitive to variation in soil moisture in the lower microwave range of the electromagnetic specrum. However, these observations alone are not immediately useful to characterize soil moisture at hillslope scales because the spatial resolutions of their respective products ranges from approximately 3 km for radar data to 40 km for radiometer data. And while RTMs for passive observation have received significant attention in the literature, the impacts of topography on modeled and observed brightness temperatures has not been well understood, historically. Moreover, while active systems can realize significantly better spatial resolution than passive observation, the forward models of microwave scatter are either limiting in their geographic applicability and/or significantly more complex in form. Hence, before microwave data can be useful for hillslope-scale soil moisture data assimilation, the significant challenge of formulating both passive and active observing systems that incorporate topographic effects on observational geometry must be overcome. This topic is dealt with in **Chapter 4: Formulation of the observing system.**

2.1.2 Land surface ecohydrology modeling: The tRIBS-VEGGIE model

Spatially distributed process models are an attractive way to estimate the spatial distribution of soil moisture. This is due, in part, to the fact that such models explicitly represent the physical, chemical, and biological processes responsible for mass, energy, and carbon exchange between the land and atmosphere as they are known. They also provide a numerical framework by which multi-scale data representing the boundary conditions (i.e., elevation from Digital Elevation Models (DEMs) and soils from SSURGO), meteorological forcings (e.g., rainfall from weather radar, temperature and solar radiation from weather stations) can be fused through the laws governing moisture redistribution to ascertain the hydrologic state of a watershed. Of particular interest to the applications outlined in the introduction are the spatial arrangement of soil moisture and biomass, and a continuous estimate of discharge at the outlet. The model used in this thesis work is the coupled Triangulated Irregular Network (TIN)-based Realtime Integrated Basin Simulator (tRIBS) [Ivanov *et al.*, 2004b,a; Vivoni *et al.*, 2004] coupled to the VEGetation Integrated Evolution model (VEGGIE) [Ivanov *et al.*, 2008a,b, 2007], hereafter referred to as tRIBS-VEGGIE. tRIBS-VEGGIE is a spatially-distributed model that resolves mass, energy and carbon balance over a watershed at the hillslope scale by representation of coupled: (1) biophysical energy processes (e.g., partitioning of input solar radiation in the canopy and soils), (2) biophysical hydrologic processes (partitioning of rainfall into interception, throughfall, plant water uptake, etc.), and (3) biochemical processes and vegetation phenology. A full treatment of the tRIBS-VEGGIE model is beyond the present review and the reader is directed to previous studies describing the development, parameterization, and confirmation of the tRIBS-VEGGIE framework [Ivanov *et al.*, 2004b,a; Vivoni *et al.*, 2004; Ivanov *et al.*, 2008a,b, 2007]. What follows in this section is a brief description of the soil moisture modeling component, the vegetation dynamics embodied by the model, and an overview of the static data required to simulate the near-surface moisture, temperature, and vegetation variables that impact microwave remote sensing.

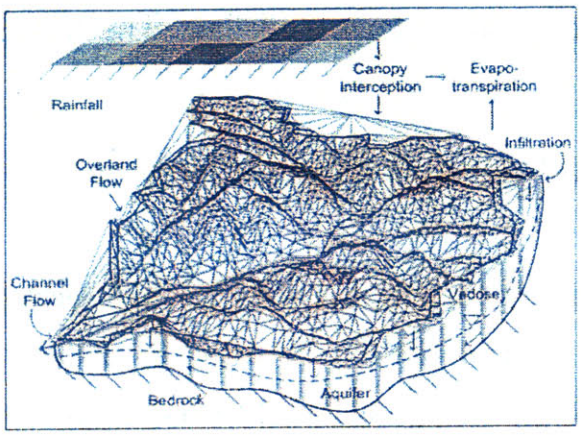


Figure 2-1: A schematic representation of the way in which the landscape is abstracted in the tRIBS framework. Reproduced from Vivoni *et al.* [2005].

The tRIBS uses a triangulated irregular network (TIN) representation of land surface topography (figure 2-1), in contrast with a square element grid representation of other distributed basin hydrology models (e.g., Wigmosta *et al.* [1994]; Downer *et al.* [2002]). A TIN representation of watershed topography can readily be constructed using DEMs such as those associated with the Shuttle Radar Topography Mission [Farr *et al.*, 2007]. The principle benefit of the TIN computation mesh is the ability to vary the resolution of the land surface according to the degree of topographic ruggedness within a basin, thereby reducing the number of computational nodes needed to simulate basin hydrologic response [Vivoni *et al.*, 2004]. Vertices of TIN triangles are centroids of the corresponding Voronoi polygon network (VPN) and represent the computational nodes where the equations mass and energy conservation are solved numerically (figure 2-2).

Spatially variable precipitation may be intercepted by the vegetation canopy, or fall directly to the ground where it may infiltrate into the subsurface or runoff the land surface as overland flow. Rainfall interception is modeled according to the canopy water balance method [Rutter *et al.*, 1971]. Intercepted water may drain to the surface through leaf dripping or stemflow following the model of Shuttleworth [1979] where it may be partitioned to infiltration and/or runoff, or intercepted precipitation may evaporate directly from the leaf surfaces following the method of Eltahir and Bras [1993]. Infiltration is modeled through an implicit finite-element

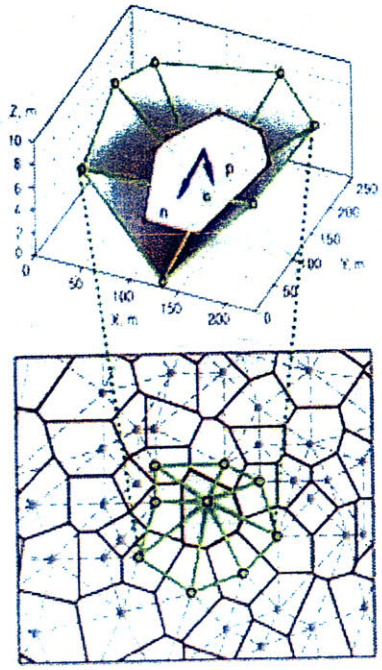


Figure 2-2: A conceptual representation of the connection between the Triangulated Irregular Network (TIN) and the Voronoi Polygon Network (VPN). The VPN constitutes the computational mesh, on which the governing equations of mass, energy, and carbon balance are solved. Adapted from the work of Ivanov *et al.* [2004b].

backward Euler time marching solution of the θ-based form of Richards equation[1] [Ivanov *et al.*, 2008a,b]. The model uses the characteristic soil water retention curve parameterization of Brooks and Corey [1964]. Runoff production in the tRIBS model can occur as saturation excess, infiltration excess, or groundwater exfiltration. Two distinct types of runoff routing are performed, hydrologic runoff routing on hillslopes and hydraulic routing in channel networks, and runoff is routed from one computational node to the next node in a steepest-descent fashion. The hydrologic runoff routing procedure parameterizes the hillslope runoff velocity as a power function of the ratio of the stream discharge at the hydrologically-nearest downstream channel node to the contributing area at that node. This velocity and a hillslope path length that is obtained from the geometry of the basin can be used to compute the travel time from any hillslope point to the hydrologically nearest stream channel. At the upstream-most channel node, the hydrologic runoff routing

[1] see a review of Richards solving methods in Celia *et al.* [1990] and van Dam and Feddes [2000]

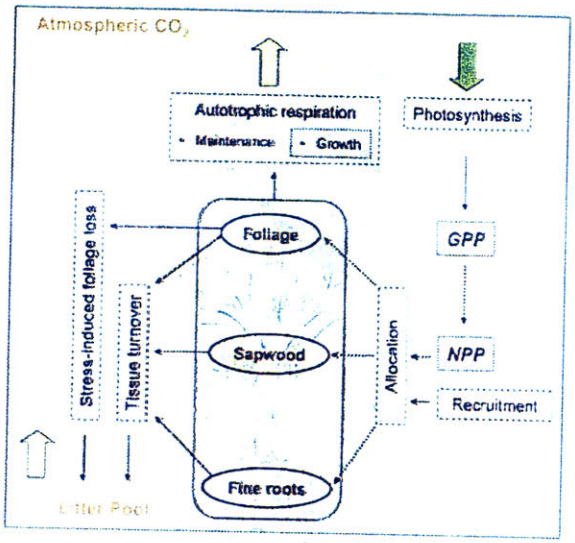


Figure 2-3: A diagram depicting the concepts of carbon cycling embodied by VEGGIE. Reproduced from the work of Ivanov *et al.* [2008a].

procedure yields a hillslope hydrograph that serves as input to the hydraulic routing scheme. These hillslope hydrographs are then routed through the channel network using a kinematic wave routing model (e.g., Goodrich *et al.* [1991]; Singh [1996]).

The dynamic vegetation component of tRIBS-VEGGIE operates on specified plant functional types (PFTs) [Bonan *et al.*, 2002a] within each Voronoi cell. For each PFT, tRIBS-VEGGIE simulates carbon fluxes by representing the processes of photosynthesis, autotrophic respiration, stress induced foliage loss, and tissue turnover. The fluxes that determine the exchange of CO_2 between the atmosphere and land surface contribute to the dynamics in three carbon pools modeled within tRIBS-VEGGIE: foliage, sapwood, and fine roots. Assimilation of CO_2 through photosynthesis is coupled to surface energy and water balance through the stomatal resistance model, which depends on the budget of longwave and solar radiation and site soil moisture throughout the rooting profile of the PFT. A conceptual map of the VEGGIE model is shown in figure 2.3.

The amount of incoming solar radiation being received at any time during the day at the land surface depends on site slope and aspect, as well as the solar azimuth angle. In this manner, tRIBS-VEGGIE simulates the spatial covariation between incident solar radiation and surface moisture, energy, and plant vigor processes, which

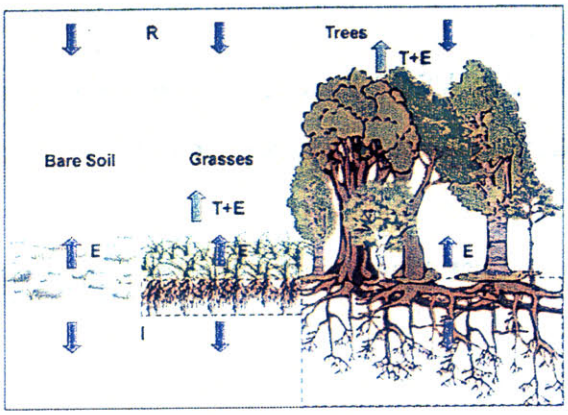


Figure 2-4: A conceptual illustration of Voronoi element scale evaporation and transpiration fluxes, partitioned by PFT. Taken from Ivanov [2006].

leads to North- and South-facing contrasts in variables such as surface temperature, canopy temperature, and available energy (figure 2-4).

Inputs to the tRIBS-VEGGIE model correspond to four categories of data: (1) hourly hydrometeorological forcings, (2) soil hydraulic and thermal properties, (3) vegetation parameters, and (4) a static elevation field representing watershed topography. Hydrometeorological forcings for tRIBS-VEGGIE include hourly: (1) precipitation, (2) sky fractional cover or incoming solar radiation, (3) air temperature, (4) dew temperature, and (5) wind speed. In this study, hydrometeorological forcings under which the moisture and vegetation state are evolved by tRIBS-VEGGIE were generated by a stochastic weather generator Ivanov *et al.* [2007]. Soil hydraulic and thermal parameters are consistent with those common to many land surface models (e.g, Liang *et al.* [1994]; Chen *et al.* [1996a]; Koster and Suarez [1996]; Peters-Lidard *et al.* [1997]; Oleson *et al.* [2004]) and available in published soil databases such as the STATSGO or SSURGO products. Soil parameters required for water- and energy-balance solution include: (1) saturated hydraulic conductivity, (2) saturation moisture content, (3) residual moisture content, (4) Brooks-Corey parameters, (5) specific volumetric heat capacity, and (6) thermal conductivity. Parameters required by the vegetation development model coincide with the C4 grass PFT parameterized in other biophysical-biochemical models (e.g., see Bonan [1995, 1996]; Sellers *et al.* [1996]; Levis *et al.* [2004]; Arora and Boer

[2005]).

The outputs of the tRIBS-VEGGIE model as they relate to remote sensing of soil moisture include spatially distributed output such as the forecast value of soil moisture at each Finite Element node within the soil profile, the surface soil temperature, the canopy temperature, Leaf Area Index, and fractional vegetation area.

2.2 Hydrologic data assimilation

Combining hydrologic models with observations to produce optimal estimates of hydrologic variables of interest has a long history in the primary literature. Much of the early work was targeted toward combining models and observations for practical applications such as flood forecasting and irrigation scheduling (e.g., Kitanidis and Bras [1980b,a]; Aboitiz *et al.* [1986]). Early work in this arena largely focused on the use of lumped parameter models for which obtaining state space representations and statistical linearization required to employ recursive filtering techniques such as the extended Kalman Filter [Schweppe, 1973; Gelb, 1974] was relatively straightforward. The Kalman Filter is a powerful estimation framework that assumes imperfect observations, with additive, white, gaussian observation errors; and linear, gaussian models whose uncertainty can be completely characterized by a mean and error variance-covariance matrix. The Kalman Filter, which is reviewed in detail in the next chapter, was the centerpiece of many early hydrologic data assimilation studies. Using a state-space formulation of the National Weather Service river forecasting model [Burnash *et al.*, 1973] that depends on basin soil moisture state, Kitanidis and Bras [1980b,a] applied the Kalman Filter to ingest observations of discharge. In a previous work, Kitanidis and Bras [1979] coupled a Kalman filtering framework to a Bayesian likelihood ratio test to detect transient errors associated with non-stationary input error structure. Aboitiz *et al.* [1986] developed a conceptual state space model describing soil moisture balance for an irrigated field, which they combined with observations of soil moisture depletion through

the Kalman Filter. Other hydrologic data assimilation works have used frameworks not based on the Kalman Filter. For instance, Seo *et al.* [2003] applied a variational data assimilation approach to update the soil moisture states of the Burnash *et al.* [1973] rainfall-runoff model and the ordinates of the unit hydrograph with observations of real time discharge and precipitation, and climatological potential evaporation values for a headwater basin. Variational assimilation methods find an optimal estimate of the model state by minimizing a cost function that usually involves some metric of error between forecast and observed quantities.

In addition to assimilation of discharge, which provides a measure of runoff production integrated in space and time, significant scientific effort has been invested in using data assimilation as a way to constrain models with remotely sensed microwave brightness temperature and radar backscatter data. Although the observed brightness and backscatter quantities are sensitive only to near-surface soil moisture, several published studies have demonstrated that these data can provide additional information about profile moisture in the context of a hydrologic model, vis-á-vis an assimilation approach. For instance, Entekhabi *et al.* [1994] developed an inverse solution to retrieve soil moisture and temperature profiles from brightness observations using a radiative transfer and a coupled moisture and heat diffusion model. Calvet *et al.* [1998] demonstrated that knowledge of surface soil moisture and temperature could be used to retrieve the root zone moisture profile. Houser *et al.* [1998] used Newtonian nudging methods, which involve post hoc corrections to the modeled state, to assimilate microwave radiometer observations and update profile soil moisture evolved through the TOPLATS model [Famiglietti and Wood, 1994b]. In a later work, Galantowicz *et al.* [1999] used the Kalman Filter methodology to assimilate brightness temperature observations into a soil hydraulic and temperature diffusion model. Hoeben and Troch [2000] also employed Kalman Filtering to assimilate microwave backscatter observations to update the soil moisture profile evolved with a linearized 1-dimensional Richards solver. The work of Reichle *et al.* [2001] provide an important counterexample, however, in their recognition of the usefulness in posing the downscaling of microwave brightness observations

as a data assimilation problem. They used a 4-D (three spatial dimensions and one temporal dimensional) variational approach to downscale radiobrightness images from resolutions of tens of kilometers to scales of a few kilometers. Fine scale spatiotemporal variation in soil moisture can be resolved using data assimilation methodologies provided that soil textural, land cover, and micrometeorological inputs exist at these scales. The work presented in this thesis is partly a conceptual extension of their work, based on the hypothesis that topography is a significant source of variation in soil moisture at the finest scales of interest (10s to 100s m). Developing a novel application of satellite data assimilation, Crow [2007]; Crow and Zhan [2007]; Crow and Bolten [2007] ingested observations from the X-band Advanced Microwave Scanning Radiometer (AMSR-E) into a relatively simple antecedent precipitation index model using the Kalman Filter and used the results to diagnose errors in satellite precipitation products.

Nonlinear representation of hydrological processes in models and the increasing use of spatially distributed models limits the feasibility of Kalman Filtering and other techniques such as some variational methods that require tangent linear models (and their adjoints) for soil moisture data assimilation at hillslope scales with models as complex as tRIBS-VEGGIE. This is largely due to numerical instabilities and the significant effort associated with linearizing complex, high-dimensional models. Algorithmic developments in the ocean modeling community, however, have significantly improved the prospects for hillslope-scale soil moisture data assimilation. Using a nonlinear quasi-geostrophic model Evensen [1994, 2003, 2004] demonstrated that Monte Carlo methods could be used to forecast error statistics and approximately estimate an error variance-covariance matrix directly, rather than evolving it forward in time as required by the Kalman Filter. This technique, known as the Ensemble Kalman Filter (EnKF) is reviewed in detail in **Chapter 3**, led to the rapid growth of ensemble-based data assimilation methods for nonlinear models with high dimensionality in the geosciences. Because the state error covariance matrix is obtained via a Monte Carlo approach, however, the EnKF algorithm requires explicit modeling of uncertainty in data input to models like tRIBS-VEGGIE

(i.e., hydrometeorological forcings and parameters). Moreover, particular care in identifying and modeling those sources of uncertainty is required to ensure that the error in the model forecast is adequately represented. The importance of proper uncertainty representation will be underscored in the next chapter.

In hydrology, ensemble-based data assimilation has become popular for estimating hydrologic state variables of interest, including profile and near-surface soil moisture, using observational data and models. Margulis *et al.* [2002] used the EnKF to combine remotely sensed soil moisture measurements from the Southern Great Plains 1997 field campaign with modeled soil moisture fields evolved by the NOAH land surface model described by [Chen *et al.*, 1996b]. Reichle *et al.* [2002] evaluated the performance of the EnKF against an optimal variational smoother for prediction of soil moisture. They also investigated the effect of non-Gaussian forecast errors and the ensemble size, or number of realizations of the soil moisture state evolved using Monte Carlo methods, on final forecast accuracy compared with soil moisture estimates unconstrained to observation. More recently, Dunne and Entekhabi [2005] argued that soil moisture estimation is more akin to a reanalysis problem and developed a hybrid estimation routine combining the EnKF operating across rainfall events with an Ensemble Kalman Smoother (EnKS) operating during drying cycles. More recently, Kim *et al.* [2007] assumed hydrologic similarity to update spatially distributed soil moisture states using an Ensemble Kalman Filter (EnKF).

While these past efforts demonstrate the utility of ensemble data assimilation for estimation of hydrological variables of interest, there are several limitations to the methods employed. For instance, the RTMs that produce predictions of observable quantities based on the modeled soil moisture state often require significant ancillary data that characterizes the vegetation canopy and the soils. Vegetation ancillary data is commonly treated as static (e.g., Crow *et al.* [2005]), although the phenology and growth of plants are clearly coupled to soil moisture dynamics while also being important to the RTM. The work of Pauwels *et al.* [2007] provides

an interesting attempt to deal with vegetation dynamics by assimilating synthetic measurements of both soil moisture and leaf area index using an EnKF. Another limitation that is important to this work is the emphasis in previous studies on spatial resolutions too coarse to account for fine scale heterogeneity in topography and soil hydraulic properties. Many make use of soil-vegetation-atmosphere transfer schemes that evolve the soil moisture state at scales that completely neglect the role of topography in the redistribution of subsurface moisture [Dickinson *et al.*, 1986; Sellers *et al.*, 1986; Bonan, 1991, 1995, 1996] or do so only in a statistical manner as a subgrid-scale parameterization [Famiglietti and Wood, 1994b,a; Peters-Lidard *et al.*, 1997; Crow and Wood, 2033]. The high dimensionality of digital terrain data-based hydrologic models is one of the primary reasons for the limited use of topography dependent models in soil moisture state estimation research. The work of van Loon and Troch [2002] provide a notable exception, although their observations consisted of geostatistically interpolated coarser-scale observations, that estimated soil moisture at spatial resolutions ranging from 20 m to 60 m for a 44 ha catchment. Another critically important element of ensemble soil moisture data assimilation is proper representation of uncertainty. Advocating for an adaptive ensemble Kalman filter approach, Reichle *et al.* [2008] demonstrated that poor characterization of input errors led to large errors in surface fluxes. Ensemble soil moisture data assimilation studies conducted over large regions frequently attribute much of the uncertainty in soil moisture to errors in the spatial arrangement and intensities of precipitation, although uncertainty in model parameters is frequently represented as well [Margulis *et al.*, 2002; Reichle *et al.*, 2002; Zhou, 2005; Dunne and Entekhabi, 2006]. In a later chapter, the importance of adequately representing uncertainty in hydrometeorological forcings and soil hydraulic and thermal properties on hillslope-scale soil moisture estimation will be demonstrated.

Finally, formulation of observing systems that relate the model estimate of soil moisture to the geophysically observed quantity is the tie that binds remote sensing of soil moisture to the model. In particular, the EnKF requires an observation operator that ingests information about the model state to produce predicted ob-

servations (e.g., see Galantowicz *et al.* [1999]; Margulis *et al.* [2002]; Reichle *et al.* [2002]; Dunne and Entekhabi [2005, 2006]). The necessity of the observing system will become clear in **Chapter 3**, and the whole of **Chapter 4** will be dedicated to formulating a system that takes as input the spatial distribution of moisture, temperature, and vegetation simulated by tRIBS-VEGGIE to yield predictions of L-band microwave brightness temperature and radar backscatter as output.

2.3 Conclusion: Required developments

This chapter concludes with a recapitulation of the developments necessary to construct an assimilation system to estimate soil moisture at hillslope scales using remotely sensed data and the tRIBS-VEGGIE model. Because it places no *a priori* restrictions of the form of the model used for assimilation and the nonlinear nature of the processes embodied by tRIBS-VEGGIE do not lend themselves well to linearization, and because it has a heritage of use in soil moisture data assimilation, the EnKF is well suited to the task at hand. Below are listed the particular components of the data assimilation system that needed to be developed by this work and the chapters of the thesis in which they are addressed:

1. Identification of a form of the ensemble Kalman Filter algorithm suited to hillslope-scale soil moisture estimation with tRIBS-VEGGIE and remotely-sensed data at scales of kilometers (**Chapter 3**),

2. Formulation of an observing system that relates the spatial distribution of near-surface soil moisture simulated by tRIBS-VEGGIE to L-band microwave measurements of brightness temperature and radar backscatter (**Chapter 4**),

3. Identification and formulation of techniques to explicitly represent uncertainty in the hydrometeorological forcings (**Chapter 5**) and soil hydraulic and thermal properties (**Chapter 6**) required as input to tRIBS-VEGGIE, and

4. A detailed uncertainty analysis assessing the relative impacts of hydrometeorological versus soil hydraulic and thermal property uncertainty on soil moisture predictability (**Chapter 7**).

CHAPTER 3

DATA ASSIMILATION MACHINERY

This chapter describes the mathematical machinery used in the soil moisture data assimilation system constructed for this work. The chapter begins with an introduction to the Kalman Filter equations, and their relevant assumptions and implications. The extension of Kalman Filtering to nonlinear and non-Gaussian problems through Monte Carlo modeling, the ensemble Kalman Filter (EnKF), is described in detail. An implementation of the EnKF that is well-suited to high dimensional and low rank problems is introduced. Since the original Kalman Filter and the EnKF pose the mathematical formulation of observing system as a linear transformation of the model state, a discussion of observing systems that nonlinearly relate the state to predicted observations is required. The final part of this chapter presents a detailed outline of the workflow required to assimilate observations of microwave brightness temperature for hillslope-scale soil moisture estimation with the tRIBS-VEGGIE model.

3.1 Vector Kalman Filtering

What follows in this section is an introduction to the Kalman filter (KF) recursive estimation algorithm. The review presented below largely follows the algorithm as outlined in Gelb [1974] and Schweppe [1973]. The KF was derived as a gener-

alization of the Weiner filter to linear stochastic systems involving non-wide-sense stationary processes, and/or observation intervals that do not allow for linear least-squares estimation through Linear Time Invariant filtering. The discrete-time KF assumes that the $L_{\mathbf{x}}$-dimensional system state vector, $\mathbf{x}$, at time $n+1$ can generically be described as a linear transition of the state at the previous time interval $\mathbf{x}[n]$, taking into account random forcings subjected to the system at time n, $\mathbf{v}[n]$, which may have dimension different from $\mathbf{x}$,

$$\mathbf{x}[n+1] = \mathbf{A}[n]\,\mathbf{x}[n] + \mathbf{B}[n]\,\mathbf{v}[n] \tag{3.1}$$

This is generally referred to as the canonical form of the state model, $\mathbf{A}[n]$ is of dimension $L_{\mathbf{x}} \times L_{\mathbf{x}}$ and referred to as the system state-transition matrix, and $\mathbf{B}[n]$ is an operator of dimension $L_{\mathbf{v}} \times L_{\mathbf{v}}$ that projects the forcings $\mathbf{v}[n]$ onto state space.

The $L_{\mathbf{y}}$-dimensional vector of observations $\mathbf{y}[n]$ is related to $\mathbf{x}[n]$ through the so-called observation equation,

$$\mathbf{y}[n] = \mathbf{H}[n]\,\mathbf{x}[n] + \mathbf{w}[n] \tag{3.2}$$

The random forcing ($\mathbf{v}[\cdot]$) and observational noise ($\mathbf{w}[\cdot]$) are assumed to be independent, zero-mean white processes with covariance matrices $\mathbf{Q}[n]$ and $\mathbf{R}[n]$, respectively. Additionally, the initial condition $\mathbf{x}[0]$ is zero-mean with covariance $\mathbf{P}$, and it is assumed that there is no correlation among $\mathbf{v}[\cdot]$, $\mathbf{w}[\cdot]$, and $\mathbf{x}[0]$.

Let $\hat{\mathbf{x}}[n|n]$ denote the linear least-squares estimate (LLSE) of the true state $\mathbf{x}[n]$ at time n, given all observations $\mathbf{y}[0]$,$\mathbf{y}[1]$,...,$\mathbf{y}[n]$. The orthogonality principle dictates that $\hat{\mathbf{x}}[n|n]$, which is optimal in a linear least-squares sense, yields an error estimate that is orthogonal to all of the data. Stated mathematically, the LLSE possesses an estimation error $\mathbf{e}[n|n] = \hat{\mathbf{x}}[n|n] - \mathbf{x}[n]$ that satisfies:

$$\mathrm{E}\left[\mathbf{e}\left[n|n\right]\mathbf{y}\left[k\right]^{\mathrm{T}}\right] = \mathrm{E}\left[\left(\hat{\mathbf{x}}\left[n|n\right] - \mathbf{x}\left[n\right]\right)\mathbf{y}\left[k\right]^{\mathrm{T}}\right] = 0, \quad k = 0, 1, \ldots, n. \tag{3.3}$$

The orthogonality principle also holds for the prediction, $\hat{\mathbf{x}}\left[n|n-1\right]$, which is derived by propagating the filtered state at time $n-1$, $\hat{\mathbf{x}}\left[n-1|n-1\right]$, forward one time step using the system equation. Hence,

$$\mathrm{E}\left[\mathbf{e}\left[n|n-1\right]\mathbf{y}\left[k\right]^{\mathrm{T}}\right] = \mathrm{E}\left[\left(\hat{\mathbf{x}}\left[n|n-1\right] - \mathbf{x}\left[n\right]\right)\mathbf{y}\left[k\right]^{\mathrm{T}}\right] = 0, \quad k = 0, 1, \ldots, n-1. \tag{3.4}$$

Because of the linearity of the system, the observation equation can be rewritten in a form to reflect the predicted observation at time n as a function of predicted state $\hat{\mathbf{x}}\left[n|n-1\right]$,

$$\hat{\mathbf{y}}\left[n|n-1\right] = \mathbf{H}\left[n\right]\hat{\mathbf{x}}\left[n|n-1\right] + \mathbf{w}\left[n\right] \tag{3.5}$$

It can be shown that choosing $\mathbf{w}\left[\cdot\right]$ to be uncorrelated with both $\mathbf{v}\left[\cdot\right]$ and $\mathbf{x}\left[0\right]$ ensures that $\mathbf{v}\left[n\right]$ is uncorrelated with $\mathbf{y}\left[0\right]$,$\mathbf{y}\left[1\right]$,...,$\mathbf{y}\left[n\right]$. $\mathbf{w}\left[n\right]$ is also uncorrelated with $\mathbf{x}\left[\cdot\right]$ because of the imposed constraint that $\mathbf{w}\left[n\right]$ is white. If the expectation of the last term in the above equation is zero (as it is assumed), then the best prediction of the data based on the model state is therefore,

$$\hat{\mathbf{y}}\left[n|n-1\right] = \mathbf{H}\left[n\right]\hat{\mathbf{x}}\left[n|n-1\right] \tag{3.6}$$

The measurement residual, or innovation, is the difference between the actual observation and the predicted observation. The innovations process, $\mathbf{z}\left[n\right]$, is a white, causally invertible, random process:

$$\mathbf{z}\left[n\right] = \mathbf{y}\left[n\right] - \hat{\mathbf{y}}\left[n|n-1\right] = \mathbf{y}\left[n\right] - \mathbf{H}\left[n\right]\hat{\mathbf{x}}\left[n|n-1\right] \tag{3.7}$$

Recasting the model equation as a mechanism to produce an unfiltered estimate at time n (the predicted state), given the posterior (filtered) state estimate at time $n-1$,

$$\hat{\mathbf{x}}[n|n-1] = \mathbf{A}[n-1]\,\hat{\mathbf{x}}[n-1|n-1] + \mathbf{B}[n-1]\,\mathbf{v}[n-1]. \tag{3.8}$$

The prior (predicted) error covariance at time n, $\mathrm{E}\left[\mathbf{e}[n|n-1]\,\mathbf{e}[n|n-1]^{\mathrm{T}}\right] = \mathbf{P}[n|n-1]$, can similarly be determined by evolving the posterior (filtered) error covariance at time $n-1$, $\mathbf{P}[n-1|n-1]$, forward under the model dynamics of the system,

$$\mathbf{P}[n|n-1] = \mathbf{A}[n]\,\mathbf{P}[n-1|n-1]\,\mathbf{A}[n]^{\mathrm{T}} + \mathbf{Q}[n] \tag{3.9}$$

At the heart of the Kalman filter is an algorithmic machinery through which the prior estimate (the predicted state), $\mathbf{x}[n|n-1]$, can be updated with new information in the form of observations at time n, $\mathbf{y}[n]$. Additionally, the prior error covariance, $\mathbf{P}[n|n-1]$, can also be updated, conditioned on the new information to arrive at a posterior error covariance $\mathbf{P}[n|n]$ that reflects the impact of new information on the uncertainty of the state. It can be shown that this posterior estimate of the state (the updated state), $\mathbf{x}[n|n]$, is the Bayes' least-squares estimate given systems that are both linear and Gaussian. The posterior estimate is simply the prior estimate plus a term that represents the product of a gain matrix and the innovation,

$$\hat{\mathbf{x}}[n|n] = \hat{\mathbf{x}}[n|n-1] + \mathbf{K}[n]\,\mathbf{z}[n], \tag{3.10}$$

where $\mathbf{K}[n]$ is the Kalman, or optimal gain, which is computed as:

$$\mathbf{K}[n] = \mathbf{P}[n|n-1]\,\mathbf{H}[n]^{\mathrm{T}}\left(\mathbf{H}[n]\,\mathbf{P}[n|n-1]\,\mathbf{H}[n]^{\mathrm{T}} + \mathbf{R}[n]\right)^{-1} \tag{3.11}$$

The prior error covariance can similarly be updated through the following,

$$\mathbf{P}[n|n] = (\mathbf{I} - \mathbf{K}[n]\,\mathbf{H}[n])\,\mathbf{P}[n|n-1]. \tag{3.12}$$

3.1.1 Implications of the Kalman Filter

In equation 3.6, if the predicted observation, $\hat{\mathbf{y}}[n|n-1]$, is identical to the actual observation, $\mathbf{y}[n]$, then the innovation $\mathbf{z}[n] = 0$ as seen in equation 3.7. This can only occur in the trivial case when the predicted state, $\hat{\mathbf{x}}[n|n-1]$, is equal to the true state, $\mathbf{x}[n]$. Of more interest are the relative influence of observational uncertainty contained in $\mathbf{R}[n]$ and the accuracy of the state estimate as described by the predicted error covariance $\mathbf{P}[n|n-1]$. As a heuristic example, following is the special case of a scalar system described by initialization,

$$x[0|-1] = 0, \text{and}, \tag{3.13}$$

$$\sigma_e^2[0|-1] = \sigma^2, \tag{3.14}$$

where $\sigma_e^2[n|n-1]$ is the error covariance at time n, given all observations through time $n-1$. For this special case the model prediction, predicted observations, and prior error covariance are of the forms:

$$\hat{x}[n|n-1] = a\hat{x}[n-1|n-1] + v[n], \tag{3.15}$$

$$y[n|n-1] = c\hat{x}[n|n-1] + w[n], \tag{3.16}$$

$$\sigma^2[n|n-1] = a^2\sigma^2[n-1|n-1] + \sigma_v^2, \tag{3.17}$$

where $w[\cdot]$ has a variance, σ_w^2, and $v[\cdot]$ has a variance, σ_v^2. In this scalar problem, the Kalman gain (also a scalar) will reduce to the following,

$$k[n] = \frac{c\sigma_e^2[n|n-1]}{c^2\sigma_e^2[n|n-1] + \sigma_w^2}. \tag{3.18}$$

Substituting the above gain term, the state update follows as,

$$\hat{x}[n|n] = \hat{x}[n|n-1] + \frac{c\sigma_e^2[n|n-1]}{c^2\sigma_e^2[n|n-1] + \sigma_w^2}\left(y[n] - c\hat{x}[n|n-1]\right), \tag{3.19}$$

and the posterior estimate of the error covariance is given by,

$$\sigma_e^2[n|n] = \sigma_e^2[n|n-1] - \frac{c\sigma_e^2[n|n-1]}{c^2\sigma_e^2[n|n-1] + \sigma_w^2} c\sigma_e^2[n|n-1] \tag{3.20}$$

Rewritten slightly,

$$\sigma_e^2[n|n] = \sigma_e^2[n|n-1] - \frac{\sigma_e^2[n|n-1]}{c + \frac{\sigma_w^2}{c^2\sigma_e^2[n|n-1]}} \tag{3.21}$$

Inspection of equation (3.19) reveals that when $\sigma_e^2[n|n-1] \gg \sigma_w^2$, the observation dominates, and the updated state approximates the inverted observation, $(1/c)y[n]$. Conversely, when $\sigma_e^2[n|n-1] \ll \sigma_w^2$, the model prediction dominates and the updated state approximates the prior estimate, $\hat{x}[n|n] \approx \hat{x}[n|n-1]$. Between these two extremes, the updated state corresponds to some intermediate weighting of the model predictions and observations, multiplied by the innovation. In terms of the uncertainty, equation 3.20 reveals that the only circumstances in which the posterior error covariance will not be reduced with respect to the prior error covariance occur when $\sigma_w^2 \gg c^2\sigma_e^2[n|n-1]$. Hence, the added value introduced by the KF algorithm is both an updated state that considers information in the form of observational data, and reduced uncertainty in the posterior state estimate due to the inclusion of those observations.

3.1.2 Summary of Kalman filter algorithm

The following is a summary outline of the Kalman Filter algorithm for state estimation with a linear model in which the state error and observational error distributions are known to be (or assumed) Gaussian:

1. Initialize the model with an initial estimate of the state, $\hat{\mathbf{x}}[0|-1]$, and initial error covariance, $\mathbf{P}[0|-1]$.

2. Propagate the initial state and error covariance forward to observation time n through equations 3.8 and 3.9, respectively.

3. As the product of the predicted state, $\hat{\mathbf{x}}[n|n-1]$, and the observational operator, $\mathbf{H}[n]$, compute the predicted observation, $\hat{\mathbf{y}}[n|n-1]$ based on equation 3.6.

4. From equation 3.7 obtain the innovation, $\mathbf{z}[n]$, as the difference between the actual observation, $\mathbf{y}[n]$, and the predicted observation.

5. Compute the Kalman gain based on the prior error covariance $\mathbf{P}[n|n-1]$, observation operator, and observational noise $\mathbf{R}[n]$.

6. Obtain the posterior (filtered) state estimate, $\hat{\mathbf{x}}[n|n]$, based by adding the product of the gain, $\mathbf{K}[n]$, and the innovation, $\mathbf{z}[n]$, to the prior estimate $\hat{\mathbf{x}}[n|n-1]$.

7. Compute the posterior error covariance, $\mathbf{P}[n|n]$, from equation 3.12.

8. Repeat steps 2-7 with the posterior state ($\hat{\mathbf{x}}[n|n]$) and error covariance ($\mathbf{P}[n|n]$) as the new initial conditions.

The above outline illustrates the two distinctive components of the KF algorithm: (1) the prediction cycle that yields the prior estimate (steps 1-2), and (2) the analysis cycle that constrains the prior estimate to new information in the form of observations (steps 3-7).

3.2 The ensemble Kalman filter

3.2.1 Original formulation

The ensemble Kalman filter (EnKF) [Evensen, 1994, 2003, 2004] provides a mathematical approximation to the KF for nonlinear models. The essence of the EnKF is that the prior error covariance matrix in equation 3.11 is replaced with an approximation that is computed from an ensemble of Monte Carlo model simulations when an observation becomes available. The ensemble members, or replicates, are produced by propagating an initial ensemble forward with the nonlinear dynamics of the model, possibly subjected to uncertain forcings or uncertain model parameters. In this case, the more generic form of the model equation (3.8) would be,

$$\hat{\mathbf{x}}\left[n|n-1\right] = \mathcal{F}\left(\hat{\mathbf{x}}\left[n-1|n-1\right], \mathbf{v}\left[n\right]\right), \tag{3.22}$$

where $\mathcal{F}(\cdot)$ is a generally nonlinear vector-valued function that propagates an initial state corresponding to the posterior state estimate at the previous analysis cycle, $\hat{\mathbf{x}}\left[n-1|n-1\right]$, under the focrings $\mathbf{v}\left[n\right]$. Dropping specific dependence on time n for sake of clarity, let $\mathbf{x}_i^f$ be an N-dimensional state vector that is an ensemble replicate corresponding to a particular Monte Carlo simulation as posed in the previous equation. The superscript f is notation used to denote that the replicate corresponds to a "first guess," or non-filtered estimate. The prior estimate of the state is embodied by the ensemble of all such replicates, the so-called first guess ensemble,

$$\mathbf{X}^f = \left[\mathbf{x}_1^f \mid \mathbf{x}_2^f \mid \ldots \mid \mathbf{x}_K^f\right]. \tag{3.23}$$

Denote the first guess ensemble mean matrix, $\bar{\mathbf{X}}$ as the matrix whose dimensions are identical to $\mathbf{X}^f$ and whose columns are the sample mean, computed through the

rows of $\mathbf{X}^f$,

$$\bar{\mathbf{X}} = \mathbf{X}^f \mathbf{1}_K, \tag{3.24}$$

where 1_K is an $K \times N$-dimensional matrix whose elements are $1/K$. The ensemble of first guess perturbations from the mean is then,

$$\mathbf{X}^{f\prime} = \mathbf{X}^f - \bar{\mathbf{X}}. \tag{3.25}$$

The prior error covariance estimate is denoted $\mathbf{P}_e^f$, and is computed directly from the matrix containing the ensemble of first guesses $\mathbf{X}^f$,

$$\mathbf{P}_e^f = \frac{\mathbf{X}^{f\prime} \left(\mathbf{X}^{f\prime}\right)^{\mathrm{T}}}{K-1}. \tag{3.26}$$

If the model describing the evolution of the state were perfect the output vector, $\mathbf{y}^t$, would be related to the true state, $\mathbf{x}^t$, through the observing system, $\mathbf{H}$, as follows,

$$\mathbf{y}^t = \mathbf{H}\mathbf{x}^t. \tag{3.27}$$

The data vector (the actual observations), $\mathbf{y}$ are assumed to be related to the true observations through additive Gaussian noise,

$$\mathbf{y} = \mathbf{y}^t + \epsilon, \tag{3.28}$$

where the vector ϵ is again zero mean with error covariance matrix, $\mathbf{R}$,

$$\mathbf{R} = E\left[\epsilon\epsilon^T\right]. \tag{3.29}$$

In ensemble-based data assimilation the uncertainty in the data are often treated

explicitly by perturbing the observations with an ensemble of K vectors ϵ_j that have a covariance matrix approximately equal to R. This yields a matrix containing an ensemble of noisy observations, Y, in which

$$\mathbf{Y} = [\mathbf{y} + \epsilon_1 \mid \mathbf{y} + \epsilon_2 \mid \ldots \mid \mathbf{y} + \epsilon_K] \tag{3.30}$$

$$\mathbf{E} = [\epsilon_1 \mid \epsilon_2 \mid \ldots \mid \epsilon_K] \tag{3.31}$$

$$\mathbf{R}_e = \frac{\mathbf{E}\mathbf{E}^{\mathrm{T}}}{K-1} \tag{3.32}$$

The topic of perturbing observations has historically been controversial in numerical data assimilation because while the observation perturbations are additive, zero mean, and Gaussian, sampling error associated with low ensemble sizes can induce bias in the observations [Whitaker and Hamill, 2002]. Although Burgers *et al.* [1998] presented a theoretical justification of perturbation of the observations, several have developed analysis schemes that do not require perturbation of the observations (e.g., see Lermusiaux and Robinson [1999]; Anderson [2001]; Bishop *et al.* [2001]; Whitaker and Hamill [2002]) Nevertheless, perturbed observation ensemble data assimilation remains widely used and is the focus here.

The nomenclature of ensemble-based data assimilation refers to the state update state (equations 3.10 and 3.11) as the analysis step, and the posterior state estimate $\mathbf{X}^a$ as the analyzed ensemble. In analogous fashion to the Kalman filter update equation the analyzed ensemble is computed as follows,

$$\mathbf{X}^a = \mathbf{X}^f + \mathbf{P}_e^f \mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\mathbf{P}_e^f \mathbf{H}^{\mathrm{T}} + \mathbf{R}_e\right)^{-1} \left(\mathbf{Y} - \mathbf{H}\mathbf{X}^f\right) \tag{3.33}$$

The EnKF is presented diagrammatically in figure 3.1. As the graphic indicates, the generic progression of the algorithm is fairly simple: (1) the nonlinear model is

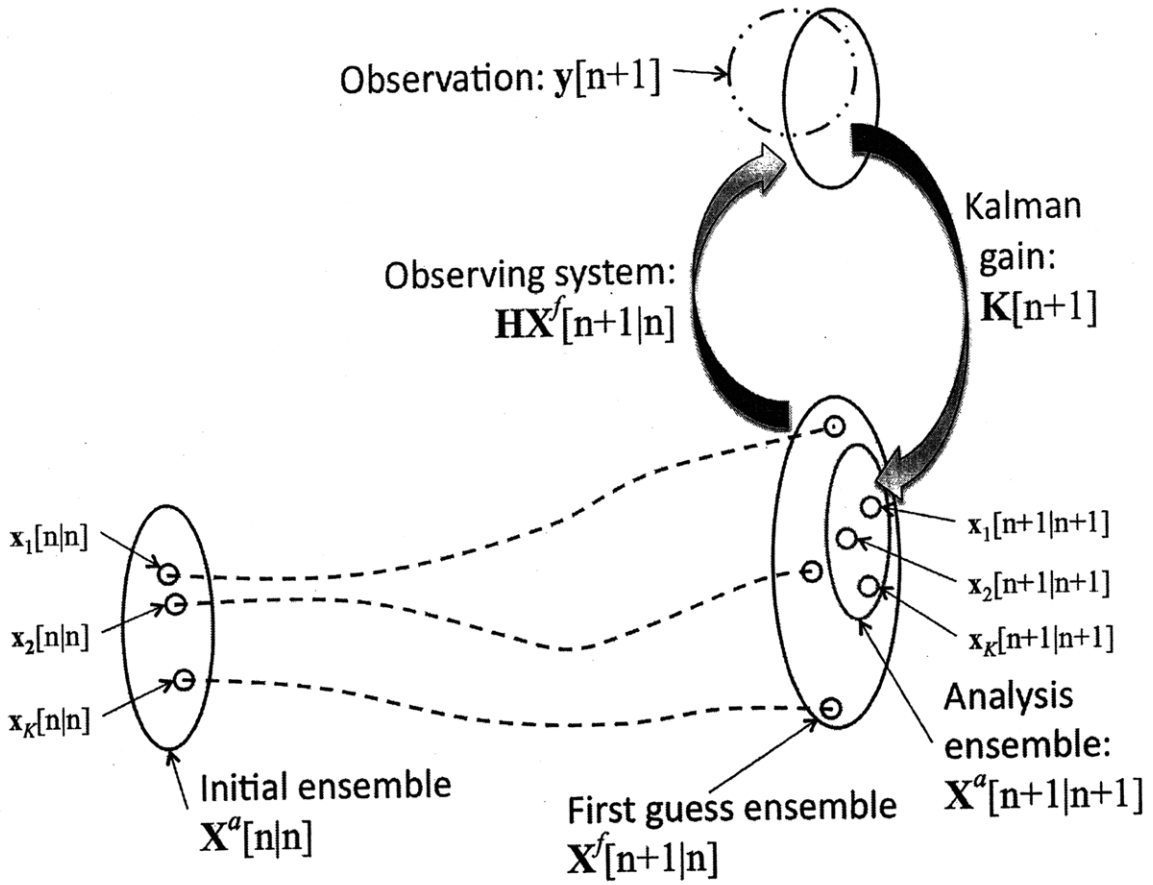


Figure 3-1: A diagrammatic representation of the ensemble Kalman Filter (EnKF). As seen, the algorithm amounts largely to a Monte Carlo simulation with explicit representation of uncertainty in the model forcings and parameters, coupled with a Kalman update procedure that substitutes an estimate of the state error covariance matrix from the ensemble itself. The updated states are then used to re-initialize the model, and the Monte Carlo simulation procedure repeats itself until the next observation becomes available.

used to propagate forward K initial states under uncertain forcings and parameters, (2) when an observation is available this ensemble is used to derive an estimate of the state error variance-covariance matrix which is then used in a Kalman update to derive a posterior estimate given the observation, and (3) the nonlinear model is then re-initialized with the K analyzed ensemble state vectors and propagated forward in time (again under uncertain forcings and parameters) until the next observation becomes available. This process can be repeated indefinitely.

In many practical applications, the relationship between the state and the observation is not a simple linear transformation. That is, the predicted observation

y cannot be posed as the matrix-vector product Hx. In such cases, the observation operator can be expressed more generally as a nonlinear operation,

$$\hat{\mathbf{y}} = \mathcal{H}(\hat{\mathbf{x}}) \tag{3.34}$$

where $\mathcal{H}$ is a vector valued function that nonlinearly transforms the $N \times 1$ state vector $\hat{\mathbf{x}}$ into the $m \times 1$ vector of predicted observations $\hat{\mathbf{y}}$.

Although rarely computed directly in practice, the analyzed state error covariance is defined as,

$$\mathbf{P}^a = \mathbf{P}_e^f - \mathbf{P}_e^f \mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\mathbf{P}_e^f \mathbf{H}^{\mathrm{T}} + \mathbf{R}_e\right)^{-1} \mathbf{H}\mathbf{P}_e^f \tag{3.35}$$

The significance of $\mathbf{P}^a$ will become apparent in the following section. Because the ensemble of analysis perturbations ($\mathbf{X}^{a\prime}$) is essentially the square root of the analysis error covariance matrix $\mathbf{P}^a$ (i.e., $\mathbf{P}^a \approx 1/(K-1)\mathbf{X}^{a\prime}(\mathbf{X}^{a\prime})^{\mathrm{T}}$), the algorithm presented below seeks to estimate $\mathbf{X}^{a\prime}$ and $\bar{\mathbf{x}}^a$ without directly constructing $\mathbf{P}_e^f$.

3.3 Implementation of a square root analysis

The matrix containing the ensemble estimate of the state error covariance, $\mathbf{P}_e^f$, is of dimension $N \times N$. For problems with large state dimensionality it can be computationally infeasible or impractical to compute $\mathbf{P}_e^f$ directly through equation 3.26. Evensen [2004] reviewed several analysis schemes that use the ensemble of first guess perturbations, $\mathbf{X}^{f\prime}$, since it is effectively the square root of $\mathbf{P}_e^f$. The following is an outline of a square root analysis scheme that largely follows the outline first set forth by Evensen [2004], which is also stable when $\mathbf{R}_e$ is of low rank. For the sake of completeness, in the meteorological literature the presented implementation is sometimes referred to as the ensemble transform Kalman filter (ETKF) (esp., Bishop *et al.* [2001]). From its definition in equation 3.33, up to

normalization by $(K-1)$, $\mathbf{P}^a$ can be written as,

$$\mathbf{X}^{a\prime}\mathbf{X}^{a\prime\mathrm{T}} = \mathbf{X}^{f\prime}\left(\mathbf{I} - \mathbf{S}^{\mathrm{T}}\mathbf{C}_e^{-1}\mathbf{S}\right)\mathbf{X}^{f\prime\mathrm{T}}, \tag{3.36}$$

where the $m \times K$ matrix S contains the ensemble of predicted observation perturbations (that is, $\mathbf{S} = \mathbf{H}\mathbf{X}^{f\prime}$) and the $K \times K$ matrix $\mathbf{C}_e$ is defined as,

$$\begin{aligned} \mathbf{C}_e &= \mathbf{S}\mathbf{S}^{\mathrm{T}} + (K-1)\,\mathbf{R}_e \\ &= \mathbf{S}\mathbf{S}^{\mathrm{T}} + \mathbf{E}\mathbf{E}^{\mathrm{T}}, \end{aligned} \tag{3.37}$$

where the matrix E is an ensemble of observation perturbations. Because matrix $\mathbf{C}_e$ is based on an ensemble approximation and is potentially close to singular, $\mathbf{C}_e^{-1}$ in equation (3.33) should be replaced by a stable pseudo-inverse, denoted by Evensen [2004] as $\mathbf{C}^{+}$. The motivation of the square root approach is to find an expression for $\mathbf{X}^{f\prime}$, and a way to find the updated ensemble mean $\bar{\mathbf{x}}^a$ without constructing the full $N \times N$ matrix $\mathbf{P}_e^f$. By inspection of equation (3.36), it can be seen that this amounts to finding some way to factor the matrix $\left(\mathbf{I} - \mathbf{S}^{\mathrm{T}}\mathbf{C}_e^{-1}\mathbf{S}\right)$ into a matrix multiplied by its own transpose. The matrix $m \times K$ matrix S can be decomposed using the singular value decomposition (SVD) as follows,

$$\mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0^T = \mathbf{S} \tag{3.38}$$

A stable pseudoinverse of S, $\mathbf{S}^{+}$ can be computed as,

$$\mathbf{S}^{+} = \mathbf{V}_0\mathbf{\Sigma}_0^{+}\mathbf{U}_0^T, \tag{3.39}$$

where the matrix $\mathbf{\Sigma}_0^{+}$ is a diagonal $K \times m$ matrix where the elements of the diagonal are defined as $\mathrm{diag}(\mathbf{\Sigma}_0^{+}) = \sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_{K-1}^{-1}, 0$, where the σ_i represent the diagonal entries of the matrix $\mathbf{\Sigma}_0$. With the stable pseudoinverse of $\mathbf{\Sigma}_0$ accessible

via the SVD, equation 3.37 can be rewritten as follows,

$$\begin{aligned} \mathbf{C}_e &= \mathbf{S}\mathbf{S}^{\mathrm{T}} + \left(\mathbf{S}\mathbf{S}^{+}\right)\mathbf{E}\mathbf{E}^{\mathrm{T}}\left(\mathbf{S}\mathbf{S}^{+}\right) \\ &= \mathbf{U}_0\boldsymbol{\Sigma}_0\left(\mathbf{I} + \boldsymbol{\Sigma}_0^{+}\mathbf{U}_0^{\mathrm{T}}\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{U}_0\boldsymbol{\Sigma}_0^{+\mathrm{T}}\right)\boldsymbol{\Sigma}_0^{\mathrm{T}}\mathbf{U}_0{}^{\mathrm{T}} \\ &= \mathbf{U}_0\boldsymbol{\Sigma}_0\left(\mathbf{I} + \mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}\right)\boldsymbol{\Sigma}_0^{\mathrm{T}}\mathbf{U}_0^{\mathrm{T}}, \end{aligned} \tag{3.40}$$

where matrix $\mathbf{X}_0$ is defined as,

$$\mathbf{X}_0 = \boldsymbol{\Sigma}_0^{+}\mathbf{U}_0^{\mathrm{T}}\mathbf{E}. \tag{3.41}$$

The matrix $\mathbf{X}_0$ can also be decomposed using the SVD to obtain,

$$\mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}_1^{\mathrm{T}} = \mathbf{X}_0. \tag{3.42}$$

Substituting equation (3.42) into equation (3.40), the matrix $\mathbf{C}_e$ can be expressed as,

$$\begin{aligned} \mathbf{C}_e &= \mathbf{U}_0\boldsymbol{\Sigma}_0\left(\mathbf{I} + \mathbf{U}_1\boldsymbol{\Sigma}_1^2\mathbf{U}_1^{\mathrm{T}}\right)\boldsymbol{\Sigma}_0^{\mathrm{T}}\mathbf{U}_0{}^{\mathrm{T}} \\ &= \mathbf{U}_0\boldsymbol{\Sigma}_0\mathbf{U}_1\left(\mathbf{I} + \boldsymbol{\Sigma}_1^2\right)\mathbf{U}_1^{\mathrm{T}}\boldsymbol{\Sigma}_0^{\mathrm{T}}\mathbf{U}_0{}^{\mathrm{T}}. \end{aligned} \tag{3.43}$$

Note that in equation (3.43) the matrix $\left(\mathbf{I} + \boldsymbol{\Sigma}_1^2\right)$ is always diagonal and therefore computing its inverse is trivial. A stable psuedoinverse of $\mathbf{C}_e$ can now be expressed as the following,

$$\begin{aligned} \mathbf{C}_e^{+} &= \left(\mathbf{U}_0\boldsymbol{\Sigma}_0^{+\mathrm{T}}\mathbf{U}_1\right)\left(\mathbf{I} + \boldsymbol{\Sigma}_1^2\right)^{-1}\left(\mathbf{U}_0\boldsymbol{\Sigma}_0^{+\mathrm{T}}\mathbf{U}_1\right)^{\mathrm{T}} \\ &= \mathbf{X}_1\left(\mathbf{I} + \boldsymbol{\Sigma}_1^2\right)^{-1}\mathbf{X}_1^{\mathrm{T}}, \end{aligned} \tag{3.44}$$

where the matrix $\mathbf{X}_1 = \mathbf{U}_0\boldsymbol{\Sigma}_0^{+\mathrm{T}}\mathbf{U}_1$. Substituting the stable psuedoinverse of $\mathbf{C}_e$, $\mathbf{C}_e^+$ into equation (3.36) for $\mathbf{C}_e^{-1}$, the square root of the ensemble estimate of the state error covariance matrix can be written as follows,

$$\begin{aligned}
\mathbf{X}^{a\prime}\mathbf{X}^{a\prime\mathrm{T}} &= \mathbf{X}^{f\prime}\left(\mathbf{I} - \mathbf{S}^{\mathrm{T}}\mathbf{C}_e^{+}\mathbf{S}\right)\mathbf{X}^{f\prime\mathrm{T}} \\
&= \mathbf{X}^{f\prime}\left(\mathbf{I} - \mathbf{S}^{\mathrm{T}}\mathbf{X}_1\left(\mathbf{I}+\boldsymbol{\Sigma}_1^2\right)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{S}\right)\mathbf{X}^{f\prime\mathrm{T}} \\
&= \mathbf{X}^{f\prime}\left(\mathbf{I} - \left[\left(\mathbf{I}+\boldsymbol{\Sigma}_1^2\right)^{-1/2}\mathbf{X}_1^{\mathrm{T}}\mathbf{S}\right]^{\mathrm{T}}\left[\left(\mathbf{I}+\boldsymbol{\Sigma}_1^2\right)^{-1/2}\mathbf{X}_1^{\mathrm{T}}\mathbf{S}\right]\right)\mathbf{X}^{f\prime\mathrm{T}} \\
&= \mathbf{X}^{f\prime}\left(\mathbf{I} - \mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2\right)\mathbf{X}^{f\prime\mathrm{T}}, \qquad (3.45)
\end{aligned}$$

with $\mathbf{X}_2 = \left(\mathbf{I}+\boldsymbol{\Sigma}_1^2\right)^{-1/2}\mathbf{X}_1^{\mathrm{T}}\mathbf{S}$. Finally, the ensemble of analyzed perturbations ($\mathbf{X}^{a\prime}$) can be computed as,

$$\mathbf{X}^{a\prime} = \mathbf{X}^{f\prime}\mathbf{V}_2\sqrt{\mathbf{I} - \boldsymbol{\Sigma}_2^{\mathrm{T}}\boldsymbol{\Sigma}_2}\,\boldsymbol{\Theta}^{\mathrm{T}}, \qquad (3.46)$$

where the matrix $\mathbf{V}_2$ arises from a singular value decomposition of the matrix $\mathbf{X}_2 = \mathbf{U}_2\boldsymbol{\Sigma}_2\mathbf{V}_2^{\mathrm{T}}$. The random orthogonal matrix Θ^{T} has the effect of spreading the variance reduction associated with the analysis across the ensemble members and Evensen [2004] recommends using it by default. A random orthogonal matrix can be constructed by using the right singular vectors from a singular value decomposition of a random $K \times K$ matrix [Evensen, 2004]. Recent studies, however, suggest that this random rotation can potentially induce a bias in the analysis ensemble [Wang *et al.*, 2004; Sakov and Oke, 2008]. Although Livings *et al.* [2008] provides some diagnostic tools to assess the degree to which multiplication of the analysis by Θ^{T} causes bias, a rigorous survey of the magnitude of this effect, particularly in geophysical models, has not yet been conducted, nor have any mitigation strategies been widely accepted.

The following summary gives the step-by-step implementation of this algorithm to compute the analyzed ensemble of perturbations $\mathbf{X}^{a\prime}$:

1. Compute the SVD of the matrix S: $\mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0^{\mathrm{T}} = \mathbf{S}$.

2. Construct the matrix $\mathbf{X}_0 = \mathbf{\Sigma}_0^{+}\mathbf{U}^{\mathrm{T}}\mathbf{E}$. The matrix $\mathbf{\Sigma}_0^{+}$ is $K \times m$ with σ_1^{-1}, σ_2^{-1},...,σ_{m-1}^{-1}, 0 on the diagonal and zeros elsewhere.

3. Compute the SVD of $\mathbf{X}_0 = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^{\mathrm{T}}$.

4. Construct the matrix $\mathbf{X}_1 = \mathbf{U}_0\mathbf{\Sigma}_0^{+\mathrm{T}}\mathbf{U}_1$.

5. Update the ensemble mean through the following equation:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{X}^{f\prime}\mathbf{S}^{\mathrm{T}}\mathbf{X}_1\left(\mathbf{I} + \mathbf{\Sigma}^2\right)^{-1}\mathbf{X}_1^{\mathrm{T}}\left(\mathbf{d} - \mathbf{H}\bar{\mathbf{x}}^f\right) \tag{3.47}$$

   This can be done efficiently through the following matrix-vector operations:

   a. Compute the vector $\mathbf{y}_0 = \mathbf{X}_1^{\mathrm{T}}\left(\mathbf{d} - \mathbf{H}\bar{\mathbf{x}}^f\right)$.

   b. Compute $\mathbf{y}_2 = \left(\mathbf{I} + \mathbf{\Sigma}_1^2\right)^{-1}\mathbf{y}_0$. Recognize that the term inside the parentheses is easily inverted because it is square and diagonal.

   c. Compute $\mathbf{y}_3 = \mathbf{X}_1\mathbf{y}_2$.

   d. Compute the vector $\mathbf{y}_4 = \mathbf{S}^{\mathrm{T}}\mathbf{y}_3$.

   e. Update the ensemble mean from $\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{X}^{f\prime}\mathbf{y}_4$.

6. Construct the matrix $\mathbf{X}_2 = \left(\mathbf{I} + \mathbf{\Sigma}_1^2\right)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{S}$.

7. Perform the SVD of $\mathbf{X}_2$, such that $\mathbf{U}_2\mathbf{\Sigma}_2\mathbf{V}_2^{\mathrm{T}} = \mathbf{X}_2$.

8. Construct the ensemble of analyzed state perturbations from
   $\mathbf{X}^{a\prime} = \mathbf{X}^{f\prime}\mathbf{V}_2\sqrt{\mathbf{I} - \mathbf{\Sigma}_2^{\mathrm{T}}\mathbf{\Sigma}_2}\mathbf{\Theta}^{\mathrm{T}}$

The ensemble of analyzed state vectors $\mathbf{X}^a$ is then determined by adding the analyzed ensemble mean ($\bar{\mathbf{x}}^a$) to each column of $\mathbf{X}^{a\prime}$. Each column. Each of the K columns of $\mathbf{X}^a$ represents a state vector with which the model can be reinitialized.

3.3.1 Treatment of nonlinear observation operator

As discussed above it may be necessary or more convenient to formulate the observing system as a nonlinear transformation of the state vector, $\mathcal{H}(\mathrm{x})$, as in equation (3.34) above. If it is not possible to linearize the observing system equations, the square root implementation of the EnKF formulation outlined above above can still be used to constrain the model estimate to the observational data. However, it requires that we assume

$$\mathbf{S} = \mathbf{H}\mathbf{X}^{f\prime} \approx \mathcal{H}(\mathbf{X}^{f}) - \mathcal{H}(\bar{\mathbf{X}}). \tag{3.48}$$

The validity of this assumption depends on the degree to which the observing system yields predicted observations that are linearly related to the state. Hence, for observing systems that yield predicted observations that exhibit only a weakly nonlinear dependence on the modeled soil moisture state the assumption is not likely overly restrictive. Alternatively, the impact of this assumption can be reduced provided that the predicted observations can be easily transformed (e.g., a logarithmic transformation) so that the observing system becomes more linear. The degree to which the microwave observing system formulated in this thesis yields predicted observations that are approximately linearly related to the soil moisture modeled by tRIBS-VEGGIE will be discussed in the following chapter.

3.4 Recasting tRIBS-VEGGIE outputs as a state vector

This chapter concludes with a description of how the square root implementation of the ensemble Kalman filter can be implemented with the tRIBS-VEGGIE model. In particular, what is required is a brief translation of how the tRIBS-VEGGIE soil moisture state can be expressed a vector.

In the soil moisture data assimilation system constructed, the state vector at the time of observation (x) is the profile soil moisture estimate for every tRIBS-

VEGGIE computational pixel in the Voronoi Polygon Network representation of the watershed under study. Since the soil profile consists of 9 finite element layers associated with 10 soil moisture values at the boundaries of each layer, the state vector can be constructed as,

$$\mathbf{x}_k = \begin{bmatrix} \theta_{k,1,1} \\ \theta_{k,1,2} \\ \vdots \\ \theta_{k,1,10} \\ \theta_{k,2,1} \\ \theta_{k,2,2} \\ \vdots \\ \theta_{k,2,10} \\ \theta_{k,N_S,1} \\ \theta_{k,N_S,2} \\ \vdots \\ \theta_{k,N_S,10} \end{bmatrix} \tag{3.49}$$

where $\theta_{k,i,j}$ is the simulated soil moisture of ensemble replicate k, at computational pixel i of N_S in the watershed, and at finite element node j of 10 in the soil profile. The first guess ensemble estimate of the state $\mathbf{X}^f$ is simply the collection of the soil moisture state from K independent tRIBS-VEGGIE simulations at the time of observation. The observing system which relates the soil moisture states to microwave brightness temperatures and radar backscatter ($\mathcal{H}(\mathrm{x})$) is formulated in the following chapter. Moreover, the specification of the observational error covariance structure (R) is a property of the observing instrument and is detailed in the following chapter as well. Treatment of input uncertainty in Monte Carlo soil moisture simulation with tRIBS-VEGGIE is dealt with in Chapters 5, 6, and 7.

In conclusion, one advantageous aspect of implementing the EnKF is that minimal modifications to the model formulation are required. The primary modification

to the tRIBS-VEGGIE code that is required in the construction of the soil moisture data assimilation outlined here is the ability to output the state of the system and restart the model with a new state vector retrieved from the Kalman update procedure when the observation is available. The Kalman update is completely modular to the tRIBS-VEGGIE modeling framework and could, in principle, be applied to any soil moisture-evolving model whose state space can be recast in the manner shown in equation (3.49).

CHAPTER 4

FORMULATION OF THE OBSERVING SYSTEM

This chapter details the formulation of two microwave observing systems that serve to transform tRIBS-VEGGIE simulations of the near-surface moisture and energy state into predicted observations. As demonstrated in the previous chapter, ensemble-based data assimilation depends on a mathematical construct to project the state vector into the observation space to be compared with data. The first observing system is a passive microwave model based on black body radiative transfer concepts. The second system introduced is a microwave radar system that models the predicted backscattered energy at a sensor depending on the degree of wetness of the reflecting soil. Effects of topography on the geometry of both observing systems, particularly as it affects the local incidence angle and polarization rotation, is studied after outlining the system of equations used for each observing system. The properties and sensitivity of each observing system are then investigated through a series of numerical experiments. These experiments are designed to distill the sensitivity of predictions of microwave brightness temperature and radar backscatter to topographic factors that influence emission/reflection and on observing system geometry.

4.1 Passive remote sensing observing system

Emission of microwave radiation from the Earth's surface is modeled through a layered single-scattering radiative transfer model [Kerr and Wigneron, 1995]. The predicted observation of brightness temperature, T_{bp} observed by the sensor at polarization p is comprised of a component from a vegetated fractional area (f_c) and a component from the bare soil ($1 - f_c$) as presented by Crow *et al.* [2005]

$$\begin{aligned} T_{bp} = f_c \Bigg\{ & T_S e_p \exp\left(-\frac{\tau_p}{\cos\theta_L}\right) + \\ & T_C(1-\omega_p)\left(1-\exp\left(-\frac{\tau_p}{\cos\theta_L}\right)\right) \\ & \cdot\left(1+r_p\exp\left(-\frac{\tau_p}{\cos\theta_L}\right)\right)\Bigg\} + (1-f_c)T_S e_p \end{aligned} \tag{4.1}$$

where T_S is the soil temperature (K), e_p is the emissivity of the soil (dimensionless), τ_p is the nadir vegetation opacity (nepers) at polarization p, θ_L is the local incidence angle (radians), T_C and T_S are the canopy and soil temperatures (K), ω_p is the single scattering albedo (dimensionless) at polarization p, and r_p is the local reflectivity of the soil ($r_p = 1 - e_p$) (dimensionless). The $-\tau_p/\cos\theta_L$ term in equation (4.1) is a theoretical representation for the vegetation opacity, taking into account the slant path through the vegetation layer Njoku and Entekhabi [1996] originally formulated by Kirdiashev *et al.* [1979] and later derived using effective medium theory by Wegmuller *et al.* [1995].

Vegetation opacity, τ_p, varies by plant type and is often modeled as a linear function of vegetation water content, V_{wc} (kg m^{-2}) in which the constant of proportionality, b_p, is a vegetation type-dependent opacity coefficient [Jackson and Schmugge, 1991; de Griend and Wigneron, 2004] that depends on polarization,

$$\tau_p = b_p V_{wc}. \tag{4.2}$$

Because this work focuses on semiarid landscapes with natural grasses, vegetation water content is estimated through an empirical linear relationship between V_{wc} and Leaf Area Index (LAI) based on field data suggested by de Rosnay *et al.* [2006] for semiarid grasses,

$$V_{wc} = 0.37LAI + 0.123. \tag{4.3}$$

The inputs to the radiative transfer model, T_S, T_C, ω, V_{wc} (LAI), b, h, and θ characterize the state of land surface soil and vegetation. These states are evolved in the spatially distributed tRIBS-VEGGIE model, which takes as input hydrometeorological forcings. The surface reflectivity, r_p, is computed as the reflectivity of an equivalent smooth surface, r_{sp}, extinguished exponentially by roughness of the soil surface as parameterized by h, which is linearly related to the root-mean-square (rms) surface height,

$$r_p = r_{sp} \exp(-h) \tag{4.4}$$

The above formulation is meant to capture the effects of soil texture, vegetation and microtopographic height variation on the reflectivity. The reflectivity of the smooth surface, r_{sp} is a function of the dielectic constant of the soil-air-water matrix and the incidence angle, θ_L. The Fresnel equations give the reflectivity of the smooth surface, r_{sp}, at both the horizontal ($p = h$) and vertical ($p = v$) polarizations [Kong, 1990]. For $p = h$,

$$r_{sh} = \left| \frac{\cos\theta_L - \sqrt{\varepsilon_{\text{eff}} - \sin^2\theta_L}}{\cos\theta_L + \sqrt{\varepsilon_{\text{eff}} - \sin^2\theta_L}} \right|^2, \tag{4.5}$$

and for $p = v$,

$$r_{sv} = \left| \frac{\sqrt{\varepsilon_{\text{eff}} - \sin^2\theta_L} - \varepsilon_{\text{eff}} \cos\theta_L}{\sqrt{\varepsilon_{\text{eff}} - \sin^2\theta_L} + \varepsilon_{\text{eff}} \cos\theta_L} \right|^2 . \tag{4.6}$$

where ε_{eff} is the effective dielectric constant of the soil-water-air medium. Estimation of the dielectric constant is also required for the active remote sensing observing system and is discussed below.

4.2 Active remote sensing observing system

Active remote sensing of soil moisture with radar is a promising complement to passive microwave radiometry [Dobson and Ulaby, 1986; Eom and Boener, 1986; Wang *et al.*, 1986; Engman, 1991; Evans *et al.*, 1992; Fung *et al.*, 1992; Oh *et al.*, 1992; Fung, 1994; Chen *et al.*, 1995; Dubois *et al.*, 1995; Altese *et al.*, 1996; Shi *et al.*, 1997; Hoeben and Troch, 2000; Njoku *et al.*, 2002; Narayan *et al.*, 2006; Thoma *et al.*, 2006; Narayan and Lakshmi, 2008]. In the lower microwave range of the electromagnetic spectrum ($\sim$ 1 GHz) the amount of energy backscattered to an observing sensor depends on the dielectric constant of the reflective soil media. Its sensitivity to soil moisture was previously discussed. Satellite observations from passive systems are of questionable use for hillslope-scale soil moisture estimation because the resolution of radiometer pixels is typically of the order of 10's of kilometers. Hillslope-scale soil moisture estimates based on brightness temperature observations at these scales would require spatial disaggregation across many orders of magnitude in scale. By contrast, relative to L-band microwave radiometers, L-band microwave radar observations have much higher spatial resolutions [Entekhabi *et al.*, 2004]. The National Aeronautic and Space Administration's (NASA's) Soil Moisture Active-Passive (SMAP) mission, tentatively scheduled for launch between 2010-2013, will release observations of active microwave backscatter at spatial resolutions of approximately 3 km. Single look backscatter observations at resolutions of 1 km may also be available.

The Integral Equation Model (IEM) [Fung, 1994; Altese *et al.*, 1996; Chen *et al.*, 1995; Hoeben and Troch, 2000] provides a conceptual framework through which the backscatter can be modeled with information about the radar sensor, the roughness features of the area, and the near-surface moisture conditions. This model is applicable for surfaces with small to medium roughness, parameterized by a root-mean-square (rms) surface height σ, and a correlation length L. Owing to its sensitivity to roughness, the IEM is most applicable for sparsely vegetated surfaces [Dubois *et al.*, 1995]. Roughness parameters are often normalized by the free space wave number k to recast the parameterization in terms of $k\sigma$ and kL. The wave number (k) is equal to $2\pi/\lambda$, where λ is the wavelength. The like polarized (e.g., hh or vv) dimensionless backscatter coefficient σ° under these conditions is expressed as,

$$\sigma^\circ_{pp} = \frac{k^2}{2} e^{-2k_z^2\sigma^2} \sum_{j=1}^{\infty} \sigma^{2j} |I^j_{pp}|^2 \frac{W^{(j)}(-2k_x, 0)}{j!}, \tag{4.7}$$

where $k_z = k\cos\theta_L$, $k_x = k\sin\theta_L$, and θ_L is the local incidence angle. Note that equation (4.7) is valid for like polarized observations (i.e., $pp = hh$ or vv). The backscattered energy is often given in units of decibels [dB], in which σ°_{pp} [dB] $= 10\log_{10}\sigma^\circ_{pp}$. Throughout the remainder of this work, backscatter will be presented in dB. The variable I^j_{pp} is computed as,

$$I^j_{pp} = (2k_z)^j f_{pp} e^{-k_z^2\sigma^2} + \frac{(k_z)^j \left[F_{pp}(-k_x, 0) + F_{pp}(k_x, 0)\right]}{2}, \tag{4.8}$$

and $W^{(j)}(\zeta, \xi)$ is the $j$the power of the surface roughness spectrum. It is related to the jth power of the surface roughness correlation function, $\rho(\zeta, \xi)$, through the Fourier transform,

$$W^{(j)}(\zeta, \xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho^j(\xi_1, \zeta_1) e^{-j\zeta\xi_1 - j\xi\zeta_1} d\xi_1 d\zeta_1. \tag{4.9}$$

Appendix 2B of Fung [1994] provides several Fourier transform pairs for the

functions $\rho(\zeta,\xi)$ and $W^{(j)}(\zeta,\xi)$. Here it is assumed that the roughness correlation is an exponential function as in Fung *et al.* [1992]; Chen *et al.* [1995] and Altese *et al.* [1996]. As pointed out in Fung [1994], although the exponential roughness correlation is frequently used in practice, it has the theoretically undesirable property of being undifferentiable at the origin. However, for practical applications, the exponential correlation function can be thought of as an approximation to a more complex function that is differentiable at the origin, examples of which are shown in Fung [1994]. For some single parameter correlation function parameterized by a correlation length L and of the form $\exp(-|\xi|/L)$, however, the jth power of the roughness spectrum can be analytically written as,

$$W^{(j)}(K) = \left(\frac{L}{j}\right)^2 \left[1 + \left(\frac{KL}{j}\right)^2\right]^{-1.5} \tag{4.10}$$

The above formulation also assumes that the roughness correlation function is isotropic and the Fourier transform in equation (4.9) reduces to a 1-dimensional transform. In equation (4.8) f_{pp} and F_{pp} are, respectively, the polarization-specific Kirchoff and complementary field coefficients. Expressions given below are taken from Fung [1994]. The Kirchoff coefficients are given by,

$$f_{vv} = \frac{2R_{\|}}{\cos\theta_L}, \tag{4.11}$$

and

$$f_{hh} = \frac{-2R_{\perp}}{\cos\theta_L}, \tag{4.12}$$

where $R_{\|}$ and $R_{\perp}$ are the Fresnel reflection coefficients in the vertical and horizontal polarizations, respectively. The Fresnel reflection coefficients are a function of θ_L, the dielectric constant of the soil-water-air medium ε, and are given by,

$$R_{\parallel} = \frac{\varepsilon \cos\theta_L - \sqrt{\varepsilon - \sin^2\theta_L}}{\varepsilon \cos\theta_L + \sqrt{\varepsilon - \sin^2\theta_L}}, \tag{4.13}$$

and

$$R_{\perp} = \frac{\cos\theta_L - \sqrt{\varepsilon - \sin^2\theta_L}}{\cos\theta_L + \sqrt{\varepsilon - \sin^2\theta_L}}. \tag{4.14}$$

The complementary field coefficients in the horizontally and vertically like-polarized states can be computed, respectively, as:

$$F_{vv}(-k_x, 0) + F_{vv}(k_x, 0) = \frac{2\sin^2\theta_L}{\cos\theta_L}\left[\left(1 - \frac{\varepsilon\cos^2\theta_L}{\varepsilon - \sin^2\theta_L}\right)(1 - R_{\parallel})^2 + \left(1 - \frac{1}{\varepsilon}\right)(1 + R_{\parallel})^2\right] \tag{4.15}$$

and,

$$F_{hh}(-k_x, 0) + F_{hh}(k_x, 0) = \frac{2\sin^2\theta_L}{\cos\theta_L}\left[4R_{\perp} - \left(1 - \frac{1}{\varepsilon}\right)(1 + R_{\perp})^2\right] \tag{4.16}$$

4.2.1 Specification of the dielectric constant ε

The dielectric constant (ε) of the soil-water-air medium is related to the relative proportions of soil, air, and water present in the matrix, and the dielectric properties dielectric properties of the individual constituents. In general, the dielectric constant of a matrix depends on both real and imaginary components,

$$\varepsilon = \varepsilon' + j\varepsilon'', \tag{4.17}$$

where the imaginary part ε'' represents the transmission loss factor, which can

be significant in highly saline soils, particularly at low moisture contents Dobson *et al.* [1985]. In estimating the dielectric constant the imaginary component is often neglected, and a bulk "effective" dielectric constant ε_{eff} is instead estimated.

In this work, two models are optionally used to compute the effective dielectric constant of the variably saturated soil-air-water medium. The first model is an empirical model based on laboratory experiments that expresses the effective dielectric constant ε_{eff} as a 3rd degree polynomial function of volumetric soil moisture θ [m^3/m^3]. The model, proposed in Topp *et al.* [1980] is given as

$$\varepsilon_{\text{eff}} = 3.03 + 9.3\theta + 146\theta^2 - 76.7\theta^3. \tag{4.18}$$

Note that while this does not explicitly account for differences between distinct soil types, it has been shown to be useful in most mineral soils. The second model of the dielectric constant used here was outlined in Friedman [1998], and is formulated based on the composite spheres model. Friedman's model is based on the assumption that the soil can be described as an array of spherical inclusions that are small with respect to the wavelength of an imposed electric potential gradient, E_0, applied at infinity (Fig. 4-1). As such, these spherical inclusions are composed of 3 concentric shells representing the air, solid, and water phases. Inclusions are assumed to be identical in the relative thicknesses of each shell representing a phase (phase-shell), although the absolute sizes of the inclusions can vary. Under these assumptions, the effective dielectric constant of the composite sphere is derived in Friedman [1998]. In Fig. 4-1 the volume fractions of the respective three concentric phase-shells are,

$$\phi_1 = \frac{R_1^3 - R_2^3}{R_1^3}, \tag{4.19}$$

$$\phi_2 = \frac{R_2^3 - R_3^3}{R_1^3}, \tag{4.20}$$

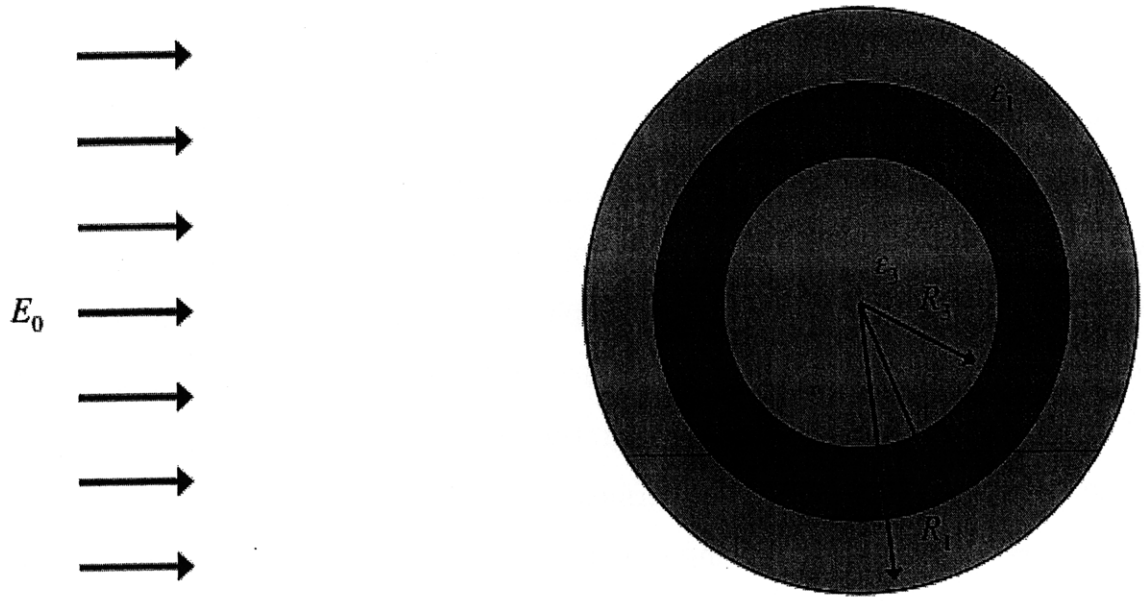


Figure 4-1: Schematic diagram of the composite sphere concept. The composite sphere consists of three concentric spherical shells, each with sizes defined by their outer radii R_i and dielectric constants defined by ε_i. Through effective medium theory the dielectric behavior of the composite sphere can be represented by a single effective dielectric constant ε_{eff} when exposed to an electrical potential gradient E_0 applied at infinity.

and,

$$\phi_3 = \frac{R_3^3}{R_1^2}. \tag{4.21}$$

The expression for the effective dielectric constant ε_{eff} of the sphere requires knowledge of two of the volume fractions described above (ϕ_2, and ϕ_3) and the dielectric constant of each phase (ε_1, ε_2, and ε_3), and is given by,

$$\begin{aligned}\varepsilon_{\text{eff}} = \varepsilon_1 + \Big(&\{3[(\phi_3+\phi_2)(\varepsilon_2-\varepsilon_1)(2\varepsilon_2+\varepsilon_3) - \phi_3(\varepsilon_2-\varepsilon_3) \\ &\cdot(2\varepsilon_2+\varepsilon_1)]\varepsilon_1\} \Big/ \Big[(2\varepsilon_1+\varepsilon_2)(2\varepsilon_2+\varepsilon_3) \\ &-2\frac{\phi_3}{\phi_3+\phi_2}(\varepsilon_2-\varepsilon_1)(\varepsilon_2-\varepsilon_3) - (\phi_3+\phi_2)(\varepsilon_2-\varepsilon_1) \\ &\cdot(2\varepsilon_2+\varepsilon_3) + \phi_3(\varepsilon_2-\varepsilon_3)(2\varepsilon_2+\varepsilon_1)\Big]\Big).\end{aligned} \tag{4.22}$$

Note that the because the above expression relies only on the volume fractions and not the absolute volume of the concentric shells, estimation of ε_{eff} can proceed with knowledge of the volumetric soil moisture and total porosity of the soil. The ordering of the phases of the composite sphere can take 6 different configurations depending on the medium assigned to each of the shells (i.e., air, water, or solid). By convention, the three-phase configuration of a composite sphere is referred to by describing the phases of the innermost to outermost shells. Friedman [1998] investigated the six possible composite sphere configurations and concluded that the water-solid-air (WSA), air-water-solid (AWS), and water-air-solid (WAS) configurations lead to unreasonably low ε_{eff} at high θ when compared to corresponding values of ε_{eff} from the model of Topp *et al.* [1980]. Of the three configurations for which ε_{eff} agrees reasonable well with the model of Topp *et al.* [1980], Friedman [1998] argues that the solid-water-air (SWA) configuration, consistent with water bound to solids at low moisture, and the air-water-solid (ASW) configuration, consistent with a soil water sheet enveloping solids and small air pockets at high moisture, are physically most consistent with hydrologic wetting-drying cycles. Through an effective medium approximation, the effective dielectric constant of the bulk medium is then estimated as a soil moisture weighted-average of the dielectric constants for the two potential composite sphere configurations (i.e., ε_{SWA} and ε_{ASW}):

$$\varepsilon_{\text{eff}} = \left[\frac{\left[f_w^{\text{SWA}} \left(\varepsilon_{\text{ASW}} - 2\varepsilon_{\text{SWA}} \right) + f_w^{\text{ASW}} \left(\varepsilon_{\text{SWA}} - 2\varepsilon_{\text{ASW}} \right) \right]^2}{16} + \frac{\varepsilon_{\text{SWA}} \varepsilon_{\text{ASW}}}{2} \right]^{1/2} - \frac{\left[f_w^{\text{SWA}} \left(\varepsilon_{\text{ASW}} - 2\varepsilon_{\text{SWA}} \right) + f_w^{\text{ASW}} \left(\varepsilon_{\text{SWA}} - 2\varepsilon_{\text{ASW}} \right) \right]}{4} \tag{4.23}$$

where $f_w^{\mathrm{SWA}} = 1 - S_e$, $f_w^{\mathrm{ASW}} = S_e$, and S_e is the effective soil saturation, defined as $S_e = (\theta - \theta_R)/(\theta_S - \theta_R)$ with θ_S and $\theta)R$ representing, respectively, the saturation and residual moisture contents of the soil [m^3/m^3].

What remains in the model outlined in Friedman [1998] is the determination of the dielectric constant of the aqueous phase (ε_W), solid phase (ε_S), and air phase (ε_A). As pointed out in Friedman [1998], estimates of the dielectric constant based on moisture alone tend to overestimate experimental measurements, because the dielectric constant of water in thin films surrounding soil particles is less than that of free water at the same frequency and temperature [Dobson *et al.*, 1985; Dirksen and Dasberg, 1993]. In Friedman [1998] this added complexity is dealt with by assuming the water phase of each composite sphere is subdivided into neighboring bound- and free-water shells. Both composite sphere configurations considered (ASW, and SWA) in this dielectric model place the soil and water phases adjacently. Therefore, the model treats the dielectric constant of the aqueous phase (ε_W) as continuously varying from a minimum value at the interface between the soil and water phases to a maximum at the external boundary of the aqueous phase. The value of ε_W is assumed decay exponentially as a function of the distance away from the soil-water interface,

$$\varepsilon_W(x) = \varepsilon_{\min} + (\varepsilon_{\max} - \varepsilon_{\min})(1 - e^{-\lambda_w x}), \tag{4.24}$$

where $\varepsilon_{\min}$ is assumed to be 5.5, and λ_w is the decay parameter taken to be 10^8 [cm^{-1}]. The bulk value of ε_W for the entire water phase is taken as the harmonic mean of equation (4.24) between the soil-water interface ($x = 0$) and the external water layer boundary ($x = d_w$),

$$\frac{1}{\varepsilon_W} = \frac{1}{d_w} \int_0^{d_w} \frac{1}{\varepsilon_W(x)} dx. \tag{4.25}$$

The upper limit of integration, d_w [m], is the thickness of the entire aqueous phase and is a function of θ, soil bulk density ρ_b [km/m^3], and specific area of the

solids S_{SA} [m^2/kg],

$$d_w = \frac{\theta}{\rho_b S_{SA}}. \tag{4.26}$$

Performing the integration in equation (4.24), the value of ε_W for the aqueous phase of each composite sphere configuration can be computed directly through

$$\varepsilon_W = \frac{d_w \varepsilon_{\max}}{d_w + \frac{1}{\lambda_w} \ln \left[\frac{\varepsilon_{\max} - (\varepsilon_{\max} - \varepsilon_{\min}) e^{-\lambda_w d_w}}{\varepsilon_{\min}} \right]}, \tag{4.27}$$

where $\varepsilon_{\max}$ is the complex dielectric constant of the free water in the medium. Consistent with the assumptions outlined above, the imaginary part of the free water dielectric constant is neglected, and $\varepsilon_{\max}$ in equation (4.27) is assumed to be dominated by the real part, $\varepsilon'_{\max}$. Since ρ_b and S_{SA} vary between soil types, equation (4.26) highlights how this model specifically incorporates information about the soil type being considered, and therefore allows for systematic variation among soil types in a way not encompassed in the model of Topp *et al.* [1980].

The real part of the dielectric constant of the free water phase $\varepsilon'_{\max}$ depends on both frequency and temperature. According to the relaxation theory first proposed by Debye [1929], the temperature and frequency dependence of $\varepsilon'_{\max}$ can be expressed as,

$$\varepsilon'_{\max}(f, T) = \varepsilon'_{\max}(\infty, T) + \frac{\varepsilon'_{\max}(0, T) - \varepsilon'_{\max}(\infty, T)}{1 + (2\pi f \tau)^2} \tag{4.28}$$

where $\varepsilon'_{\max}(\infty, T)$ is the high frequency limit of the dielectric constant of water (assumed to be 5.5 as in Friedman [1998]), $\varepsilon'_{\max}(0, T)$ is the temperature-dependent static (zero frequency) dielectric component of water, f is the frequency [Hz], and τ is the relaxation time of water (taken as 10^{-11} s). The temperature dependence of the static dielectric constant of water $\varepsilon'_{\max}(0, T)$ is computed through an empirical relationship based on the temperature T [°C] Weast [1983],

$$\varepsilon'_{\max}(0,T) = 78.54[1 - 4.579 \times 10^{-3}(T-25) + 1.19 \times 10^{-5}(T-25)^2 - 2.8 \times 10^{-8}(T-25)^3]. \quad (4.29)$$

To reiterate, in this implementation the dielectric model of Friedman [1998] it is assumed that the imaginary part of the dielectric constant (ε''), which represents a transmission loss factor, is negligible. However, as discussed in Dobson *et al.* [1985] and Friedman [1998], the imaginary part of the moist soil dielectric constant is sensitive to the ionic concentration within the aqueous phase. The assumption of a negligible loss factor is, therefore, potentially restrictive in areas with soils that are particularly nutrient-rich (e.g., heavily salinized soils, wetlands, some irrigated lands, etc.). Procedures to estimate the imaginary part of the dielectric constant exist in the remote sensing literature (e.g., Dobson *et al.* [1985]), and future addition of a model to estimate ε'' would make the soil dielectric model outlined here more general. Incorporation of such a model would require introduction of additional soil parameters related to the chemical reactivity of the soil to capture salinity effects at low moisture contents.

To summarize estimation of the dielectric constant with the model proposed by Friedman [1998]:

1. The maximum dielectric constant of the aqueous phase is computed at the frequency and temperature of interest (equations 4.28 and 4.29).

2. The bulk dielectric constant of the aqueous phase is computed based on the thickness of the water shell (equations 4.26 and 4.27).

3. Based on θ and the total porosity, the volume fractions of each phase are computed, and the dielectric constant for both composite sphere configurations (ASW and SWA) is computed (equation 4.22).

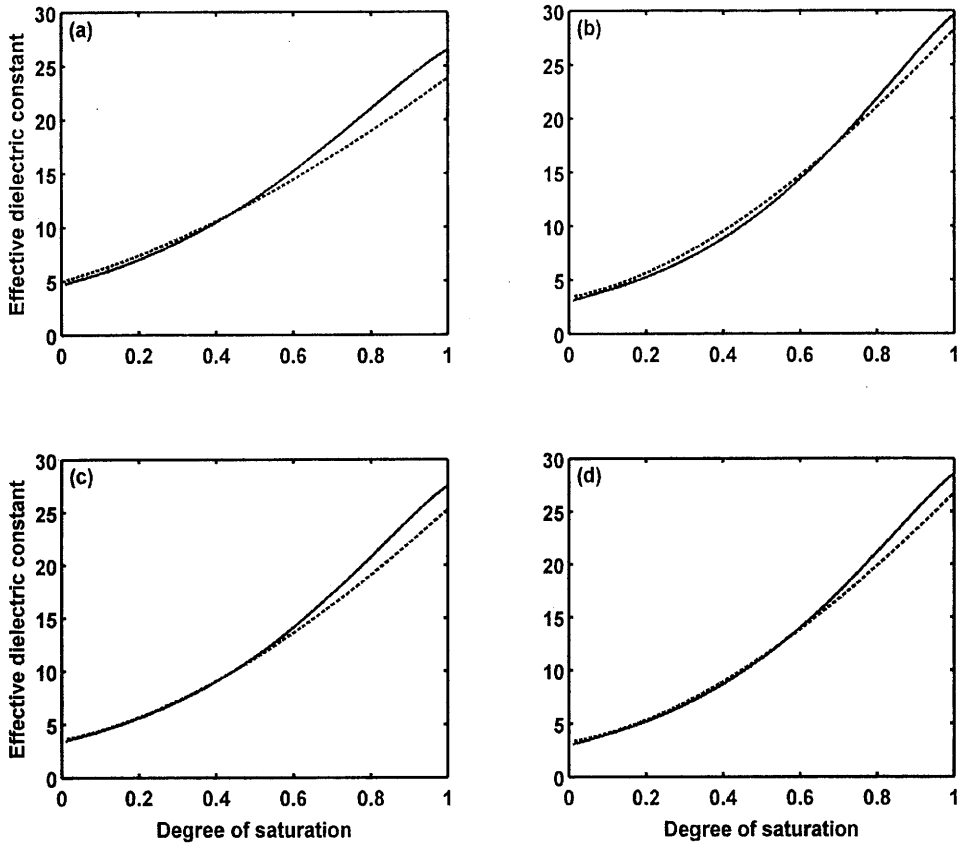


Figure 4-2: Effective dielectric constant versus degree of saturation for the model of *Topp* [1980] (dashed line) and *Friedman* [1998] (solid line) for: (a) clay, (b) loam, (c) loamy sand, and (d) sand soils.

4. The soil-moisture weighing functions for each composite sphere configurations are computed and the bulk dielectric constant of the soil-water-air medium are computed (equation 4.23).

The dielectric behavior of the two models outlined above is compared in Figure 4-2. As depicted, even though the model of Topp *et al.* [1980] contains no explicit representation of soil textural variation, it reasonably well fits the more sophisticated model of Friedman [1998]. However, the Topp *et al.* [1980] model is consistently below the Friedman [1998] for all soil types near saturation.

4.3 Dependence of local incident angle on topography

At a location at which the emission of microwave radiation is modeled, R, the position of an observing satellite relative to the reference location can be expressed as a function of the azimuth (ζ_S) and zenith (δ_S) angles from the reference to the satellite (figure 4-3). By convention, ζ_S is defined as the angle made between the locally-horizontal line connecting R with true North and the locally-horizontal line connecting R with the sub-satellite point (the point on the Earths surface directly beneath the satellite), positive clockwise. δ_S is the angle between a line originating at R in the $-g$ direction and the line connecting the R and the satellite. For a sloping surface, the local incidence angle that the observing satellite makes with the reference location is a function of the local topographic slope (α_∇) and aspect (ζ_∇) (figure 4-3). By convention, ζ_∇ is defined as the angle in the locally-horizontal plane that the local direction of steepest descent makes with true North. α_∇ is the local gradient with respect to the horizontal in the direction of steepest descent. Following the work of Mätzler and Standley [2000] the local incidence angle, θ, can be computed directly knowing ζ_S, ζ_∇, δ_S, and α_∇ through the spherical law of cosines,

$$\cos\theta_L = \cos\alpha_\nabla \cos\delta_S + \sin\alpha_\nabla \sin\delta_S \cos(\zeta_S - \zeta_\nabla) \tag{4.30}$$

Further, as depicted by Mätzler and Standley [2000], surface topography also results in a rotation of the linear polarization by an angle φ that can be calculated as,

$$\sin\varphi = \sin(\zeta_S - \zeta_\nabla)\sin\alpha_\nabla / \sin\theta_L. \tag{4.31}$$

Following the formulation outlined in Mätzler and Standley [2000], the horizontally-polarized reflectivity in the reference frame of the observing sensor, R_h, can be determined as a function of the linear polarization rotation angle, φ,

$$R_h = R_{h,l}\cos^2\varphi + R_{v,l}\sin^2\varphi. \tag{4.32}$$

Similarly, the vertically-polarized reflectivity in the observing sensor reference frame, R_v, can be determined as,

$$R_v = R_{v,l}\cos^2\varphi + R_{h,l}\sin^2\varphi. \tag{4.33}$$

The reflectivities (R_p) in equations (4.32) and (4.33) are a function of the dielectic constant of the soil-air-water matrix, ϵ_{eff} and the local incidence angle, θ_L, but are computed in a slightly different manner for the passive and active observing systems. For the passive observing system $R_{h,l} = r_{sh}$ and $R_{v,l} = r_{sv}$ from equations (4.5) and (4.6), respectively. Similarly, for the active system $R_{h,l} = R_{\|}$ and $R_{v,l} = R_{\perp}$ from equations (4.14) and (4.15), respectively.

At a locally-flat location ($\alpha_\nabla = 0°$), $\theta_L = \delta_S$, and the local incidence angle is independent of the azimuth angle to the satellite, ζ_S. Studies examining large areas typically assume θ_L equals δ_S, which is approximately equal to the off-nadir look angle of the observing satellite (e.g., Crow *et al.* [2005]). Equation (4.30) describes how the local incidence angle can be directly computed with variables that can be obtained from rudimentary terrain analyses of widely available DEMs, such as the Shuttle Radar Topography Mission dataset [Farr *et al.*, 2007].

Hillslope-scale predictions of brightness temperature and radar backscatter can be spatially aggregated to the ground resolution of the observing sensor by weighing each pixel according to its contribution to the radiation received at the sensor. The contribution of a pixel to the bulk observation at the ground resolution of the sensor depends on the solid angle $\Omega = A\cos\theta/r^2$, where A is the surface area of the pixel, and r is the distance from the pixel to the radiometer antenna Mätzler and Standley [2000]. Within a single pixel at the sensor ground resolution, r will not vary substantially and in this work is assumed uniform. Therefore, the predicted observation of microwave brightness temperature or radar bacscatter observed at

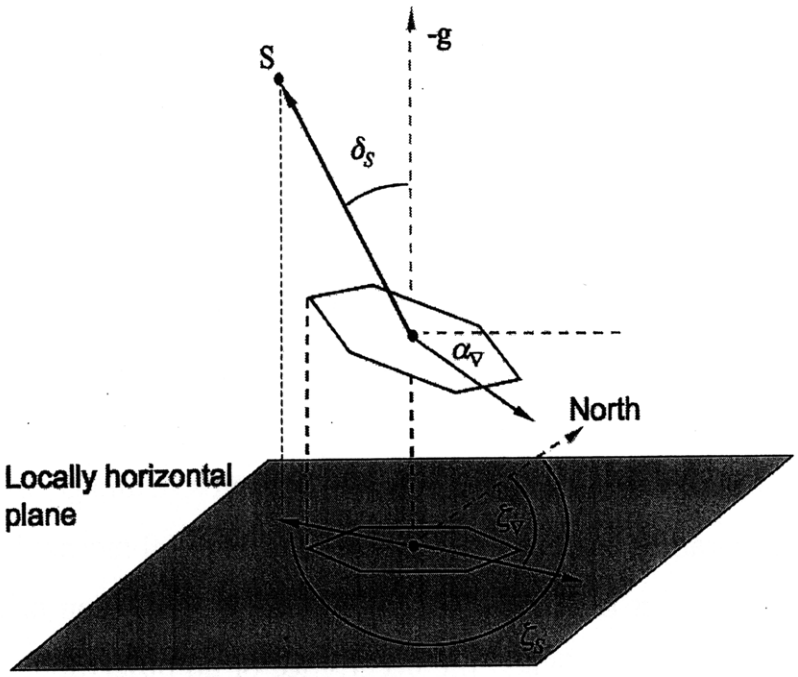


Figure 4-3: Definition of topographic slope, α_∇, aspect, ζ_∇, and relation to satellite sky position, S, as characterized by the azimuth angle to the satellite, ζ_S, and the zenith angle to the satellite, δ_S.

the sensor ($\overline{T}_{bp}$ or $\overline{\sigma}^\circ_{pp}$, respectively) due to emission or reflection from the visible pixels within a single satellite pixel can be computed as,

$$\overline{T}_{bp} = \frac{\sum_{j=1}^{n} T_{bp,j}\Omega_j}{\sum_{j=1}^{n} \Omega_j}, \tag{4.34}$$

in the case of the passive observing system, or

$$\overline{\sigma}^\circ_{pp} = \frac{\sum_{j=1}^{n} \sigma^\circ_{pp,j}\Omega_j}{\sum_{j=1}^{n} \Omega_j}, \tag{4.35}$$

in the case of the active system. In equations (4.34) and (4.35) n is the number of pixels within a single satellite pixel that are visible at given satellite azimuth (ζ_S). It should be noted that the effects of local viewshed on visible sky in modeling

emission of microwave radiation from the surface are not considered. Incorporation of viewshed effects on visible sky is a straightforward extension of the present work: emission of microwave radiation is simply not modeled at computational elements obscured from view of the satellite by surrounding terrain.

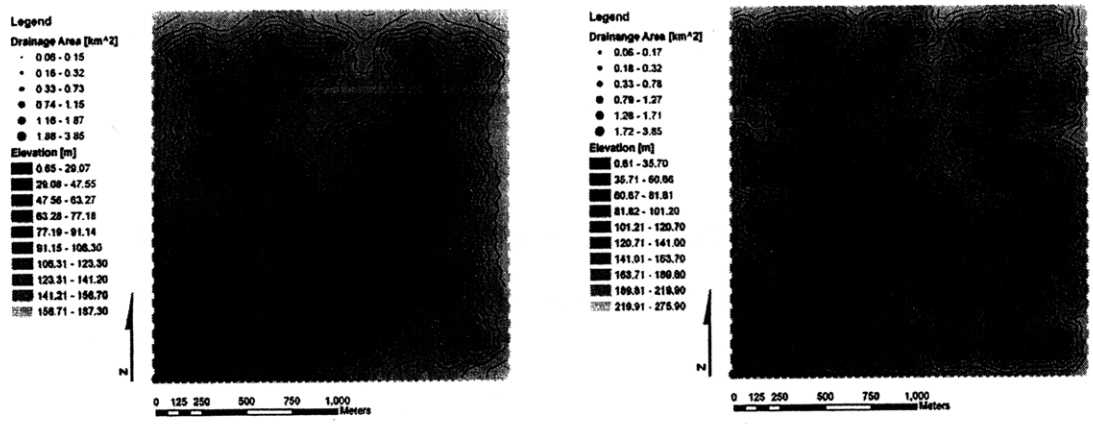
Determination of the local incidence and polarization rotation angles is predicated upon knowledge of δ_S and ζ_S. For low Earth orbiting satellites, observations are discrete but regularly occurring events that depend on characteristics of the satellite such as orbital height, inclination, and footprint size and geometry. Hence, determination of δ_S and ζ_S requires simulation of the orbital mechanics of the observing satellite to detect the temporal occurrence of observation of a particular point on the Earth's surface. When the occurrence of an observation is detected, then δ_S and ζ_S can be computed knowing the relative positions of the point being observed and the location of the satellite. **Appendix A** provides an outline of such a simulation framework that can be used to determine when a location is being observed and subsequently determine δ_S and ζ_S.

4.4 Properties of the passive observing system

4.4.1 Passive observing system: Organization of hillslope-scale brightness temperatures

This section provides an analysis of the properties of the passive observing system. The purpose of this analysis is to diagnose the sensitivity of the predicted observations to variation in factors affecting emission/reflection of microwave energy and the topographic effects on observing geometry. As an experimental location for the coupled hydrology-vegetation model, they used two synthetic landscapes. Both domains were generated with a physically-based landscape evolution model [Tucker and Bras, 1998; Tucker *et al.*, 2001b,a], which evolves local elevation as a function of sediment influx and outflux. Local tectonic uplift and input of sediment eroded

from upstream account for the local influx of sediment, and the export of sediment is derived from local erosion. The local change in elevation is computed as the divergence of sediment influx and export. Equilibrium topography was reached in each of the domains when the volume of sediment input through tectonic uplift to the watershed is balanced by the export of sediment from the watershed. The two simulated domains correspond to two different dominating erosional mechanisms. The diffusive erosion terrain assumes that slope-dependent processes (e.g., soil creep) are the primary mechanism of erosion locally, resulting in shallower slopes, longer hillslopes and lower topographic relief (range in watershed elevation) (figure 4-4(a)). The fluvial erosion terrain assumes that the primary mechanism of local erosion is shear stress above some threshold imparted by surface runoff, resulting in higher slopes, shorter hillslopes, and greater topographic relief (figure 4-4(b)).



(a) (b)

Figure 4-4: Two catchments used as computational domains for the tRIBS-VEGGIE and generated by a physically-based landscape evolution model with (a) slope-dependent diffusive erosion resulting in landscape less dissected by channels with longer hillslopes and shallower slopes (average and standard deviation in slope is 0.231 m/m and 0.103 m/m, respectively), and (b) overland flow-dependent fluvial erosion resulting in more channel dissection of the landscape, shorter hillslopes, and higher slopes (average and standard deviation in slope is 0.500 m/m and 0.192 m/m, respectively). Black circles represent the channel network and are proportional in size to upstream contributing area (km^2). Black lines are elevation contours and are drawn at 10 m intervals in the diffusion domain, and 20 m intervals in the fluvial domain.

Ivanov *et al.* [2008b] use the diffusive erosion and fluvial erosion synthetic terrains as the geometric constructs for several multi-year simulations to develop patterns of soil moisture and vegetation biomass consistent with field-based observations. They confined their investigation to consider only the behavior of C4 grasses and climatology consistent with the Sevilleta National Wildlife Refuge (SNWR). The meteorological forcings consistent with SNWR were generated with the stochastic weather generator of Ivanov *et al.* [2007]: hourly rainfall, temperature, pressure, sky cloud-cover fraction, and relative humidity. From the simulated outputs of Ivanov *et al.* [2008b], the following variables were used as input to the radiative transfer model: soil moisture in the top 25 mm of the soil column (θ), soil temperature (T_S), canopy temperature (T_C), and LAI. Instantaneous values of the land surface state inputs to equations (4.1), (4.3) and (4.18) are obtained from a tRIBS-VEGGIE model simulation corresponding to a hypothetical midday on August 14 with no rain or clouds. Vegetation water content, V_{wc}, was computed from equation (4.3) with the simulated LAI, and the dielectric constant from equation (4.18) with the simulated soil moisture. The range of these variables for the two domains is reported in Table 4-1.

In the semiarid-grassland considered, simulated fractional vegetation cover is approximately 34%, maximum LAI is approximately 0.94, and maximum vegetation height is approximately 0.24 m. Undisturbed natural grasses are primarily oriented vertically, which could lead to significant polarization effects. However, several studies have found that these polarization effects are small in grasslands similar to the ones considered here [Crow *et al.*, 2005; Merlin *et al.*, 2005; Vall-Ilossera *et al.*, 2005], and the polarization effects on τ_p, b_p, and ω_p are therefore neglected. However, it should be noted that Schwank *et al.* [2005] conclude that anisotropic vegetation models are most appropriate in soil moisture retrieval algorithms because of the anisotropic nature of real canopies. Importantly, in a series of field experiments in forested areas Guglielmetti *et al.* [2008] concluded that significant canopy cover can significantly attenuate the moisture signal from the ground surface.

Table 4.1: Range of land surface state inputs to radiative transfer model for diffusive and fluvial erosion domains.

	Diffusive erosion landscape		Fluvial erosion landscape	
Variable	Mean	Range	Mean	Range
T_S [K]	324.68	318.82-327.22	321.66	310.13-327.29
T_C [K]	314.77	311.48-316.32	313.01	306.67-316.43
f_c [m^2/m^2]	0.342	0.311-0.378	0.337	0.265-0.390
θ [m^3/m^3]	0.0291	0.0287-0.0297	0.0291	0.0286-0.0299
LAI [m^2/m^2]	0.799	0.711-0.904	0.783	0.586-0.940
h [-]	0.10			
b [-]	0.10			
ω [-]	0.05			

As a way of developing some intuition about the degree to which topographic effects on incidence angle may impact modeled brightness temperatures, the sensitivity of the modeled microwave brightness temperature to incidence angle is investigated. Using the spatial average values of T_S, T_C, f_c, and LAI shown for the diffusive landscape in Table 4.1 together with assumed values of effective saturation of 0.2, 0.3, 0.5, 0.7 and 0.9[1] the horizontally and vertically polarized brightness temperatures are computed at incidence angles varying from 0° to 90° off nadir (figure 4-5(a)). Along with a histograms of incidence angles in the diffusive and fluvial erosion domains produced by assuming $\zeta_S = 150°$ and $\delta_S = 40°$ and computing θ_L through equation (4.30) (figure 4-5(b)), the dependence of T_{bp} on θ_L gives a first order expectation of the range of hillslope-scale variation in brightness temperatures.

The dependence of brightness temperature on incidence angle reveals that for $65° \leq \theta_L < 90°$, the brightness temperature decreases rapidly in the vertical polarization as θ_L approaches 90° for surface conditions corresponding to mean values in Table 4.1 (see figure 4-5(a)). Since T_{bh} decreases with increasing θ_L for given surface states and θ_L less than approximately 83°, hillslopes with low θ_L would be

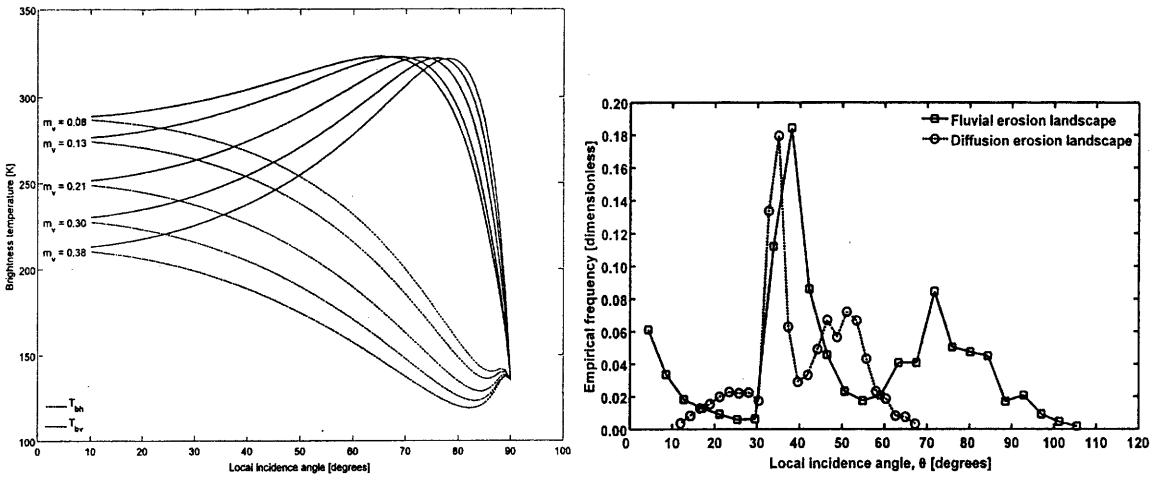
[1]Effective saturation, S_e, is related to the volumetric moisture content θ as $S_e = (\theta - \theta_R)/(\theta_S - \theta_R)$, where θ_S and θ_R are the saturation and residual moisture contents, respectively. For the loam soil considered here and in Ivanov *et al.* [2008b], it is assumed that θ_S equals 0.434 and θ_R equals 0.027

expected to exhibit higher T_{bh} (figure 4-5(a)). At θ_L greater than approximately 85° for given surface state T_{bh} increases slightly as θ_L approaches approximately 88° and then decreases again as θ_L approaches 90°. However T_{bh} decreases with increasing θ for all θ_L. By contrast, for given surface states T_{bv} increases with increasing θ_L to a maximum value of T_{bv} that occurs between values of θ_L between 65° and 80° depending on the surface state, and then decreases as θ_L increases toward a value θ_L equals 90° (figure 4-5(a)). At θ_L greater than approximately 80° T_{bv} increases with moisture at given θ_L, whereas at θ_L less than approximately 65° T_{bv} decreases with increasing moisture at given θ_L.

Moreover, for $65° \leq \theta_L < 90°$, increasing volumetric soil moisture is associated with increasing T_{bv}. In the diffusive erosion domain, few hillslopes exhibit $\theta_L \geq 65°$ (figure 4-5(b)) and T_{bv} increases monotonically with θ_L in this domain for given surface states. In the fluvial erosion domain, by contrast, a significant number of hillslopes exhibit $\theta_L \geq 65°$ (figure 4-5(b)). Hence, for the assumed satellite position the fluvial domain contains areas with θ such that T_{bv} decreases with increasing volumetric soil moisture (i.e., $\theta_L \leq 65°$) and areas with θ_L such that T_{bv} increases with increasing volumetric soil moisture (i.e., $\theta_L \geq 65°$).

In the diffusive erosion domain, few hillslopes exhibit $\theta_L \geq 65°$ (figure 4-5(b)) and the relationship between T_{bv} and θ_L in the diffusive domain is monotonic for given surface states. In the fluvial erosion domain, by contrast, a significant number of hillslopes exhibit $\theta_L \geq 80°$ (figure 4-5(b)). Hence, for the assumed satellite position the fluvial domain contains pixels with θ_L such that T_{bv} decreases with increasing θ (i.e., $\theta_L \leq 65°$) and pixels with θ_L such that T_{bv} increases with increasing θ (i.e., $\theta_L \leq 65°$).

The local incidence angle, θ_L, is computed with equation (4.13) at each computational node in the domains given the local slope (α_∇) and aspect (ζ_∇), and assumed values for satellite azimuth (ζ_S) and satellite zenith (δ_S). In the following spatially distributed examples the satellite zenith angle is assumed to be equal to 40°, which is consistent with conceived soil moisture sensing platforms (e.g., Crow



(a) (b)

Figure 4-5: (a) Brightness temperature as a function of incidence angle in the horizontal (dashed lines) and vertical (full lines) polarizations at moisture levels corresponding to effective soil saturation of 0.2, 0.3, 0.5, 0.7 and 0.9, and (b) empirical frequency distributions of incidence angle (θ_L), assuming $\zeta_S = 150°$ and $\delta_S = 40°$ for fluvial (full lines marked by squares) and diffusive (dashed lines marked by diamonds) erosion domains.

et al. [2005]). Furthermore, a value of $\zeta_S = 150°$ is assumed, which is close to the azimuth that would be encountered during the ascending limb of a soil moisture sensing satellite in a Sun synchronous orbit. The assumed values of ζ_S and δ_S imply that hillslopes with aspects oriented toward the observing sensor (South- to Southeast-facing hillslopes in this case) possess the lowest values of θ_L within the domain. Conversely, hillslopes with aspects oriented away from the sensor (North- to Northwest-facing hillslopes) would possess highest values of θ_L.

For the two synthetic domains, spatially varying inputs to the radiative transfer model represent instantaneous values of local T_S, T_C, V_{wc}, h, and θ simulated by tRIBS-VEGGIE during a rain-free day in mid August. Again, the range of variability of each state variable input to the radiative transfer scheme for both domains is given in Table 4.1. The spatial organization in near-surface soil moisture, soil temperature, canopy temperature, and vegetation height and abundance (as modulated by the effect of LAI on vegetation optical thickness) impacts the spatial distribution of hillslope-scale brightness temperatures in a significant and consistent manner (figure 4-6). In the diffusive erosion landscape, the range of modeled

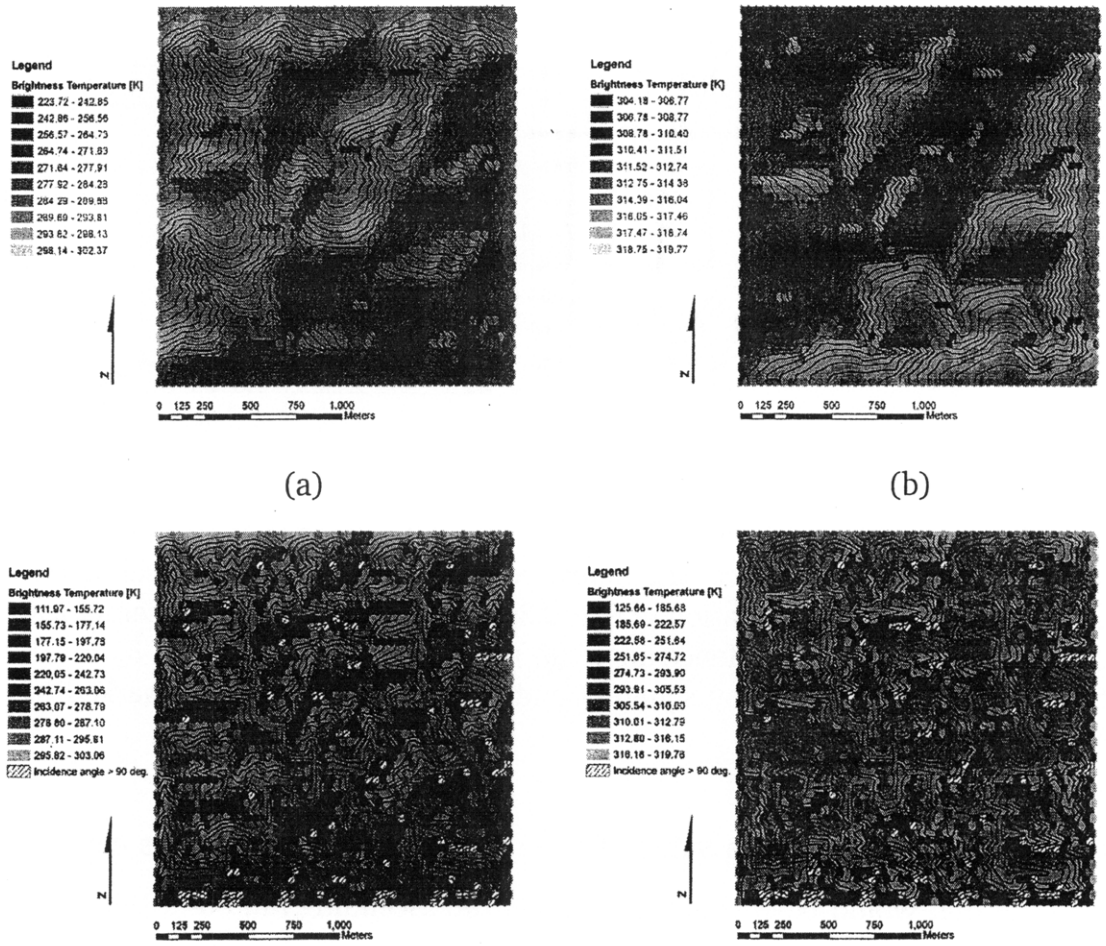


Figure 4-6: Spatial distribution of brightness temperature assuming $\zeta_S = 150°$ and $\delta_S = 40°$ with land surface states at each pixel evolved by the tRIBS+VEGGIE model for (a) diffusion-dominated terrain in the horizontal polarization, (b) diffusion dominated terrain in the vertical polarization, (c) fluvial erosion-dominated terrain in the horizontal polarization, and (d) fluvial erosion-dominated terrain in the vertical polarization.

brightness temperatures varies from approximately 224 to 302 K in the horizontal polarization (figure 4-6(a)) and from approximately 298 to 320 K in the vertical polarization (figure 4-6(b)). Computational pixels with West- and North-facing aspects exhibit lower T_{bh} relative to South- and East-facing pixels. Alternatively, South-and East-facing pixels tend to exhibit lower T_{bv} than pixels facing North and West. Similar patterns to those seen in the diffusive erosion domain can be seen in the fluvial erosion domain, though the scale of spatial variation in modeled bright-

ness temperatures is finer than in the diffusive erosion domain and the range of variability in brightness temperature higher. There exist computational elements within the fluvial erosion domain that cannot be observed because the local value of θ would be greater than 90° given the assumed ζ_S and δ_S, and local values of α_∇ and ζ_∇. In the fluvial erosion landscape, the range of modeled brightness temperatures varies from approximately 121 to 317 K in the horizontal polarization (figure 4-6(c)) and from approximately 117 to 320 K in the vertical polarization (figure 4-6(d)). Similar to the diffusive erosion domain, North- and West-facing hillslopes in the fluvial erosion domain are associated with the lowest values of T_{bh}. Explaining the spatial distribution of T_{bv} in the fluvial erosion domain is more difficult since the distribution of θ within the domain results in areas where T_{bv} increases with θ, T_S, and T_C and areas where T_{bv} decreases with increasing θ, T_S, and T_C. Spatial patterns of hillslope-scale T_{bh} and T_{bv} in both domains are associated with both topography-controlled variation on incidence angles and surface states.

Topographic controls on the distribution of incoming solar radiation lead to tendencies of South- and Southeast-facing hillslopes to exhibit higher soil temperatures, lower moisture and lower vegetation biomass (in the Northern hemisphere). The tendency for lower moisture and higher soil temperatures in these areas would lead to higher values of T_{bh} for given θ. On the other hand, North- and Northwest-facing hillslopes are exposed to less incident solar radiation and therefore tend to exhibit lower daytime soil and canopy temperatures, higher soil moisture, and greater vegetation biomass than South-facing hillslopes. Therefore, topographic gradients in surface states act to enhance North-South contrasts in hillslope-scale T_{bh} compared with topographic effects on incidence angle alone for the assumed ζ_S and δ_S.

To diagnose the relative impacts of spatial variability in surface states input to the RTM for the assumed sensor sky position, T_{bh} and T_{bv} are recomputed at each computation node using the locally-derived value of θ and φ, but with the spatially-averaged values of T_S, T_C, f_c, θ, LAI, and h reported in Table 4.1. The maps of

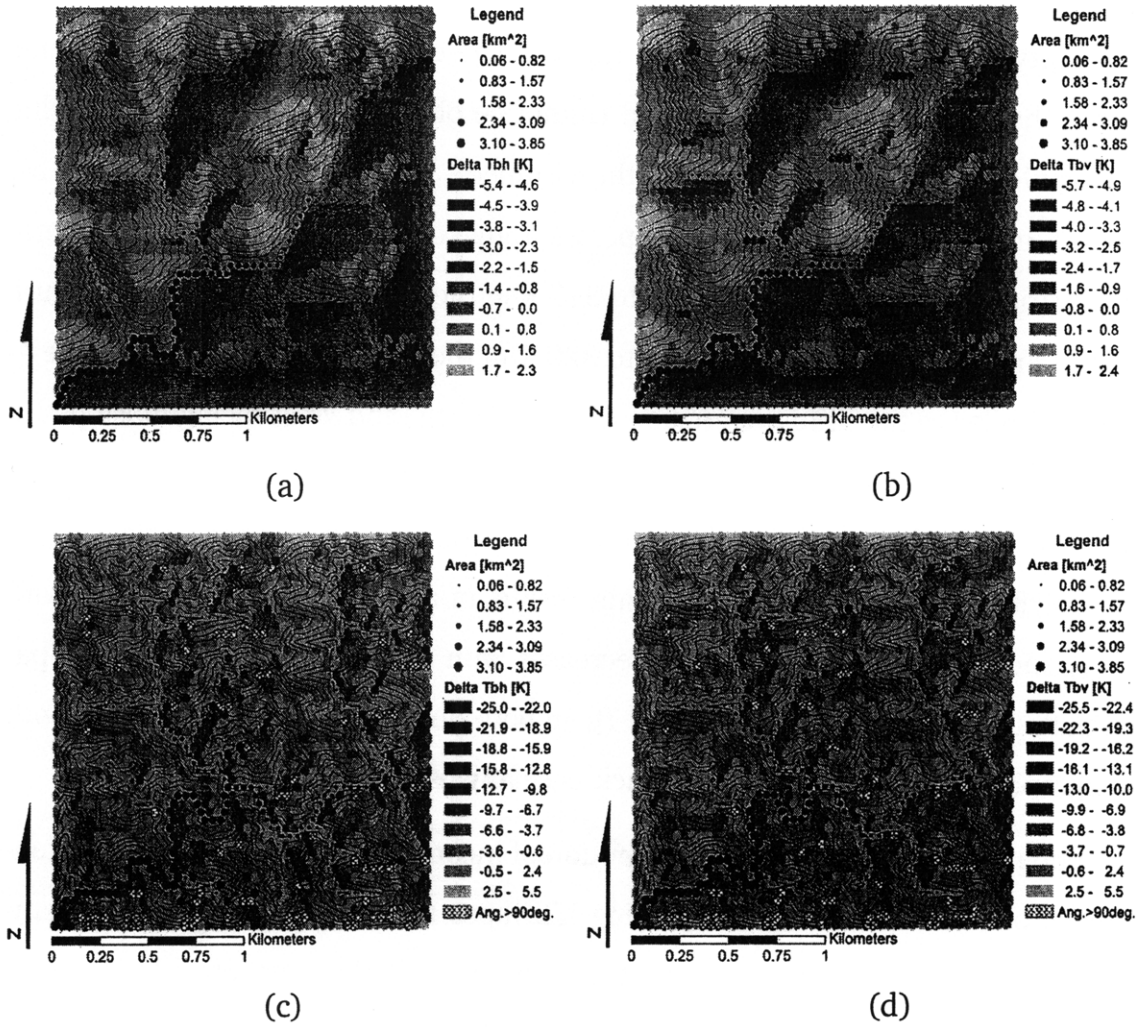


Figure 4-7: Spatial distribution of difference between simulated brightness temperatures using spatially-varied and spatially-averaged surface states assuming ζ_S = 150° and δ_S = 40° for (a) diffusive erosion terrain in the horizontal polarization, (b) diffusive erosion terrain in the vertical polarization, (c) fluvial erosion terrain in the horizontal polarization, and (d) fluvial erosion terrain in the vertical polarization.

T_{bh} and T_{bv} computed with spatially-averaged RTM inputs are denoted T_{bh}^{sa} and T_{bv}^{sa}, respectively. Denoting the brightness temperatures in both polarizations computed from spatially-distributed RTM inputs (i.e., those presented in figure 4-6) as T_{bh}^{sv} and T_{bh}^{sv}, respectively, figure 4-7 shows the impact of hillslope-scale heterogeneity in surface states on hillslope-scale microwave radiation emission (i.e., $(T_{bh}^{sv} - T_{bh}^{sa})$ and $(T_{bv}^{sv} - T_{bv}^{sa})$). In the diffusive erosion domain, the impact of spatial heterogeneity in the surface states results in values of $(T_{bh}^{sv} - T_{bh}^{sa})$ ranging from approximately -5.4

to +2.3 K (figure 4-7(a)), while in the fluvial erosion domain spatial heterogeneity leads to values of $(T_{bh}^{sv} - T_{bh}^{sa})$ ranging from −25.0 to +5.5 K (figure 4-7(c)). West- and North-facing pixels exhibit the lowest values of $(T_{bh}^{sv} - T_{bh}^{sa})$ in both domains, while in the fluvial erosion domain West-facing pixels demonstrate the lowest values of $(T_{bh}^{sv} - T_{bh}^{sa})$. Conversely, South- and Southeast-facing pixels exhibit the highest values of $(T_{bh}^{sv} - T_{bh}^{sa})$ in both the diffusive and fluvial erosion domains. Values of $(T_{bv}^{sv} - T_{bv}^{sa})$ range from approximately −5.7 to +2.4 K in the diffusive erosion domain (figure 4-7(b)), and approximately −25.5 to +5.5 K in the fluvial erosion domain (figure 4-7(d)). Pixels with aspects ranging West to North demonstrate the lowest values of $(T_{bv}^{sv} - T_{bv}^{sa})$ in the respective domains. In the fluvial erosion domain a few North-facing pixels (primarily clustered around pixels with $\theta \geq 90°$) exhibit positive values of $(T_{bv}^{sv} - T_{bv}^{sa})$, suggesting such pixels possess high θ associated with T_{bv} that increases with higher θ, and lower T_S, and T_C that predominate in these areas. It is important to note that for this particular set of experiments, the spatial patterns in $(T_{bh}^{sv} - T_{bh}^{sa})$ and $(T_{bv}^{sv} - T_{bv}^{sa})$ in figure 4-7 likely arise due mostly to spatial variation in soil and canopy temperature rather than soil moisture and optical thickness. For the simulated conditions, near-surface soil moisture is relatively low and LAI ranges from approximately 0.6 to 0.9.

4.4.2 Passive observing system: Sensitivity of aggregate brightness temperature to satellite sky position

In the previous section significant hillslope-scale variation in microwave brightness temperature was attributed to the effects of topography on both the spatial organization of factors affecting emission and the angles describing the observational geometry. Here the sensitivity of spatially-aggregated brightness temperature to satellite azimuth angle (ζ_S) is assessed in order to diagnose the potential effects of topography at the sensor.

The spatial distribution of horizontally- and vertically-polarized microwave bright-

ness temperatures was computed through the RTM using the previously described instantaneous surface states evolved by tRIBS-VEGGIE for both computational domains. In this portion of the sensitivity analysis, it is assumed that the satellite viewing angle (δ_S) is 40°, while the azimuth angle to the satellite (ζ_S) is allowed to vary from 1° to 360°, in ζ_S increments of 1°. Local incidence angle and polarization rotation are computed through equations (4.30) and (4.31), respectively, at every computational node in both the diffusive and fluvial erosion domains. For each domain, this yields 360 spatial maps of horizontally- and vertically-polarized brightness temperature derived from spatially distributed land surface states and denoted T_{bh}^{sv} and T_{bv}^{sv}, respectively. Each of these hillslope-scale brightness temperature maps are spatially aggregated, weighing each pixel according to its contribution to the radiation received at the sensor, through equation (4.34).

For comparative purposes, two cases are considered in which brightness temperatures at the sensor are modeled assuming θ, T_S, T_C, and LAI are spatially uniform and equal to the mean value illustrated in Table I for each domains. These two cases are meant to capture potentially important hypothetical microwave emission modeling scenarios within a consistent land surface modeling environment: (1) a coarse-scale land surface model augmented with high-resolution (e.g., 30 m) digital elevation data to encompass topographic effects on observational geometry, and (2) a coarse-scale land surface model neglecting topographic effects on observational geometry. In the first case, topographic effects on θ and φ are included in modeling the dependence of horizontally- and vertically-polarized brightness temperature at the sensor on ζ_S. These brightness temperatures modeled at the sensor are denoted $\overline{T}_{bh}^{sa}$ and $\overline{T}_{bv}^{sa}$, respectively, and are computed by substituting T_{bp}^{sa} for T_{bp}^{sv} in the summation on the right hand side of equation (4.34). In the second case, topographic effects are neglected $\theta = \delta_S = 40°$ is assumed. The modeled brightness temperatures at the sensor, which does not vary with satellite azimuth (ζ_S), are denoted $\overline{T}_{bh,\text{flat}}$ and $\overline{T}_{bv,\text{flat}}$, in this case.

For the diffusive erosion domain (e.g., rolling hills with relatively shallow slopes),

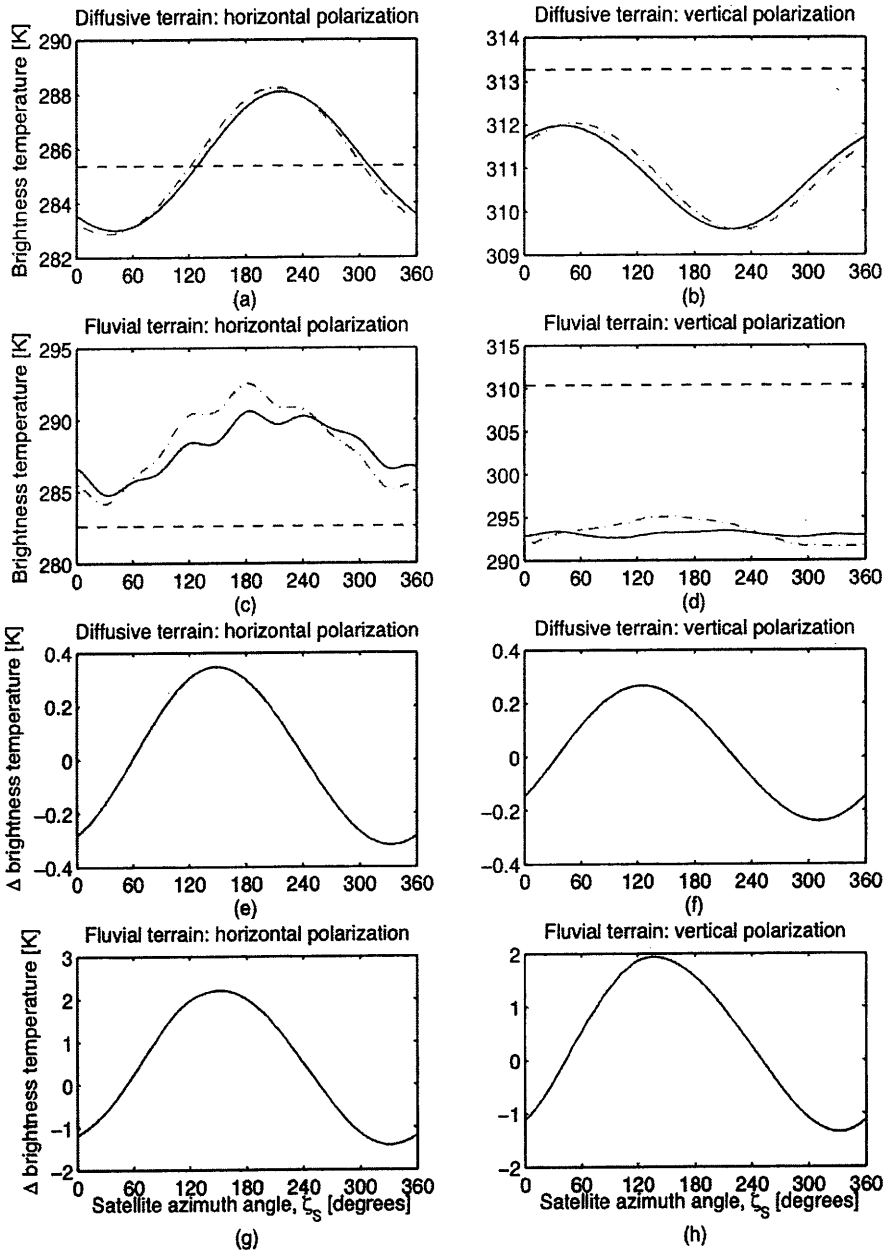


Figure 4-8: Spatial mean values of brightness temperatures as a function of satellite azimuth angle, for (a) diffusion-dominated terrain in the horizontal polarization, (b) diffusion dominated terrain in the vertical polarization, (c) fluvial erosion-dominated terrain in the horizontal polarization, and (d) fluvial erosion-dominated terrain in the vertical polarization. $(\overline{T}_{bp}^{sv} - \overline{T}_{bp}^{sa})$ as a function of satellite azimuth angle (ζ_S), for (e) diffusion-dominated terrain in the horizontal polarization, (f) diffusion dominated terrain in the vertical polarization, (g) fluvial erosion-dominated terrain in the horizontal polarization, and (h) fluvial erosion-dominated terrain in the vertical polarization.

$\overline{T}_{bh}^{sv}$, $\overline{T}_{bh}^{sa}$, $\overline{T}_{bv}^{sv}$ and $\overline{T}_{bv}^{sa}$ vary in a sinusoidal fashion with azimuth angle to the satellite (ζ_S) (figures 4-8(a),(b)). This sinusoidal variation brightness temperature modeled at the sensor with ζ_S arises because the assumed value of δ_S (40°) is greater than the maximum slope in the domain. For any value of ζ_S this leads to a distribution of θ within the diffusive erosion domain that leads to the inclusion of every computational node in aggregation of the pixel-scale brightness temperatures. This contrasts with the results from the fluvial erosion landscape (e.g., rugged hills with relatively steep slopes), which exhibits a more variable relationship between aggregated brightness temperatures and ζ_S (figures 4-8(c),(d)). Because the value of δ_S equals 40° is less than the maximum slope in the fluvial erosion domain, groups of pixels depending on ζ_S will exhibit θ greater than or equal to the upper 90° limit. At any value of ζ_S, those pixels with $\theta \geq 90°$ are not included in aggregation of pixel-scale brightness temperatures because they cannot be observed by the sensor.

Another important contrast between the diffusive and fluvial erosion terrains lies in the amplitude of variability in aggregated brightness temperatures with ζ_S. In the horizontal polarization, $\overline{T}_{bh}^{sv}$ varies from a low of about 283 K at ζ_S of approximately 40° to a maximum of approximately 288 K at approximately ζ_S equal to 220°, a range of 5 K in the diffusive erosion landscape (figure 4-8(a)). In the fluvial erosion domain $\overline{T}_{bh}^{sv}$ ranges from a low of around 284 K at ζ_S equal to approximately 40°, and a maximum of approximately 292 K at ζ_S equal to 180°, an amplitude of about 8 K for conditions of spatially varying surface states (figure 4-8(c)). In comparison, the corresponding amplitude of variation $\overline{T}_{bh}^{sa}$ is approximately 5 K (figure 4-8(c)). The impact of spatial heterogeneity in surface states, as illustrated in the more rugged domain by the difference between $\overline{T}_{bh}^{sv}$ and $\overline{T}_{bh}^{sa}$ (figure 4-8(c)) is close to the sensitivity of the radiometer for many values of ζ_S and greater than the sensitivity of many operational and planned microwave radiometers at a few particular values of ζ_S [Kerr *et al.*, 2001; Entekhabi *et al.*, 2004].

The $\overline{T}_{bv}^{sv}$ in the diffusive erosion landscape exhibits a maximum of about 312 K at ζ_S equal to approximately 40° and a minimum slightly more than 309.5 K at ζ_S

equal to about 220°, corresponding to a range in aggregated $\overline{T}_{bv}^{sv}$ of approximately 2.5 K (figure 4-8(b)). By comparison, $\overline{T}_{bv}^{sv}$ in the fluvial erosion domain exhibits a range of approximately 3 K (slightly greater than the radiometer sensitivity), with an approximate minimum of 292 K at ζ_S near 330° and an approximate maximum of 295 K at near 150° (figure 4-8(d)).

Comparing the modeled brightness temperatures at the sensor that include topographic effects on incidence angle and polarization rotation ($\overline{T}_{bh}^{sv}$ and $\overline{T}_{bh}^{sa}$) to the modeled brightness temperatures that neglect topographic effects on incidence angle and polarization rotation, $\overline{T}_{bh,\text{flat}}$ is approximately 285.5 K in the rolling topographic domain (figure 4-8(a)) and 282.5 K in the rugged topographic domain (figure 4-8(c)). In the diffusive domain the value of $\overline{T}_{bh,\text{flat}}$ is approximately the midpoint of variation in $\overline{T}_{bh}^{sv}$ and $\overline{T}_{bh}^{sa}$ with ζ_S (figure 4-8(a)), while $\overline{T}_{bh,\text{flat}}$ is at least 1.5 K less than the minimum values of $\overline{T}_{bh}^{sv}$ and $\overline{T}_{bh}^{sa}$ in the fluvial domain (figure 4-8(c)). In the vertical polarization, $\overline{T}_{bv,\text{flat}}$ is more than 1 K higher than the maximum values of $\overline{T}_{bv}^{sv}$ and $\overline{T}_{bv}^{sa}$ in the diffusive domain (figure 4-8(b)) and at least 15 K higher than the maximum value of $\overline{T}_{bv}^{sv}$ in the fluvial domain (figure 4-8(d)).

The difference in modeled brightness temperatures observed at the sensor ($\overline{T}_{bh}^{sv} - \overline{T}_{bh}^{sa}$ and $\overline{T}_{bv}^{sv} - \overline{T}_{bv}^{sa}$) as a function of ζ_S is shown explicitly in figures 4-8(e)-(h). These plots highlight illustrates the impact of hillslope-scale organization in factors affecting microwave emission at the observing sensor. The fluvial domain (figures 4-8(g) and 4-8(h)) is more sensitive to hillslope-scale soil moisture, soil temperature, and vegetation variation than is the diffusive domain (figures 4-8(e) and 4-8(f)). At ζ_S near 330°, $\overline{T}_{bh}^{sv}$ is approximately 0.3 K lower than $\overline{T}_{bh}^{sa}$ in the diffusive domain (figure 4-8(e)), and $\overline{T}_{bh}^{sv}$ is approximately 1.5 K lower than $\overline{T}_{bh}^{sa}$ in the fluvial erosion domain (figure 4-8(g)). Conversely, when the satellite is situated to the South of the landscape (ζ_S near 150°), $\overline{T}_{bh}^{sv}$ is warmer than $\overline{T}_{bh}^{sa}$: approximately 0.3 K and 2.0 K in the diffusive (figure 4-8(e)) and fluvial (figure 4-8(g)) domains, respectively. The ζ_S corresponding to the minimum and maximum values of $(\overline{T}_{bh}^{sv} - \overline{T}_{bh}^{sa})$ are accounted for by the hillslope-scale organization and correlation of soil moisture, soil

temperature, and vegetation biomass. By contrast, $\overline{T}_{bv}^{sv}$ differs from $\overline{T}_{bv}^{sa}$ by at most about 0.25 K in the diffusive erosion domain (figure 4-8(f)) and $(\overline{T}_{bv}^{sv} - \overline{T}_{bv}^{sa})$ is at a maximum when at ζ_S near 130° and a minimum at ζ_S near 300°. Meanwhile, $\overline{T}_{bv}^{sv}$ is up to 2 K greater that and nearly 1.5 K lower than $\overline{T}_{bv}^{sa}$ in the more topographically rugged domain (figure 4-8(h)). The non-symmetry in figure 4-8(h) about 0 K is due to the fact that T_{bv} exhibits non-monotonic dependence on θ when θ is greater than approximately 65°.

Figure 4-8 depicts two important findings of this study: (1) the difference between $\overline{T}_{bp}^{sv}$ and $\overline{T}_{bp}^{sa}$ demonstrates the role of covariation in the land surface factors affecting microwave emission at hillslope scales in the modeled brightness temperature at the sensor, and (2) the difference between $\overline{T}_{bp,\text{flat}}$ and $\overline{T}_{bp}^{sa}$ illustrates the impact of hillslope-scale topography on modeled brightness temperature at the sensor vis-á-vis its influence on observational geometry.

4.4.3 Effects of hillslope-scale heterogeneity at the scale of planned radiometers

The influence of hillslope-scale topography on local incidence angle over real terrain is also studied. The case considered is a 43 x 59 km area within North-central New Mexico, which contains the Sevilleta National Wildlife Refuge and Long-term Ecological Research station. The DEM was obtained from the Sevilleta Spatial Database (see `http://sevilleta.unm.edu/data/archive/gis/`), derived from interferometric synthetic aperture radar at a resolution of 10 m square. Elevation within the area ranges from 1403 to 2802 m msl.

Local incidence angle (θ_L) is computed through equation (4.30) at every DEM cell within imposed 10, 25, and 40 km square windows meant to represent the size of a single radiometer pixel based on the topographic slope (α_∇) and aspect (ζ_∇) at each DEM cell and assuming $\zeta_S = 150°$ and $\delta_S = 40°$. This is a simplifying assumption, which could be relaxed by computing ζ_S and δ_S at each DEM pixel

knowing the location of the sub-satellite point on the Earths surface and the orbital altitude of the sensor. Empirical histograms of local incidence angle within each of the three area-aggregations are then computed.

The empirical histograms shown in figure 4-9 illustrate that for each area-aggregation, both slope and aspect control the distribution of incidence angle. The influence of aspect on the histograms is reflected on frequency peaks in figure 4-9. Undoubtedly, the fact that there are four peaks in the histogram of θ_L arises because aspect angle (cardinal direction of maximum gradient) can take only eight values on a rectangular grid. However, the presence of a coherent topographic structure in any study domain (such as the North-South oriented ridges present in the study areas) will lead to distinct ranges of incidence angles that are encountered more frequently, irrespective of the terrain model (i.e., regular grid spacing vs. irregular mesh). The frequency at which particular ranges of incidence angles are encountered within the area-aggregation does not change substantially with increasing spatial aggregation (figure 4-9).

4.5 Properties of the active observing system

This section provides a sensitivity analysis of the properties of the active observing system. The sensitivity analysis begins with an investigation to determine the sensitivity of L-band radar backscatter to both moisture content and a categorical soil type in the two copolarized states (σ°_{hh} and σ°_{vv}). Sensitivity of microwave backscatter to incidence angle and soil moisture for a particular soil type is then studied. In addition to diagnosing the sensitivity of the predicted observations to factors affecting emission/reflection of microwave energy and the topographic effects on observing geometry, the sensitivity of the active system to surface roughness has been well-studied [Dobson and Ulaby, 1986; Wang *et al.*, 1986; Evans *et al.*, 1992; Shi *et al.*, 1997]. Therefore, a set of experiments to understand the sensitivity of predictions of microwave backscatter to assumptions about surface

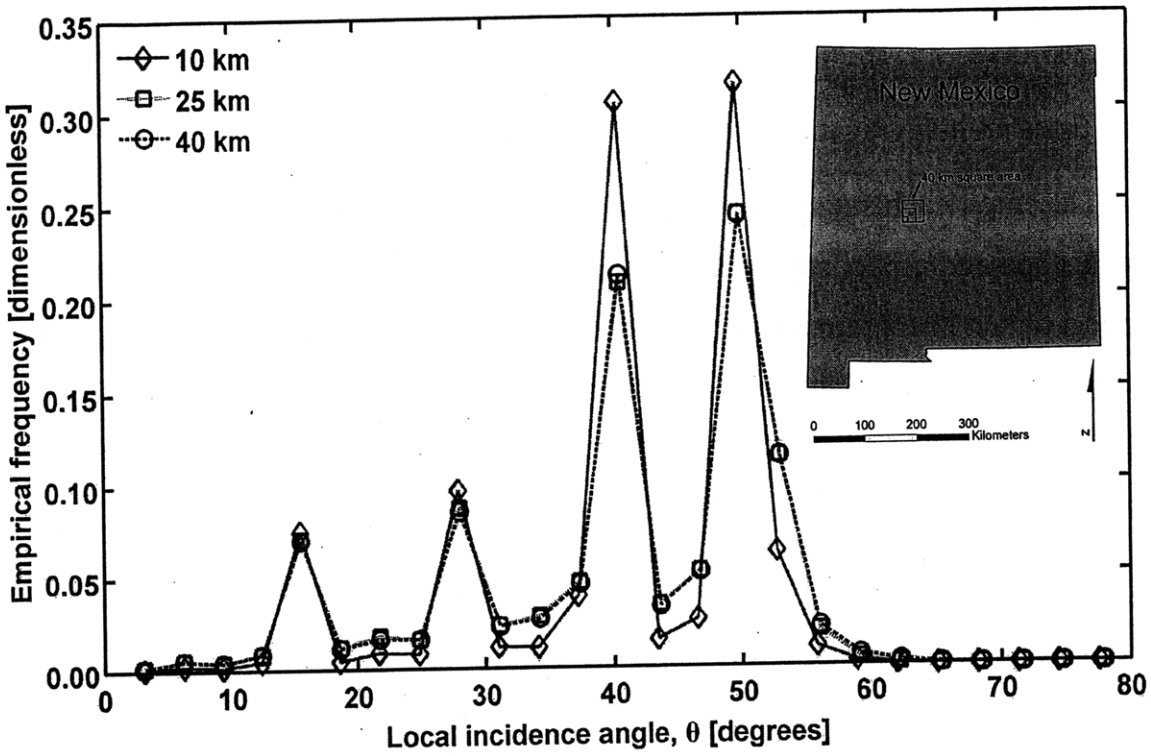


Figure 4-9: Empirical frequency distributions of local incidence angles computed from a 10 m digital elevation model for square domains of size 10 (full line marked with diamonds), 25 (dashed line marked with squares), and 40 (dotted line marked with circles) km, assuming a satellite azimuth angle of 150° and a zenith angle of 40°.

roughness is performed. Through experiments that use tRIBS-VEGGIE to simulate a spatially varying soil moisture field, the role of topography in the organization of hillslope-scale predictions of microwave backscatter is analyzed. The sensitivity of spatially-averaged microwave backscatter predictions to satellite sky position is investigated to determine the potential impact of hillslope-scale topographic variation at the scale of observation. Finally, the concept of temporal radar backscatter change, which has been found to be related to the corresponding temporal moisture change [Njoku *et al.*, 2002; Narayan *et al.*, 2006; Narayan and Lakshmi, 2008], is briefly investigated to discern the degree to which topography may affect inference of moisture change from temporal backscatter change.

4.5.1 Active observing system: moisture and incidence angle sensitivity

To investigate the sensitivity of predicted L-band microwave backscatter to moisture and soil properties, the IEM was used to simulate backscatter as a function of a broad range of soil moisture values for four generic soil types: (1) clay, (2) loam, (3) loamy sand, and (4) sand. The dielectric constant model used was that of Friedman [1998], and the properties of the categorical soil types used in this experiment are shown in Table 4-2. Each soil type is associated with particular values of saturation moisture content (θ_S), residual moisture content (θ_R), bulk density (ρ_b), and specific surface area (S_{SA}). Values of θ_S and θ_R were taken from a sensitivity analysis in Ivanov *et al.* [2008a,b]. The value of ρ_b was approximated by assuming all solid materials had a specific gravity equal to quartz 2.65 and then calculating $\rho_b = 2.65(1-\theta_S)$ for each categorical soil type. S_{SA} was estimated using the pedotransfer function of Banin and Amiel [1969], which approximates S_{SA} as a function of the clay fraction of the soil as $S_{SA} = 5.78 \times \%\text{clay} - 15.064$. Clay fractions assumed for each soil type were 0.65 (clay), 0.20 (loam), 0.08 (sandy loam) and 0.05 (sand). For each of the four soil types σ°_{pp} is plotted against degree of saturation ($S_e = (\theta - \theta_R)/(\theta_S - \theta_R)$) in figure 4-9.

In both copolarized states, the amount of backscattered energy increases monotonically with increasing moisture. As seen if figure 4-10, at values of S_e above approximately 0.6, the predicted radar backscatter does not differ substantially between soil types. By contrast, at low effective moisture contents (e.g., $S_e < 0.3$) there generally exists substantial variation predicted backscatter (in both copolarized states) that is associated with differences in soil type. This between soil type

Table 4.2: Soil hydraulic parameters for each categorical soil type

Soil type	θ_S [m^3/m^3]	θ_R [m^3/m^3]	ρ_b [km/m^3]	S_{SA} [m^2/kg]
Clay	0.385	0.090	1630	330.64
Loam	0.434	0.027	1500	100.54
Loamy sand	0.401	1587	1200	28.29
Sand	0.417	0.020	1545	13.84

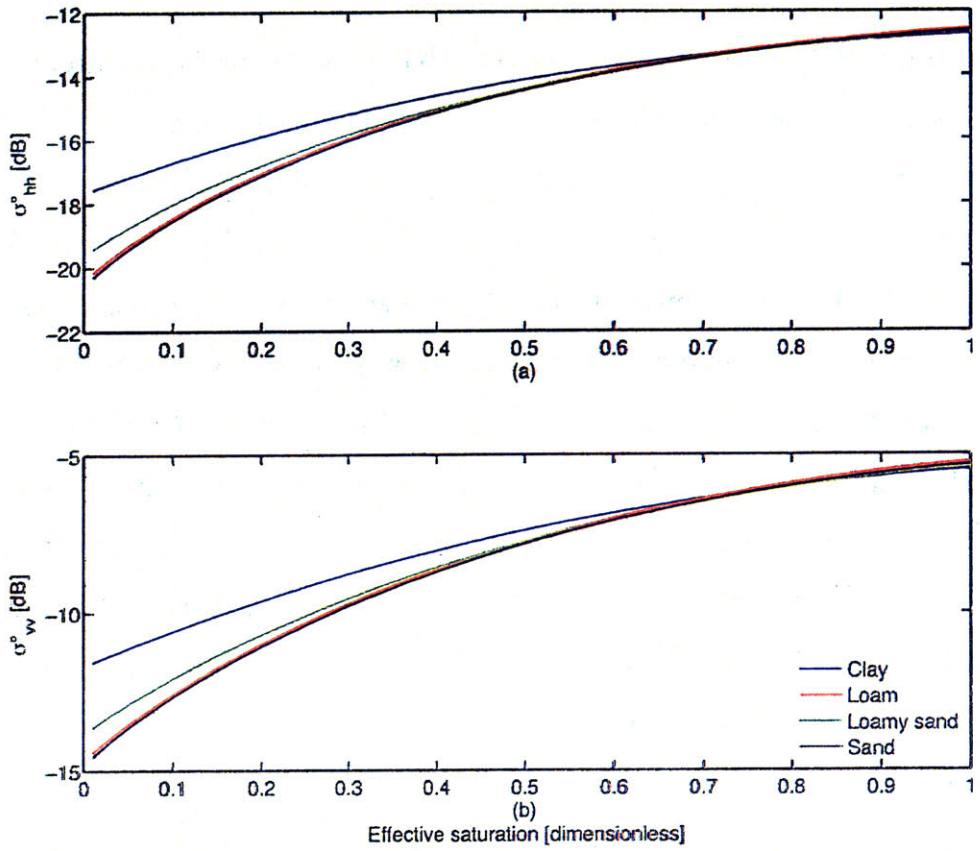


Figure 4-10: Backscatter coefficient σ_0 versus degree of saturation in the hh polarization (top panel) and vv polarization (bottom panel) for four soil textural classes.

variation in microwave backscatter suggested by the IEM at low moisture contents is suggestive of a potential application of microwave backscatter data in sparsely vegetated and dry areas (e.g., semiarid grasslands and deserts). Differences in radar backscatter in similarly-vegetated and dry locations should, in theory be due to differences in residual moisture content and soil roughness. This conclusion is not, however, universally true as sand and loam soils exhibit very similar values of σ°_{pp} across the entire range of effective saturation. It is important to note that in this particular analysis the soil roughness has been treated identically between soil types, with the rms surface height assumed to be $\sigma = 2.5$ cm and the correlation length assumed to be $L = 10$ cm. This is, however, not likely to be the case and it is plausible that differences in characteristic roughness scales between soil types could enhance differences in dry backscatter between soil type. Using microwave

radar backscatter data as a way of potentially classifying soils in dry areas, therefore, depends on the relative impacts of residual moisture contents and roughness on microwave backscatter, and the degree to which between soil type variation in these soil parameters is distinct from within soil type variation.

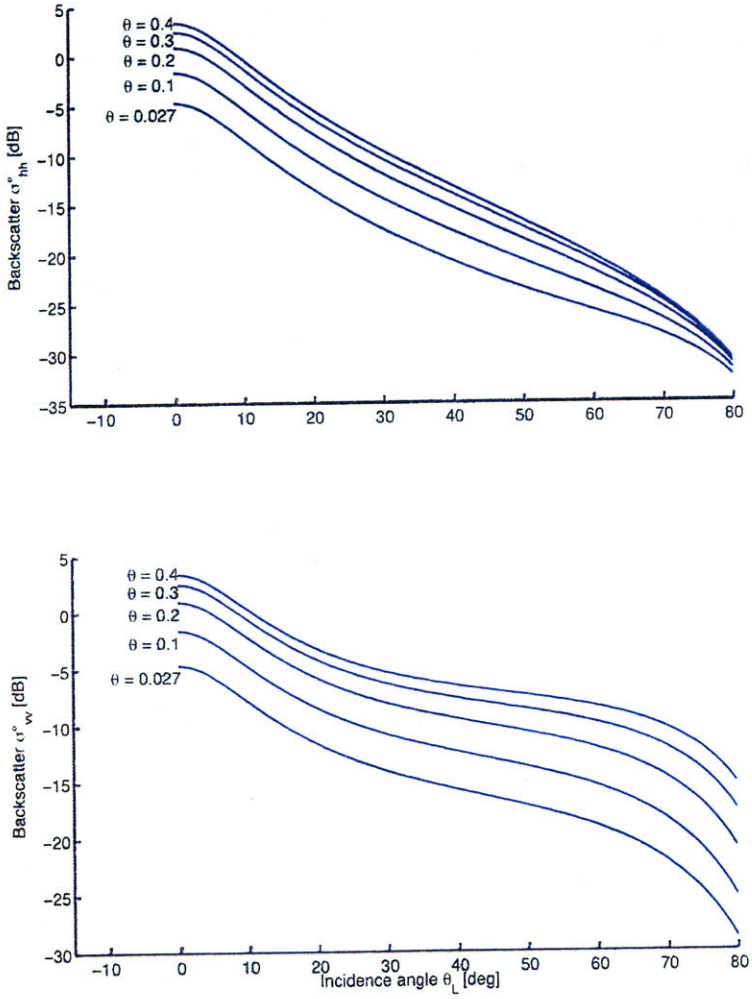


Figure 4-11: Backscatter coefficient σ_0 versus local incidence angle θ_L in the: (a) hh and (b) vv polarizations for an indicated range of volumetric moisture contents θ.

Potential impacts of topography vis-á-vis its effect on local incidence angle are highlighted by plotting predicted microwave backscatter in the two copolarized states as a function of local incidence angle θ_L and for a range of volumetric mois-

ture contents, θ (figure 4-11). The assumed soil type if figure 4-11 corresponds to the loamy soil, whose parameters are displayed in Table 4.2. As illustrated in figure 4-11, and consistent with the idea that backscatter is a measure of energy received at a sensor, σ°_{pp} decreases monotonically as incidence angle increases. Beyond $\theta_L \approx 80°$, the backscattered energy decreases rapidly to values that are likely below the detection limit of a radar receiving are therefore not shown. Although the decrease in backscatter is monotonic as θ_L increases, in the vertically copolarized state the relationship between σ°_{vv} and θ_L is more nonlinear than the corresponding relationship between σ°_{hh} and θ_L. In the horizontally copolarized state, however, the degree of nonlinearity in the relationship between σ°_{hh} and θ_L is a function of the moisture content. As the soil becomes dryer, the σ°_{hh} versus θ_L becomes more nonlinear. Also of interest is how the dynamic range of σ°_{pp}, that is the difference between σ°_{pp} in the wettest and driest states, varies as a function of θ_L. As seen in figure 4-11, the behavior of the dynamic range of σ°_{pp} versus θ_L is different in the horizontally and vertically copolarized states. In the vertically copolarized state, the dynamic range of σ°_{vv} increases as the incidence angle increases (figure 4-11b). Conversely, in the horizontally copolarization the dynamic range of σ°_{hh} decreases as θ_L increases (figure 4-11a). This is a potentially important characteristic of the active observing system as it pertains to soil moisture estimation and observation in topographically variable regions. It implies that in steeper portions of the landscape that are associated with higher θ_L, horizontally copolarized observations are less sensitive to variation in moisture relative to the vertically copolarized observations.

4.5.2 Active observing system: parameter sensitivity

As outlined above, the IEM requires specification of a roughness power spectrum and associated correlation function. This portion of the chapter is dedicated to an investigation of the sensitivity of predicted microwave backscatter observations to surface roughness. Note that it is assumed in the literature that the roughness is a property of the soil surface itself, although not necessarily a time-invariant property

[Choudhury *et al.*, 1979; Evans *et al.*, 1992; Shi *et al.*, 1997; Dobson and Ulaby, 1986; Wang *et al.*, 1986]. Assuming a loam soil with characteristics presented in Table 4.2, the parameters of the roughness power spectrum (σ and L) are varied from $1.0 \leq \sigma \leq 10.0$ cm and $5 \leq L \leq 50$ cm. In an effort to diagnose the degree to which the roughness effect on backscatter depends on soil moisture, a range of volumetric moisture contents is assumed, from dry conditions ($\theta = 0.027$ m^3/m^3) to wet conditions ($\theta = 0.434$ m^3/m^3).

For a given soil moisture, the horizontally copolarized backscatter is more sensitive to σ than L (figure 4-12). For a given value of L, variation in σ can cause variation in σ°_{hh} of up to approximately 30 dB. For given values of σ, however, variation in L is associated with variation in σ°_{hh} of approximately 10 dB at most. As figures 4-12(a)-(f) also indicate, the sensitivity of σ°_{hh} to moisture content increases as soil moisture increases. For dry conditions, the sensitivity of σ°_{hh} to roughness is approximately 20 dB at, while for wet conditions the sensitivity is closer to 30 dB. For a given value of σ, increasing the moisture content tends to decrease the sensitivity of σ°_{hh} to L.

By comparison, the vertically copolarized backscatter is relatively more sensitive to σ and L than is the horizontally copolarized backscatter (figure 4-13). For a given value of L, variation in σ can cause variation in σ°_{vv} of up to at least 30 dB. Variation in L results in a variation in σ°_{vv} of approximately 10 dB at most, for given values of σ. Generally speaking, similar to the case for σ°_{hh}, increasing the volumetric moisture content tends to increase the sensitivity of σ°_{vv} to roughness (figures 4-13(a)-(f)). Interestingly, however, as moisture increases the predicted backscatter tends to become less sensitive to roughness, provided that σ is below approximately 6 cm. In a similar fashion to σ°_{hh}, increasing the moisture content tends to decrease the sensitivity of σ°_{vv} to L at given σ.

This sensitivity analysis indicates that the microwave backscatter predicted by the IEM is fairly sensitive to surface roughness. At a given moisture content, the range of predicted values of σ°_{pp} as a function of soil roughness parameters can be of

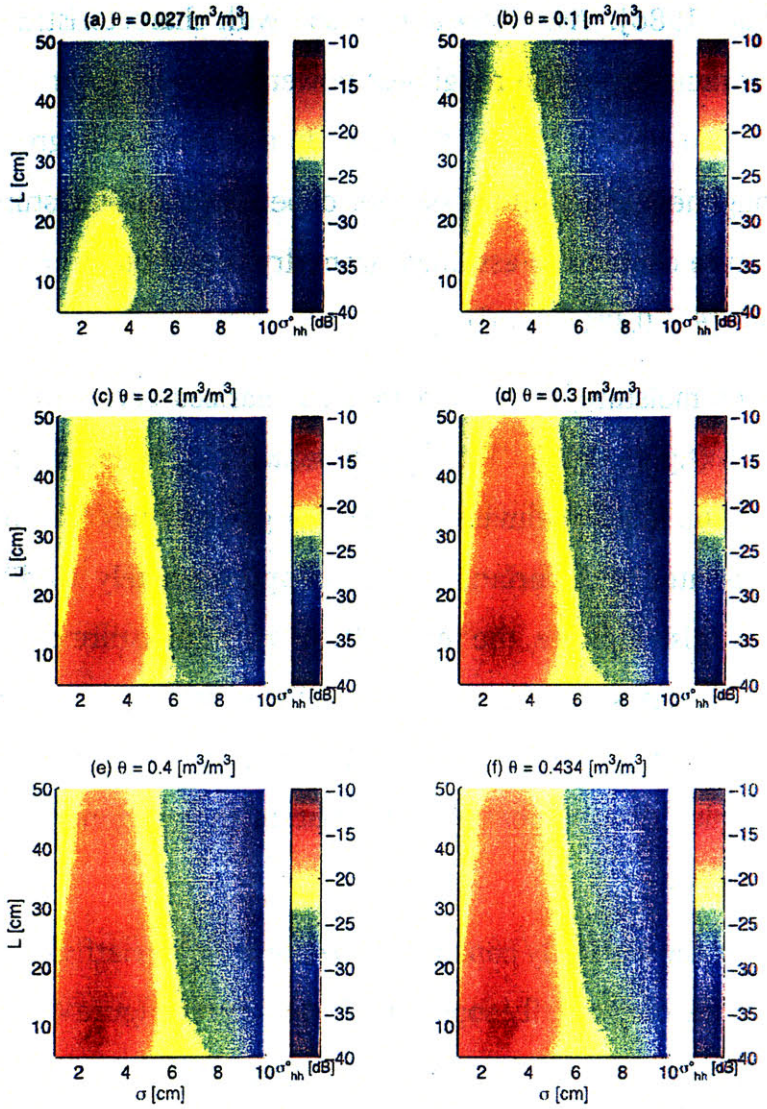


Figure 4-12: Horizontally copolarized microwave backscatter (σ°_{hh}) [dB] as a function of the parameters of the assumed form of the surface roughness power spectrum: the surface rms height (σ) on the x-axis and the linear correlation scale L on the y-axis. Each plot corresponds to a different value of volumetric soil moisture: (a) 0.027, (b) 0.100, (c) 0.200, (d) 0.300, (e) 0.400, and (f) 0.434 m^3/m^3.

similar magnitude to the range of dynamic range of predicted backscatter associated with variation in soil moisture for given roughness conditions. In other words, to diagnose the relative contributions of soil moisture and surface roughness to microwave backscatter, either roughness or soil moisture must be accurately known.

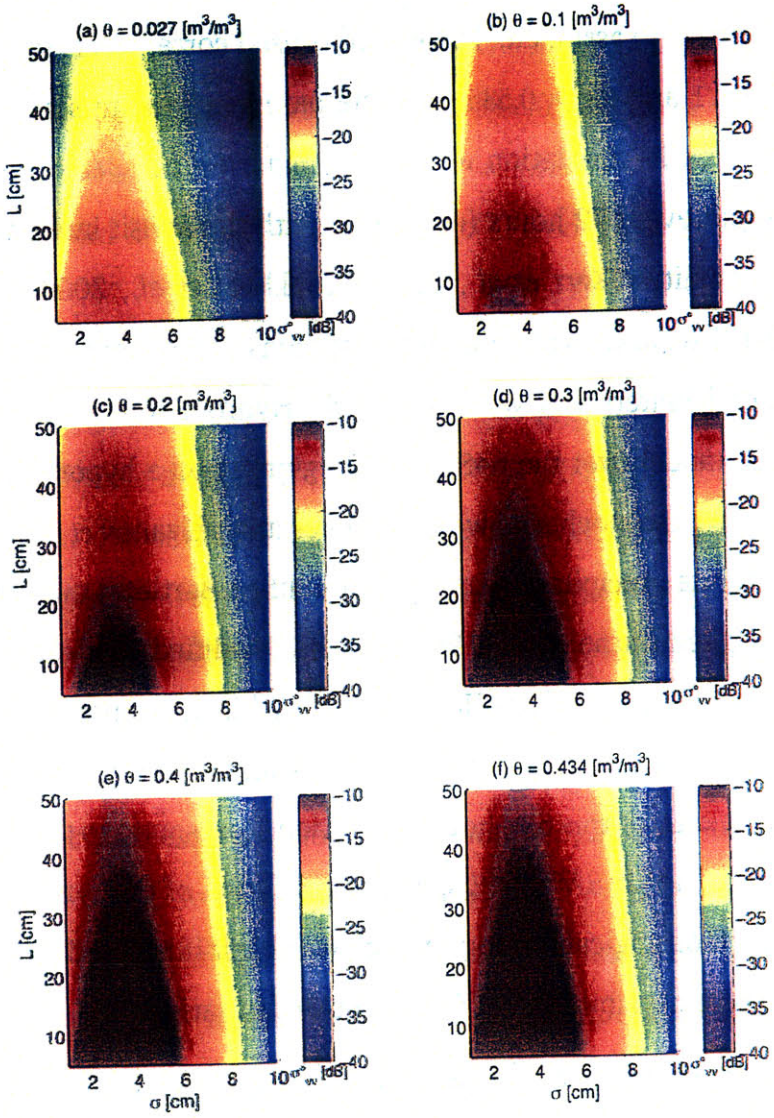


Figure 4-13: Vertically copolarized microwave backscatter (σ°_{vv}) [dB] as a function of the parameters of the assumed form of the surface roughness power spectrum: the surface rms height (σ) on the x-axis and the linear correlation scale L on the y-axis. Each plot corresponds to a different value of volumetric soil moisture: (a) 0.027, (b) 0.100, (c) 0.200, (d) 0.300, (e) 0.400, and (f) 0.434 m^3/m^3.

4.5.3 Active observing system: organization of hillslope-scale radar backscatter

The diffusive erosion domain was used again to examine the spatial organization of microwave backscatter modeled at the hillslope scale. Because the IEM is

best suited to simulation of microwave backscatter in sparsely vegetated conditions, bare soil conditions were assumed. The soil type used corresponds to a sandy loam soil with $\theta_R = 0.121$ and $\theta_S = 0.535$. For the bare sandy loam soil considered, a 648 hour tRIBS-VEGGIE simulation was performed and the spatial distribution of soil moisture output every 72 hours (consistent with the revisit interval of planned remote sensing satellites [Kerr *et al.*, 2001; Entekhabi *et al.*, 2004]) at 0900 local time. Semiarid to arid hydroclimatic conditions (consistent with the Walnut Gulch Experimental Watershed) were assumed. Hydrometeorological forcings to the tRIBS-VEGGIE model over the 648 hours, beginning on a hypothetical July 1st, were generated using the stochastic weather generator of Ivanov *et al.* [2007], with alterations to generate precipitation using the modified Bartlett-Lewis precipitation model (e.g., Rodriguez-Iturbe *et al.* [1987, 1988]). Modified Bartlett-Lewis parameters for Tucson were obtained from Hawk [1992], while the remaining parameters for the stochastic weather generator were taken from Ivanov *et al.* [2007]. Initial soil moisture conditions for this 648 hour simulation correspond to the soil moisture state from the end of a tRIBS-VEGGIE spin-up simulation of the same 648 hour sequence of hydrometeorological forcings, which was itself initialized with a spatially uniform soil moisture distribution of 10% effective saturation.

As mentioned in **Chapter 2**, the presence of vegetation is a potentially complicating factor in retrieving soil moisture from microwave backscatter data (e.g., Dubois *et al.* [1995]). The change detection approach, however, argues that temporal change between successive microwave backscatter observations is due predominantly to changes in soil moisture [Njoku *et al.*, 2002; Narayan *et al.*, 2006; Narayan and Lakshmi, 2008]. The current sensitivity analysis, therefore, seeks to examine both: (1) the influence of the spatial distribution of soil moisture on hillslope-scale microwave backscatter as predicted by the IEM in bare soil conditions during a single observation, and (2) the degree to which temporal change in backscatter between two successive observations reflects temporal change in soil moisture during the intervening time. Consistent with these dual objectives, the spatial distributions of soil moisture at 216 and 288 hours into the simulation were used as input the

microwave backscatter model. These time instances were chosen because they represent the soil moisture state at two successive potential observation times and span a significant rainfall event. The snapshot of the spatial distribution of soil moisture at 216 hours is toward the end of an approximately 180 hour drydown period, and the soil is dry. During the 72 hour period between the two soil moisture snapshots a sequence of rainfall events deliver approximately 12 mm of precipitation in a 10 hour period. Therefore the soil moisture conditions at 288 hours are comparatively wet with respect to the conditions at 216 hours.

The hillslope-scale distribution of microwave backscatter is computed in both the horizontally and vertically copolarized states for the dry (hour 216) and wet (hour 288) soil moisture distributions. Similar to the analysis of modeled hillslope-scale brightness temperatures above, the assumed value of satellite azimuth angle (ζ_S) is 150° and the assumed satellite zenith angle (δ_S) is 40°. Given the distribution of incidence angles in the diffusive domain (figure 4-5(b)) and the generic behavior of microwave backscatter as a function of local incidence angle (θ_L), a relatively wide range of microwave backscatter can be expected at hillslope scales. In contrast to the relationship between spatial patterns in brightness temperature in different polarizations, because σ°_{pp} decays with increasing θ_L for both copolarized states the hillslope-scale organization of backscatter should be consistent between polarizations. Specifically, North- and Northwest-facing hillslopes, which possess higher θ_L at the assumed ζ_S and δ_S, would exhibit lower values of σ°_{hh} and σ°_{vv} than South- and Southeast-facing hillslopes with the same soil moisture. On the other hand, because the microwave brightness temperature depends on both soil moisture (which impacts emissivity) and soil temperature, while microwave backscatter exhibits relatively little sensitivity to land surface temperature[2], the impact of topographic covariation in factors affecting reflection of incident microwaves may be more important in the passive microwave observing system.

For the dry conditions (hour 216) σ°_{hh} varies from -61.7 to -14.2 dB (figure 4-

[2]The impact of temperature on microwave backscatter is the dependence of the static dielectric constant of water on temperature in equation 4.29.

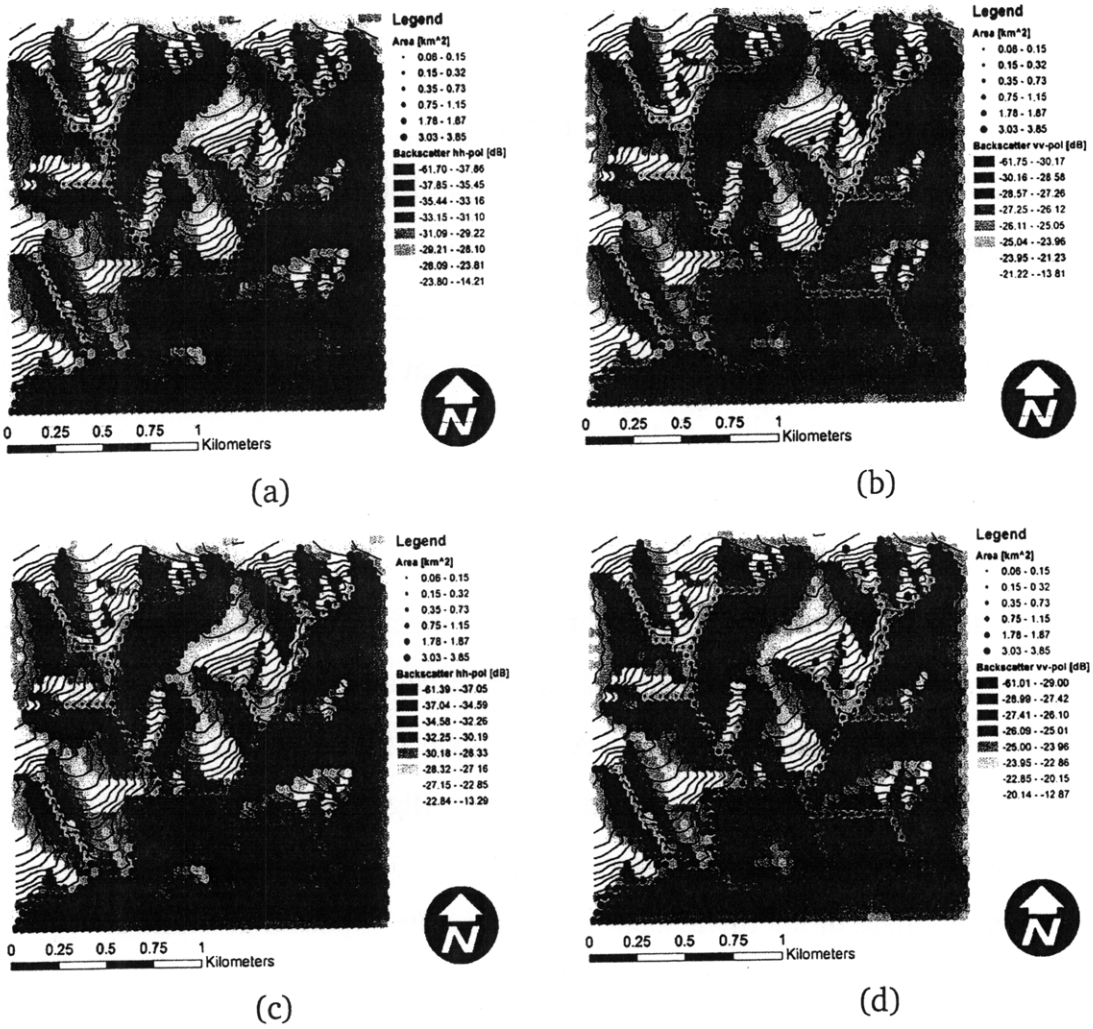


Figure 4-14: Spatial distribution of backscatter coefficient assuming $\zeta_S = 150°$ and $\delta_S = 40°$ with land surface states at each pixel evolved by the tRIBS-VEGGIE model for (a) dry conditions in the horizontally copolarized state, (b) dry conditions in the verically copolarized state, (c) wet conditions in the horizontally copolarized state, and (d) wet conditions in the verically copolarized state. Note: intervals in the legend are expressed as quantiles to better visualize spatial structure in backscatter.

14(a)) while σ°_{vv} varies from -61.7 to -13.8 dB (figure 4-14(b)). By contrast, for the wet conditions (hour 288) σ°_{hh} varies from -61.4 to -13.3 dB (figure 4-14(c)) while σ°_{vv} varies from -61.0 to -12.9 dB (figure 4-14(d)). As expected, for both copolarized states and hydrologic conditions considered, the hillslope-scale distribution of microwave backscatter largely follows the expected trends discussed above. The lowest simulated values backscatter coefficients are found on North- to West-facing slopes (figure 4-14). It should be noted, however, that several relatively steep East-

facing slopes in also exhibit relatively low backscatter, indicating that despite an azimuth angle which is ostensibly favorable to viewing under the assumed values of ζ_S and δ_S, that those portions of the terrain are sufficiently steep to exhibit relatively large values of θ_L. It should also be noted that some of the modeled backscatter coefficients are possibly below the detection limit of the radar sensor, and therefore do not contribute meaningfully to the observation at the satellite. Although it can be argued that these pixels should be filtered from the spatial aggregation as presented in equation (4.35), it is also true that the solid angle Ω (weighting term in equation (4.35)) corresponding to these pixels is small because these pixels are associated with high θ_L. Hence, even when pixels with low σ°_{pp} caused by high θ_L are retained in the aggregation, their contribution is likely insignificant for the diffusive domain. For other domains, however, in which the hillslope length approach the ground resolution of the sensor it is plausible that even the spatially aggregated predicted microwave backscatter would be below the detection threshold of the sensor.

The hillslope-scale temporal backscatter change varies from approximately -0.6 to 1.8 dB in the horizontally copolarized state (figure 4-15(a)) and approximately -0.2 to 2.3 dB in the vertically copolarized state over the three day interval (figure 4-15(b)). The precision of the 1.26 GHz SMAP radar is reported to be 0.85 dB meaning that some of the backscatter change values shown in figures 4-15(a) and 4-15(b) are within the noise sensitivity of the radar sensor. The spatial pattern of temporal backscatter change in the diffusive erosion landscape is reversed in the horizontally and vertically copolarized states. In the horizontal copolarized state, lower values of $\Delta\sigma^\circ_{hh}$ are seen in the North- to West-facing hillslopes, while the higher values of $\Delta\sigma^\circ_{hh}$ are observed in the South- and East-facing hillslopes (figure 4-15(a)). By comparison, lower values of $\Delta\sigma^\circ_{vv}$ are seen in the South- and East-facing hillslopes, as well as in the valley bottoms, while the higher values of $\Delta\sigma^\circ_{vv}$ are largely confined to North- to West-facing slopes (figure 4-15(b)). Defining the total backscatter temporal changes as $\Delta\sigma_{\text{total}} = \Delta\sigma^\circ_{hh} + \Delta\sigma^\circ_{vv}$, the complementary nature of the spatial variability in $\Delta\sigma^\circ_{hh}$ seen in figure 4-15(a) and $\Delta\sigma^\circ_{vv}$ seen in figure 4-15(b) can be seen more clearly in figure 4-15(c). The lowest values of $\Delta\sigma_{\text{total}}$ are

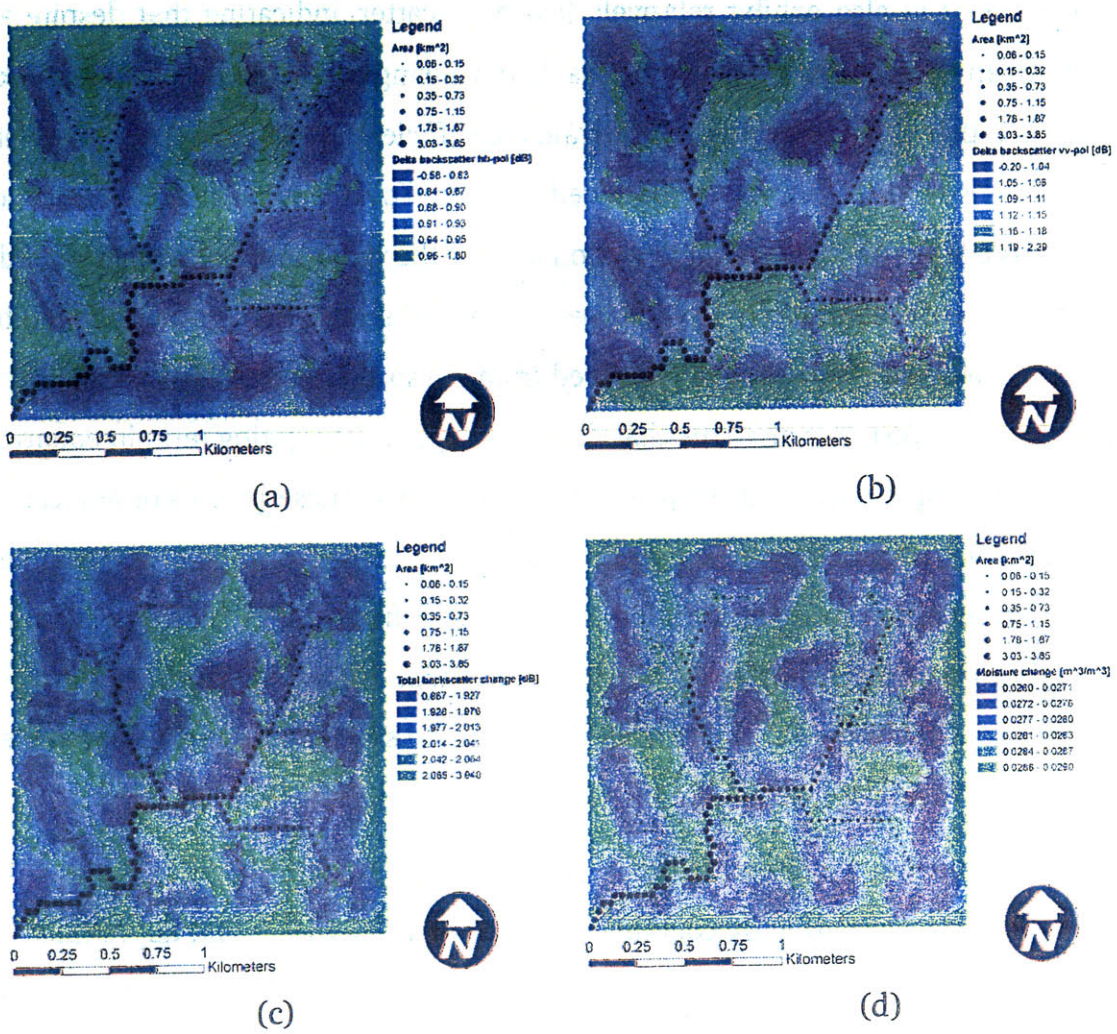


Figure 4-15: Spatial distribution of 3 day change in backscatter coefficient, encompassing a significant precipitation event, assuming $\zeta_S = 150°$ and $\delta_S = 40°$ with land surface states at each pixel evolved by the tRIBS-VEGGIE model in the (a) horizontally- and (b) vertically-copolarized states, and (c) the cumulative backscatter change. For comparison, the (d) spatial distribution of moisture change is shown. Note: intervals of backscatter and moisture change are expressed as quantiles.

largely seen in topographically convergent portions of the landscape: hollows and valley bottoms (figure 4-15(c)). The highest values of $\Delta\sigma_{\text{total}}$, on the other hand, are observed in the topographically divergent portions of the terrain: the ridges. When the hillslope-scale distribution of $\Delta\sigma_{\text{total}}$ is compared to the hillslope-scale variation in the total moisture change between hours 216 and 288, it is seen that $\Delta\sigma_{\text{total}}$ largely follows the same pattern as the moisture change (figure 4-15(d)). The areas of smallest temporal soil moisture change correspond mostly to the hol-

lows in the landscape, particularly those at the heads of the valley network while the areas of greatest temporal moisture change are the topographic ridges (figure 4-15(d)). However, it is important to notem that the values of moisture change in the near surface during the 3 day interval are also relatively small, with moisture change varying spatially from 0.026 to 0.029 m^3/m^3. An interesting contrast between $\Delta\sigma_{\text{total}}$ and the temporal moisture change is in the valley network, where $\Delta\sigma_{\text{total}}$ is consistently lower than in the surrounding hillslopes (figure 4-15(c)), while the corresponding moisture change in the valley network is typically low but more variable throughout the network (figure 4-15(d)). These results indicate that, indeed, the temporal change in radar backscatter (as expressed via $\Delta\sigma_{\text{total}}$) seems to be driven mostly by chances in moisture. Moreover, the hillslope-scale distribution of temporal backscatter change as captured by $\Delta\sigma_{\text{total}}$ largely captures the corresponding temporal change in moisture. This provides some evidence, through the relatively complex IEM, that the temporal change in backscatter between two successive observations is closely related to the change in moisture, as has been reported in the literature [Njoku *et al.*, 2002; Narayan *et al.*, 2006; Narayan and Lakshmi, 2008].

4.5.4 Active observing system: sensitivity of aggregate backscatter to satellite sky position

In a similar fashion to the passive observing system, the sensitivity of the backscatter predicted at the sensor (i.e., spatially aggregated through equation 4.35) to the satellite sky position are investigated. The spatial distribution radar backscatter in both the horizontally- and vertically-copolarized states was computed through the IEM using the previously described instantaneous near-surface soil moisture conditions simulated by tRIBS-VEGGIE for the diffusion erosion domains at hours 216 and 288 of the simulation. Again, consistent with the HYDROS/SMAP sensor characteristics the off-nadir satellite viewing angle (δ_S) is assumed to be 40°, while the azimuth angle to the satellite (ζ_S) is allowed to vary from 1° to 360°, in

ζ_S increments of 1°. Local incidence angle and polarization rotation are computed through equations (4.30) and (4.31), respectively, at every computational node in both the diffusive and fluvial erosion domains. The 360 spatial maps of σ°_{hh} and σ°_{vv} this process yields are aggregated to the 2 × 2 km domain. The spatially-aggregated horizontally- and vertically-copolarized backscatter coefficients are denoted $\overline{\sigma^\circ_{hh}}$ and $\overline{\sigma^\circ_{vv}}$, respectively.

For comparative purposes, two cases are considered in which brightness temperatures at the sensor are modeled assuming θ is spatially uniform and equal to the mean value at hours 216 and 288 of the simulation. These two cases are meant to capture potentially important hypothetical microwave emission modeling scenarios within a consistent land surface modeling environment: (1) a coarse-scale land surface model augmented with high-resolution (e.g., 30 m) digital elevation data to encompass topographic effects on observational geometry, and (2) a coarse-scale land surface model neglecting topographic effects on observational geometry. In the first case, topographic effects on θ and φ are included in modeling the dependence of horizontally- and vertically-polarized brightness temperature at the sensor on ζ_S. These brightness temperatures modeled at the sensor are denoted $\overline{\sigma}^{\circ sa}_{hh}$ and $\overline{\sigma}^{\circ sa}_{vv}$, respectively, and are computed by substituting the hillslope-scale backscatter coefficients estimated with the spatially uniform value of θ for σ°_{pp} in the summation on the right hand side of equation (4.35). In the second case, topographic effects are neglected $\theta_L = \delta_S = 40°$ is assumed. The modeled brightness temperatures at the sensor, which does not vary with satellite azimuth (ζ_S), are denoted $\overline{\sigma}^\circ_{hh,\text{flat}}$ and $\overline{\sigma}^\circ_{hh,\text{flat}}$, in this case.

For the dry conditions (hour 216 of the simulation), $\overline{\sigma}^\circ_{hh}$ varies from a minimum of approximately -30.9 dB at a ζ_S of approximately 40° to a maximum of approximately -29.3 dB at ζ_S ranging from approximately 150° to 280° (figure 4-16(a)). By contrast, $\overline{\sigma}^\circ_{vv}$ achieves a minimum of approximately -25.8 dB at a ζ_S of around 40° and a maximum of approximately -29.3 dB at ζ_S ranging from approximately 150° to 280° (figure 4-16(b)). Under the wet conditions associated with hour 288

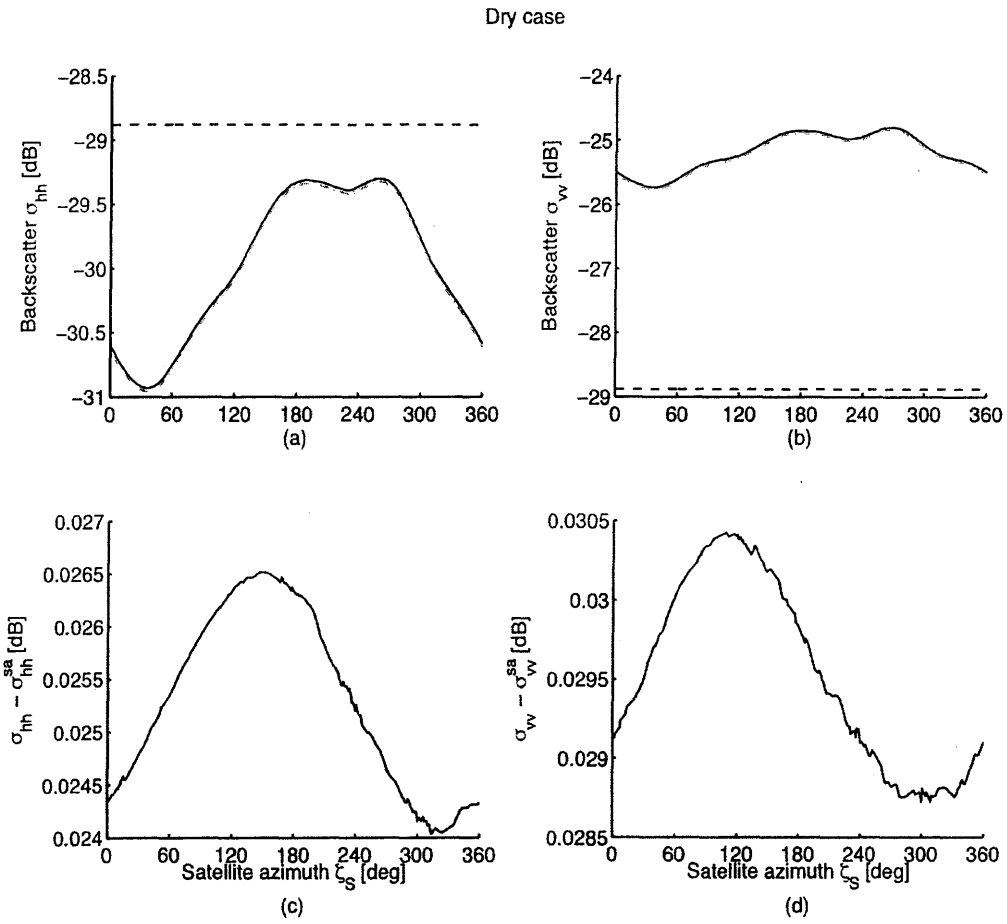


Figure 4-16: For the dry conditions of hour 216 of the simulation, the predicted backscatter at the observing sensor as a function of ζ_S in the (a) horizontally and (b) vertically copolarized domains. Plots show $\overline{\sigma}^{\circ}_{pp}$ (black solid line), $\overline{\sigma}^{\circ sa}_{pp}$ (gray dash-dotted line), and $\overline{\sigma}^{\circ}_{pp,\text{flat}}$ (black dashed line). The difference between $\overline{\sigma}^{\circ}_{pp}$ and $\overline{\sigma}^{\circ sa}_{pp}$ is shown for the (c) horizontally- and (d) vertically-copolarized states.

of the simulation, $\overline{\sigma}^{\circ}_{hh}$ varies from a minimum of approximately -30.0 dB at a ζ_S of approximately 40° to a maximum of approximately -28.5 dB at ζ_S ranging from approximately 150° to 280° (figure 4-17(a)). The corresponding values of $\overline{\sigma}^{\circ}_{vv}$ under the wet conditions are a minimum of approximately -24.6 dB at a satellite azimuth of about 40° and a maximum of approximately -23.7 dB at satellite azimuth angles ranging from approximately 150° to 280° (figure 4-17(b)).

The difference between $\overline{\sigma}^{\circ}_{hh}$ and $\overline{\sigma}^{\circ sa}_{hh}$ is maximized (in an absolute sense) at ζ_S of approximately 150° for the dry conditions of hour 216 of the simulation, although the range of variation in $(\overline{\sigma}^{\circ}_{hh} - \overline{\sigma}^{\circ sa}_{hh})$ is small at all ζ_S (figure 4-16(c)). The

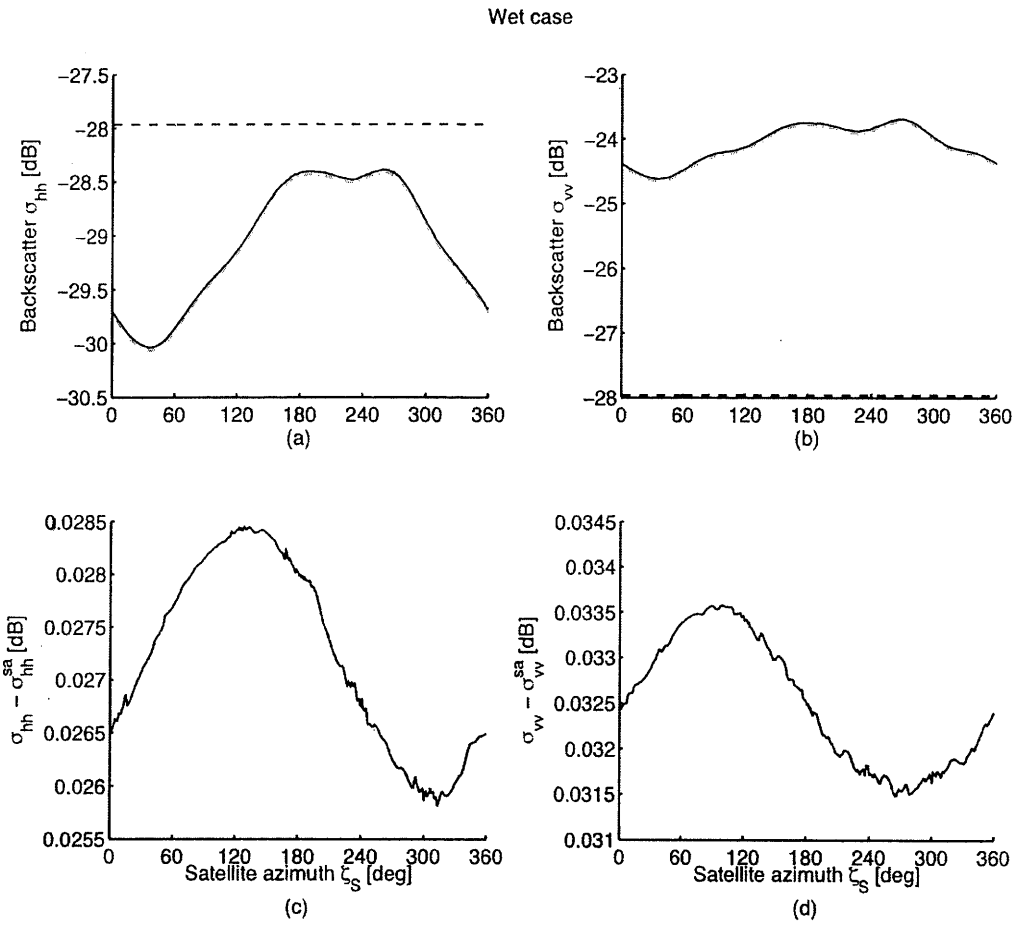


Figure 4-17: For the wet conditions of hour 288 of the simulation, the predicted backscatter at the observing sensor as a function of ζ_S in the (a) horizontally and (b) vertically copolarized domains. Plots show $\overline{\sigma}^{\circ}_{pp}$ (black solid line), $\overline{\sigma}^{\circ sa}_{pp}$ (gray dash-dotted line), and $\overline{\sigma}^{\circ}_{pp,\text{flat}}$ (black dashed line). The difference between $\overline{\sigma}^{\circ}_{pp}$ and $\overline{\sigma}^{\circ sa}_{pp}$ is shown for the (c) horizontally- and (d) vertically-copolarized states.

range of variability of $(\overline{\sigma}^{\circ}_{hh} - \overline{\sigma}^{\circ sa}_{hh})$ is approximately 0.0240 dB at a minimum and 0.0265 dB at maximum (figure 4-16(c)). The behavior of $(\overline{\sigma}^{\circ}_{vv} - \overline{\sigma}^{\circ sa}_{vv})$ as a function of ζ_S is similar to that of $(\overline{\sigma}^{\circ}_{hh} - \overline{\sigma}^{\circ sa}_{hh})$ (figure 4-16(d)). However, the range of variability is slightly greater, with $(\overline{\sigma}^{\circ}_{vv} - \overline{\sigma}^{\circ sa}_{vv})$ exhibiting a minimum of slightly less than 0.0290 dB and a maximum of less than 0.0305 dB (figure 4-16(d)). Similarly small values of $(\overline{\sigma}^{\circ}_{hh} - \overline{\sigma}^{\circ sa}_{hh})$ are also seen for the wet conditions of hour 288 (figure 4-17(c)). Under wet conditions $(\overline{\sigma}^{\circ}_{hh} - \overline{\sigma}^{\circ sa}_{hh})$ varies between about 0.0260 dB when ζ_S is approximately 320° and approximately 0.0285 dB when ζ_S is about 120°. For the wet conditions the variation of $(\overline{\sigma}^{\circ}_{vv} - \overline{\sigma}^{\circ sa}_{vv})$ as a function ζ_S is slightly different than the corresponding relationship between $(\overline{\sigma}^{\circ}_{hh} - \overline{\sigma}^{\circ sa}_{hh})$ and ζ_S. A maximum of

($\overline{\sigma}^{\circ}_{vv} - \overline{\sigma}^{\circ sa}_{vv}$) approximately equal to 0.0335 dB is achieved at ζ_S of approximately 100°, while the minimum of about 0.0315 dB is realized at ζ_S of approximately 270°.

Comparing backscatter values predicted at the sensor, and which include topographic effects on observation geometry ($\overline{\sigma}^{\circ}_{hh}$ and $\overline{\sigma}^{\circ sa}_{hh}$) to the modeled brightness backscatter that neglects topographic effects, $\overline{\sigma}^{\circ}_{hh,\text{flat}}$ is approximately -28.4 dB for the dry conditions of hour 216 (figure 4-16(a)) and -27.9 dB for the wetter conditions of hour 288 (figure 4-17(a)). For the dry conditions the value of $\overline{\sigma}^{\circ}_{hh,\text{flat}}$ is approximately 0.5 dB more than the maximum values of either $\overline{\sigma}^{\circ}_{hh}$ or $\overline{\sigma}^{\circ sa}_{hh}$ (figure 4-16(a)), while $\overline{\sigma}^{\circ}_{vv,\text{flat}}$ is approximately 3 dB less than both either $\overline{\sigma}^{\circ}_{vv}$ or $\overline{\sigma}^{\circ sa}_{vv}$ (figure 4-16(b)). Similarly, for the wet conditions of hour 288 $\overline{\sigma}^{\circ}_{hh,\text{flat}}$ is approximately 0.5 dB more than the maximum values of either $\overline{\sigma}^{\circ}_{hh}$ or $\overline{\sigma}^{\circ sa}_{hh}$, and $\overline{\sigma}^{\circ}_{vv,\text{flat}}$ is almost 3.5 dB less than both either $\overline{\sigma}^{\circ}_{vv}$ or $\overline{\sigma}^{\circ sa}_{vv}$ (figure 4-17(b)). Figure 4-16 depicts two important aspects of the active observing system: (1) hillslope-scale variation in soil moisture is less significant than hillslope-scale topographic affects on observing geometry in setting the predicted microwave backscatter observed at the sensor, and (2) predicted microwave backscatter measured at the sensor are significantly different when topographic effects on observing geometry are included, compared to when the incidence angle is assumed equal to the satellite off-nadir look angle.

While the analysis of the sensitivity of backscatter at a spatial scale consistent with observation to satellite viewing orientation would seem to indicate that the impact of spatial heterogeneity in soil moisture is most pronounced when the sensor views the landscape from either the Southeast or Northwest, the relatively small magnitude of ($\overline{\sigma}^{\circ}_{hh} - \overline{\sigma}^{\circ sa}_{hh}$) suggests that these contrasts do not influence the observation at the scale of the sensor. Although topographic effects on the observational geometry can lead to significant sensitivity of the observed backscatter to satellite azimuth (as evidenced in figures 4-16(a) and (b)), the spatial organization of soil moisture has comparatively less effect on aggregate backscatter (as seen in figures 4-16(c) and (d)). Given that the SMAP radar sensitivity is nominally 0.85 dB, the

predicted microwave radar observations for bare soil conditions at the satellite sensor do not materially vary when a spatial average value of soil moisture is used, as long as the topographic effect on observing geometry at hillslope scales is preserved. This conclusion underscores a fundamental difference between the active and passive microwave observing systems. Because it is formulated as a black body radiative transfer problem, the modeling of emission of microwave energy from moist soils in the passive observing system depends both on the soil moisture effect on emissivity and the impact of surface energy balance on the physical temperature of the soil. The impact on microwave brightness temperature of the hillslope-scale covariation of soil moisture and surface temperature tend to enhance topographic contrasts in brightness. In the Northern hemisphere for example, South- and Southeast-facing hillslopes tend to exhibit lower soil moisture which tends to increase emissivity, while also possessing a higher soil temperature which directly enhances microwave brightness. Conversely, North- and Northwest-facing slopes exhibit higher moisture which suppresses emission and lower physical temperatures which also tends to decrease the brightness. The active observing system, by comparison, has only a weak dependence on temperature. Therefore, for bare soils the spatial organization of microwave backscatter at hillslope scales will largely reflect only the distribution of soil moisture at that scale and the influence of topography on observing geometry, which (through equations 4.30 through 4.33) is a function of the sky position of the satellite. The presence of vegetation, provided that the spatial distribution of vegetation type and biomass exhibits some topographic dependence, will undoubtedly influence the hillslope-scale distribution of microwave radar backscatter. Moreover, hillslope-scale vegetation effects will also likely tend to enhance topographic contrasts in radar backscatter as well. For instance, North- and Northwest-facing slopes tend to have greater vegetation biomass (and in many semiarid locations greater abundance of woody species) and higher soil moisture contents. Both of these covarying factors tend to decrease backscattered energy. The converse could be argued for South- and Southeast-facing slopes. Another degree of complexity that is potentially important, but neglected here, is the role of

hillslope-scale organization of roughness in the modeled microwave backscatter. Little information exists that can be used to infer how roughness varies at hillsope scales and covaries with attributes of topography such as aspect and slope. However, since backscatter is particularly sensitive to roughness, and because the parametric representations of roughness are meant to capture microtopographic variation, roughness is undoubtedly related to geomorphic processes responsible for erosion and deposition of sediment on the landscape. Hence, characterizing the hillslope-scale distribution of geomorphic process dominance would potentially be helpful in constraining how the roughness that influences the microwave radar observation varies at the same scale.

4.6 Conclusions

This chapter has outlined the formulation to two observing systems. The first is a passive microwave observing system that takes the outputs of the tRIBS-VEGGIE model and, given a satellite sky position in the form of an azimuth and zenith angle, predicts the microwave brightness temperature at the observing sensor. The second takes as input the spatial distribution of soil moisture simulated by the tRIBS-VEGGIE model and (again depending on satellite sky position) predicts the microwave backscatter measured at the receiving sensor. Both of these systems share a dependence on topography that explicitly represents the effects of slope and topographic aspect on the local incidence and polarization rotation angles. In the following chapters, these observing systems will be used both to generate synthetic observations of a hidden soil moisture state simulated by tRIBS-VEGGIE, and as the observation operator that projects the ensemble of tRIBS-VEGGIE state vectors into observation space during the data assimilation experiments. The development of these observing systems involved several developments that are at the core of the original contribution of this thesis work. Hence, some important implications of the formulation of these observing systems are discussed in the conclusion.

CHAPTER 5 | MODELING UNCERTAINTY IN FORCINGS

5.1 Introduction

Hydrometeorological variables such as precipitation, incoming radiation, and windspeed are the dynamic forcings that drive the temporal evolution of soil moisture. For the real-time applications outlined in the introduction of the thesis, an estimate of the current spatial distribution of soil moisture and a forecast of how that distribution will change over some known or imposed time horizon (referred to as the lead time or forecast window) are critical. Hence, the nature of the hydrometeorological data required as input to hydrological models, such as tRIBS-VEGGIE, in order to simulate the spatial distribution of soil moisture is twofold: (1) recent historical hydrometeorological data are required to simulate the contemporary spatial distribution of soil moisture, and (2) forecasts of the time evolution of hydrometeorological variables required as input to the model are necessary to produce the corresponding forecasts of the spatial distribution of soil moisture during the lead time.

Both historical and forecast hydrometeorological data are subject to uncertainties that are unique to each data source, which when interpreted by a hydrology model leads to uncertainty in the simulated soil moisture. Assimilation of observational data in the form of microwave brightness temperature and radar backscatter

serves to counteract the influence of uncertainty in hydrometeorologic drivers on the simulated distribution of soil moisture. As outlined in **Chapter 3**, the Monte Carlo-based data assimilation system employed in this work requires explicit treatment of uncertainty in the hydrometeorological forcings required as input to the tRIBS-VEGGIE model. The required forcings for tRIBS-VEGGIE include hourly: (1) precipitation, (2) sky fractional cover or incoming solar radiation, (3) air temperature, (4) dew temperature, and (5) wind speed. This chapter outlines the suite of techniques used to treat uncertainty in these hydrometeorological forcings. For the purpose of the following discussion, it is useful to broadly classify these drivers into precipitation forcings and thermodynamic forcings (solar radiation, air and dew temperature, and wind speed). Precipitation controls the rate and volume of moisture input to the soil column, while the thermodynamic drivers affect the evaporative demand of the atmosphere and the energy available for the evaporation process. What follows is a discussion of the varying sources of uncertainty in both historical and forecast hydrometeorological forcings. The modeling of precipitation uncertainty in space and time is then described. An outline of the treatment of uncertainty in the thermodynamic forcings then follows. Finally, the treatment of hydrometeorological forcings using the outlined techniques in the context of the ensemble data assimilation system is discussed.

5.2 Sources of uncertainty

As discussed above, historical hydrometeorological data and quantitative weather forecasts are necessary for a simulation approach to real-time applications that depend on contemporary knowledge and forecasts of soil moisture at hillslope scales. In this section, uncertainties in each kind of hydrometeorological data source are briefly discussed.

5.2.1 Uncertainty in historical data

Historical data are often retrieved from a network of meteorological stations, such as the network maintained by the Federal Aviation Administration (FAA), or archives of data from multiple networks like the one maintained by the National Oceanic and Atmospheric Administration's (NOAA) National Climatic Data Center (NCDC). These point-scale stations typically collect data related to precipitation, winds peed, temperature, humidity, and solar radiation, at temporal resolutions ranging from 5-minutes to daily. While these networks and archives are critical resources for hydrologic modeling, they nevertheless introduce uncertainty into predictions of soil moisture when they are supplied to hydrological models as input. The most basic form of uncertainty associated with these data is the instrument precision and accuracy. While an in-depth discussion of the data collection techniques for the hydrometeorological variables of interest is beyond the scope of this thesis, it is important to recognize that there is uncertainty associated with the instrumentation used to collect these data. Details regarding the accuracy and resolution of particular sensors are most often reported by the instrument manufacturer and/or the maintainer of the station or network (e.g., see Brock *et al.* [1995] for a review of the stations comprising the Oklahoma Mesonet), and to thoroughly treat the uncertainty associated with measurements from a particular hydrometeorological station it is necessary to ascertain the details of the station's instrument configuration. Furthermore, when characterizing uncertainty in hydrometeorological variables from automated data collection networks and/or archive meta-databases comprised of information from multiple networks, there may be an additional layer complexity associated with heterogeneity in sensor packages between stations. This implies an added degree of caution and attention to detail that must be taken in assigning uncertainty characteristics to historical hydrometeorological data.

As alluded to above, many historical weather station data have temporal resolutions poorer than the one hour time-step required by models like tRIBS-VEGGIE. Moreover, some rudimentary weather stations (particularly those in developing

countries) may only report daily weather statistics or cumulative measures such as minimum and maximum temperatures, cumulative precipitation, and dew temperature at specified times. Employing these data for hydrologic simulation with continuous hydrologic forecasting models like tRIBS-VEGGIE requires temporal disaggregation of these data to hourly time scales. Temporal disaggregation schemes can be either deterministic or stochastic, and produce hourly estimates of the required hydrometeorological variables through a blend of the the historical weather data, climatic information, and the time of day. Using a mix of stochastic and deterministic techniques for a watershed in Texas, Debele *et al.* [2007] disaggregated hydrometeorological observations of daily cumulative rainfall; daily maximum and minimum air temperature; and daily average wind speed and humidity. Because their daily historical data consisted of aggregated hourly data, they were able to assess the accuracy of the various methods they employed. Because the evolution of hydrometeorological variables at hourly timescales depends on micrometeorological and synoptic meteorological phenomena, these disaggregation methods are inherently error prone. Thus, while temporal disaggregation of historical weather observations to hourly timescales may in some instances be required for hydrologic simulation, it will nevertheless introduce uncertainty into the hydrometeorological forcings that will ultimately lead to uncertainty in the simulated soil moisture distribution.

Another source of uncertainty in historical data from weather stations that is particularly important in spatially distributed soil moisture simulation is associated with the lack of hydrometeorological information in portions of the watershed that are spatially removed from weather stations. Computational pixels in tRIBS-VEGGIE and similar spatially distributed simulation models require local estimates of the required hydrometeorological variables in order to simulate the soil moisture response. Yet, with a few notable exceptions corresponding to well-gauged experimental watersheds, automated weather collection networks are relatively sparse. Many techniques have been applied historically to estimate hydrometeorological variables, particularly precipitation, at locations in a watershed that are

relatively far from weather stations in the collection network. The most simple of these techniques, assigns a location the hydrometeorological forcings of the nearest weather station to every computational pixel in the watershed. This nearest neighbor approach has the effect of discretizing the watershed into Thiessen polygons that delineate the area of influence of each weather station. More complex algorithms, such as Kriging and Artificial Neural Networks [Hung *et al.*, 2008], have also been applied to interpolate hydrometeorological conditions between weather stations. Local hydrometeorological conditions are the result dynamically coupled and nonlinear interactions between the land surface and atmosphere. Regardless of the complexity of the interpolation procedure, statistically interpolated estimates of hydrometeorological variables at locations between weather stations are almost invariably in error. When the hydrologic simulation model experiences hydrometeorological forcings that are in error, the simulated distribution of soil moisture correspondingly will be uncertain.

In addition to point-scale weather station data, remotely sensed observations from spaceborne satellites and ground-based radar stations are becoming increasingly useful as sources of spatially distributed historical hydrometeorological data. For instance, the Advanced Microwave Sounding Unit-A (AMSU-A) and AMSU-B sensors have been used to retrieve global atmospheric humidity and temperature profile estimates [Rosenkranz, 2001], data from the Moderate Resolution Imaging Spectroradiometer (MODIS) has been used to derive Photosynthetically Active Radiation at the surface [Laake and Sanchez-Azofeifa, 2005], and the scatterometer aboard the first European remote sensing satellite (ERS-1) has been used to derive wind vectors [Freilich and Dunbar, 1993]. Quantitative precipitation estimation (QPE) receives much of the emphasis in remote sensing of hydrometeorological phenomena. Both ground- and space-based techniques have been exploited for QPE, and the data that these instruments provide is widely used in the hydrologic sciences. For instance, the NASA's Tropical Rainfall Measurement Mission (TRMM) provides both active and passive observations of rainfall in Tropical latitudes. The TRMM 3B42 algorithm, in particular, provides a 3-hourly average

precipitation estimate at spatial scales of 0.25° that combines passive and active microwave observations from the TRMM satellite, infrared data from other satellites, and rain gage data [Huffman *et al.*, 1995]. The related Global Precipitation Measurement (GPM) mission is scheduled to provide similar precipitation products globally. Among remotely sensed precipitation data, ground-based Doppler weather radar are among the most widely used in the United States. The National Weather Service's (NWS) Next-Generation Radar (NEXRAD) network of doppler weather radars provides data that can be processed into QPEs at temporal resolutions as fine as 15 minutes and spatial scales of approximately 2 km [Grassotti *et al.*, 2003].

Uncertainties arising from remotely sensed hydrometeorological data typically fall into two varieties: (1) uncertainties associated with the spatial and temporal resolution of remotely sensed products, and (2) uncertainties associated with inversion algorithms that retrieve estimates of hydrometeorological variables from geophysically observable quantities. Like data from point-scale weather stations, remotely sensed hydrometeorological products often are associated with spatial and temporal resolutions that are inconveniently coarse from the perspective of hillslope-scale soil moisture modeling. While the hydrometeorological information provided by weather station data is highly localized, each pixel of a remotely sensed hydrometeorological product is representative of spatially-averaged conditions within that pixel. In many circumstances, the coarse resolution of such products tends to smooth out important sub-pixel variability. The 0.25° and 3 hr resolution of the TRMM 3B42 product is a prime example. The spatial resolution of 3B42 may smooth out significant spatial variability in rainfall rates in areas of steep topography associated with orographic forcing, while the temporal resolution of the product may smooth rainfall rates associated with short-duration, high-intensity convection-dominated rainfall. These smoothing tendencies may hide potentially important spatiotemporally localized extreme rainfall that would substantially affect localized soil moisture conditions. As in the case of interpolation of weather station data, spatial and temporal disaggregation of remotely sensed products will introduce errors in the hydrometeorological forcings, leading to uncertainty in the

simulated soil moisture response.

Retrieval from the geophysically observed quantities leads to significant uncertainties in the remotely sensed estimates of hydrometeorological variables. As the above outline of current and planned remote sensing products indicates, microwave microwave, visible, and infrared technologies are commonly used to observe hydrometeorological phenomena. Similar to the remotely sensed soil moisture data discussed in **Chapter 2**, approximate empirical or physically-based retrieval algorithms are used to estimate relevant hydrometeorological variables from observations of observable quantities like brightness temperature, backscatter, Doppler shift, and reflectance. Of particular importance in hydrological modeling is the uncertainty in the transformation of radar and radiometer data into QPEs. While the sources of uncertainty associated with ground-based weather radar are often distinct from satellite retrieval uncertainties (e.g., ground clutter, the orientation of the radar, storm, and watershed, etc.) [Sharif *et al.*, 2002; Borga *et al.*, 2006], previous work investigating QPE retrieval errors have revealed that it is not uncommon for precipitation estimates from NEXRAD reflectance data to be in error by a factor of 2 or more [Sharif *et al.*, 2004]. In a study using a mesoscale meteorological to generate a synthetic true rainfall field together with a stochastic radar simulator, Sharif *et al.* [2002] showed that uncertainties of this magnitude can have substantially increase error in predicted hydrological variables, such as discharge when propagated through a simulation model. Because rainfall serves as a mass balance input constraint on the hydrologic system, it receives significant attention in the treatment of uncertainties in this chapter.

5.2.2 Uncertainty in weather forecasts

To issue forecasts of the spatial distribution of soil moisture out to some lead time in the future, it is necessary to have quantitative predictions of the hydrometeorological variables required by a simulation model over the same period of time. From a physically-based perspective, the coupled nature of the land surface

and atmosphere should suggest a modeling framework that captures coevolution of the land surface and atmospheric states during the forecast window. While the high-dimensionality of simulation models like tRIBS-VEGGIE makes such a framework computationally expensive, significant effort is underway to couple tRIBS-VEGGIE to the Weather Research Forecasting (WRF) model. In absence of such a framework, it is necessary to use quantitative weather forecasts to directly force a simulation model. Forecasting centers such as the European Center for Medium Range Weather Forecasting (ECMWF) and the National Center for Environmental Prediction (NCEP) in the United States provide quantitative ensemble forecasts with 10 days lead time. It is important to note that these forecast products are associated with a particular suite of coupled land-ocean-atmosphere models and that using forecast hydrometeorological variables to force a different land surface model, such as tRIBS-VEGGIE, implies a degree of conceptual mismatch that may lead to uncertainties in the predictions of the hydrological variables of interest. Furthermore, the need to produce sometimes global forecasts based on a network of current and recent-past hydrometeorological observations implies that the spatial resolution of forecast products are often too coarse to resolve spatial variability in hydrometeorological variables that may be significant in terms of hillslope-scale soil moisture prediction (e.g., ECMWF forecasts have a 40 km resolution in space [Persson and Grazzini, 2005]). The more significant source of uncertainty in numerical weather forecasts arises due to the nonlinear and chaotic nature of the atmosphere, which is sensitively dependent to uncertainty in initial conditions. Hence, substantial uncertainty in hydrometeorological forecasts is often encountered at longer lead times, that may translate to significant uncertainty in simulated soil moisture and discharge response when propagated through a simulation model. The uncertainty is quantitatively represented by many forecasting centers (such as ECMWF and NCEP) through ensemble forecasts, and individual ensemble replicates are frequently made available in addition to quantitative metrics of ensemble consensus and uncertainty. Spatial disaggregation of coarse forecast products will introduce additional uncertainties beyond the inherent uncertainties associated with the non-

linear chaotic nature of the weather models.

5.3 Modeling uncertainty

The goal of this portion of the work is to capture some degree of the conceptual uncertainties outlined above by employing relatively simple quantitative tools that can quickly produce ensembles of the hydrometeorological variables required as input to tRIBS-VEGGIE at hourly timescale.

To represent uncertainty in the hydrometeorological forcings, three stochastic models are used. Emphasis is placed on representing uncertainty in the spatiotemporal character of rainfall because it imposes significant constraints on soil moisture. Two coupled stochastic models are used to model temporal and spatial uncertainty in precipitation. The first is the Modified Bartlett-Lewis (MBL) model that generates time series of hourly point-scale rainfall. The second is a stochastic cascade model that is used both to introduce noise into the point-scale rainfall generated with the MBL model and to disaggregate it in space. The third model is a stochastic weather generator developed by Ivanov *et al.* [2007], which is used to simulate point-scale thermodynamic drivers at hourly timescales required by tRIBS-VEGGIE. Because the thermodynamic variables derived by the stochastic weather generator are conditioned on the hourly occurrence of rainfall, the hydrometeorological forcings simulated by this suite of stochastic models is completely internally consistent. That is, the occurrence of rainfall in any given hour impacts the corresponding thermodynamic drivers in well-defined ways. Figure 5-1 presents a flow chart of how these models are used to simulate uncertain hourly hydrometeorological forcings. Brief introductions to each of these models follow below. This hydrometeorological uncertainty modeling framework is applied extensively in **Chapter 7** in an set of experiments designed to diagnose the relative contributions of uncertainty in the forcings and soil parameters required by tRIBS-VEGGIE to uncertainty in the simulated soil moisture state. Several different scenarios in which these

stochastic models are used to simulate ensembles of uncertain hydrometeorological inputs to tRIBS-VEGGIE will be described there.

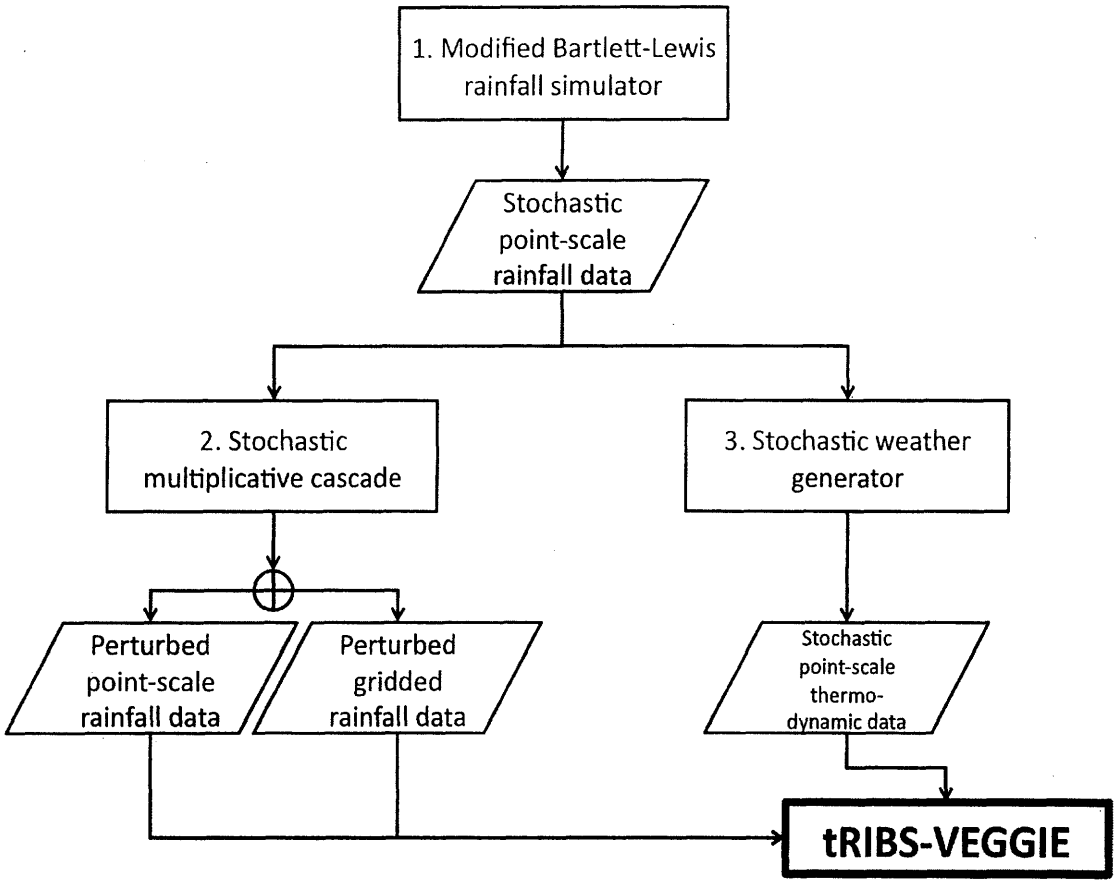


Figure 5-1: A schematic representation of how 3 stochastic models are used to generate uncertain hydrometeorological forcings for the tRIBS-VEGGIE model. The Modified Bartlett-Lewis model is used to generate time series of hourly rainfall. This hourly rainfall time series is then input to the multiplicative cascade model which is used both to perturb point scale data and to disaggregate the MBL-derived rainfall series in space. The MBL-derived rainfall time series is also given as input to the stochastic weather generator of Ivanov *et al.* [2007] to derive stochastic thermodynamic forcings required as input to tRIBS-VEGGIE.

5.3.1 Modeling uncertainty: Modified Bartlett-Lewis model

The MBL model stochastically partitions time into rain and non-rain events. The arrival of storm events is treated as a marked Poisson process [Rodriguez-Iturbe *et al.*, 1988, 1987], and the time between successive storm events are assumed

independent and exponentially distributed, and parameterized by a storm arrival rate, λ. During a finite window of time after the beginning of a storm event, the MBL approach models the arrival of individual storm cells. For a particular storm event, this window of time of cell-genesis is of a random duration that follows an exponential distribution, parameterized by the rate parameter γ. Individual cells arrive independently, and the time between the arrival of successive cells is random and exponentially distributed, with a rate parameter β. Individual storm cells are of random duration and intensity. Both cell duration and intensity are mutually independent and are assumed to follow exponential distributions. Cell durations are parameterized by the rate parameter η, while storm cell intensity is parameterized by the mean intensity, μ_x. The average rainfall rate during some finite interval (i.e., one hour) is the integral of the active storm cell intensities during that interval, normalized by the duration of the interval. Figure 5-2 is a schematic representation of the Bartlett-Lewis storm-cell arrival concept for one storm event. Individual cell and integrated storm intensities are shown.

A considerable amount of effort has gone into estimation of parameters of the MBL model from historical rain gauge data. While a detailed overview of parameter estimation techniques for the MBL model is not given here, the work of Islam *et al.* [1990] provides a detailed assessment of parameter estimation techniques as well as relevant statistical metrics that can be used to estimate and validate parameters. Hawk [1992] performed an analysis of weather stations throughout the continental United States, and provides estimates of the MBL parameters for each of these weather stations. Consistent with the semiarid and arid conditions to which tRIBS-VEGGIE is best suited, the MBL parameters for Tucson International Airport, Arizona (IATA code TUS) estimated by Hawk [1992] are used for point-scale rainfall generation.

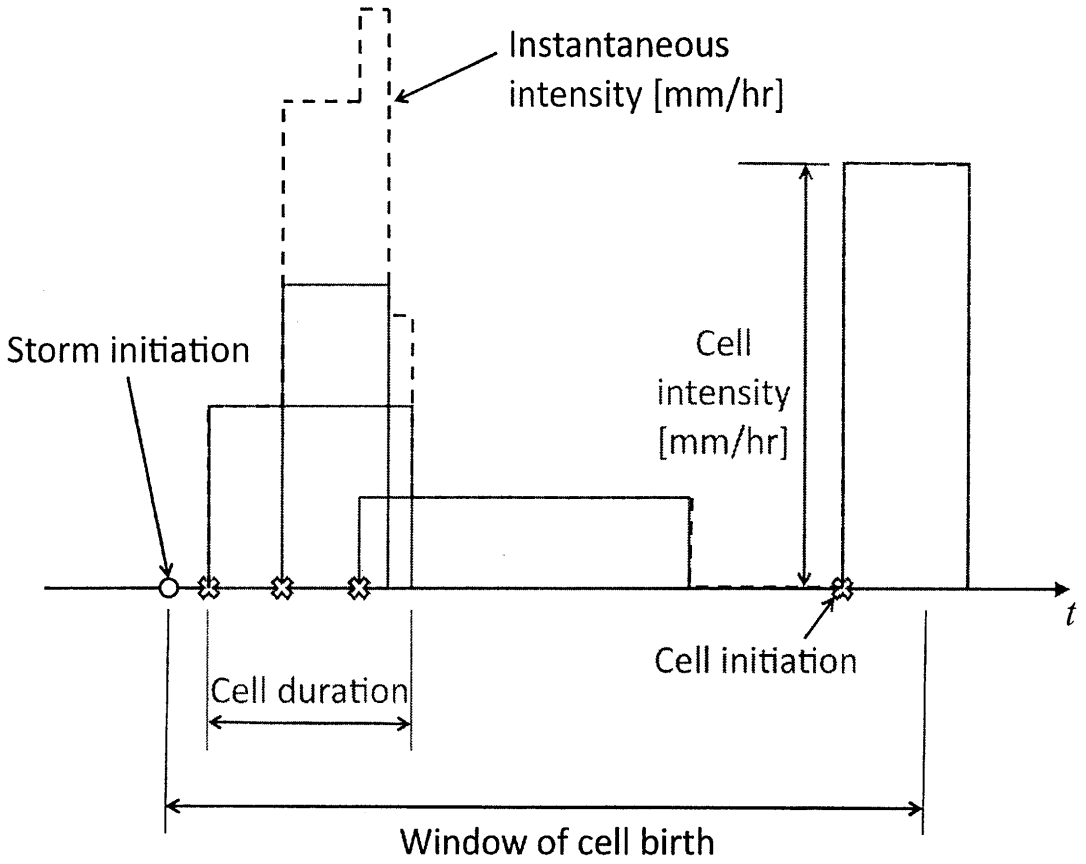


Figure 5-2: A schematic representation of the Bartlett-Lewis storm cell concept. The storm outline is shown black dashed lines, while individual storm cells are outlined in gray. The arrival of a storm event (black circle) is followed by the arrival of individual cells (gray hollow x's) during some finite window of random duration during which cell birth may occur. The total instantaneous rainfall intensity at any time during the storm is the summation of intensities of the active cells at that time (gray lines).

5.3.2 Modeling uncertainty: stochastic multiplicative cascade

The multiplicative cascade model used here was originally developed by Mandlebrot [1974] to model intermittent turbulent fields, and was later adopted by others (e.g., Schertzer and Lovejoy [1987]; Over and Gupta [1996]; Menabde *et al.* [1997] and Menabde *et al.* [1999]) to parsimoniously reproduce much of the multiscaling phenomena often observed in radar reflectance data. The underlying assumption of these multiplicative cascades is the validity of the assertion that turbulence is a phenomenologically self-similar cascade of energy across at least a

range of scales [Richardson, 1922; Schertzer and Lovejoy, 1987]. While some have attempted to provide a rigorous theoretical grounding for many variations of the multiplicative cascade framework (e.g., Lovejoy and Mandlebrot [1985]; Schertzer and Lovejoy [1987]; Over and Gupta [1996]; Deidda [2000], others argue that the strongest reasoning for use of a cascade downscaling process is the fact that empirically it seems to reproduce observed variations in radar-rainfall data [Menabde *et al.*, 1997].

The version of the multiplicative cascade used here is the same as the one outlined in Menabde *et al.* [1999],Menabde *et al.* [1997] and Ferraris *et al.* [2003]. It partitions the rainfall volume, π_0, in a square pixel of linear dimension L into rainfall volumes within four square pixels with linear dimension $L/2$ and volumes π_i. This cascade process repeats J times, leaving 4^J square pixels with linear dimension $L/2^J$. At cascade step J, the volume in any (i) of the 4^J child pixels at step J of the cascade, $\pi_i(J)$, is given by,

$$\pi_i(J) = \frac{\pi_0}{4^J} \prod_{j=1}^{J} W_{k(j)}. \tag{5.1}$$

A schematic representation of the multiplicative cascade process is shown in Figure 5-2. In this implementation of the multiplicative cascade, the multiplicative weight at step j of the cascade and at a child pixel indexed by i, $W_{k(j)}$, takes one of two values, W_1 or W_2, occurring with probability p and $(1-p)$, respectively. These weights depend on a set of parameters and the cascade step j:

$$W_1 = \left[1 + (w_1 - 1)2^{-jh}\right] \tag{5.2}$$

occurs with probability p, while

$$W_2 = \left[1 + (w_2 - 1)2^{-jh}\right] \tag{5.3}$$

occurs with probability $(1-p)$. The parameters w_1, w_2 and h in equations (5.2)

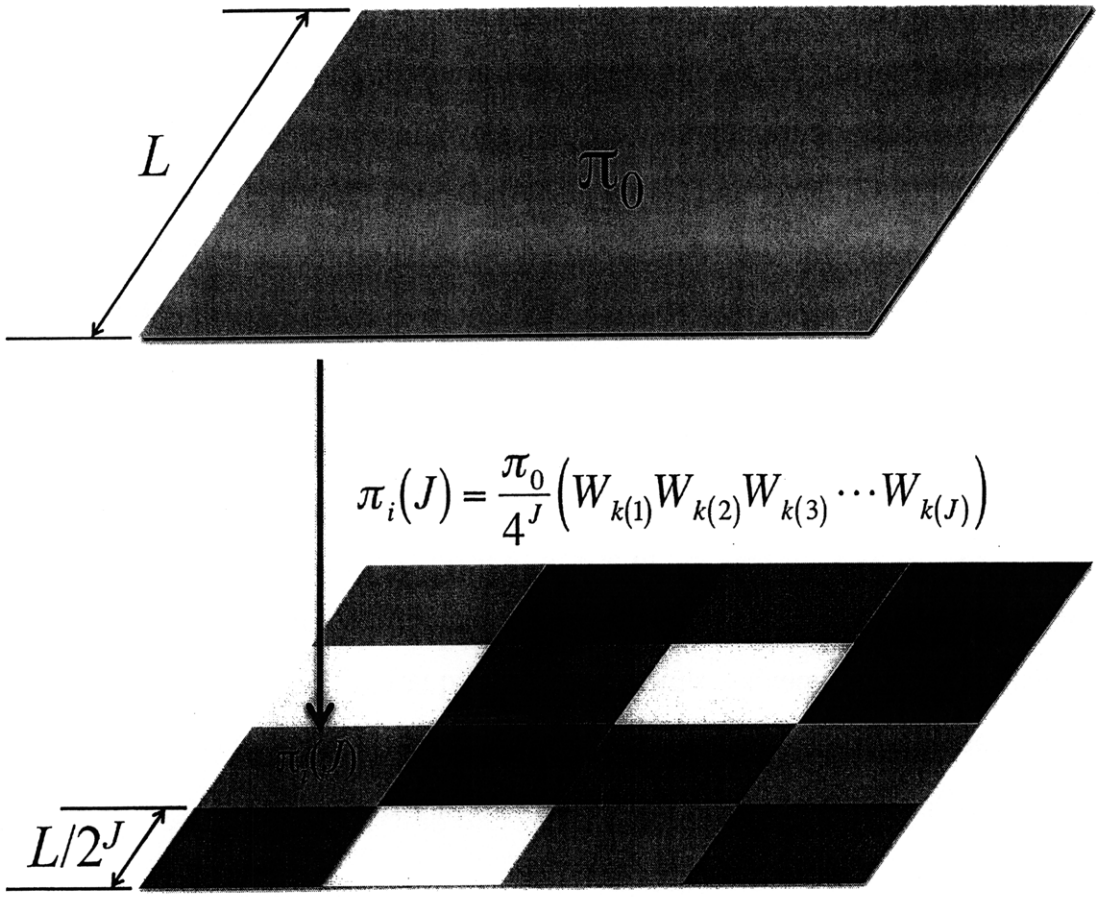


Figure 5-3: A schematic representation of application of the mutiplicative cascade model used in this work to spatial disaggregation of rainfall. The volume rainfall at some spatial pixel indexed by i in the j^{th} step of the cascade, π_i, is the cumulative product of the volume of rainfall at the coarsest scale of the cascade, π_0 and the j random weights, $W_{k(j)}$ realized at each step in the cascade.

and (5.3) are estimated from radar rainfall images in a way that preserves some statistical attributes of the data. In particular, it is common to aggregate radar rainfall images to coarser scales, and then use the cascade to downscale back to the native resolution using a broad range of parameter combinations (e.g., Ferraris *et al.* [2003]). There are some theoretical arguments that suggest the parameter h should be close to 1/3 [Menabde *et al.*, 1997]. Otherwise, parameters are typically fit on the basis of the degree to which the downscaled rainfall images reproduce first- and second-order statistics, as well as fractal dimensions of the observed data. Constraints on the parameters can be found to ensure that the multiplicative cascade is mass conservative in the ensemble sense. This means that the spatial pattern of a

given rainfall image can be changed using the multiplicative cascade disaggregation process without changing the spatial average precipitation volume. Ensemble-sense mass conservation requires that the expectation of the weights be unity,

$$E[W] = 1. \tag{5.4}$$

Expanding the expectation based on the binary definition of the weights outlined above, gives,

$$E[W] = pW_1 + (1-p)W_2 = 1, \tag{5.5}$$

and substituting the definitions of W_1 and W_2 above into equation gives,

$$E[W] = p\left[1 + (w_1 - 1)2^{-jh}\right] + (1-p)\left[1 + (w_2 - 1)2^{-jh}\right] = 1. \tag{5.6}$$

Rearranging the terms in this expression, it is straightforward to show that for equation 5.4 to be true the parameters w_1 and w_2 must be related to the probability, p, of weight W_1 occurring in the following way,

$$w_2 = \frac{1}{1-p} - w_1\frac{p}{1-p}. \tag{5.7}$$

Imposing a mass balance constraint, therefore, eliminates one parameter (either w_1 or w_2), reducing the number of independent parameters to three (w_1 or w_2, h, and p). Consistent with the findings of Menabde *et al.* [1997], the parameter h is assumed to be 0.3. Because the spatial extent of the watersheds considered is relatively small (e.g., approximately 10 km) the remaining independent parameters of the cascade (w_1 and p)are tuned based on heuristic criteria that seeks to ensure that a broad range of multiplicative weights could be realized at any given simulated radar pixel. The values of the parameters used are shown in Table 5-1.

The values of the multiplicative weights W_1 and W_2 at each step of the cascade,

Table 5.1: Parameters of the multiplicative cascade

Parameter	Value
J	5
h	0.3
p	0.6
w_1	0.4
w_2	1.9

for $j = 1, 2 \ldots 5$ are graphically depicted in Figure 5-4. The multiplicative weight that determines the rainfall volume in the i^{th} pixel of the downscaled radar field at $J = 5$ steps into the cascade will be the product of one of the two multiplicative weights chosen through a random draw (with W_1 occurring 6 times in 10) at each step of the cascade. The maximum multiplicative weight that can be realized in this procedure, which occurs if W_2 is drawn at each step of the cascade, is approximately 7.50. The minimum realizable weight is approximately 0.12.

To generate a single ensemble replicate of spatiotemporally varying rainfall over a watershed, the MBL model is used to generate a time series of rainfall at one hour intervals, and the multiplicative cascade model is used to partition the hourly rainfall volume in space. The cascade model is also employed in this thesis to introduce uncertainty into the rainfall time series generated by the MBL model for point-scale and small catchment scale simulations. Using the a time series of simulated rainfall derived from the MBL model, a time series of multiplicative weights, equal in length to the rainfall time series, is generated using the multiplicative cascade with the parameters shown in Table 5-1. The noise-corrupted point-scale time series of rainfall is the hour-by-hour product of the rainfall and the multiplicative weight. For the data assimilation experiments in this thesis, an ensemble of spatiotemporally varying rainfall is generated by: (1) generating a synthetic true time series of rainfall simulated by the MBL model, (2) perturbing the true rainfall time series using the multiplicative cascade, and (3) spatially disaggregating the perturbed rainfall time series in space using the multiplicative cascade, and (4) repeating steps (2) and (3) until the desired ensemble size is achieved. This procedure assumes that the initiation of storm events is relatively well known at the watershed scale, while the

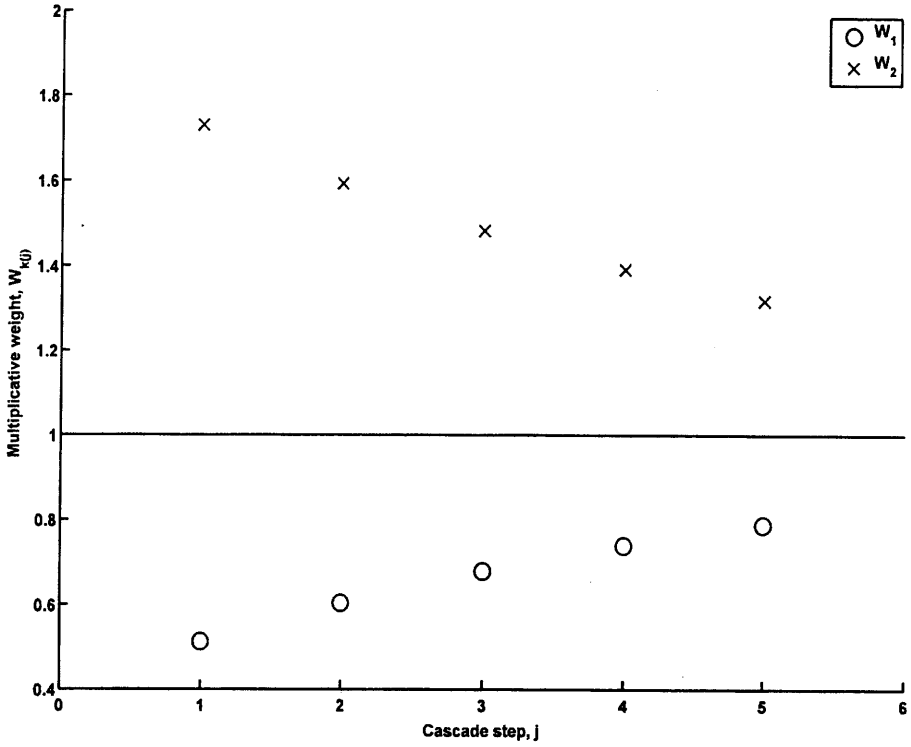


Figure 5-4: Values of the multiplicative weights, W_{1j} (blue circles) and W_{2j} (red x's), at each step in the multiplicative cascade. The multiplicative weight that determines the volume of rainfall in pixel i of the disaggregated rainfall field will be the product of one weight chosen by a random draw at each step of the cascade.

rainfall volumes are the primary source of uncertainty in rainfall data. What follows is a discussion of the simulation of the remainder of the hydrometeorological forcings required as input to tRIBS-VEGGIE.

5.3.3 Modeling uncertainty: thermodynamic forcings

As described in the introduction to this chapter, uncertainty in the thermodynamic forcings of the hydrologic system leads to uncertainty in the atmospheric evaporative demand, thereby affecting uncertainty in soil moisture. This section provides a brief description of the stochastic weather generator formulated by Ivanov *et al.* [2007], to which the interested reader is referred for additional details. The original implementation of the weather generator began by simulating hourly precipitation using the Poisson Rectangular Pulse stochastic point-scale rainfall model, which is similar to but simpler than the MBL model. This was modified slightly so that hourly rainfall data simulated by the MBL model discussed above could be used as input to a stochastic weather generator. The weather generator simulates the time evolution of sky cloudiness fraction, conditioned by the hourly rainfall time series, using the stochastic model of Curtis and Eagleson [1982]. Based on the Julian day and hour being simulated, the sky position of the Sun is calculated and the amount of incoming solar radiation at the top of the atmosphere is determined. The amount of solar radiation in two frequency bands (visible and near-infrared) is then computed as a function of the solar radiation input to the top of the atmosphere and atmospheric attenuation factors that are a function of the cloudiness fraction and precipitable water in the atmospheric column. Air temperature is computed as the sum of a deterministic term that depends on an estimate of the incoming longwave radiation and a stochastic perturbation term that follows a first-order Markov process. Dew temperature is simulated on a daily basis and is a function of the minimum air temperature during the day and an estimate of the evaporative fraction during the day. Some assumptions about potential evaporation rates are required to estimate the daily evaporative fraction and the interested reader is referred to

the work of Ivanov *et al.* [2007] for further detail. Wind speed is simulated as a first-order autoregressive process, taking into account the positive skew in the distribution of hourly wind speeds typically observed. To ensure consistency with the parameters of the MBL model, the parameter estimates that Ivanov *et al.* [2007] derived for Tucson International Airport (TUS) in Tucson, Arizona are used to simulate the thermodynamic forcings.

The stochastic weather generator is used in this work to ensure that all thermodynamic forcings are self-consistent, and consistent with the rainfall forcings. More broadly, the approach to modeling uncertainty in the hydrometeorological forcings required as input to the tRIBS-VEGGIE model is substantially less complex than using a numerical weather prediction model to generate ensembles of rainfall fields, and adequately captures dynamics of spatiotemporal evolution of rainfall for relatively small watersheds.

CHAPTER 6

MODELING UNCERTAINTY IN SOIL PROPERTIES

As described in Chapter 3, the EnKF retrieves an estimate of the "first guess" state error covariance matrix directly from the ensemble of model simulations. Along with the observational error covariance matrix, this ensemble estimate of the state error covariance matrix estimate determine the gain, and therefore the relative contributions of the model and observation in the updated state estimate. Inadequate treatment of uncertainty in the model and its inputs can result in first guess error covariance estimates that are unrealistically low. As a consequence, the state update machinery of the EnKF interprets an artificially low state error covariance as a high model confidence, leading to a relative mistrust of the observations and minimal propagation of information from the observations to the model forecast. This idea is shown diagrammatically in figure 6-1. In this representation, as the ensemble replicates are propagated forward they tend to converge to a relatively limited area in state space, which represents an attractor state of the system (figure 6-1). The ensemble state error covariance approximated from this relatively small area of state space will correspondingly small. Although the ensemble of predicted observations may be substantially different than the data, thus implying significant innovations, the unrealistically small approximation of the state error covariance will limit the amount of information in the observations propagating back to the state.

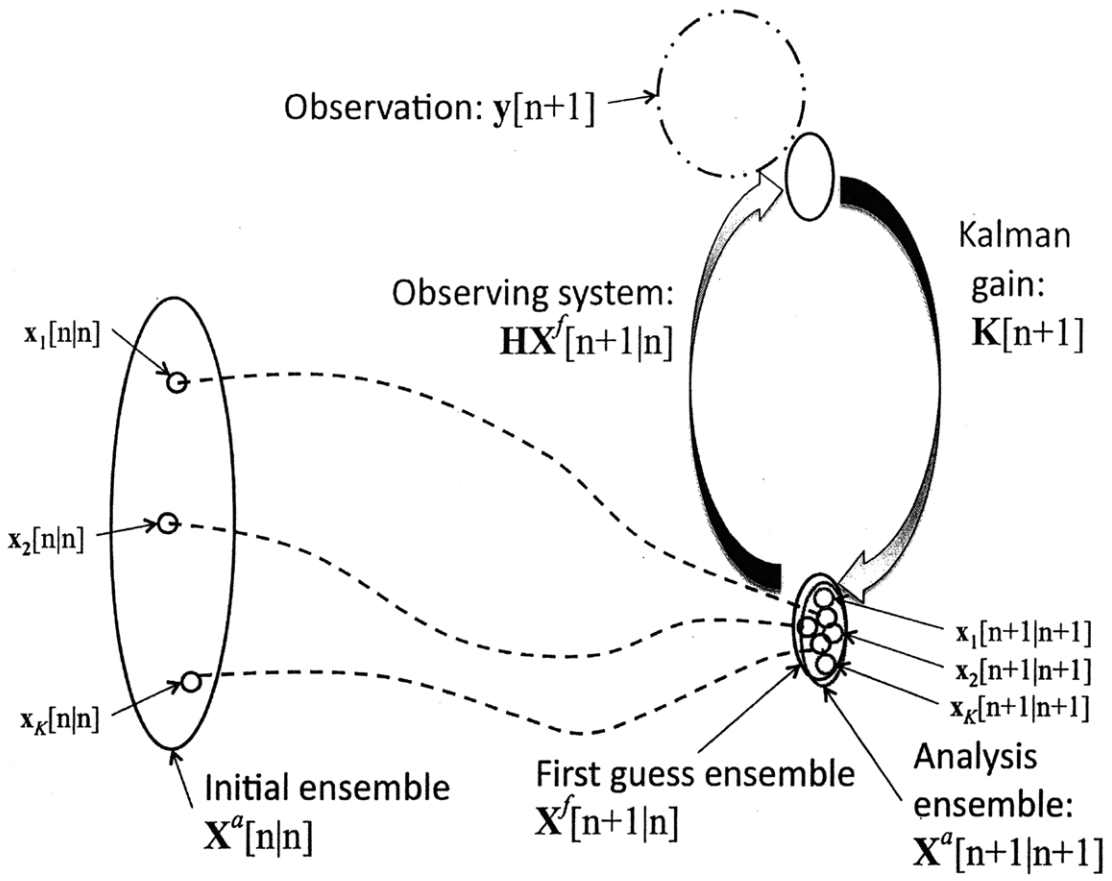


Figure 6-1: A schematic representation of unrealistically low model error resulting from inadequate representation of sources of uncertainty in the model and its forcings.

The particular worry in the context of ensemble data assimilation with the tRIBS-VEGGIE model specifically, and all dynamic soil moisture models generally, is unrealistically small error covariance in the soil moisture state arising from unrealistic or inadequate representation in either the hydrometeorological forcings or the model itself. Model uncertainties can be further subdivided into several additional categories: (1) loss of fidelity to the physical processes of moisture redistribution associated with model abstraction and formulation of the governing equations, (2) simplification, spatiotemporal discretization, and parameterization of the governing equations, and (3) imperfect characterization and knowledge of the properties of the heterogeneous and anisotropic porous media.

The latter of these model uncertainties has been the topic of a significant amount

of research in hydrologic science literature. The thrust of the majority of these efforts has been to efficiently estimate the soil properties in parameter specification and calibration exercises. Early employed Kalman Filtering procedures, augmenting the state vector with some of the model parameters[1], to simultaneously estimate the model state and parameters. This approach to parameter estimation, commonly referred to as automatic calibration, should ideally converge to the best parameter and state estimate through continued assimilation of observations. Hino [1973] applied the Kalman Filter to estimate the ordinates of the unit hydrograph for a lumped basin discharge model for which sequential observations were available. Kitanidis and Bras [1979] investigated the stability of parameters estimated through a Kalman filtering framework with increasing numbers of observations. In an interesting investigation of structural model errors, Restrepo-Posada and Bras [1982] used a three linear reservoir model to generate synthetic observations of discharge and then used the augmented state vector approach to estimate the parameters for a two linear reservoir model as a way of investigating the effects of structural model error.

More recently, a substantial body of work has been devoted to global optimization approaches (see e.g., Duan *et al.* [1992]; Gupta and Sorooshian [1994]; Yapo *et al.* [1998]; Boyle *et al.* [2000]). These global optimization approaches, which minimize an often multi-objective cost function that penalizes deviations between predicted observations and data to arrive at some pareto-optimal parameter estimate, may be advantageous when using complex and high-dimensional hydrologic models because augmented state vector approaches may be relatively more computationally expensive. Vrugt *et al.* [2004] demonstrated the feasibility of estimating soil hydraulic properties for a distributed model in an agricultural water district through an inverse modeling framework using observations of plot-scale tile drainage. Minasny and Field [2005] also used inverse modeling with Generalized Likelihood Uncertainty Estimation (GLUE) to estimate soil hydraulic properties and their uncertainty with observations of evaporation. The idea underlying the param-

[1]see Schweppe [1973] and Gelb [1974] for a review of state augmentation

eter estimation portion of the combined parameter-state estimation problem is that the model can be used to solve for the parameter set that minimizes estimation errors between model outputs and observations.

Uncertainty in soil properties has increasingly received attention in ensemble-based soil moisture data assimilation. For instance, Margulis *et al.* [2002] and later Dunne and Entekhabi [2006] introduce random variation in soil hydraulic and vegetation properties between ensemble replicates as a way in including effects of parameter uncertainty. Vrugt *et al.* [2005] attempt to exploit the strengths of both data assimilation and global optimization strategies to estimate hydrologic model states and parameters in an ensemble-based framework that they term simultaneous optimization and data assimilation (SODA). Further, Moradkhani *et al.* [2005] investigated a particle filtering approach, a form of ensemble-based data assimilation, to address the problem of the combined state and parameter estimation problem for a lumped conceptual hydrology model. Hence, in ensemble-based hydrologic data assimilation, the importance of representing uncertainty in model parameters is being increasingly recognized.

This chapter, therefore, deals with the treatment of the uncertainty in soil hydraulic and thermal properties (SHTPs). The goal is not to perform parameter estimation, but rather to ensure that uncertainty in SHTPs is represented sufficiently well to ensure that ensemble estimates of the state error covariance are not unrealistically low. Extension of this work to parameter estimation will, however, be briefly discussed as a potential future research objective in the concluding chapter of the thesis. It is a straightforward heuristic exercise to deduce a simple, yet feasible, case in which failure to represent uncertainty in SHTPs may lead to an unrealistically low estimate of the soil moisture state error covariance in an ensemble-based data assimilation system. Specifically, using a single set of SHTPs to characterize a particular soil type in an ensemble tRIBS-VEGGIE simulation would imply a single constant value for residual moisture content, θ_R. Assuming that the timing and duration of rainfall events is well known, but storm-event volumes and rainfall rates

uncertain, the timing of wetting and drying cycles would be similar across all of the ensemble replicates. During storm events the uncertainty in soil moisture, particularly in the near-surface where it is most related to remotely sensed geophysical observations, would be controlled largely by the variance in rainfall rates and volumes. Immediately after rainfall cessation, uncertainty in soil moisture will begin to decrease as individual ensemble replicates dry at rates that are dictated by the soil moisture content at the end of the storm, vis-á-vis the dependence of hydraulic conductivity on soil moisture. As the drying cycle progresses, the uncertainty in soil moisture will asymptotically approach zero because each ensemble replicate is parameterized by the identical residual moisture content. Hence, for circumstances in which the arrival and duration of storms is well known but the precipitation volumes and rates are uncertain, and to the extent that the model is parameterized by some residual moisture content that constitutes an absolute minimum value of soil moisture, toward the end of each drying cycle the uncertainty in soil moisture will be small. This conceptual argument is demonstrated through a single point-scale tRIBS-VEGGIE ensemble simulation in figure 6-2. With a single set of SHTPs, well-known rainfall timing and duration, and uncertain rainfall rates and volumes, the variance in near-surface (i.e., top 10 cm) soil moisture increases during rainfall events and decays to a small value as the drying cycle progresses (figure 6-2).

An argument can be made that low soil moisture uncertainty is consistent with the physical expectation of very little near-surface moisture toward the end of a drying cycle, irrespective of the soil. However, while that argument may be more physically realistic, the underlying criticism reflects potential deficiencies in the model abstraction and parameterization. Moreover, the potentially most important role the residual soil moisture parameter plays in hydrologic simulation is to affect the infiltration rate at the beginning of the storm. Finally, the residual moisture content parameter is meant to capture the fact that most natural soils do retain some small amount of moisture in pore spaces not connected to the continous pore-network, such as intra-aggregate spaces, even under intense natural drying conditions [Hillel, 1998].

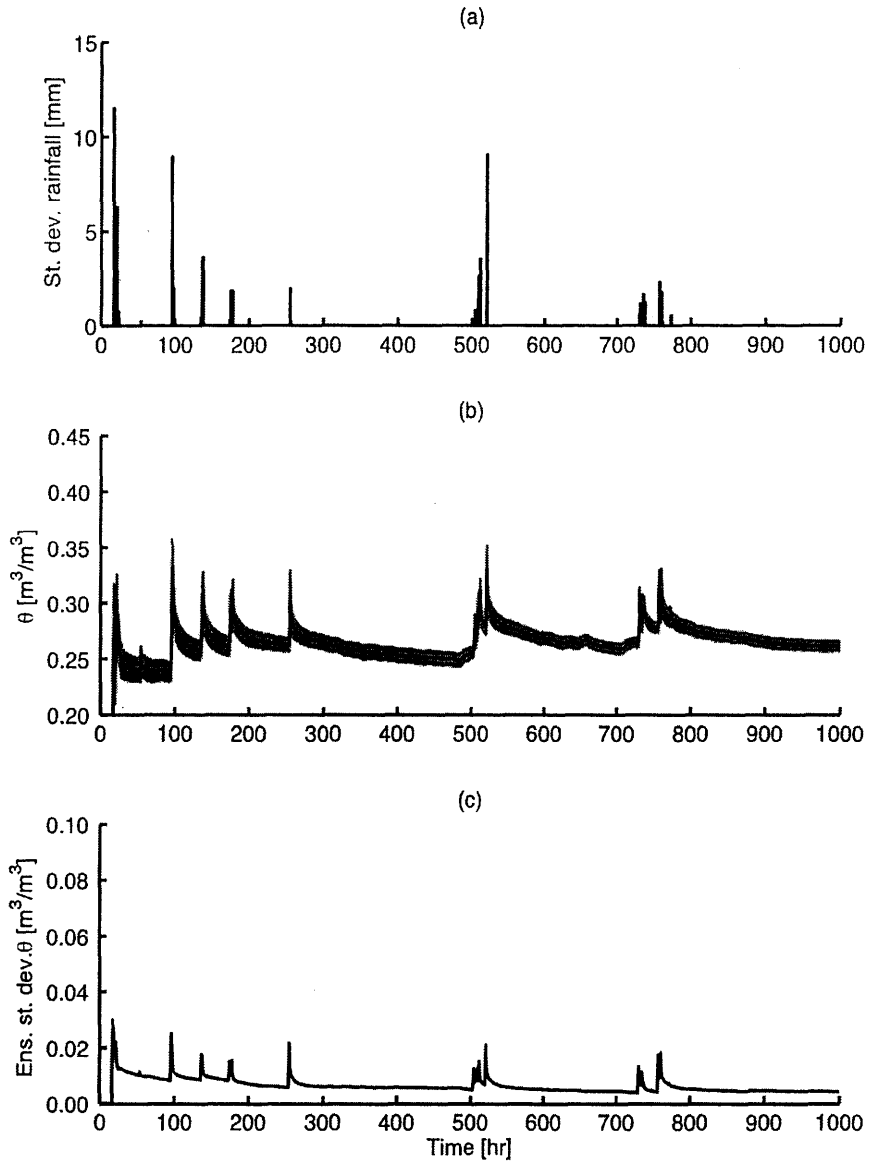


Figure 6-2: Results of a tRIBS-VEGGIE 100 member "open loop" ensemble simulation in which uncertainty in rainfall timing and duration, and SHTPs are not represented. Rainfall rates are simulated for each rainfall replicate using the multiplicative approach outlined in Chapter 5. Shown are the: (a) ensemble standard deviation of rainfall used to force the ensemble of simulations, (b) time series of ensemble mean soil moisture (black line) [m^3/m^3] and the time series of one ensemble standard deviation above and below the ensemble mean (gray area) [m^3/m^3], and (c) the time series of the ensemble standard deviation in soil moisture [m^3/m^3].

The remainder of this chapter begins with a general introduction to the treatment of parameter uncertainty. First the concept of random sampling is presented, followed by an overview of the Latin Hypercube sampling (LHS) algorithm. A simple conceptual example that generates a finite number of samples of two uncer-

tain parameters that are required as input to an unspecified model is employed to demonstrate the implications of each sampling scheme. The impact of the random and LHS techniques are then compared and contrasted through the use of two simple, but illustrative models that again require two uncertain parameters as input. The Restricted Pairing (RP) technique, which augments the benefits of the LHS approach with correlation control, is then introduced. The application of the RP technique to sampling the SHTPs required as input to tRIBS-VEGGIE is then outlined, along a meta-analysis of a widely-used SHTP database to establish marginal distributions and among-parameter correlations. A detailed analysis of implications of the RP algorithm as applied to sampling of the SHTPs on ensemble soil moisture simulation with tRIBS-VEGGIE concludes the chapter.

6.1 Random sampling

For reasons discussed in **Chapter 2** and above, Monte Carlo techniques are powerful for uncertainty analysis and data assimilation when the model being used is nonlinear, and/or of high state dimension. Under these circumstances analytical uncertainty or sensitivity analyses are often labor intensive or numerically unstable. Highly nonlinear models may also demonstrate sensitive dependence to input parameters, and depending on the application it may be desirable to identify regions of the parameter or state space that, when input to the model, have significant consequences to model outputs. Therefore, in Monte Carlo application the way in which uncertain parameters are sampled is of significant importance.

The most elementary technique for obtaining an ensemble of uncertain model inputs is to repeatedly sample each of the uncertain input variables from its appropriate marginal distribution through random number generation until some desired sample (ensemble) size is reached. Suppose that some arbitrary model, $\mathbf{y} = f(\mathbf{\Pi}, \mathbf{x})$, produces a vector of outputs $\mathbf{y}$ as a generally nonlinear function of a vector of input parameters $\mathbf{\Pi}$ and some input state vector $\mathbf{x}$. Further suppose that

the input parameter vector Π is a collection of uncertain inputs,

$$\Pi = [\Pi_1, \Pi_2, \Pi_3, \Pi_4, \ldots, \Pi_n] \tag{6.1}$$

Assume that the uncertain variables, Π_j, are independent of one another so that the marginal behavior of the Π_j are completely specified by their respective marginal distributions,

$$\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \ldots, \mathcal{D}_n \tag{6.2}$$

Let the desired ensemble size consist of m samples of each of the n uncertain parameters (i.e., an ensemble consisting of m replicate samples of the input vector Π). The ensemble of input parameter vectors is obtained by sampling sampling each of the n marginal distributions ($\mathcal{D}_j$) m times. For example, assume that the n marginal distributions of the Π_j uncertain parameters follow a gaussian distribution with mean μ_j and variance σ_j^2. The algorithm used to generate m samples of the n-dimensional input parameter vector Π can be summarized as follows:

Algorithm 1:

For $i = 1 \ldots m$

  For $j = 1 \ldots n$

    1. Generate a number u from $U[0,1]$
    2. Obtain Π_{ij} by solving $\Phi^{-1}(u, \mu_j, \sigma_j^2)$, where $\Phi(\cdot)$ is the inverse of the normal distribution with parameters μ_j and σ_j^2

  End

End

To demonstrate one of the potential implications of this most elementary pa-

rameter sampling scheme, a simple example to sample $m = 5$ pairs of uncertain model parameters, $\Pi = [U, V]$, is shown in figure 6-3. The variable U is assumed to be gaussian distributed with $\mu = 5$ and $\sigma = 2.5$ (figure 6-3(a)), while V is assumed to be uniformly distributed on [-10,10] (figure 6-3(b)). Figures 6-3(a) and (b) are illustrated in a manner to underscore the commonly-used random number generation techniques used to sample the respective parameter replicates (e.g., see Press *et al.* [1992]). Specifically, for each input parameter and sample a random deviate uniformly distributed on the interval [0,1] was generated and assumed to be a realization of the appropriate cumulative density function (cdf), which was then inverted either analytically or through some numerical approach to obtain the corresponding realization of the parameter U or V. The location of these input pairs $[U_i, V_i]$ in the joint parameter space illustrate the possibility that the paired random samples of U_i and V_i can cluster within a relatively confined portion of the parameter space (figure 6-3(c)).

Hence, in this example where the number of samples is relatively small, the sampled parameters to be input to the arbitrary model, $\mathbf{y}$, do not characterize well the full range of plausible values each respective parameter can take. In the limit of a large m, the histogram of the sampled variables U and V would reasonably well reproduce the marginal distribution from which they were drawn ($\mathcal{D}_U$ and $\mathcal{D}_V$, respectively), and the feasible region of parameters would be more broadly represented. However, the degree to which the structures of the marginal distributions are well-preserved and the feasible parameter space well-represented by a finite number of samples generally diminishes as the sample size decreases. For the input variable U in particular, at small m the sampled values are likely to cluster in the vicinity of the mode of U using the outlined random sampling approach. More generally, at small sample sizes there is no *a priori* guarantee that the feasible parameter region will be broadly represented when using random sampling as outlined here. In Monte Carlo-based uncertainty assessment and data assimilation this random sampling technique is potentially problematic when: (1) the computational costs of the model limit the sample size of uncertain inputs that can feasibly

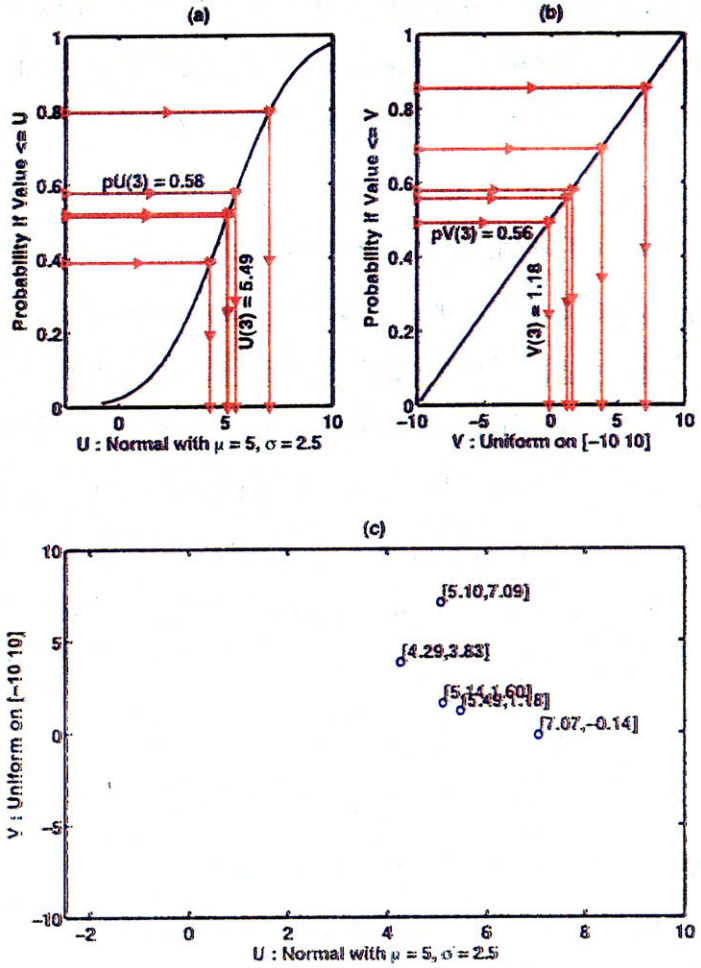


Figure 6-3: An example of random sampling applied to generation of two uncertain model input parameters. Random values corresponding to probabilities are sampled from a uniform distribution on [0,1], and then mapped to corresponding variable values through the cdf of the marginal distributions of the uncertain inputs. (a) U is sampled from a normal distribution with $\mu_U = 5$ and $\sigma_U = 2.5$, (b) V is sampled from a uniform distribution on [-10,10], and the values of U and V are paired to arrive at the $m = 5$ samples of Π.

be simulated, and/or (2) low probability realizations of model input parameters can cause model outcomes that are of high consequence. The following section details the Latin Hypercube approach to sampling uncertain model inputs, which was developed precisely to cope with these two conditions.

6.2 The Latin Hypercube sampling algorithm

Latin Hypercube sampling (LHS) is a technique to sample uncertain inputs from their respective marginal distributions in a manner that ensures that the marginal distribution of each input variable as broadly sampled as possible given a limited sample size, while maintaining some degree of randomness [Helton and Davis, 2003]. In this manner LHS was developed to ensure that in a Monte Carlo uncertainty analyses, a model would experience parameter values that are of low probability, but (depending on the nature of the model) potentially high consequence. The LHS algorithm originated in the need for reliable and consistent Monte Carlo-based uncertainty assessments of nuclear reactors, an application for which low probability model outcomes are of disproportionately high supposed consequence [McKay *et al.*, 1979]. More recently, the use of LHS has found its way to landscape ecology [Xu *et al.*, 2005] and groundwater hydrology modeling applications [Gwo *et al.*, 1996; Sohrabi *et al.*, 2002].

The thrust of the LHS algorithm is to provide a numerical framework which samples the marginal distributions of the uncertain model parameters, $\mathcal{D}_j$, as broadly as possible while maintaining some degree of randomness, subject to the constraint of sampling the $\mathcal{D}_j$ only m times. The algorithm achieves this by partitioning the cdf representing the marginal distribution of variable Π_j into m distinct and non-overlapping subspaces of equal probability mass and subsequently sampling one value of Π_j from each of the m subspaces. Again suppose that m samples of the n uncertain parameters that follow gaussian distributions parameterized by a mean μ_j and variance σ_j^2 are to be generated. The LHS algorithm provides for sampling m replicates of the n-dimensional input parameter vector $\mathbf{\Pi}$ in the following manner:

Algorithm 2:

1. Partition the range of probabilities [0,1] into m subspaces of equal intervals

For $i = 1 \ldots m$

For $j = 1 \ldots n$

2. Generate a number u from a distribution that is uniform over the interval of the ith of m subspaces
3. Obtain Π_{ij} by solving $\Phi^{-1}(u, \mu_j, \sigma_j^2)$, where $\Phi(\cdot)$ is the inverse of the normal distribution with parameters μ_j and σ_j^2

End End

4. Randomly permute the m random samples for each of the n input parameters.

To compare and contrast LHS with random sampling, consider again the example from the previous section in which $m = 5$ replicates of the uncertain parameter vector $\mathbf{\Pi} = [U, V]$ are sampled, assuming U is normally distributed with $\mu = 5$ and $\sigma = 2.5$ and V is assumed uniformly distributed on the interval [-10,10]. As illustrated in figures 6-3(a) and (b), the cdfs of U and V are partitioned into 5 distinct and non-overlapping subspaces, each with a probability mass of 1/5 = 0.2. Within each of these subspaces, a random number uniformly distributed between the lower and upper probability bounds within the subspace is generated, and the underlying values of U and V obtained by inverting the appropriate cdf (figure 6-4(a) and (b)). The 5 samples of U and V are then paired randomly to avoid systematic association of low/high probability values of U with low/high probability values V (figure 6-4(c)). The LHS scheme in effect partitions the joint parameter space into boxes from which values are drawn, its construction ensuring that at least one pair will be drawn from each of the $m = 5$ strata. Comparing figures 6-3 and 6-4, the input vectors $\mathbf{\Pi} = [U, V]$ sampled using the LHS technique are drawn more broadly from

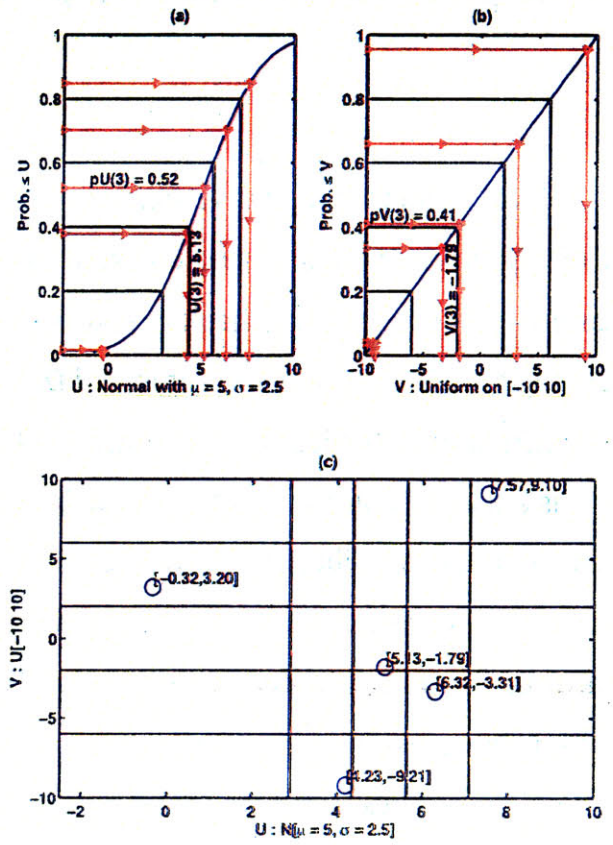


Figure 6-4: A simple example using LHS to generate uncertain model inputs: (a) U is sampled from a normal distribution with $\mu = 5$ and $\sigma = 2.5$, (b) V is sampled from a uniform distribution on [-10,10], and (c) the values of U and V are paired randomly to arrive at the $m = 5$ samples of x.

the two marginal distributions, $\mathcal{D}_U$ and $\mathcal{D}_V$. As a consequence, any sensitivities of the model outputs y to less frequently occurring parameter values are more likely to be captured when the input parameter vectors Π are generated using the LHS technique when the sample size m is small. Moreover, as will be demonstrated in the following section, input parameter vectors constructed using LHS lead to more reproducible probabilistic behavior in the model outputs when limitations on the sample size exist.

6.3 Example: A comparison of LHS and Simple Random Sampling

In this section the LHS and random sampling techniques are compared in terms of the reproducibility of cdfs from outputs of two different models. In this comparison the two models are each nonlinear mathematical relationships that contain thresholds and were used in a similar analysis by Helton and Davis [2003]. The first model is a monotonic function and requires two parameters $\Pi = [U, V]$ as input, while the second model is a non-monotonic function of the input parameter vector Π. For this example, the marginal distributions for U and V are the same for both model 1 and model 2. U is assumed to be uniformly distributed on the interval [1.0, 1.5], while V is assumed uniformly distributed on [0.0, 1.0]. The impacts of sample size, m, are also investigated in an attempt to diagnose the limitations on each sampling technique. Sample sizes considered for both models are $m = 25$, 50, and 100. To compare each sampling technique on the basis of consistency in the probabilistic behavior of the model outputs ($f_1(U, V)$ and $f_2(U, V)$), both sampling technique were repeated ten times for each model and the cdfs of the model outputs estimated using the Weibull plotting procedure.

6.3.1 Model 1: Monotonic with thresholds

The first model used is a monotonic function in $\Pi = [U, V]$, with thresholds. The model is nonlinear in the input parameters, with quadratic and cross-product terms. The formula for this mode is given as:

$$f_1(U, V) = U + V + UV + U^2 + V^2 + U \min[\exp(3V), 10] \tag{6.3}$$

Again, U is assumed to be uniformly distributed on [1.0, 1.5], while V is assumed uniformly distributed on [0.0, 1.0]. Realizations of the uncertain inputs $\Pi = [U, V]$ are generated using the LHS algorithm for $m = 25$, 50 and 100. The

m realizations of Π are then input to model 1 and the cdf of the output computed using the Weibull plotting formula. This process is similarly repeated, instead using the random sampling approach to generate the m realizations of Π. This process of generating m realizations of Π, inputting them to model 1, and estimating the cdf is repeated ten for both sampling techniques.

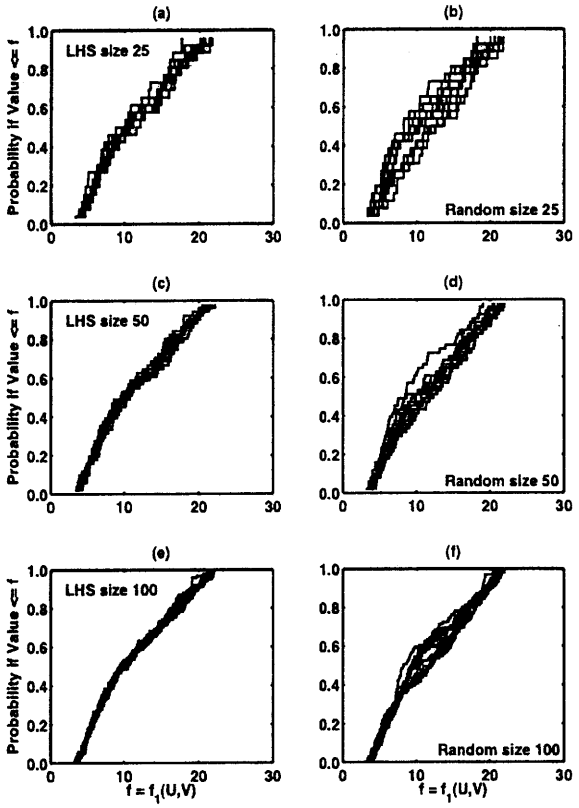


Figure 6-5: Comparison of cdfs computed with output of the monotonic model: (a) LHS-derived parameters with $m = 25$, (b) random sampling derived parameters with $m = 25$, (c) LHS-derived parameters with $m = 50$, (d) random sampling derived parameters with $m = 50$, (e) LHS-derived parameters with $m = 100$, and (f) random sampling derived parameters with $m = 100$.

The LHS-derived parameters clearly lead to more reproducible cdfs from the model output for $m = 25$ and 50 (figures 6-5(a)-(d)). However, the contrast in cdf reproducibility between the LHS and random sampling approaches diminishes as m increases further (figure 6-5(e),(f)). For $m = 25$, generating parameters using random sampling leads to discrepancies of up to 0.3 in the cdf estimated from the

model output, even using this relatively simple monotonic model with thresholds (figure 6-5(b). The corresponding discrepancy in the estimated value of the cdf in the case of LHS-derived parametersis substantially smaller (figure 6-5(a)). Furthermore, even when $m = 100$ there remains substantial dispersion in the cdf computed from the model output when the parameters were sampled using the random sampling scheme (figure 6-5(f)). By contrast, at $m = 100$ LHS-generated parameters lead to significantly more consistent cdfs, even clarifying some of the more subtle structures in the cdf hidden at smaller m (i.e., the break in slope at approximately $f_1(U,V) = 16$) (figure 6-5(e)). Hence, even in circumstances in which computational burden does not necessarily limit the size of m, the LHS may provide some benefit when it is necessary to finely resolve the probabilistic behavior of the model predictions. However, the more noticeable benefit of the LHS algorithm is that the cdf of the model output is more reproducible at smaller m when uncertain input parameters are generated using LHS.

6.3.2 Model 2: Non-monotonic with thresholds

The second model used is non-monotonic in $\mathbf{\Pi} = [U,V]$, and also contains thresholds. Having both quadratic and cross-product terms, this model is also non-linear in model inputs. The marginal distributions and analytical techniques outlined for model 1 above, are also used for model 2. Model 2 is formulated in the following manner:

$$f_2(U,V) = U + V + UV + U^2 + V^2 + Ug(V) \tag{6.4}$$

where,

$$\begin{aligned} h(V) &= (V - 11/43)^{-1} + (V - 22/43)^{-1} + (V - 33/43)^{-1} \\ g(V) &= h(V) \text{ if } |h(V)| < 10 \\ g(V) &= 10 \text{ if } h(V) \geq 10 \end{aligned}$$

$$g(V) \quad = -10 \text{ if } h(V) \leq -10. \tag{6.5}$$

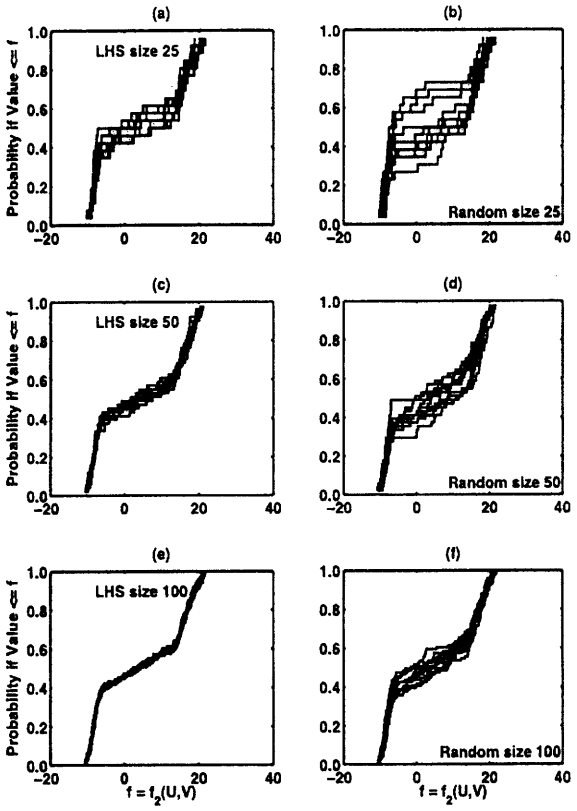


Figure 6-6: Comparison of cdfs computed with output of the non-monotonic model: (a) LHS-derived parameters with $m = 25$, (b) random sampling derived parameters with $m = 25$, (c) LHS-derived parameters with $m = 50$, (d) random sampling derived parameters with $m = 50$, (e) LHS-derived parameters with $m = 100$, and (f) random sampling derived parameters with $m = 100$.

As was the case with the monotonic model, in the case of the non-monotonic model the LHS approach leads to reproducible model cdf, even at $m = 25$ (figures 6-6(a),(b)). Apparent at $m = 100$, the non-monotonicity and non-linearity of model 2 leads to three regimes in the structure of its cdf (figures 6-6(e),(f)). Non-exceedance probability rises relatively quickly at low values of f_2, and then rises less quickly at moderate f_2. At approximately $f_2 = 15$, the non-exceedance probability again increases rapidly with increasing f_2. For all m, the cdfs estimated from the model outputs derived from LHS-generated input parameters capture bet-

ter the three regimes in the cdf of f_2 (figures 6-6(a),(c),(e)). For $m = 25$, the cdfs of model outputs derived from parameters generated using random sampling exhibit differences of between 0.3 and 0.5 over a broad range of values of $f_2(U, V)$ (figure 6-6(b)). In contrast, the corresponding difference in non-exceedance probability in cdfs estimated from outputs of model 2 produced with LHS-derived inputs is comparatively smaller (figure 6-6(a)). Substantial variation in the cdfs of model 2 output exist even at $m = 50$ when the inputs are derived from the random sampling approach (figure 6-6(d)). Compared with $m = 50$, at $m = 100$ variation in cdfs estimated from model outputs is substantially reduced when the inputs are generated using random sampling (figure 6-6(f)). However, this reduction in cdf variation at high m still does not lead to the level of reproducibility achieved when model inputs are generated using LHS (figure 6-6(e)). Similar to the conclusion for model 1, analysis using this slightly more complex non-monotonic, non-linear model with thresholds reveals that LHS-derived parameters lead to more reproducible cdfs in model output, which are capable of illuminating distinct regimes in the cdf of the model output, at much smaller m than parameters generated through random sampling.

6.4 Control correlation among uncertain model inputs

There may be physical reasons to suspect or empirical data to suggest that uncertain parameters required as input to a model are correlated, or are otherwise not independent. As is frequently the case the n-dimensional joint probability distribution exactly characterizing the joint behavior of the n input parameters may not be known or accessible. Although fully characterizing the joint behavior between the n input parameters may be infeasible, it may be nevertheless useful to impose correlation among the m samples of the n input parameters.

Suppose the m samples of the n uncertain input parameters Π_{ij} are collected into an $m \times n$ matrix, $\mathbf{\Pi}$,

$$\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \dots & \Pi_{1n} \\ \Pi_{21} & \Pi_{22} & \dots & \Pi_{2n} \\ \vdots & \vdots & & \vdots \\ \Pi_{m1} & \Pi_{m2} & \dots & \Pi_{mn} \end{bmatrix} \tag{6.6}$$

The goal is to sample the Π_j in a way that the correlation structure of the matrix Π approximates an $n \times n$ correlation matrix C, which is known *a priori*. The restricted pairing (RP) method, first proposed by Iman and Conover [1982], provides a straightforward way to sample uncertain input parameters from their respective marginal distributions using LHS, while imposing a desired correlation structure on the ensemble of realizations of the uncertain inputs, Π. The underlying objective of the RP algorithm is to "rearrange" the elements of Π within each column so that its correlation structure approximates C. The elements of the matrix C represent correlation coefficients between any two of the input parameters Π_j,

$$\mathrm{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix} \tag{6.7}$$

The two most common ways of expressing correlation between two variables are through the Pearson correlation coefficient and the Spearman rank correlation coefficient. Both the Pearson correlation coefficient and the Spearman rank correlation coefficient take values between -1 and 1. For any two variables, denoted x_1 and x_2, the Pearson correlation coefficient represents the strength of a linear relationship between x_1 and x_2. A negative Pearson correlation coefficient implies an inversely linear relationship between the variables. Alternatively, the Spearman rank correlation coefficient measures the strength of a monotonic relationship between the variables x_1 and x_2. Similarly to the case for the Pearson correlation coefficient, a negative value of the Spearman rank correlation coefficient indicates an inversely

monotonic relationship between x_1 and x_2. Mathematically, the Pearson correlation coefficient between the variables x_1 and x_2 can be computed as,

$$r_{x_1x_2} = \frac{\sum_{i=1}^{m}(x_{i1} - \bar{x_1})(x_{i2} - \bar{x_2})}{\left[\sum_{i=1}^{m}(x_{i1} - \bar{x_1})^2\right]^{1/2}\left[\sum_{i=1}^{m}(x_{i2} - \bar{x_2})^2\right]^{1/2}}, \tag{6.8}$$

where,

$$\begin{aligned} \bar{x_1} &= \frac{1}{m}\sum_{i=1}^{m} x_{i1}, \text{ and,} \\ \bar{x_2} &= \frac{1}{m}\sum_{i=1}^{m} x_{i2}. \end{aligned} \tag{6.9}$$

In a similar fashion, the Spearman rank correlation coefficient between the variables x_1 and x_2 can be computed as follows,

$$R_{x_1x_2} = \frac{\sum_{i=1}^{m}(R(x_{i1}) - \bar{R}(x_1))(R(x_{i2}) - \bar{R}(x_2))}{\left[\sum_{i=1}^{m}(R(x_{i1}) - \bar{R}(x_1))^2\right]^{1/2}\left[\sum_{i=1}^{m}(R(x_{i2}) - \bar{R}(x_2))^2\right]^{1/2}}, \tag{6.10}$$

where the operator $R(\cdot)$ denotes the rank transformation of the argument, taking values from $1 \ldots m$ in ascending order, and $\bar{R}(x_1) = \bar{R}(x_2) = (m+1)/2$.

The RP algorithm as presented below represents a mathematical framework to perform the rearrangement of the elements of Π, to achieve a correlation structure that approximates the *a priori* known Spearman correlation matrix C. The algorithm begins by constructing the $m \times n$ matrix, V whose columns contain the m van der Waerden scores,

$$\mathbf{V} = \begin{bmatrix} \Phi^{-1}(1/(m+1)) & \Phi^{-1}(1/(m+1)) & \dots & \Phi^{-1}(1/(m+1)) \\ \Phi^{-1}(2/(m+1)) & \Phi^{-1}(2/(m+1)) & \dots & \Phi^{-1}(2/(m+1)) \\ \vdots & \vdots & & \vdots \\ \Phi^{-1}(m/(m+1)) & \Phi^{-1}(m/(m+1)) & \dots & \Phi^{-1}(m/(m+1)) \end{bmatrix} \tag{6.11}$$

where $\Phi^{-1}(\cdot)$ is the inverse operator of the standard normal distribution. The elements of the n identical columns of $\mathbf{V}$ are standard normal variates with probability $i/(m+1)$ for $i = 1, 2, 3, \dots, m$. The algorithm then proceeds as follows:

1. By column, randomly permute the elements of $\mathbf{V}$ to arrive at the matrix $\mathbf{V}^*$

2. Using the above equation, compute the Spearman rank correlation matrix imposed by the random permutation in step 1, $\mathbf{R}_{VV}$.

3. Compute the $n \times n$ lower triangular matrix $\mathbf{Q}$ as the Cholesky factorization of the square symmetric, positive definite matrix $\mathbf{R}_{VV}$. The Cholesky factorization solves $\mathbf{R}_{VV} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}}$.

4. Compute the $n \times n$ lower triangular matrix $\mathbf{P}$, through the Cholesky factorization of the known matrix $\mathbf{C}$, such that $\mathbf{C} = \mathbf{P}\mathbf{P}^{\mathrm{T}}$.

5. Compute the $n \times n$ matrix $\mathbf{S} = \mathbf{P}\mathbf{Q}^{-1}$, where $\mathbf{Q}^{-1}$ is the inverse of matrix $\mathbf{Q}$. The matrix $\mathbf{S}$ is a linear operator that has the effect of removing the correlation structure of $\mathbf{K}_{VV}$ and replacing it with that of $\mathbf{C}$.

6. Now construct the $m \times n$ matrix $\mathbf{S}^* = \mathbf{V}\mathbf{S}^{\mathrm{T}}$. The matrix $\mathbf{S}^*$ now has approximately the desired Spearman rank correlation structure of $\mathbf{C}$.

7. Transform the standard normal variates contained in $\mathbf{S}^*$ back into normalized ranks (i.e., variables ranging in values from $1/(m+1)$ to $m/(m+1)$) by operating on each element with the forward operator of the standard normal distribution, $\Phi(\cdot)$. Denote this matrix $\mathbf{V}^{**}$.

8. By column, invert the appropriate marginal distributions with the elements of $\mathbf{V}^{**}$ as the arguments to arrive at the $m \times n$ matrix $\mathbf{\Pi}^{*}$. The matrix $\mathbf{\Pi}^{*}$ contains the m-dimensional ensemble of realizations of the n uncertain inputs to the model and has a Spearman rank correlation structure that is approximately C. That is, the m rows of $\mathbf{\Pi}^{*}$ are realizations of the inputs to the model, that in an ensemble-sense are described by the rank correlation structure in C.

While both random sampling, LHS, and RP assume that the marginal distributions $\mathcal{D}_j$ are known, the RP algorithm additionally requires information about the among parameter correlation, in the form of the matrix C. The Spearman correlation matrix C is ideally based on a statistical analysis of databases containing information about the parameters. More generally, the algorithm will produce correlated parameters sampled using LHS for any valid (i.e., square, symmetric, and positive definite) correlation matrix C. The example below will demonstrate that care must be exercised in establishing the marginal distributions of the Π_j and the correlation matrix C, because the parameter marginal distributions ($\mathcal{D}_j$) together with the imposed correlation (C) impact the joint behavior between the parameters.

6.4.1 Example: Implementation of the RP algorithm

As discussed above, the restricted pairing algorithm first suggested by Iman and Conover [1982] provides a means of inducing a desired rank correlation among samples of uncertain input parameters. In this example, 1000 realizations of two uncertain parameters of an arbitrary model are generated. These two parameters will be denoted as X_1 and X_2 in this example. The variable X_1 is assumed to be normally distributed with zero mean and unit variance and X_2 is assumed to be uniform on the interval [-4,4]. Using the RP algorithm as enumerated above, the RP algorithm employs the Spearman rank correlation matrix to project a random permutation of the inverse van der Waerden scores (which themselves are standard

normal variates) into a space with the desired correlation structure. The elements of the resulting matrix are then mapped to the standard normal cdf, and through inverting the appropriate marginal distributions of X_1 and X_2, a matrix of $m = 1000$ samples of the $n = 2$ inputs with the desired correlation structure is obtained. For illustrative purposes, six off-diagonal Spearman rank correlation coefficients, $C_{X_1X_2} = 0.00$, 0.25, 0.50, 0.90, and 0.99, are assumed.

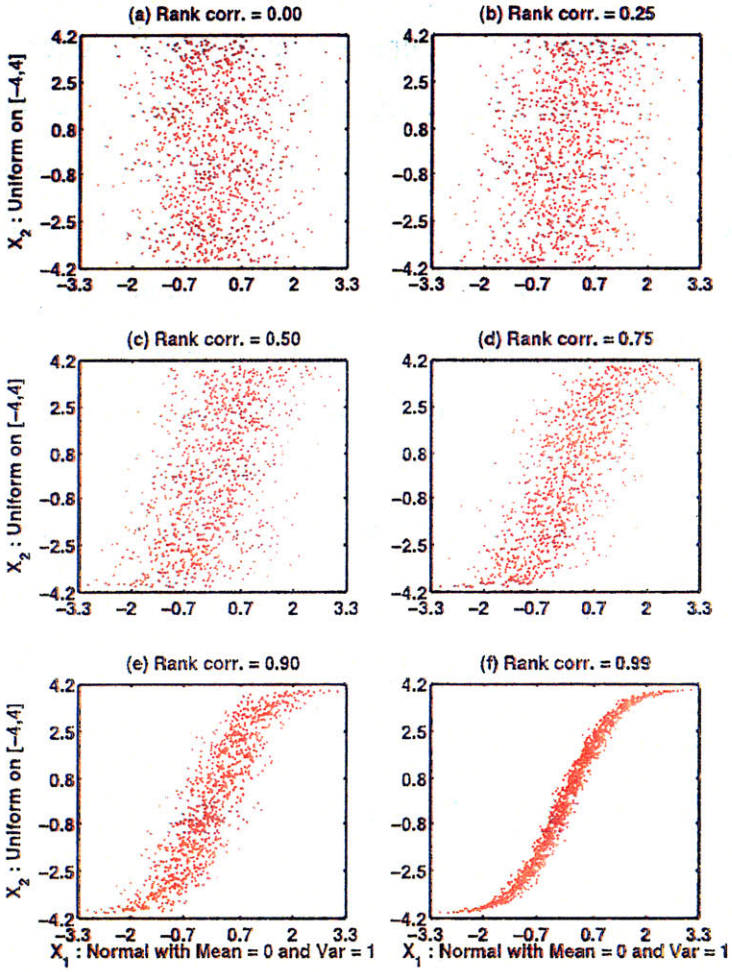


Figure 6-7: One thousand random draws of two random variables, X_1 which is assumed normal with zero mean and unit variance and X_2 which is assumed uniform on [-4,4]. The restricted pairing algorithm is used to impose a Spearman correlation coefficient between X_1 and X_2 of (a) 0.00, (b) 0.25, (c) 0.50, (d) 0.75, (e) 0.90, and (f) 0.99.

At small values of $C_{X_1X_2}$, little correspondence between X_1 and X_2 is seen (fig-

ure 6-7(a)). As $C_{X_1X_2}$ increases, the monotonic nature of the correlation between X_1 and X_2 becomes more apparent (figures 6-7(b), (c), and (d)). At high values of $C_{X_1X_2}$ (e.g., $C_{X_1X_2} = 0.90$ and 0.99), a non-linear relationship between between X_1 and X_2 is evident, whereas at lower values of $C_{X_1X_2}$ the use of a linear relationship to describe the tendency of the behavior between X_1 and X_2 is not entirely implausible. The nature of the relationship between X_1 and X_2 at high $C_{X_1X_2}$ is imposed entirely by the assumed marginal distributions of X_1 and X_2.

An important implication of this simple example is that caution must be exercised when selecting the n marginal distributions of the input parameters to ensure that when possible, the assumed marginal distributions and correlation matrices reproduce the structure of observed tendencies between individual pairs of parameters. Stated differently, because of the use of the Spearman correlation matrix in the RP algorithm, the algorithm itself only imposes monotonic correlation rather than a more restrictive form of correlation (e.g., linear as with Pearson's correlation matrix). Hence, while the RP algorithm ensures monotonic relationships at given correlation levels between input parameters, poor selection of the marginal distributions of input parameters may result in misrepresentation of the nature of the monotonicity between any pair of input parameters.

6.5 Applying the RP algorithm to ensemble-based soil moisture estimation

LHS and RP have been shown to be useful tools to generate ensembles of uncertain input parameters under constraints of limited sample size. In the context of ensemble soil moisture data assimilation, limitations on sample size arise from the numerical complexity of the tRIBS-VEGGIE model and the computational burden associated with Monte Carlo integration of tRIBS-VEGGIE. Moreover, there is some physical reason to suspect that infrequently occurring SHTPs may have important consequences for the dynamics of soil moisture. For example, loamy soils

with very low saturated hydraulic conductivity may be encountered infrequently but are of high consequence to soil moisture and runoff generation, because they easily saturate, where they occur. Furthermore, process ecohydrology models such as tRIBS-VEGGIE require parameters that ideally represent physical attributes of the soil column or are directly measurable. There is also some physical justification to believe that SHTPs may be correlated to each other. For example, residual moisture content is always less than saturation moisture content and total porosity. Hence, the benefits of the RP and LHS algorithms outlined above make these techniques attractive as a means of generating SHTPs for tRIBS-VEGGIE ensemble simulations. The added difficulty of implementing the RP algorithm to generate SHTP inputs for tRIBS-VEGGIE lies in the need to characterize the marginal behavior of the input parameters and establish their correlation structure.

This section details the application of the RP methodology to generate the SHTPs required as input to tRIBS-VEGGIE. It is important to note here that once a single realizations of the parameters required for simulation with tRIBS-VEGGIE are supplied to the model, they remain static during the entire simulation. However, the SHTPs vary across the ensemble and each ensemble replicate has a different realization of the required SHTPs. It is further assumed throughout that the only a categorical soil texture (e.g., loam, sand, etc.) is reliably known about the system being modeled. What follows is a brief description of the SHTPs required by the tRIBS-VEGGIE model. Following this description, efforts to define appropriate marginal distributions and estimate Spearman correlation matrices for 9 categorical soil types through a statistical analysis of a widely used soil meta-database is discussed. After discussion of soil database meta-analysis, a series of numerical experiments using tRIBS-VEGGIE are undertaken to investigate the sensitivity of the statistical behavior of soil moisture ensembles to both the technique used to sample the SHTPs (i.e., random sampling versus RP) and ensemble size. Specifically, similarly to the previous examples, the reproducibility of ensemble soil moisture statistics is studied as the ensemble size as well as the SHTP generation technique changes.

6.5.1 Summary of soil parameters required as input to tRIBS-VEGGIE

The process mechanisms represented in tRIBS-VEGGIE that depend on SHTPs are briefly described here: infiltration of precipitation in variably saturated soils, ground heat flux, and bare soil latent and sensible heat fluxes. Infiltration of water into the soil is modeled using a one-dimensional Richards equation for a sloped surface that allows for lateral gravitational drainage. The lower boundary condition of the model is a flux boundary condition, consistent with the assumption of significant depth to the saturated zone in the semiarid environment for which the model is currently most applicable. Moisture in the finite element soil column can vary between the input residual volumetric moisture content (m^3/m^3), θ_R, and the volumetric moisture content at saturation (m^3/m^3), θ_S. tRIBS-VEGGIE uses the Brooks-Corey model [Brooks and Corey, 1964] to characterize the relationship between volumetric moisture content, θ, hydraulic conductivity, $K(\theta)$, and soil matric potential, $\psi(\theta)$. The Brooks-Corey parameterization requires specification of a hydraulic conductivity at saturation [cm/hr], K_S, the pore distribution index parameter (dimensionless), λ_p, and the air entry pressure [mm], ψ_b.

Ground heat flux in the tRIBS-VEGGIE model is calculated through the method outlined by Wang and Bras [1999], which is based on a numerical solution to the one-dimensional heat diffusion equation with constant heat diffusivity. The solution to the heat diffusion equation proposed by Wang and Bras [1999] is based on the recent history of soil surface temperatures, and requires specification of the volumetric thermal conductivity and heat capacity of the soil. Both the thermal conductivity and heat capacity depend on the moisture state at the time of calculation, and therefore require specification of soil-specific thermal parameters as input. Computation of the soil heat capacity is moisture-dependent linear combination of the input heat capacity of the soil solid materials, $C_{s,\text{solids}}$ ($Jm^{-3}K^{-1}$), θ_S, the heat capacity of liquid water, and the moisture state in the near-surface. Moisture-dependent calculations of thermal conductivity in tRIBS-VEGGIE are based on the method suggested by Farouki [1981] and require specification of the thermal conductivity of

the dry soil ($Jm^{-1}s^{-1}K^{-1}$), $k_{s,\text{dry}}$, and saturated soil ($Jm^{-1}s^{-1}K^{-1}$), $k_{s,\text{sat}}$.

Latent and sensible heat fluxes from the bare soil in tRIBS-VEGGIE are computed through a resistance formulation, in which independent resistances to latent and sensible heat flux are calculated. The gradient between air temperature and soil skin temperature drives sensible heat flux, while the gradient between atmospheric humidity and air humidity in the near-surface pore space drives latent heat flux. Humidity in the pore spaces near the soil surface, in turn, depends on the soil skin temperature. In this formulation, the latent heat flux depends on the soil matric potential and moisture state in the near-surface, and on the input parameters θ_S and θ_R. It should also be noted, that sensible and latent heat fluxes also indirectly depend on the soil thermal properties outlined above due because each flux component depends on the soil skin temperature.

The SHTPs required as input to the tRIBS-VEGGIE model and considered as uncertain in the present study are summarized in Table 6.1.

Table 6.1: Soil hydraulic and thermal properties required by tRIBS-VEGGIE

Symbol	Description
K_S	Hydraulic conductivity at saturation [mm/hr]
θ_R	Residual moisture content [m^3/m^3]
θ_S	Saturation moisture content [m^3/m^3]
λ_p	Brooks-Corey pore distribution index [dimensionless]
ψ_b	Brooks-Corey air entry pressure [mm]
$k_{s,\text{dry}}$	Thermal conductivity of dry soil [J/(msK)]
$k_{s,\text{sat}}$	Thermal conductivity of saturated soil [J/(msK)]
$C_{s,\text{solids}}$	Volumetric heat capacity of soil solids [$J/(m^3K)$]

6.5.2 Meta-analysis of a soil database

The soils data used in this study are a meta-database from 3 soil surveys [Rawls and Brakensiek, 1985; Ahuja *et al.*, 1989; Leij *et al.*, 1996]. These data have previously been analyzed by Schaap and Leij [1998], and underlie the ROSETTA software issued by the U.S. Department of Agricultures Salinity Laboratory. This meta-

database contains 2134 analyzed records, 1309 of which posses a measurement of saturated hydraulic conductivity (K_S). The parameters measured for each record are summarized in Table 6.2. Note that the meta-database used by Schaap and Leij [1998] present parameter values for the van Genuchten-Mualem soil water retention curve, whereas the tRIBS-VEGGIE model requires Brooks-Corey parameters. Rawls *et al.* [1993], however, provides equations expressing equivalence between the parameters of the van Genuchten and Brooks-Corey soil water retention functions. Furthermore, the meta-database used by Schaap and Leij [1998] does not contain values of $C_{s,\text{solids}}$, $k_{s,\text{sat}}$, and $k_{s,\text{dry}}$. Therefore, $C_{s,\text{solids}}$ and $k_{s,\text{dry}}$ are estimated given the sand and clay fraction in the meta-database using an empirical relationship presented Bonan [1996], and $k_{s,\text{sat}}$ is computed as a function of $k_{s,\text{dry}}$ and θ_S using an empirical relationship presented by Peters-Lidard *et al.* [1998]. While $C_{s,\text{solids}}$ and $k_{s,\text{dry}}$ are fit to beta distributions, they are perfectly anticorrelated with one another. Because $k_{s,\text{sat}}$ depends on $k_{s,\text{dry}}$ and θ_S, no distribution is fit.

Table 6.2: Parameters in the metadatabase of Schaap and Leij [1998]

Parameter	Description
% Clay	Percent clay by mass [percent]
% Sand	Percent sand by mass [percent]
% Silt	Percent silt by mass [percent]
ρ_b	Bulk density - not used [g/cm^3]
K_S	Hydraulic conductivity at saturation [mm/hr]
θ_R	Residual moisture content [m^3/m^3]
θ_S	Saturation moisture content [m^3/m^3]
α	van Genuchten fitting parameter [dimensionless]
n_p	van Genuchten fitting parameter [dimensionless]

Each record was assigned to a soil textural class based on its reported sand, silt and clay fraction. The number of records within each textural class ranges from 3 (silt) to 334 (sandy loam) and is shown in Table 6.3. Of the 1309 records with K_S data, 9 textural classes are represented by at least 20 records and are set apart for further analysis. In our analysis of these 9 textural classes we assume that the

within-class ranges of parameter values and correlation structure characterize the ensemble behavior of each textural class. In light of this assumption, within each of the 9 selected textural classes we fit marginal distributions to each parameter and computed the Spearman correlation matrices, as required to generate uncertain replicates of the SHTPs necessary to simulate soil moisture with the tRIBS-VEGGIE model.

Table 6.3: Summary of soil textures in database

Soil texture	Number of records in database	Number of records with K_S data
Clay	94	63
Sandy clay	10	8
Silty clay	29	14
Sandy clay loam	181	135
Silty clay loam	92	42
Clay loam	142	56
Sandy loam	514	334
Loam	252	119
Silt loam	327	135
Sand	342	277
Loamy sand	141	123
Silt	10	3
Total	**2134**	**1309**

We assumed that the log-transformed hydraulic conductivity data is normally distributed based on previous studies of hydraulic conductivity distributions. Further, the residual moisture content (θ_R) data exhibited a significant number of records possessed θ_R equals zero. We treated the marginal distribution of θ_R as a mixed discrete-continuous distribution, with an atom of probability at 0 with mass equal to the empirical frequency of occurrence of $\theta_R = 0$, and a two-parameter beta distribution for non-zero values of θ_R. Initial candidate distributions for the remainder of the parameters (θ_S, λ_p, ψ_b, $k_{s,dry}$, $k_{s,sat}$, $C_{s,solids}$) were the gamma, two parameter beta, and exponential distributions. The chosen distribution for each parameter was based both on the significance of computed Kolmogorov-Smirnoff (KS) goodness-of-fit statistics and visible comparison between the fit marginal dis-

tributions and empirical histograms of each parameter. According to these criteria, the two-parameter beta distribution was chosen to represent the uncertainty in the remaining parameters. In the context of the soil properties considered here, the beta distribution is particularly advantageous because it is defined over a finite interval and can therefore constrain parameters to realistic values. The method of moments was used to estimate the beta distribution parameters a and b. Parameter estimates of the fit marginal distributions values for each soil property in the 9 textural classes of the database are given in **Appendix B**. Computed Spearman rank correlation matrices for each of the 9 considered textural classes are shown in **Appendix B**. Figure 6-8 depicts an example of fit marginal distributions plotted together with the empirical histograms of Brooks-Corey air entry pressure by soil textural class.

6.5.3 Ensemble behavior of the sampling method

To isolate the impact of SHTP uncertainty, the simulations presented here were conducted at the spatial scale of a single element (e.g., point). We also seek to assess the influence of SHTP uncertainty on soil moisture ensemble behavior independently of vegetation effects. Thus, unvegetated conditions are assumed. To contrast the two SHTP sampling techniques in the context of ensemble soil moisture modeling, we present the results of a single experiment in figure 6-9. Using each SHTP sampling algorithm independently, 100 combinations of soil parameter inputs to tRIBS-VEGGIE were generated. The near-surface soil moisture (top 10 cm) response during a 1000 hr period was then simulated, assuming initial soil moisture conditions corresponding to 10% effective saturation (defined as $S_e = [\theta - \theta_R]/[\theta_S - \theta_R] = 0.10$). Both ensembles were subjected to the same hydrometeoroloigcal forcings and the rainfall (figure 6-9(a)). The ensemble mean soil moisture response appears virtually identical for the two SHTP sampling approaches (figure 6-9(b),(c)). However, several soil moisture replicates evolved with random sampling-generated soil parameters appear to be physically unlikely (fig-

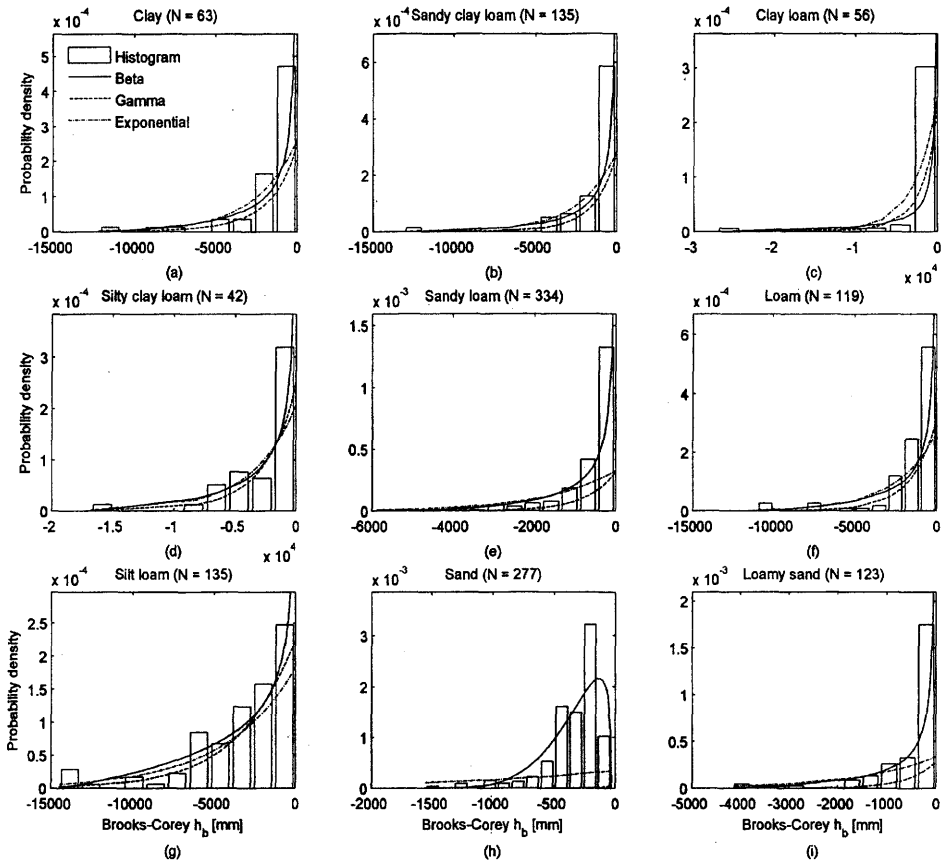


Figure 6-8: Histograms of Brooks-Corey air entry pressure and fit trial distributions for (a) clay, (b) sandy clay loam, (c) clay loam, (d) silty clay loam, (e) sandy loam, (f) loam, (g) silt loam, (h) sand, and (i) loamy sand.

ure 6-9(c)). Specifically, a few replicates seem to saturate after the first rainfall event and remain saturated throughout the rest of the simulation. Furthermore, some replicates evolved with random sampling-generated soils demonstrate large increases in near-surface soil moisture during interstorm periods, an implausibly large sensitivity to evaporative forcing (figure 6-9(c)).

The ensemble standard deviation in soil moisture, a measure of soil moisture uncertainty, generally responds similarly in time (figure 6-9(d)). However, ensemble standard deviation is higher for the experiment in which soil parameters were generated using the random sampling approach. This is because the random sampling algorithm generates soil parameter combinations that lead to physically implausible soil moisture behavior in the context of the tRIBS-VEGGIE model. The overarching

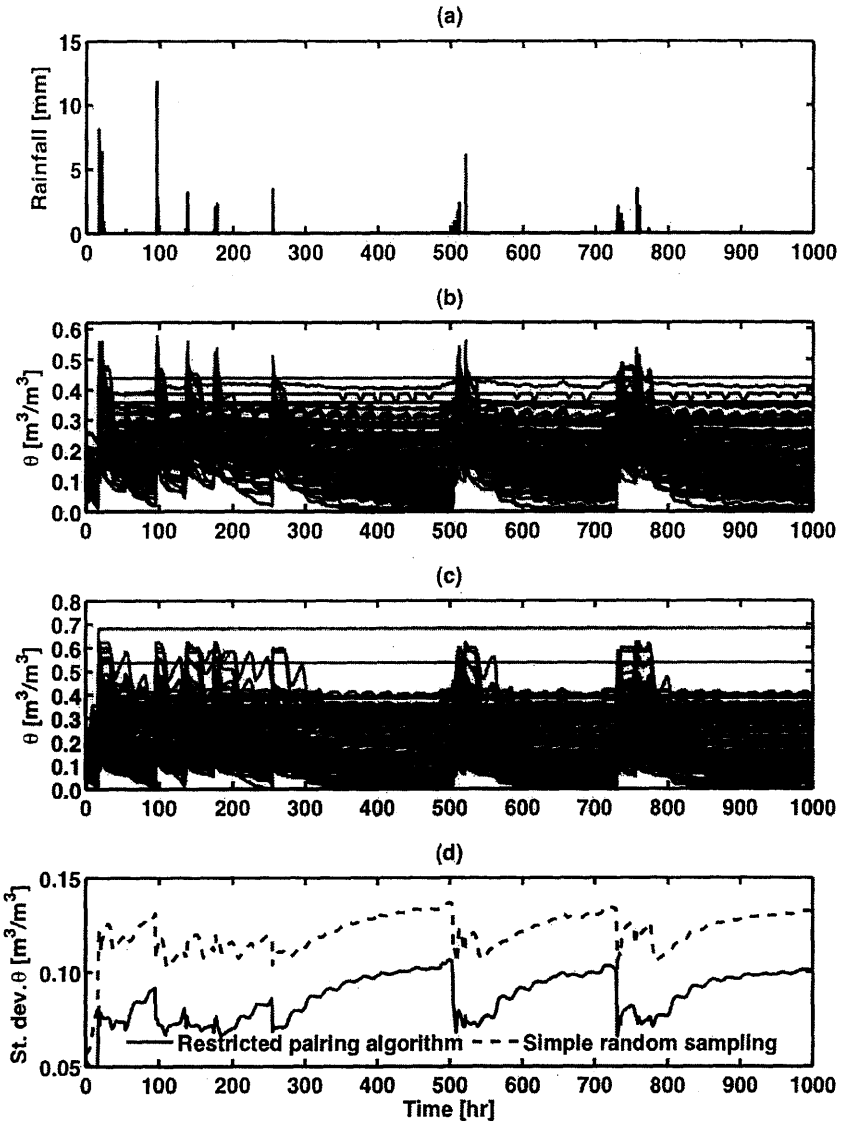


Figure 6-9: The time evolution of: (a) rainfall used to drive the tRIBS-VEGGIE model, (b) soil moisture in the top 10 cm [m^3/m^3] for the simulations in which SHTPs were generated using the RP technique, (c) soil moisture in the top 10 cm [m^3/m^3] for the simulations in which SHTPs were generated using the random sampling technique, and (b) the standard deviation in soil moisture [m^3/m^3]. In (b) and (c), gray lines depict individual ensemble replicates while the black line depicts the ensemble mean. In (d) the gray dotted line shows ensemble in which SHTPs were generated using random sampling while the black solid line indicate the ensemble in which they were generated using the RP technique.

objective of this portion of the work is to investigate the degree to which the ensemble estimate of mean and variance in near-surface soil moisture vary depending on: (1) the technique by which SHTP uncertainty is represented and, (2) the number of samples or size of the ensemble. To this end, adequate assessment of the sensitivity of each sampling technique to ensemble size requires producing sufficiently many ensemble first- and second-order statistics, across a range of ensemble sizes, to quantify estimator variances. We vary the ensemble size, K, geometrically from 24 to 210 (i.e., from 16 to 1024 replicates). In this set of simulations we consider three soil textures: loam, sandy loam, and clay. For each ensemble size (K) and soil textural class we generate 20 independent ensemble parameter combinations, each consisting of K SHTP combinations, using both the RP and random sampling techniques. Each replicate in all of the simulations is subjected to the same hydrometeorological forcings for a period of 1000 hrs, and the soil moisture state is not constrained to observations at any point during the simulation (i.e., soil moisture ensemble simulations are open loop). The rainfall time series used to drive the model is the same time series depicted in figure 6-9(a). Initial soil moisture conditions again correspond to 10% effective saturation. This set of simulations yields 20 time-evolving ensemble estimates of mean and variance in near-surface soil moisture for each ensemble size, K.

Figure 6-10 depicts the time evolution of 20 estimates of ensemble mean (figure 6-10(a)) and ensemble standard deviation (figure 6-10(b)) soil moisture, each computed over 64 replicates[2]. For this particular ensemble size, ensemble mean soil moisture is estimated consistently using either RP or random sampling to generate SHTPs input to the model (figure 6-10(a)). However, the estimate of ensemble standard deviation in soil moisture varies when random sampling is used to sample SHTPs, to a greater extent than the estimate of ensemble standard deviation in soil moisture when RP is used, for this ensemble size (figure 6-10(a)). This suggests that, for this particular ensemble size (64), the estimate of ensemble soil moisture variance is sensitive to the particular combination of soil parameters sam-

[2]Similar plots for the other ensemble sizes considered are provided in **Appendix B**

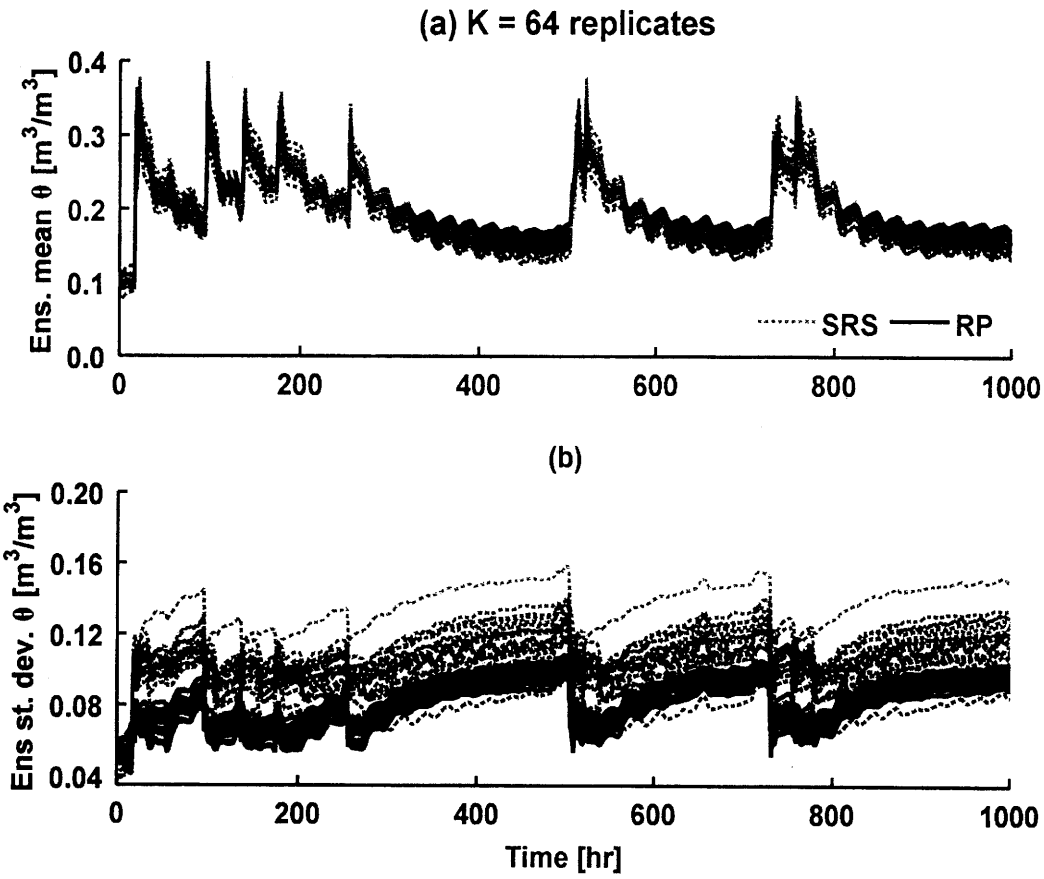


Figure 6-10: The time evolution of: (a) soil moisture in the top 10 cm [m^3/m^3], and (b) the standard deviation in soil moisture [m^3/m^3]. Gray dashed lines show ensembles in which SHTPs were generated using random sampling while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

pled when random sampling is used to generate the soil parameters required by the model. In the context of ensemble data assimilation, the goal is to achieve the minimum number of SHTP samples (ensemble replicates) for which the ensemble statistics, particularly mean and variance, are independent of the actual parameter values used. We are therefore interested in the consistency with which ensemble soil moisture mean and variance are estimated as the ensemble size changes.

For each ensemble size, a sample variance in the ensemble estimates of mean and variance in soil moisture is computed from the 20 independent ensembles. Specifically, the relevant statistics are,

$$\hat{s}^2\left(\hat{\mu}_\theta(t)\right) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\mu}_{\theta i}\left(t\right) - \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_{\theta i}\left(t\right)\right)^2, \tag{6.12}$$

and,

$$\hat{s}^2\left(\hat{\sigma}_\theta(t)\right) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\sigma}_{\theta i}\left(t\right) - \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{\theta i}\left(t\right)\right)^2, \tag{6.13}$$

where $N = 20$ ensembles, and $\theta_{ij}(t)$ is the soil moisture for replicate k ($1 \leq k \leq K$) of ensemble i ($1 \leq i \leq N = 20$) at time t. In equation 6.12 $\hat{\mu}_{\theta i}\left(t\right)$ is the estimate of the mean soil moisture for ensemble i at time t, and is defined as,

$$\hat{\mu}_\theta\left(t\right) = \frac{1}{K}\sum_{k=1}^{K}\theta_k\left(t\right). \tag{6.14}$$

Similarly, in equation 6.13 $\hat{\sigma}^2_{\theta i}\left(t\right)$ is the estimate of the standard deviation in soil moisture for ensemble i at time t, and is defined as,

$$\hat{\sigma}^2_{\theta i}\left(t\right) = \frac{1}{K}\sum_{k=1}^{K}\left(\theta_k\left(t\right) - \hat{\mu}_{\theta i}\left(t\right)\right)^2. \tag{6.15}$$

In theory both $s^2\left(\hat{\mu}_\theta(t)\right)$ and $s^2\left(\hat{\sigma}_\theta(t)\right)$ should be zero for infinitely large ensembles, regardless of how the uncertainty in the parameters is represented. That is, if the multi-dimensional joint distribution of SHTPs was represented in perfect detail, then there would be no variance in the estimate of any soil moisture statistic. At small ensemble sizes, however, $s^2\left(\hat{\mu}_\theta(t)\right)$ and $s^2\left(\hat{\sigma}_\theta(t)\right)$ can both be appreciably different from zero. For given hydrometeorological forcings, the rate at which $s^2\left(\hat{\mu}_\theta(t)\right)$ and $s^2\left(\hat{\sigma}_\theta(t)\right)$ decrease as ensemble size increases can highlight tradeoffs between computational burden due to increased ensemble size and the associated decrease in the variance of the ensemble estimates of mean and variance in soil moisture. Comparing the values of $s^2\left(\hat{\mu}_\theta(t)\right)$ and $s^2\left(\hat{\sigma}_\theta(t)\right)$ for the RP and random sampling approaches to representing SHTP uncertainty at a given ensemble size

and during different times in the wetting-drying cycle can demonstrate the potential benefits of more careful treatment of parameter uncertainty under ensemble size constraints.

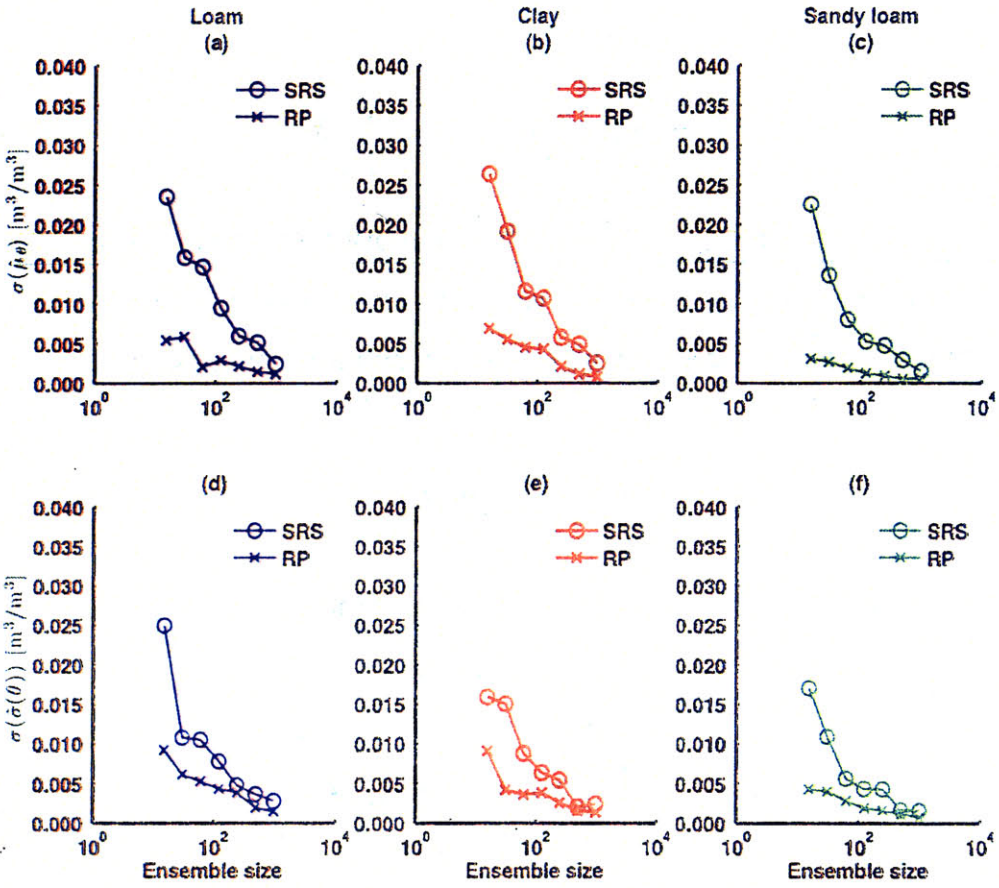


Figure 6-11: At 750 hours (just after cessation of rainfall) the top row of plots shows the standard deviation in the ensemble mean soil moisture estimate across 20 ensembles as a function of ensemble size for (a) loam, (b) clay, and (c) sandy loam soils. The bottom row shows the standard deviation in the ensemble estimate of standard deviation in soil moisture across 20 ensembles as a function of ensemble size for (d) loam, (e) clay, and (f) sandy loam soils.

The behavior of $s^2\left(\hat{\mu}_\theta(t)\right)$ versus ensemble size (figures 6-11(a)-(c)) and $s^2\left(\hat{\sigma}_\theta(t)\right)$ (figures 6-11(d)-(f)) versus ensemble size are shown at $t = 771$ hr into the simulation. This point in the simulation corresponds to a significant rain event. For all soil textures, using RP to generate soil parameters input to tRIBS-VEGGIE yields a lower value of $s^2\left(\hat{\mu}_\theta(t)\right)$ (figures 6-11(a)-(c)) at all ensemble sizes. When comparing the techniques, the difference in $s^2\left(\hat{\mu}_\theta(t)\right)$ is greatest at the smallest ensemble size, and

is relatively insignificant at $K = 1028$. Although $s^2(\hat{\mu}_\theta(t))$ is relatively low at the minimum ensemble size (16) when RP is used to sample SHTPs, it decreases only modestly as ensemble size increases. Similarly, using RP to generate soil parameters input to tRIBS-VEGGIE yields a lower value of $s^2(\hat{\sigma}_\theta(t))$ at all ensemble sizes for all soil textures (figures 6-11(d)-(f)). The difference in $s^2(\hat{\sigma}_\theta(t))$, when comparing the two techniques, is again greatest at the smallest ensemble size, but is small by ensemble sizes of $K = 512$ for clay and sandy loam soils, and $K = 256$ for loam soils. When using random sampling to generate soil parameters, doubling or quadrupling the ensemble size from the minimum 16 yields much more consistency in the estimate of ensemble mean and variance in soil moisture.

Similar results during a significant dry spell in the rainfall record ($t = 1000$ hr) are shown in figure 6-12. Conclusions are largely the same, however, note that at $K = 16$ for clay soils using random sampling-generated soil parameters actually results in a lower value of $s^2(\hat{\sigma}_\theta(t))$ when compared to using RP-generated parameters (figures 6-12(d)-(f)). This may be an issue of sampling error as only 20 independent ensembles were used to compute $s^2(\hat{\mu}_\theta(t))$ and $s^2(\hat{\sigma}_\theta(t))$.

To conclude this section, the results indicate that when computational resources limit the size of the soil moisture ensemble, using a sampling technique that: (1) samples low probability but potentially high consequence combinations of soil parameters, and (2) imposes correlation known or believed to exist among those parameters can potentially result in more consistent estimation of ensemble mean and variance in soil moisture. This conclusion is of importance to the hillslope-scale soil moisture data assimilation addressed in this thesis, and more broadly Monte Carlo simulation with models like tRIBS-VEGGIE. It demonstrates that careful attention to the way in which uncertainty in the SHTPs required as input to tRIBS-VEGGIE is represented can lead to a significant reduction of computational burden. This reduction in computational costs is realized because when generating realizations of the uncertain SHTPs the first- and second-order ensemble statistics (i.e., those used in the Kalman update step) can be consistently reproduced at lower ensemble sizes

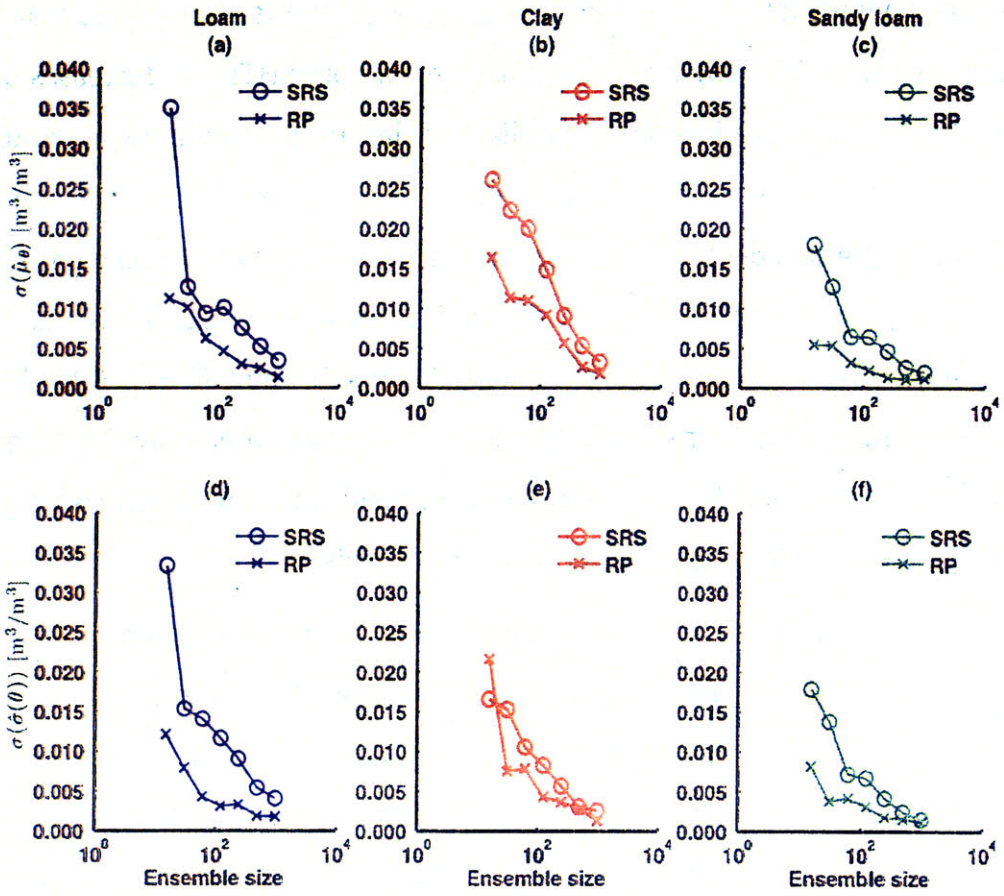


Figure 6-12: At 100 hours (during a significant dry-down) the top row of plots shows the standard deviation in the ensemble mean soil moisture estimate across 20 ensembles as a function of ensemble size for (a) loam, (b) clay, and (c) sandy loam soils. The bottom row shows the standard deviation in the ensemble estimate of standard deviation in soil moisture across 20 ensembles as a function of ensemble size for (d) loam, (e) clay, and (f) sandy loam soils.

when SHTPs are generated using RP, compared with when they are generated using random sampling. Further implications of the findings presented in this chapter are discussed in the concluding chapter of the thesis.

CHAPTER 7

SENSITIVITY ANALYSIS OF SOIL MOISTURE

An important component of this thesis seeks to understand the relative impact of the sources of uncertainty in soil moisture prediction, particularly uncertainty arising from imperfect measurement or forecast of hydrometeorological processes and inadequate knowledge about soil hydraulic and thermal properties (SHTPs), on soil moisture predictability. Understanding the degree to which uncertainty in the forcings and parameters of a model independently propagates to the state and observable quantities related to the state informs the economic construction of data assimilation systems. Specifically, it may possible to decrease computational burden by neglecting uncertainty in particular forcings or parameters to which predictibility of soil moisture is fairly insensitive. Because microwave observations are sensitive to moisture only in the few centimeters near the surface, the impact of uncertainty in hydrometeorological and parameter uncertainty on near-surface soil moisture is of particular interest in this thesis. This chapter details a series of numerical experiments designed constitute a sensitivity analysis to quantitatively assess the degree to which uncertainty in hydrometeorological forcings and SHTPs propagate to the soil moisture state.

Through analysis of a suite of point-scale Monte Carlo experiments, the relative impact of uncertainties in precipitation forcings, thermodynamic forcings (sky cloud cover fraction, air and dew temperature, solar radiation, windspeed), and

SHTPs on soil moisture uncertainty are investigated. Each Monte Carlo experiment corresponds to a particular and potentially realistic scenario of assumptions regarding treatment of uncertainty in the precipitation forcings, thermodynamic forcings, and SHTPs. As such, these point-scale experiments are designed to test multiple hypotheses regarding the predominant sources of uncertainty in local soil moisture predictability. From this series of point-scale Monte Carlo experiments, conclusions about the relative impacts of the targeted sources of uncertainty on the temporal evolution of uncertainty in soil moisture are drawn, which are used to set the treatment of uncertainty in the remainder of the thesis. Based on these conclusions, a Monte Carlo experiment is performed to assess the spatial distribution of uncertainty in soil moisture at particular instances in time during the simulations in a semiarid to arid experimental watershed.

This chapter is organized first to focus on the point-scale exercises to determine the impacts of the targeted sources of uncertainty on soil moisture predictability, then on the spatially distributed assessment of soil moisture uncertainty.

7.1 Point-scale soil moisture temporal sensitivity

The goal of the point-scale sensitivity analysis is to diagnose which of the potentially uncertain model inputs (the thermodynamic forcings, precipitation forcings, and SHTPs) most significantly influences uncertainty in soil moisture. A suite of scenarios is designed to assess the nature of and degree to which uncertainty in each of the identified model inputs impacts soil moisture uncertainty. These scenarios represent realistic assumptions about sources of input uncertainty in which uncertainties in thermodynamic forcings, precipitation, and SHTPs are considered both together and independently of one another. Comparing the soil moisture response to the assumed conditions for each scenario provides insight into the model inputs to which soil moisture is most sensitive, which is instructive in construction of the data assimilation system. For each scenario, this approach requires an

ensemble of hydrometeorological forcings in which the thermodynamic forcings are consistent with the precipitation forcings (e.g., precipitation is associated with cloudiness, which, in turn, impacts incoming solar radiation and air temperature). Hence, the stochastic weather generator discussed in **Chapter 5** is used to derive thermodynamic forcings based on realizations of point-scale rainfall simulated from the Bartlett-Lewis rainfall simulator also introduced in **Chapter 5**. Uncertainty in SHTPs is represented using the Restricted Pairing approach outlined in **Chapter 6**, and throughout the point-scale sensitivity analysis the soil texture is assumed to be a loam. Below is a detailed discussion about the prevailing assumptions regarding the sources of model input uncertainty and the corresponding soil moisture response for the scenarios considered.

7.1.1 Scenario 1: Deterministic hydrometeorological forcings, deterministic SHTPs

A 30-day realization of hourly hydrometeorological forcings is simulated using the MBL rainfall model and the stochastic weather generator [Ivanov *et al.*, 2007], with MBL and weather generator parameters corresponding to those derived for Tucson International Airport, Arizona, USA by Hawk [1992] and Ivanov *et al.* [2007], respectively. These forcings are chosen to represent the synthetic observations of hydrometeorological variables at a hypothetical weather station in Walnut Gulch Experimental Watershed (WGEW), Arizona, USA in July. A single set of SHTPs, assuming a loam soil, is chosen randomly from an ensemble of loam SHTP sets generated using the Restricted Pairing (RP) methodology discussed in **Chapter 6** to describe the soils at the hypothetical weather station. This realization of the hydrometeorological forcings and SHTPs (referred to here as the "observed" forcings and parameters) are supplied as input to the tRIBS-VEGGIE model to establish an "observed" soil moisture response against which an ensemble of soil moisture responses, simulated under the varying combinations of forcing and parameter uncertainty that characterize the scenarios below, can be compared. The initial soil

moisture was assumed to be 10% effective saturation (S_e).

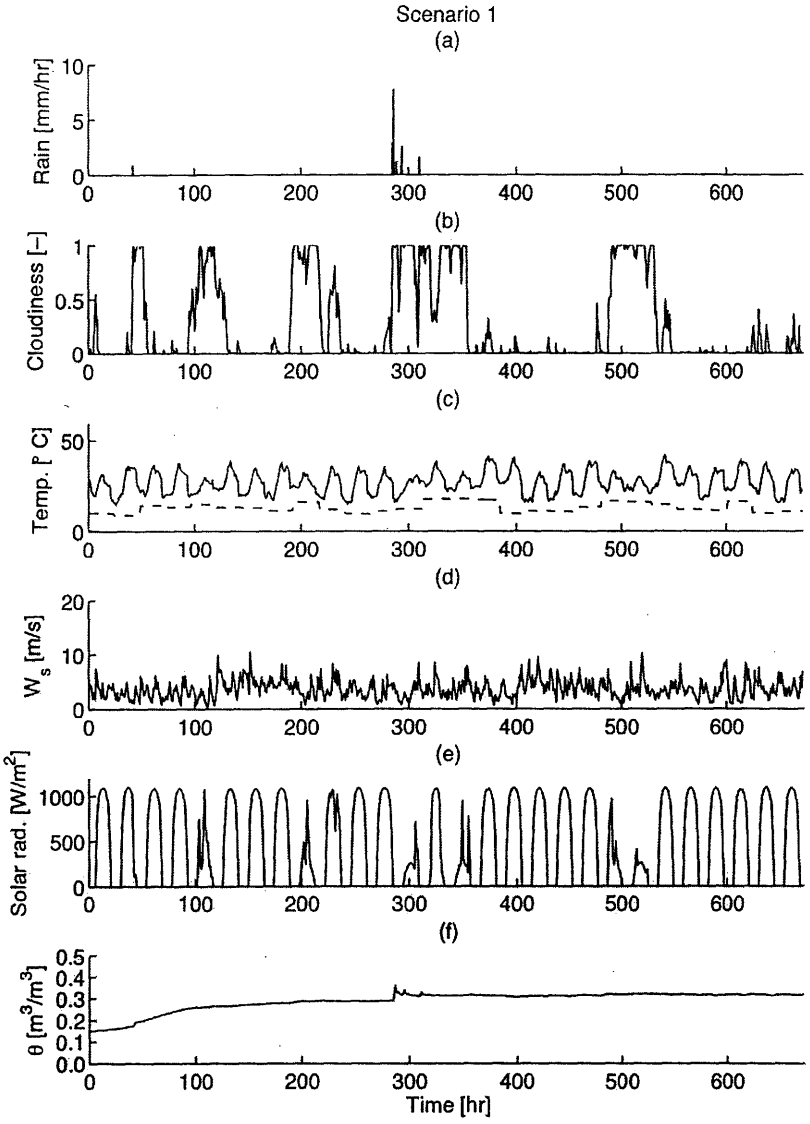


Figure 7-1: Simulated observed (a) rainfall [mm/hr], (b) sky cover fraction [dimensionless], (c) air (solid lines) and dew (dashed lines) temperature [°C], (d) wind speed [m/s], (e) incoming solar radiation [W/m^2], and (f) soil moisture response [m^3/m^3] during the 30 day simulation

Figure 7-1 depicts the observed hydrometeorological forcings and corresponding soil moisture response in the top 10 cm of the soil column as simulated by tRIBS-VEGGIE. In the simulated forcings there are essentially two periods of rainfall. The first is a relatively small event with an intensity slightly less than 2 mm/hr approx-

imately 45 hours after the beginning of the simulation (figure 7-1(a)). The second storm event occurs at approximately 300 hours into the simulation, with rainfall rates varying from approximately 2 mm/hr to 4 mm/hr during an approximately 40 hour period of time, with some intermittency (figure 7-1(a)). Interestingly, in addition to the two rainfall events, during which there is persistent cloudiness, there are two periods of fair weather overcast conditions from approximately 100 to 120 hours and again from approximately 180 to 230 hrs, as well and periods of fair weather overcast from approximately 490 and 550 hours into the simulation (figure 7-1(b)). As would be expected, periods of overcast conditions are frequently associated with depressed air temperatures (figure 7-1(c)), higher relative humidity as seen by the distance between the air and dew temperatures (figure 7-1(c)), and decreased incoming solar radiation (figure 7-1(d)). Based on the observed SHTPs, the soil moisture in the top 10 cm primarily responds to the rainfall events (figure 7-1(e)). Soil moisture in the near surface increases during the first approximately 100 hrs of the simulation, reflecting the anomalously high (negative) bubbling head and low pore distribution index that characterize the soil properties drawn as the observation.

7.1.2 Scenario 2: Uncertain rainfall, deterministic thermodynamic forcings, uncertain SHTPs

In this scenario, uncertainty in SHTPs and the hourly rainfall volumes observed during the 30 day time series are assumed to predominate as sources of uncertainty in the soil moisture predicted by tRIBS-VEGGIE. To represent uncertainty in the rainfall forcings, an ensemble of rainfall time series (64 replicates) was obtained by repeatedly and independently perturbing the observed rainfall forcings generated in **Scenario 1** using the multiplicative cascade model described in **Chapter 5**. Uncertainty in SHTPs was represented using the Restricted Pairing (RP) technique described in **Chapter 6** to generate an ensemble of soil parameters (64 replicates) required as input to tRIBS-VEGGIE, assuming a loamy soil. Thermodynamic forcings

in this experiment correspond to the observed thermodynamic forcings described in **Scenario 1**, and are treated as invariant across the ensemble. Note that the outlined treatment of rainfall uncertainty implicitly assumes that the timing and duration of rainfall observed at the hypothetical gauge is accurate. Hence, across the ensemble of soil moisture responses to the applied rainfall forcings, it is expected that the temporal occurrence of wetting and drying events will be consistent while the magnitude of the response will be governed by: (1) the amount of rainfall input to the soil column, and (2) the rate at which the soil (given antecedent moisture conditions) can infiltrate or evaporate water given the observed atmospheric conditions.

The ensemble mean near-surface soil moisture response, as estimated across an ensemble of tRIBS-VEGGIE simulations, gradually increases to a value of approximately 0.15 m^3/m^3 after the beginning of the simulation, achieving an equilibrium after approximately 100 hrs (figure 7-2(f)). Ensemble mean near-surface soil moisture responds to a cluster of rainfall events at approximately 280 hrs and decays gradually back to an equilibrium by approximately 400 hrs into the simulation (figure 7-2(f)). The uncertainty in the predicted soil moisture response, as measured by the ensemble standard deviation in near-surface soil moisture, is also shown as a shaded band encompassing one ensemble standard deviation above and below the ensemble mean in figure 7-2(f). Ensemble standard deviation in near-surface soil moisture gradually increases to about 0.10 m^3/m^3 after the beginning of the simulations, and basically remains constant after approximately 120 hrs (figure 7-2(f)). The occurrence of rainfall at approximately 280 hrs diminishes ensemble standard deviation in soil moisture slightly, but increases gradually as the soil dries out (figure 7-2(f)). Note that the diurnal nature of the atmospheric forcings are evident as mild undulations in the ensemble mean near-surface soil moisture, as well as the region encompassing the ensemble mean plus and minus one ensemble standard deviation (figure 7-2(f)).

The results presented in figure 7-2 suggest that uncertainty in SHTPs is the predominant source of uncertainty in the soil moisture predicted by tRIBS-VEGGIE in

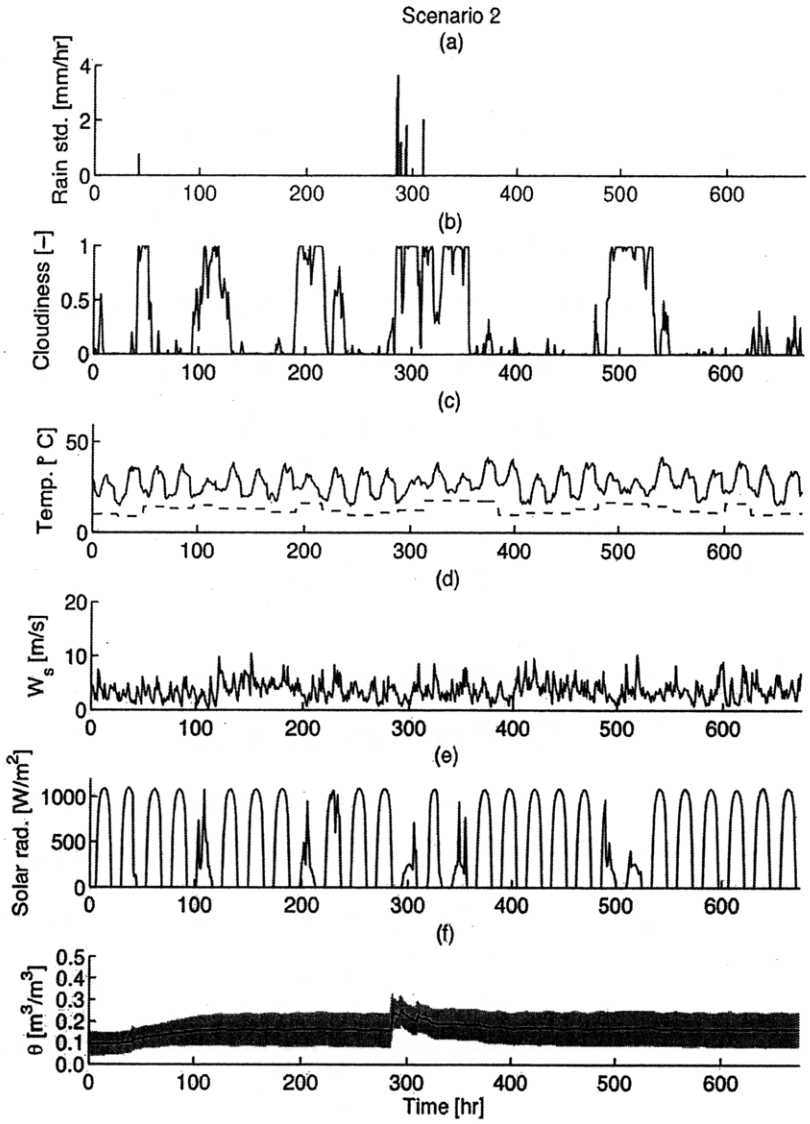


Figure 7-2: Simulated (a) ensemble standard deviation of rainfall [mm/hr], (b) observed sky cover fraction [dimensionless], (c) observed air (solid lines) and dew (dashed lines) temperature [°C], (d) observed wind speed [m/s], (e) observed incoming solar radiation [W/m^2], and (f) ensemble mean (black line) and range of one standard deviation in each direction outside the ensemble mean (green area) near-surface soil moisture response [m^3/m^3] during the 30 day simulation.

response to the uncertain forcings and parameters supplied as input. This is evidenced by the fact that the uncertainty in soil moisture associated with the initial conditions (of 10% effective saturation) is nearly as large as the uncertainty at the end of the simulation and at the end of the drying cycle which ends with initiation

of rainfall at approximately 280 hrs (figure 7-2(f)). This uncertainty in predicted soil moisture response in the presence of uncertain SHTPs after significant time has elapsed since the end of rainfall can be understood as an imperfect knowledge of the residual moisture content of the assumed loamy soil. This uncertainty in residual moisture content also impacts the initial response of the soil to rainfall as well as bare soil evaporation, since the uncertain residual moisture content to which the soil relaxes during extended drying periods directly impacts the hydraulic conductivity (whose parameters K_S, λ_p, and ψ_b are also uncertain).

7.1.3 Scenario 3: Uncertain hydrometeorological forcings, uncertain SHTPs

Scenario 3 treats rainfall and thermodynamic forcings, as well as the SHTPs required as input to tRIBS-VEGGIE, as uncertain. The rainfall ensemble developed for **Scenario 2** and described above is used as the rainfall forcing in this experiment. Uncertain thermodynamic forcings are obtained by supplying each of the 64 rainfall replicates from **Scenario 2** as input to the stochastic weather generator, leading to an ensemble of thermodynamic forcings comprised of 64 replicates. SHTPs input to tRIBS-VEGGIE in this experiment are those generated for an assumed loamy soil in **Scenario 2.**

Figure 7-3 depicts the generated hydrometeorological forcings and ensemble soil moisture response simulated by tRIBS-VEGGIE. Again, there are two periods of time associated with rainfall, a relatively short event with a relatively small variance in rainfall rate occurring approximately 48 hrs into the simulation, and a longer-lasting sequence of rain events with more significant variance in hourly rainfall rates that occurs beginning at approximately 280 hrs into the simulation (figure 7-3(a)). While rainfall occurs the ensemble mean sky cover fraction is unity (figure 7-3(b)). In the ensemble mean sense, fair weather skies are mostly clear, although sky cover fraction through the ensemble of thermodynamic forcings simulated by the stochastic weather generator range from completely clear to overcast conditions at all times during the simulation (figure 7-3(b)). The ensemble behavior of the sky

cover fraction process is reflected in the corresponding time series of simulated incoming solar radiation, which shows fairly broad spread about the ensemble mean incoming solar radiation (figure 7-3(e)). The temporal occurrence of rainfall is the only mechanism that appreciably alters the ensemble mean solar radiation during the simulation period (figure 7-3(e)). The simulated ensemble mean air temperature mostly follows expected climatological trends, decreasing somewhat when rainfall occurs (figure 7-3(c)). Uncertainty in air temperature is presented in the light gray area in figure 7-3(e), which encompasses the area within one ensemble standard deviation above and below the ensemble mean hourly air temperature, closely follows the diurnal temperature cycle. Uncertainty in dew temperature is presented in the dark gray area in figure 7-3(e), encompassing the area within one ensemble standard deviation above and below the ensemble mean hourly dew temperature[1], is fairly narrowly distributed about the ensemble mean dew temperature. The ensemble mean wind speed (figure 7-3(d)) is essentially constant at slightly less than 5 m/s, while the ensemble standard deviation in wind speed is approximately 2.5 m/s (figure 7-3(d)). The stochastic weather generator of Ivanov *et al.* [2007] treats the wind speed as a first-order autoregressive process with a random initial condition, which is reflected in the lack of temporal structure in the behavior of the simulated ensemble wind speed.

Interestingly, the ensemble soil moisture response simulated by tRIBS-VEGGIE for this scenario of hydrometeorological forcings and SHTPs is virtually indistinguishable from the ensemble soil moisture response simulated for **Scenario 2**. This suggests that any additional uncertainty in the soil moisture response arising due to the addition of uncertainty in the thermodynamic forcings is negligible when compared to uncertainty in soil moisture arising from imperfect quantification of the SHTPs required by tRIBS-VEGGIE.

[1]The stochastic weather generator of Ivanov *et al.* [2007] simulates only a daily dew temperature

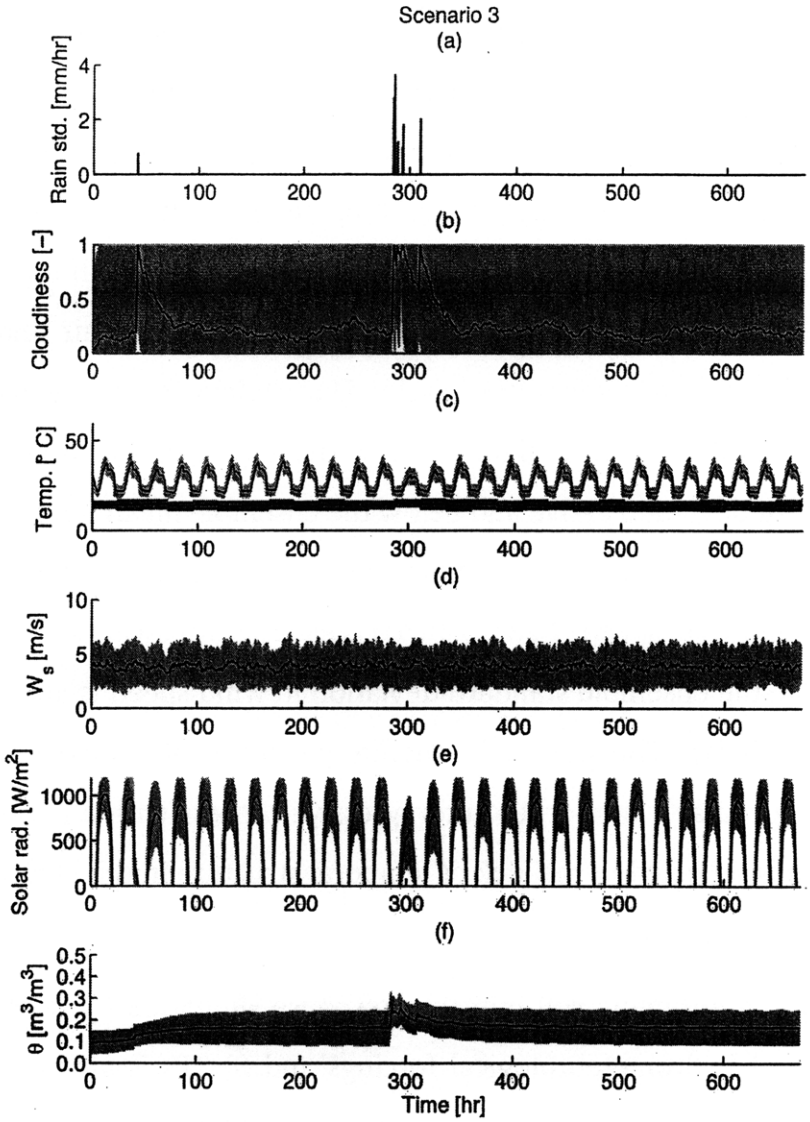


Figure 7-3: Simulated (a) ensemble standard deviation of rainfall [mm/hr], (b) ensemble mean sky cover fraction [dimensionless] (black) and range (light gray area), (c) air and dew temperature [°C], (d) wind speed [m/s], (e) incoming solar radiation [W/m^2], and (f) near-surface soil moisture response [m^3/m^3]. In (c)-(f), ensemble means are in black and areas encompassing one ensemble standard deviation above and below the mean in light gray. In (c), mean dew temperature [°C] is the black line dark gray shaded area encompassing one standard deviation above and below the mean.

7.1.4 Scenario 4: Deterministic rainfall, uncertain thermodynamic forcings, uncertain SHTPs

This scenario attempts to isolate the impact of uncertainty in thermodynamic forcing independently of uncertainty in the rainfall forcing. Therefore, the rainfall forcing used in this scenario is treated as deterministic and corresponds to the observed hourly rainfall time series generated in **Scenario 1**. The uncertain thermodynamic forcings generated with the stochastic weather generator of Ivanov *et al.* [2007] for **Scenario 3** are used again in this ensemble scenario to represent the uncertainty in the thermodynamic forcings input to tRIBS-VEGGIE. The uncertain SHTPs used in this scenario correspond to the ensemble of soil parameters generated with the RP methodology outlined in **Chapter 6** and used to parameterize an assumed loam soil in **Scenario 2** and **Scenario 3**.

The rainfall used to drive the tRIBS-VEGGIE model in this scenario is shown in figure 7-4(a), and is identical to the rainfall forcing shown in figure 7-1(a). The uncertain thermodynamic forcings supplied as input to tRIBS-VEGGIE in this scenario are depicted in figures 7-4(b)-(e), and are identical to the corresponding thermodynamic forcings used in **Scenario 3** shown in figures 7-3(b)-(e). Again, the simulated near-surface soil moisture response to the assumed conditions for this scenario is remarkably similar to the results shown for **Scenarios 2 and 3** (figure 7-4(f)). This again underscores the conclusion that imperfect knowledge of the SHTPs required to parameterize the soil column most greatly impacts the uncertainty in the simulated near-surface soil moisture response.

7.1.5 Scenario 5: Uncertain rainfall timing, uncertain thermodynamic forcings, uncertain SHTPs

This scenario investigates the degree to which the uncertainty in the simulated near-surface soil moisture response to uncertain hydrometeorological forcings is sensitive to a prevailing assumption in **Scenario 2** through **Scenario 4**: that the

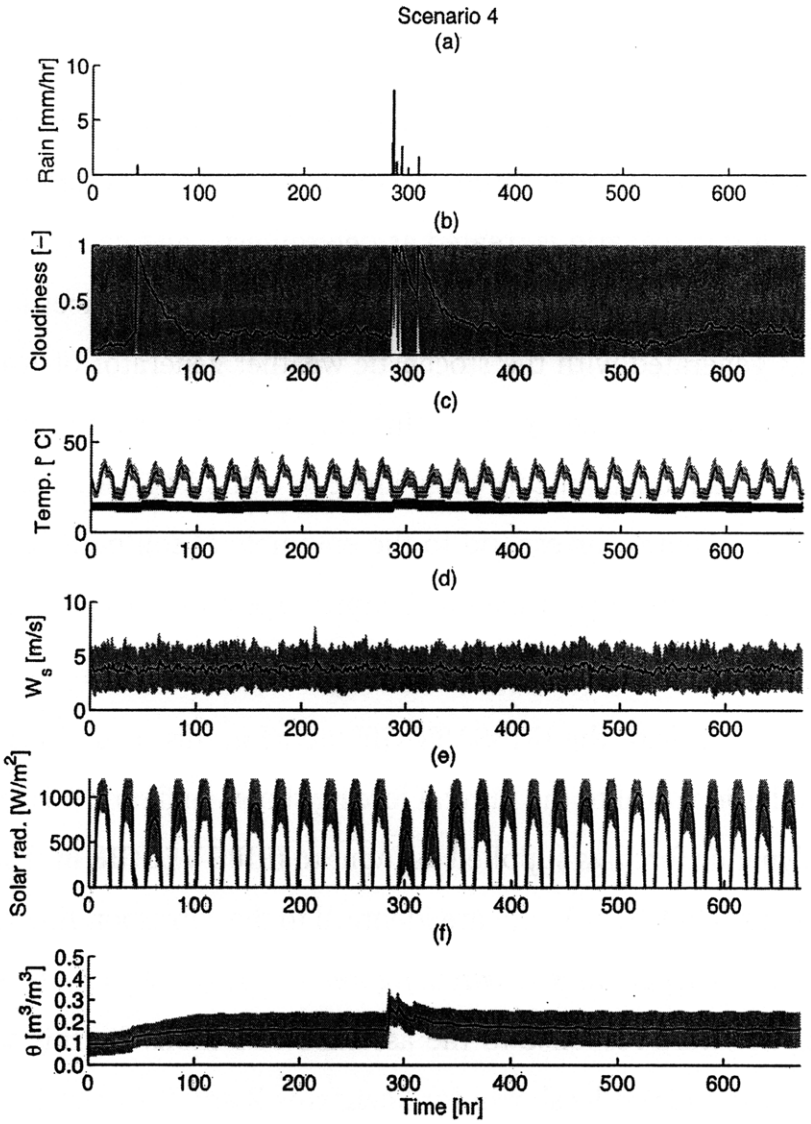


Figure 7-4: Simulated (a) observed rainfall [mm/hr], (b) ensemble mean sky cover fraction [dimensionless] (black) and range (light gray area), (c) air and dew temperature [°C], (d) wind speed [m/s], (e) incoming solar radiation [W/m^2], and (f) near-surface soil moisture response [m^3/m^3]. In (c)-(f), ensemble means are in black and areas encompassing one ensemble standard deviation above and below the mean in light gray. In (c), mean dew temperature [°C] is the black line dark gray shaded area encompassing one standard deviation above and below the mean.

timing and duration of rainfall events is relatively well constrained. Therefore, in this scenario rainfall exhibit arrival rates, durations, and rainfall intensities known only in a statistical sense through the parameters of the MBL model estimated for

Tucson, AZ by Hawk [1992]. Conceptually, this would correspond to modeling soil moisture in a location with either no or extremely sparse monitoring of hydrometeorological variables, or using hydrometeorological forcings simulated by a numerical weather prediction model at lead times long enough that the forecast exhibits little to no accuracy regarding the occurrence of rainfall.

To simulate the rainfall forcings in this scenario, the MBL model is used to generate an ensemble (with 64 replicates) of completely independent hourly rainfall time series of 30 days in duration[2]. Thermodynamic forcings corresponding to each to each simulated hourly rainfall time series were obtained by supplying the rainfall time series to the stochastic weather generator. In this scenario, the assumed uncertain SHTPs correspond to those used in **Scenario 2** through **Scenario 4** generated by the RP technique for assumed loamy soil conditions.

As depicted in figure 7-5(a), this scenario is associated with uncertain rainfall that occurs at nearly all times of the simulation period. This is reflected in the corresponding thermodynamic forcings shown in figures 7-5(b)-(e). Through the ensemble, cloud sky cover fraction ranges from completely clear conditions to completely overcast at all times during the simulation (figure 7-5(b)). The ensemble mean of the simulated air temperature reflects the the climatological diurnal cycle, while the area encompassed by one ensemble standard deviation above and below the ensemble mean captures the ensemble variability in rainfall rate at any given time (figure 7-5(c)). The lack of substantial temporal variation in the ensemble mean dew temperature and the associated area encompassing one ensemble standard deviation above and below the ensemble mean reflects the primacy of climatology on air humidity at any given time (figure 7-5(c)). Similar to **Scenarios 3 and 4**, the ensemble behavior of wind speed again reflects the first-order autoregressive process treatment of wind speed in the weather generator (figure 7-5(d)). The treatment of uncertainty in rainfall in this scenario has a pronounced affect on the simulated incoming solar radiation (figure 7-5(e)). Because rainfall occurs

[2]As opposed to **Scenario 2 and 3** in which the rainfall ensemble was constructed by perturbing the observed rainfall forcings simulated for **Scenario 1** using the multiplicative cascade.

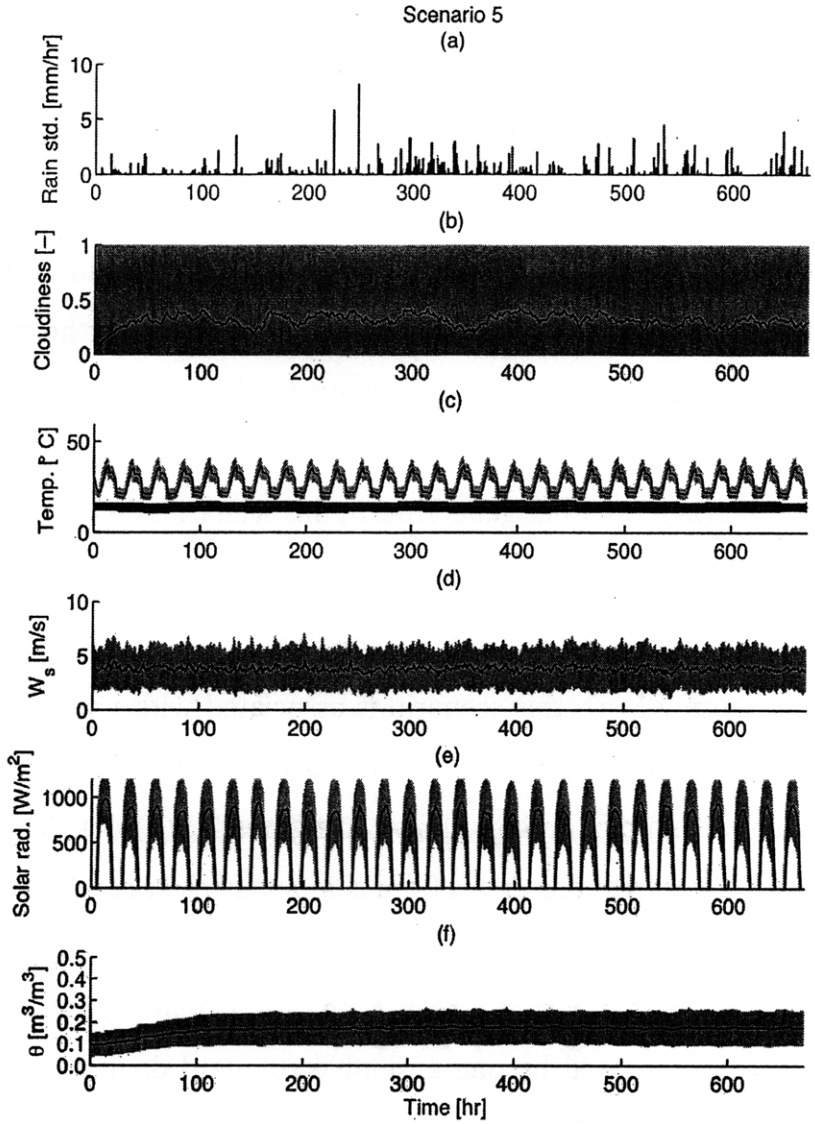


Figure 7-5: Simulated (a) ensemble standard deviation of rainfall [mm/hr], (b) ensemble mean sky cover fraction [dimensionless] (black) and range (light gray area), (c) air and dew temperature [°C], (d) wind speed [m/s], (e) incoming solar radiation [W/m^2], and (f) near-surface soil moisture response [m^3/m^3]. In (c)-(f), ensemble means are in black and areas encompassing one ensemble standard deviation above and below the mean in light gray. In (c), mean dew temperature [°C] is the black line dark gray shaded area encompassing one standard deviation above and below the mean.

within the ensemble at nearly every time during the simulation, the ensemble mean incoming solar radiation is suppressed, relative to the simulated incoming solar ra-

diation in **Scenarios 3 and 4** (figure 7-5(e)). Furthermore, this broad uncertainty in the occurrence of rainfall and sky cover fraction at all times of the simulation leads to relatively little temporal variation in the daily maximum ensemble mean incoming solar radiation(figure 7-5(e)). The broad uncertainty in sky cover fraction at all times during the simulation and described above also leads to an uncertainty in incoming solar radiation that is correspondingly broad at all daytime hours of the simulation (figure 7-5(e)). As one might expect, the corresponding simulated near-surface soil moisture response to this scenario of forcings and SHTPs leads to an ensemble behavior relatively devoid of temporal structure (figure 7-5(f)). With little consensus within the rainfall ensemble about the temporal occurrence of rainfall, the associated ensemble near-surface soil moisture response, as depicted by the ensemble mean and the area encompassing one ensemble standard deviation above and below the ensemble mean, provides virtually no reflection of the occurrence of individual rainfall events (figure 7-5(f)). The only temporal structure evident in figure 7-5(f), are the mild undulations associated with the diurnal cycles of incoming solar radiation and air temperature that influence soil moisture through the atmospheric evaporative demand.

It is important to note that the treatment of the uncertainty temporal occurrence of rainfall is at the opposite extreme of the assumptions made in the previously considered scenarios. While those scenarios treat the arrival and duration of rainfall events as perfectly known with uncertain hourly rainfall volumes, this scenario assumes the arrival, duration, and temporal intensities of storms are known only in a climatological sense. In reality, when simulating a spatially distributed soil moisture response to historical or forecast rainfall events, a more fair treatment of uncertainty in hydrometeorological forcings probably lies in between these two extreme assumptions. For instance, in simulating the spatial distribution of soil moisture response to historical data, there are likely periods during which there is significant confidence in the lack rainfall occurrence, while uncertainties in the timing and duration of rainfall events can arise from inadequate spatial resolution of rainfall products or errors in radar-rainfall retrieval algorithms. Similarly, in simulating a

spatially distributed soil moisture response to a numerical weather forecast, particular mesoscale phenomena (e.g., anti-cyclonic high pressure systems) are likely to be associated with weather forecasts that exhibit significant confidence in the non-occurrence of rainfall. On the other hand, significant uncertainty can exist in the temporal arrival of rainfall-producing phenomena (e.g., cyclonic lows), even though the forecast may exhibit a strong consensus that such an event is likely to occur.

7.1.6 Additional scenarios 2A-5A: deterministic SHTPs

A corresponding set of scenarios was designed to deduce the degree to which uncertainty in the rainfall and thermodynamic forcings lead to uncertainty in the near-surface soil moisture response simulated by tRIBS-VEGGIE in the absence of uncertainty in SHTPs. The goal of these additional scenarios was to verify the conclusions reached above that inadequate knowledge of the soil parameters required as input to tRIBS-VEGGIE is the most significant source of uncertainty in the simulated near-surface soil moisture. For **Scenarios 2A-5A**, treatment of the uncertainty in the thermodynamic and rainfall forcings for each scenario matched the corresponding assumptions made in **Scenarios 2-5**, except that the observed SHTPs generated for **Scenario 1** were used and were treated as deterministic. Results from this suite of scenarios (comparisons between **Scenarios 2 and 2A** through **Scenarios 5 and 5A** are shown in Appendix C) indicate that uncertainty in the near-surface soil moisture simulated by tRIBS-VEGGIE, as measured by the ensemble standard deviation, in response to uncertain hydrometeorological forcings exhibits relatively little uncertainty as long as uncertainty in SHTPs is neglected. While these results do indicate that the uncertainty in soil moisture as simulated by tRIBS-VEGGIE is more sensitive to uncertainty in thermodynamic forcings than uncertainty in hourly rainfall volumes (when the timing and duration of rainfall is assumed well known; **Scenarios 2A-4A**), the uncertainty in the soil moisture response when SHTPs are treated as deterministic is insignificant relative to the above scenarios in which

SHTPs are assumed uncertain.

7.1.7 Conclusion to point-scale sensitivity analysis

The results of this point-scale sensitivity analysis are important in the context of the data assimilation problem this work addresses. As described in **Chapter 6**, because the ensemble Kalman Filter uses an estimate of the variance retrieved from an Monte Carlo-derived ensemble representation of the model state, unreasonably low estimates of variance in the state (soil moisture in this case) is associated with artificially high confidence in the model estimate. This can lead to relatively little propagation of information from the observations to the model states, potentially causing the filter to diverge from the true state. This point-scale sensitivity analysis suggests that adequately characterizing uncertainty in the SHTPs required as input to tRIBS-VEGGIE is critical to avoid the problem of unreasonably low variance in the ensemble estimate of the state.

It is important to note at this point that while imperfect knowledge of the soil parameters is seemingly the dominant source of uncertainty in the simulated soil moisture response to hydrometeorological forcings, in the scenarios outlined above representing uncertainty in the forcings (rainfall and thermodynamic) results in little additional computational burden. In these scenarios because of the relatively simple stochastic models used to generate the forcings for tRIBS-VEGGIE the only burden incurred arises from the additional storage needed to record the simulated stochastic hydrometeorological forcings, which is typically not the limiting computational resource. However, in ensemble simulation and data assimilation scenarios in which the hydrometeorological forcings are represented via a physically-based numerical weather prediction model, computational burden associated with generation of the forcings may be substantial. These results seem to indicate that in such scenarios, emphasis on adequately characterizing uncertainty in SHTPs is most critical in efforts to adequately represent the spatiotemporal structure of uncertainty in soil moisture.

Because of the relatively small burden of representing the uncertainty in the hydrometeorological forcings, and in order to maintain a degree of realism consistent with the data available to characterize the parameter and hydrometeorological forcings input to tRIBS-VEGGIE, uncertainty in SHTPs, rainfall, and thermodynamic forcings are explicitly represented in the remainder of the experiments in this thesis.

7.2 Spatially distributed soil moisture uncertainty

This section describes an ensemble open loop experiment in which uncertainty in SHTPs, rainfall, and thermodynamic forcings are explicitly modeled. The purpose of this experiment is to diagnose the ways in which landscape features, particularly the spatial organization of soil units and topography, contribute to uncertainty in soil moisture. Below is a description of the experimental setup, assumptions, and statistical analyses used to perform this sensitivity analysis. Following this description, the results of the spatially distributed soil moisture sensitivity analysis are presented, and the implications for the data assimilation problem addressed by this thesis are discussed.

7.2.1 Description of watershed and assumptions of experiments

This spatially distributed sensitivity analysis is set in the Walnut Gulch Experimental Watershed (WGEW) in Tombstone, Arizona, USA. Established by the U.S. Department of Agriculture (USDA) in 1958, WGEW is approximately 150 km^2 in area, and is located at a transition between the Chihuahuan and Sonoran deserts Moran *et al.* [2008]. The semiarid watershed is associated with a mean annual temperatures of 17.7 °C, while mean annual precipitation is 312 mm Moran *et al.* [2008]. The North American Monsoon System (NAMS) brings approximately 60% of the mean annual precipitation in the months of July through September, and summertime rainfall events tend to be localized and of high intensity Moran *et al.*

[2008].

Recommendations of Vivoni *et al.* [2005] were followed to fit an efficient Triangulated Irregular Network (TIN) mesh from a U.S. Geological Survey 30 m Digital Elevation Model for the WGEW area using the ArcInfo Geographic Information System (GIS) package developed by ESRI. Attempts to delineate floodplain features following the techniques used by Vivoni *et al.* [2005] revealed little discernable floodplain structure, and the final TIN mesh is comprised of 19,447 computational nodes.

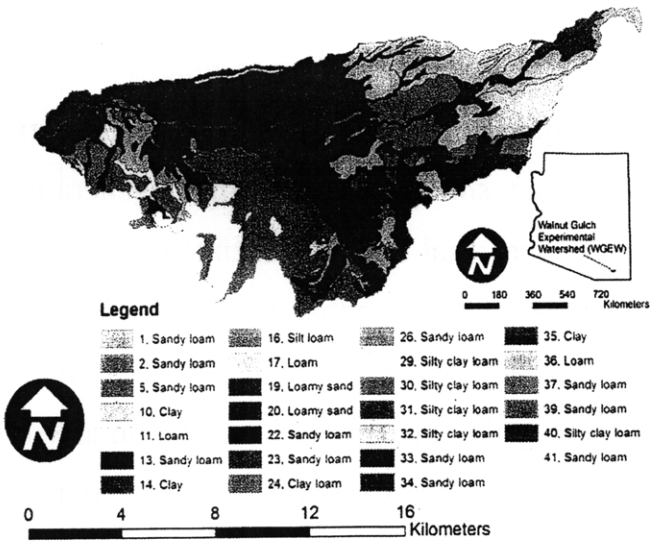


Figure 7-6: Map of assumed soil textures in the Walnut Gulch Experimental Watershed (WGEW) in Arizona, USA. The spatial distribution of soil units is provided from the Soil Survey Geographic (SSURGO) database.

The spatial distribution of soil units (i.e., areas of relatively homogeneous soil texture) was obtained from the Soil Survey Geographic (SSURGO) database maintained and published by the USDA. Investigation of the surface soil textural classes in WGEW as recorded in the SSURGO database reveals several soil textural classes outside of the nine classes used in development of the Restricted Pairing (RP)-based SHTP generation technique described in **Chapter 6**. The majority of SSURGO textural classes within WGEW that do not strictly conform to one of the nine classes found in the database from which uncertain SHTPs are frequently only slightly different from one of the nine classes. For instance, one of the non-conforming

soil textural classes found in the SSURGO database for WGEW is a "stony sandy loam." While the distinction between a "stony sandy loam" and a sandy loam may be important in the context of the soil hydrology, in order to facilitate the use of the RP-based SHTP generation scheme the non-conforming soil textural classes in the SSURGO database for WGEW were reclassified to one of the nine soil textural classes found in the database from which the RP-based scheme was derived. When possible, the non-conforming SSURGO soil was reclassified to the textural class most similar to one of the nine classes. Hence, a "stony sandy loam" would be reclassified as a "sandy loam." Figure 7-6 provides a spatial map of the reclassified soil units. Using the RP-based SHTP sampling scheme, 1024 potential realizations of the soil parameters required as input by tRIBS-VEGGIE were generated based on the assumed soil unit map shown in Figure 7-6. In generating the ensemble of SHTPs, soil units sharing a textural class but distinct in their spatial location were treated as independent to reflect the fact that individual soil units and complexes in the SSURGO database represent regions sharing relatively homogeneous pedogenic origins and processes. Hence, in any particular realization of SHTPs obtained from the RP-based sampling scheme, it is very unlikely that two soil units bearing the same textural class would also share the same parameter values.

To obtain an ensemble characterization of hydrometeorological forcings that is internally consistent, the stochastic approach to generating hourly hydrometeorological forcings for the tRIBS-VEGGIE model described in **Chapter 5** is used in this sensitivity analysis. The simulated hydrometeorological forcings correspond to the month of August because, being associated with the presence of the NAMS, it is critical in the annual water budget in the semiarid WGEW. In this sensitivity analysis, it is assumed that both the rainfall and thermodynamic forcings are uncertain. Because WGEW is extensively monitored, it is reasonable to assume that the temporal arrival of rainfall events is relatively well known. Consistent with these assumptions, the Modified Bartlett Lewis model is used to generate a single time series of hourly rainfall forcings over a period of 27 days for WGEW. The stochastic multiplicative cascade is then used both to perturb this simulated hourly rainfall time

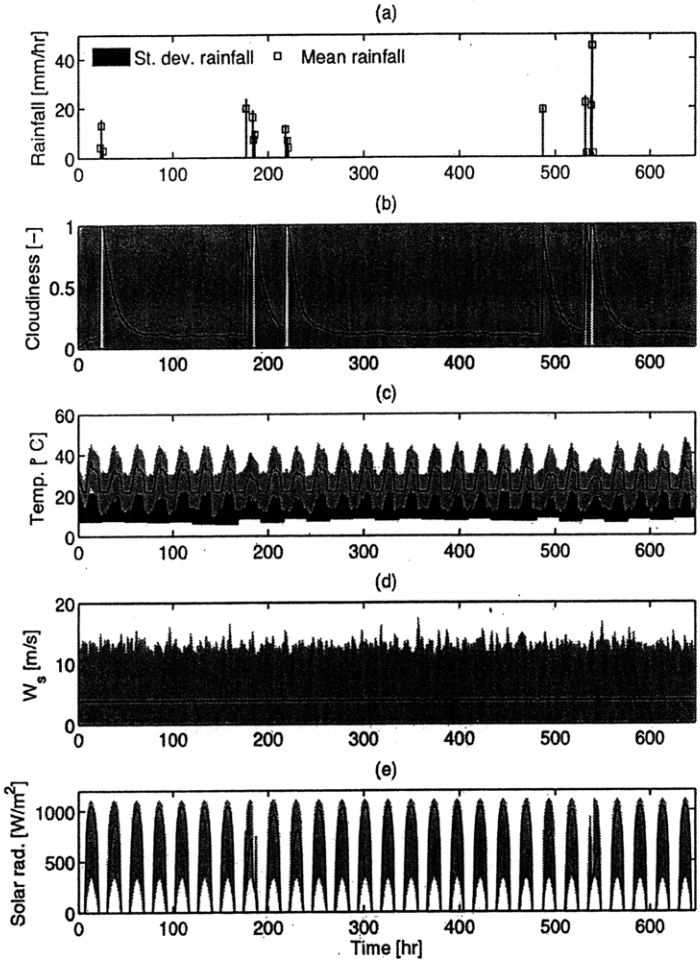


Figure 7-7: Simulated (a) ensemble standard deviation of rainfall [mm/hr], (b) ensemble mean sky cover fraction [dimensionless] (black) and range (light gray area), (c) air and dew temperature [°C], (d) wind speed [m/s], and (e) incoming solar radiation [W/m^2]. In (c)-(e), ensemble means are in black and areas encompassing one ensemble standard deviation above and below the mean in light gray. In (c), mean dew temperature [°C] is the black line dark gray shaded area encompassing one standard deviation above and below the mean.

series and to disaggregate the hourly rainfall pulses in space. Consistent with the availability of NEXRAD radar-rainfall products over WGEW, the assumed spatial resolution of the disaggregated hourly rainfall pulses in 4 $\times$ 4 km [Klazura and Imy, 1993]. There are six periods of rainfall during the 27 day-long simulation, and hourly rainfall volumes are associated with substantial variability through the ensemble (figure 7-7(a)). There is one large rainfall at approximately 550 hrs into

the simulation with an ensemble mean rainfall rate of approximately 42 mm/hr and an ensemble standard deviation of approximately 50 mm/hr (figure 7-7(a)). The range of simulated hourly rainfall totals were qualitatively compared against historical rainfall data published for WGEW [Goodrich *et al.*, 2008] and found to be in reasonable agreement.

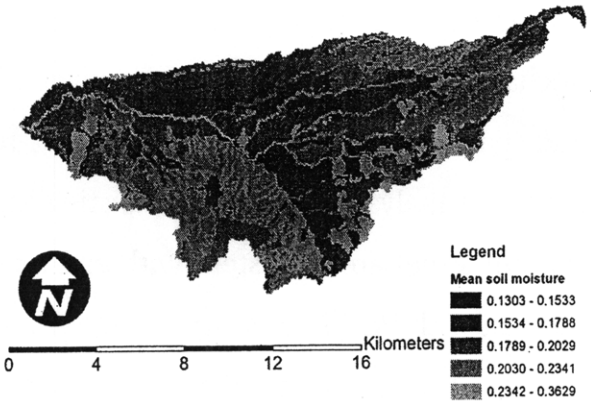
The stochastic weather generator is used to simulate the 1024 realizations of the thermodynamic forcings by supplying each of the 1024 realizations from the ensemble of hourly rainfall time series as input. Simulated rainfall events are, as expected, associated with overcast conditions in the ensemble mean sense and the stochastic weather generator produces a broad diversity of fair weather cloudiness conditions (figure 7-7(b)). Rainfall occurrence is associated with a depression in both the ensemble mean and maximum air temperatures, although there is broad ensemble variability in air temperatures throughout the simulation (figure 7-7(c)). As expected, the ensemble mean wind speed exhibits virtually no temporal structure and a skew toward lower wind speeds, consistent with the treatment of hourly wind speeds in the weather generator (figure 7-7(d)). The temporal dynamics of incoming solar radiation during rain-free days reflects the ensemble mean behavior of fair weather cloudiness, with the ensemble mean exhibiting a bias toward the higher end of the simulated range of incoming solar radiation (figure 7-7(e)). Consistent with the occurrence of rainfall and the associated overcast conditions, during rainfall the ensemble mean and maximum values of incoming solar radiation are suppressed (figure 7-7(e)).

The initial conditions for each replicate of the 1024 replicate open loop sensitivity analysis were derived from a 240 hr open loop "spin up" simulation. For each realization of the spin up, one realization of SHTPs was paired with a corresponding realization of hydrometeorological forcings. Initial soil moisture conditions for the spin up were set to 10% effective saturation throughout WGEW, and each realization of the spin up forced with the final 10 days (240 hr) of the simulated hydrometeorological forcings. Maintaining the pairing between realizations of SHTPs and

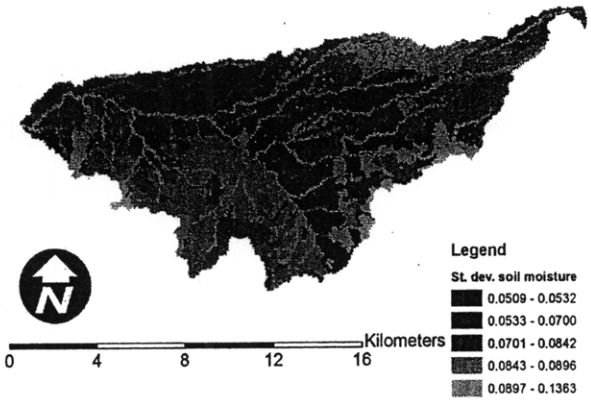
hydrometeorological forcings assigned for the spin up, each replicate of the sensitivity analysis experiment was then initialized with the corresponding soil moisture distribution at the end of the 240 hr spin up simulation, and forced with the full 27 days (648 hrs) of stochastic hydrometeorological forcings. Although vegetation is a potentially important part of the water cycle in WGEW, bare soil conditions are assumed in this sensitivity analysis. Obtaining realistic spatial distributions of multiple plant functional types that are consistent with and reflect the uncertainty in SHTPs through simulation with VEGGIE remains a significant challenge. Moreover, although it would be an attractive alternative to using the dynamic vegetation model to simulate vegetation states, obtaining estimates of the spatial distribution of the vegetation state from remote sensing data is the subject of ongoing research that is beyond the scope of this thesis work. This sensitivity analysis is nevertheless a useful exercise for examining the relative influence of soils, topography, and hydrometeorological forcings on uncertainty in the spatial distribution of soil moisture.

7.2.2 Results and discussion: first- and second-order behavior

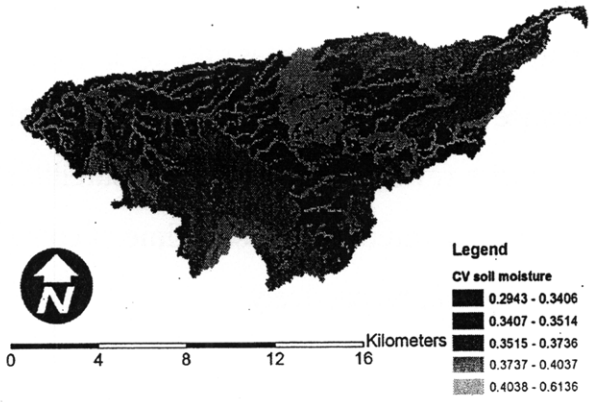
This section describes the results of the open loop sensitivity analysis experiment for WGEW. In the context of the soil moisture data assimilation problem addressed by this thesis, of principal importance is the ensemble behavior of the full state vector (soil moisture at all computational nodes and layers in the model), and of soil moisture in the near-surface and throughout the soil column at temporal intervals and local times consistent with revisit times of planned satellites such as SMOS and SMAP. The appropriate temporal interval and local time of consideration, therefore, is assumed to be 72 hrs at 9 A.M., respectively.



(a)

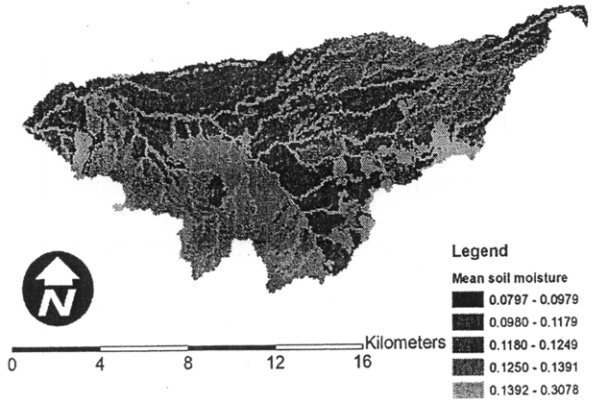


(b)

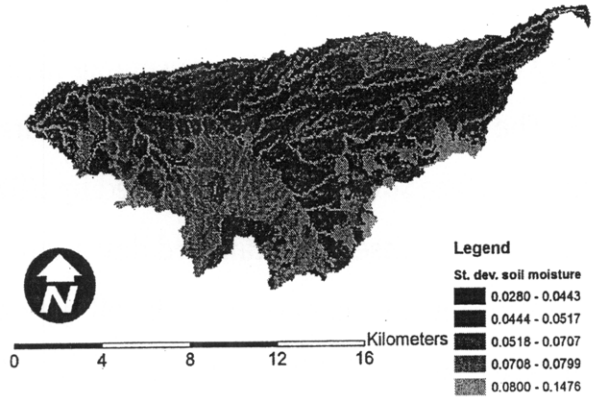


(c)

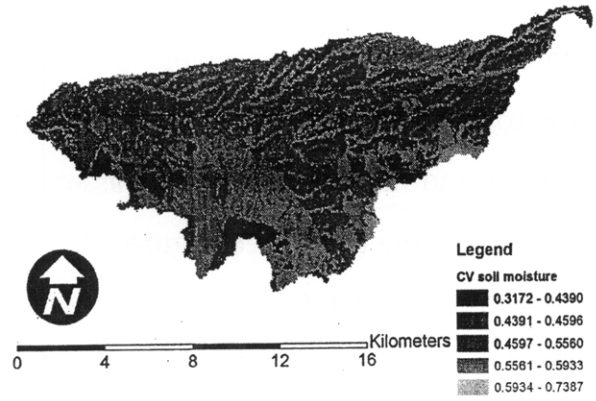
Figure 7-8: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 144 hrs into the 1024 replicate open loop simulation.



(a)



(b)



(c)

Figure 7-9: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 144 hrs into the 1024 replicate open loop simulation.

Integrating the ensemble soil moisture state over particular depths allows for visualization of the spatial distribution of the sample statistics describing the first- and second-order behavior of the ensemble. Because L-band microwave remote sensing platforms are sensitive to soil moisture in approximately the top 10 cm, it is useful to investigate the dominant components in near-surface soil moisture anomalies. Furthermore, the soil moisture integrated over the entire soil column provides a metric that facilitates visualization of the spatial distribution of moisture in the unsaturated zone throughout the WGEW.

At 144 hours into the simulation, near-surface ensemble mean soil moisture in WGEW largely reflects a combination of the heterogeneity in soil types, as well as the spatial organization of the channel and valley network (figure 7-8(a)). The channel network is mostly associated with higher ensemble mean near-surface soil moisture (figure 7-8(a)). The spatial variation in soil types results in sharp contrasts in the ensemble mean near-surface soil moisture predicted by tRIBS-VEGGIE (figure 7-8(a)). The spatial distribution of ensemble standard deviation in near-surface soil moisture also largely follows contrasts associated with soil type heterogeneity and, to a lesser extent, topography (figure 7-8(b)). Interestingly, soil types associated with higher ensemble mean near-surface soil moisture also seem to be associated with higher ensemble variance in near-surface soil moisture. Visualizing the spatial distribution of the ensemble coefficient of variation, which is produced by normalizing the ensemble standard deviation by the ensemble mean, leads to a much more prominent depiction of the role of topography in ensemble variation in near surface soil moisture, with the channel and valley bottoms being associated with the highest values of coefficient of variation in the watershed (figure 7-8(c)). This suggests that while soil type heterogeneity imposes significant spatial variation in ensemble first- and second-order statistics, for a given soil type topography plays the predominant role in the uncertainty in near-surface moisture. Spatial distributions of ensemble mean, standard deviation, and coefficient of variation in near-surface soil moisture for all other time 72 hr time intervals in the open loop simulation largely conform to the conclusions discussed here and are shown in their entirety **Appendix C.**

The ensemble mean profile-integrated soil moisture, by constrast, reflects a combination of the heterogeneity in soil types and topography, although the role of topography is more pronounced in this case (figure 7-9(a)). Similar to the ensemble mean near-surface moisture, the channel network is associated with higher ensemble mean profile soil moisture (figure 7-9(a)). Again sharp contrasts associated with heterogeneity in soil types can be seen in the spatial distribution of ensemble mean profile soil moisture simulated by tRIBS-VEGGIE (figure 7-9(a)). Similarly, the spatial distribution of ensemble standard deviation in profile soil moisture also largely follows contrasts associated with soil type heterogeneity and topography, with higher ensemble variance in profile soil moisture associated with channel and valley bottoms (figure 7-9(b)). The ensemble coefficient of variation shows the a distinct topographic impression, with channel and valley bottoms being associated with the higher values of coefficient of variation in the watershed (figure 7-9(c)). Again, the spatial distribution of ensemble mean, standard deviation, and coefficient of variation in profile-integrated soil moisture for all other time 72 hr time intervals in the open loop simulation bear similar patterns to those described here and are presented in **Appendix C**.

One interesting commonality between the near-surface and profile-integrated soil moisture distributions shown in figures 7-8 and 7-9, respectively, lies in the observation that channel and valley bottoms are typically associated with higher ensemble variance in soil moisture. This is likely a result of the uncertainty in soil hydraulic properties, particularly residual and saturation moisture content and saturated hydraulic conductivity. At topographic highs (i.e., ridgelines), uncertain rainfall is partitioned into infiltration and runoff at uncertain rates that depend on the soil textural class and the uncertain antecedent moisture. Because of the lateral redistribution and runon mechanisms in tRIBS-VEGGIE, both runoff and infiltrated water be redistributed downslope, to a topographically lower pixel that is also exposed to uncertain incoming rainfall and uncertain soil moisture conditions. In this manner, topography has the effect of amplifying uncertainty in precipitation, runoff, and lateral redistribution in a downstream direction, resulting in higher

ensemble variance in the topographic lows of the domain. While this reasoning may be valid for semiarid regions in which the water table is at significant depth beneath the surface and channels are mostly ephemeral, in more humid regions groundwater-surface water interaction would curtail this amplification. This is because the downslope boundary condition of saturated conditions would limit the amount of upslope runoff that could re-infiltrate downslope and suppress subsurface lateral redistribution.

7.2.3 Results and discussion: rank and empirical orthogonal functions of moisture anomalies

Another important facet of this sensitivity analysis, particularly in the context of the EnKF, is the rank of the ensemble approximation of the state error covariance matrix, $\mathbf{P}_e$. The rank of $\mathbf{P}_e$ can be computed directly by finding the nonzero singular values of the matrix containing the soil moisture state anomalies. The left singular vectors associated with the nonzero singular values represent the linearly independent modes within the matrix of ensemble soil moisture anomalies. Therefore, if the rank of $\mathbf{P}_e$ is smaller than the minimum of either the state dimensionality (N) or number of ensemble replicates (K), the covariance structure of the state space can be captured by fewer ensemble replicates and $\mathbf{P}_e$ is said to be rank deficient. Collecting the K ensemble replicates of the N-dimensional state vector into a matrix Θ and subtracting the matrix $\bar{\Theta}$, which contains N copies of the ensemble mean gives a matrix of soil moisture state anomalies, $\tilde{\Theta}$. Performing a singular value decomposition on $\tilde{\Theta}$ leads to,

$$\tilde{\Theta} = \mathrm{USV}^{\mathrm{T}}, \tag{7.1}$$

in which the $N \times N$ matrix U contains the left singular vectors or empirical orthogonal functions (EOFs), S is the $N \times K$ diagonal matrix containing the singular values, and the $K \times K$ matrix V contains the right singular vectors. The nonzero

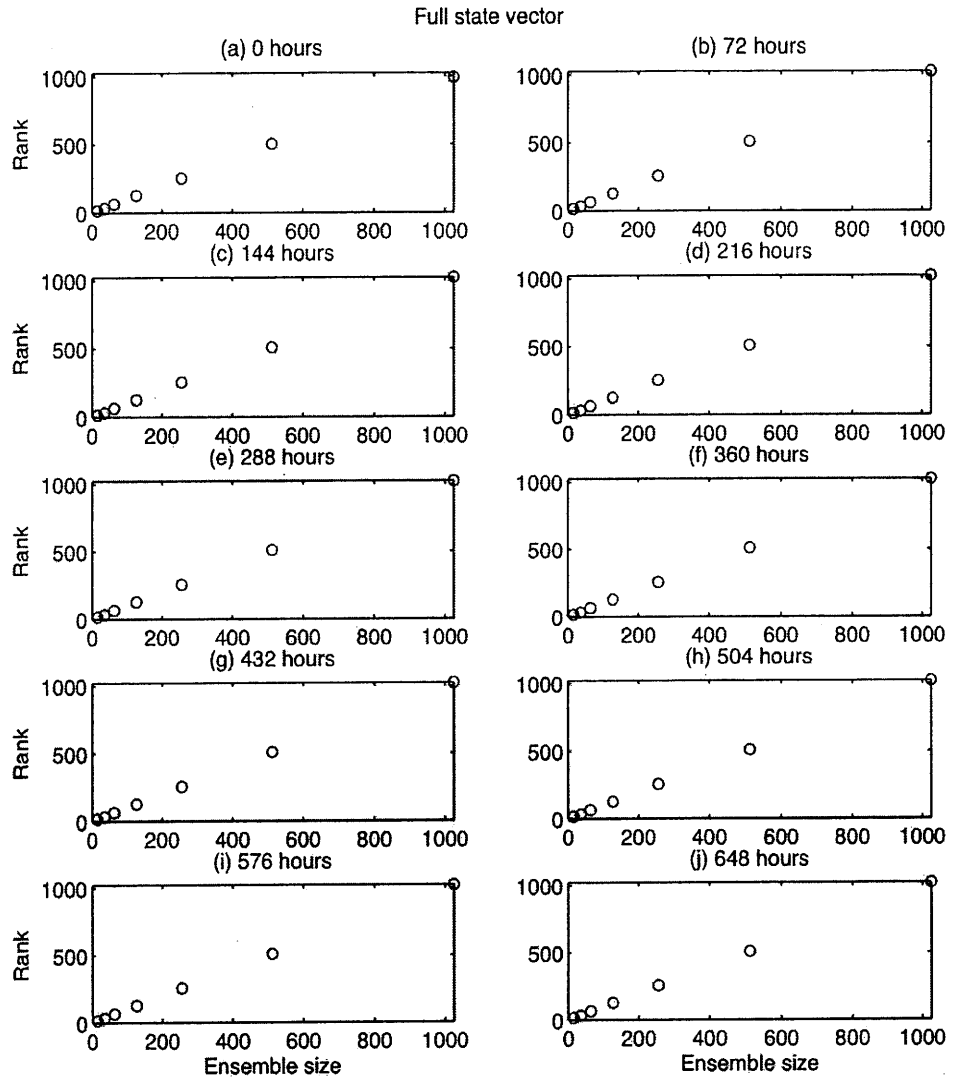


Figure 7-10: Rank versus ensemble size at (a) 0 hrs, (b) 72 hrs, (c) 144 hrs, (d) 216 hrs, (e) 288 hrs, (f) 360 hrs, (g) 432 hrs, (h) 504 hrs, (i) 576 hrs, (j) 648 hrs into the simulation.

singular values of $\tilde{\Theta}$ contained on the diagonal of the matrix S are the square roots of the corresponding eigenvalues of the matrix $\tilde{\Theta}\tilde{\Theta}^{\mathrm{T}} \sim \mathbf{P}_e$ [Strang, 1986]. Similarly, the EOFs contained in the matrix U are the eigenvectors of the matrix $\mathbf{P}_e$ [Strang, 1986]. Hence, the columns of U contain correspond to spatial mappings of the eigenmodes that describe the ensemble of soil moisture anomalies, while the squares of the corresponding nonzero singular values are proportional to the variance described by each singular vector. For the WGEW system under consideration, the size of the state vector is 194,470 $\times$ 1 (19,447 pixels with 10 soil layers each), and N will, for all practical purposes, always be greater than the ensemble size

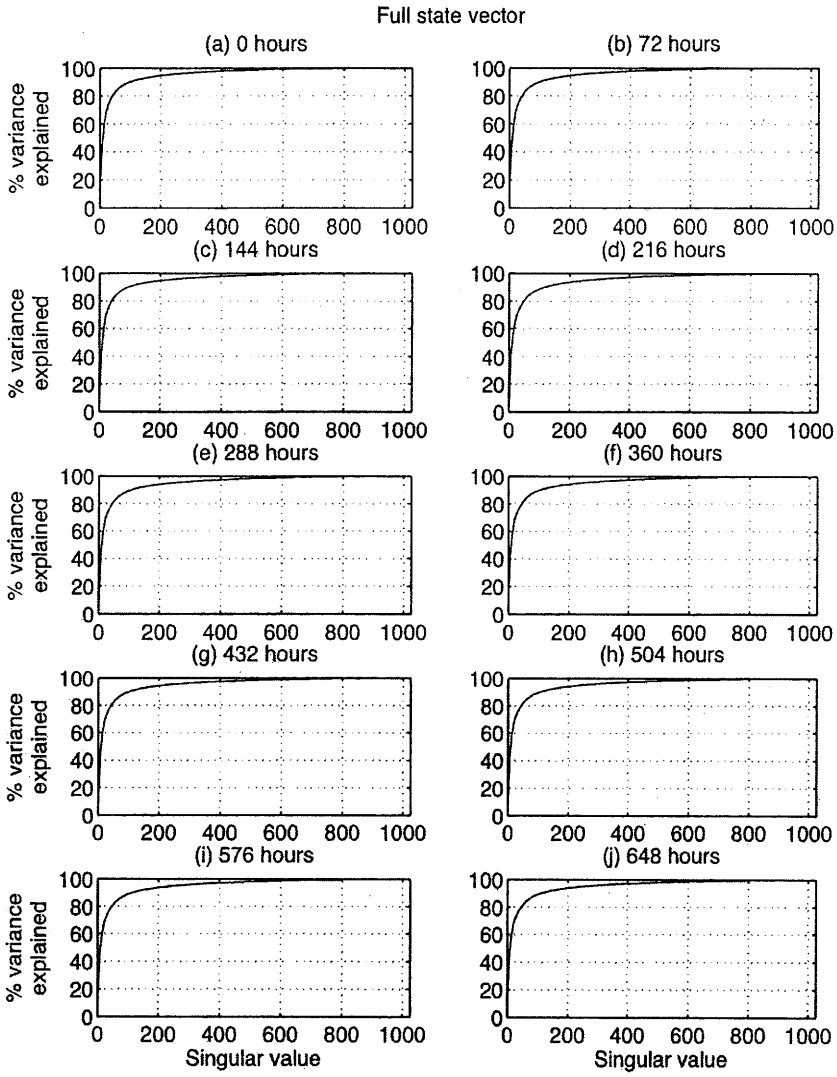


Figure 7-11: Spectrum of singular values (a) 0 hrs, (b) 72 hrs, (c) 144 hrs, (d) 216 hrs, (e) 288 hrs, (f) 360 hrs, (g) 432 hrs, (h) 504 hrs, (i) 576 hrs, (j) 648 hrs into the simulation.

(K). Thus, the maximum rank of $\tilde{\Theta}$ is K. However, should the rank of $\tilde{\Theta}$ plateau at some ensemble size K less than 1024, it would imply that $\tilde{\Theta}$ possesses fewer than 1024 degrees of freedom and that the variance in $\tilde{\Theta}$ could be explained by the first rank($\tilde{\Theta}$) singular vectors (EOFs).

The rank of the matrix $\tilde{\Theta}$ was determined for ensemble sizes, K, ranging from 16 to 1024 by factors of 2 at every 72 hrs during the simulation. Rather than performing additional ensemble simulations for K less than 1024, K state vector replicates were sampled from among the full 1024 replicates at each time interval

and the rank of $\tilde{\Theta}$ determined using the `rank` function in MATLAB[3]. For ensemble sizes up to 512, this process of sampling from among the full 1024 replicates to compute the rank of $\tilde{\Theta}$ was repeated 20 times, and the average rank over those 20 repetitions computed. For every time interval during the 72 hour simulation, the rank of $\tilde{\Theta}$ was found to increase linearly and with a slope of 1 with K (figure 7-10). This implies that the methods used to introduce uncertainty into the predicted soil moisture state (perturbation of hydrometeorological forcings and soil parameters) are such that the ensemble representation captured in $\tilde{\Theta}$ is of full rank, through the largest ensemble considered (1024 replicates). Furthermore, it suggests that the total variance captured through an ensemble representation of soil moisture, up to at least 1024 replicates, is limited only by the computational resources to add ensemble replicates.

To diagnose an economic ensemble size (i.e., one that minimizes computational resources while maximizing the variance captured by the ensemble representation of uncertainty in soil moisture), it is useful to investigate the singular value spectrum. As stated above, the squared nonzero singular values of $\tilde{\Theta}$ are equal to the nonzero eigenvalues of the ensemble approximation of the state error covariance matrix, and are proportional to the relative variance explained by the corresponding EOF/eigenvector. At each 72 hour interval, the cumulative percent variance explained versus singular value number shows that approximately the first 50 EOFs dominate the description of the variance in soil moisture anomalies since the cumulative percent variance explained is at or near 90% at singular value numbers of approximately 50 (figure 7-11).

Visualizing the cumulative percent of variance explained as a function of time and in response to the uncertain rainfall forcings reveals that at least 80% of the variance in soil moisture anomalies is captured by the first 50 EOFs at all times, and at least 88% of the variance is explained by the first 100 EOFs (figures 7-12(e) and (f)). Interestingly, the rainfall event at approximately 220 hrs results in a decrease

[3]MATLAB computes the rank of matrix A by computing the number of singular values above some threshold close to zero.

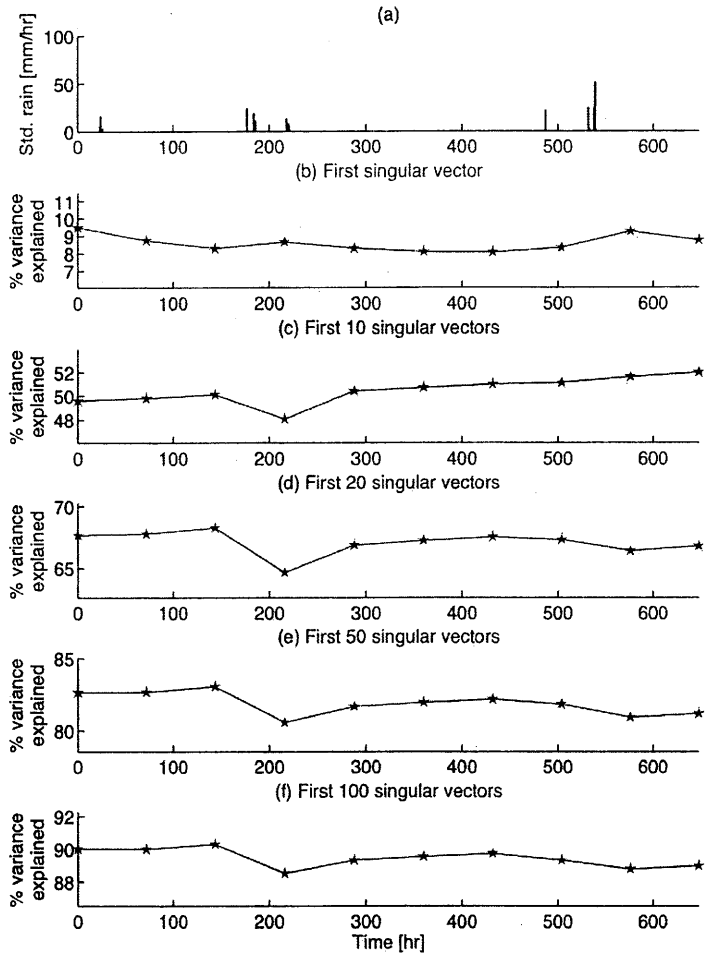


Figure 7-12: Time evolution of (a) the standard deviation in the hourly rainfall forcings, and the cumulative variance in soil moisture anomalies explained by the: (b) first EOF, (c) first 10 EOFs, (d) first 20 EOFs, (e) first 50 EOFs, and (f) first 100 EOFs.

in the soil moisture anomaly variance explained by the first 10 EOFs, while the first EOF is relatively insensitive to the rainfall forcings (figures 7-12(c) through (f)). A closer inspection of the variance in soil moisture anomalies explained by the first four EOFs reveals that they capture approximately 9%, 7%, 7% and 5% of the variance in soil moisture anomalies at all times, respectively (figure 7-13). While the variance explained by the first EOF seems to increase slightly in response to rainfall forcings (figure 7-13(a)), EOFs 2-4 demonstrate little sensitivity to rainfall forcings (figure 7-13(b) through (e)).

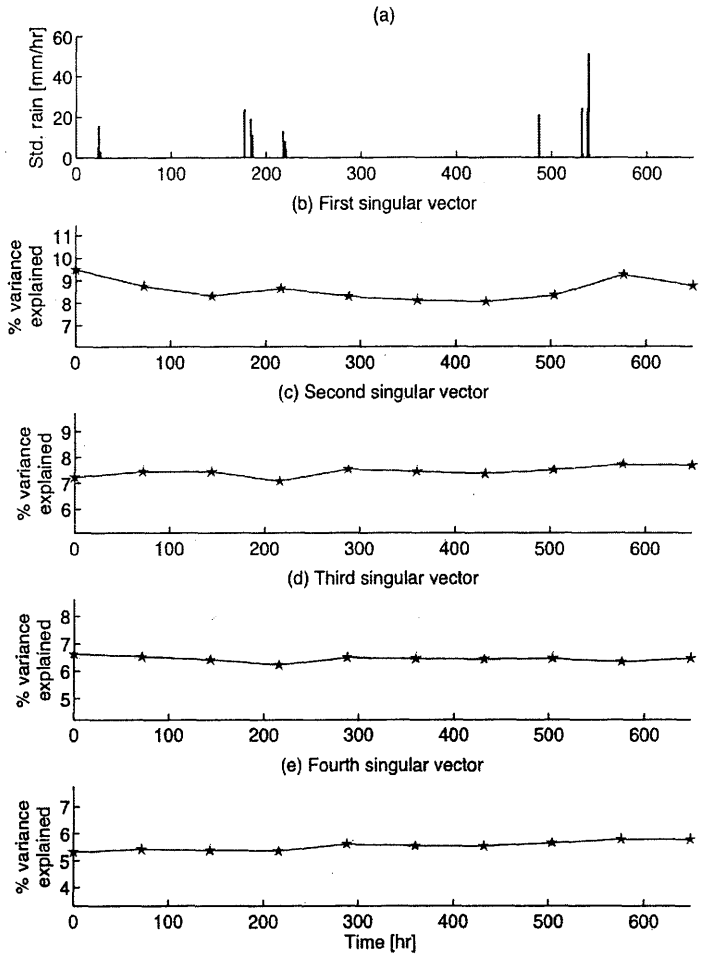


Figure 7-13: Time evolution of (a) the standard deviation in the hourly rainfall forcings, and the variance in soil moisture anomalies explained by the: (b) first, (c) second, (d) third, and (e) fourth EOFs.

Integrating the ensemble soil moisture state over particular depths and performing a singular value decomposition on the column-integrated soil moisture anomalies also allows for visualization of the spatial distribution of the EOFs, and investigation of the amount of variance described by those EOFs. Therefore, it is useful to consider soil moisture anomalies in the the top 10 cm as well as integrated over the entire soil column. The first 20 EOFs cumulatively account for at least 84% and 85% of the variance in near-surface and profile soil moisture anomalies, respectively (figure 7-14(d)). The first 50 EOFs, by contrast capture at least 92% and

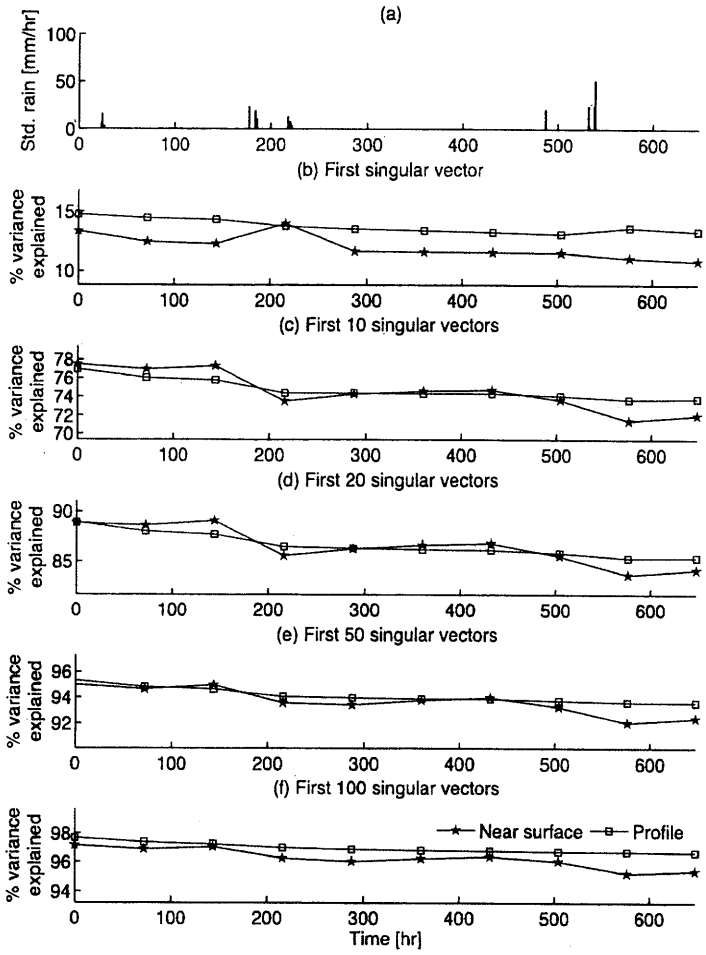


Figure 7-14: Time evolution of (a) the standard deviation in the hourly rainfall forcings, and the cumulative variance in near surface (blue lines marked with stars) and profile-integrated (red lines marked with squares) soil moisture anomalies explained by the: (b) first EOF, (c) first 10 EOFs, (d) first 20 EOFs, (e) first 50 EOFs, and (f) first 100 EOFs.

94% of the variance in near-surface and profile soil moisture anomalies, respectively (figure 7-14(e)). Interestingly, the first 10 EOFs describe 6% less variance in near-surface soil moisture anomalies at the end of the simulation, steadily declining relative to the beginning of the simulation (figure 7-14(c)). This suggests that the first 10 EOFs become less powerful descriptors of the variance in near-surface soil moisture anomalies as the simulation progresses. Similar, albeit less severe, trends are observed in both variance explained in near-surface and profile soil moisture

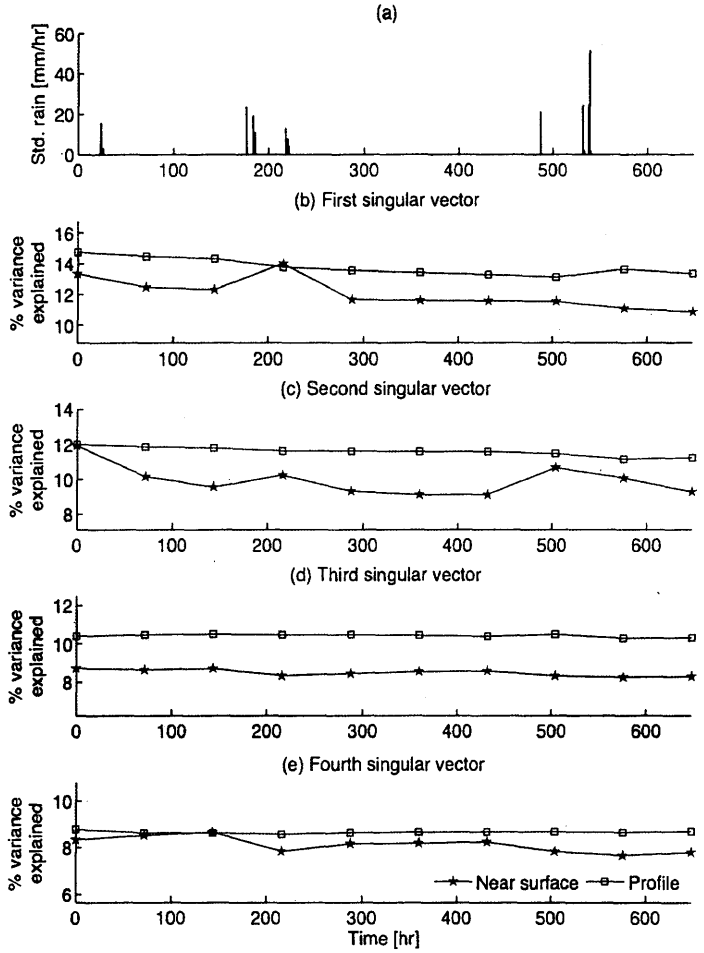


Figure 7-15: Time evolution of (a) the standard deviation in the hourly rainfall forcings, and the variance in near surface (blue lines marked with stars) and profile-integrated (red lines marked with squares) soil moisture anomalies explained by the: (b) first, (c) second, (d) third, and (e) fourth EOFs.

anomalies as a function of time in figure 7-14. Investigating the first four EOFs of near-surface and profile-integrated soil moisture reveals that the first two EOFs of near-surface soil moisture anomalies are somewhat sensitive to rainfall occurrence, where as the first two EOFs of profile-integrated soil moisture anomalies are relatively insensitive to rainfall (figure 7-15(b) and (c)). The first EOF of profile-integrated soil moisture anomalies describes approximately 14% of the variance, while the second, third, and fourth describe approximately 12%, 10%, and 8% of the variance, respectively (figures 7-15(c)-(d)). By contrast the first four EOFs of

near surface soil moisture anomalies describe approximately 12%, 10%, 8% and 8% of the variance in the ensemble (figures 7-15(c) through (e)). At maximum, the first EOF of near-surface soil moisture anomalies describes approximately 14% of the variance in the ensemble and is associated approximately with the rainfall event that occurs at 220 hrs (figure 7-15(b)). The second EOF of near surface soil moisture anomalies describes at maximum 12% of the variance in the ensemble at 0 hrs, and almost 12% again at 500 hrs, perhaps in response to the rainfall event that occurs at approximately 480 hrs (figure 7-15(c)).

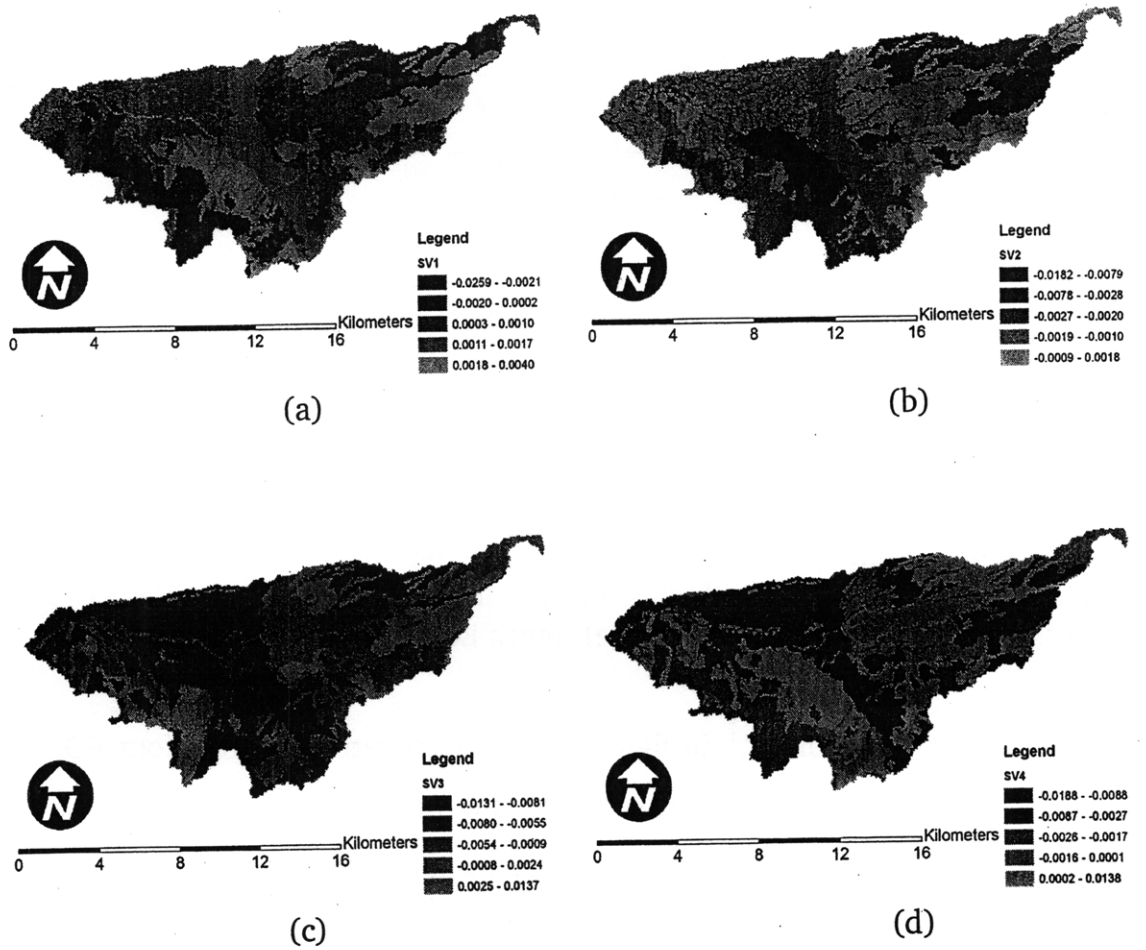


Figure 7-16: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture anomalies at 144 hr into the simulation. They explain 12, 10, 8.3, and 8.3 percent of the variance in soil moisture, respectively.

The spatial distribution of the first four EOFs of near-surface soil moisture anoma-

lies is shown in figure 7-16. The first near-surface soil moisture anomaly EOF (which describes 12% of the variance in near-surface soil moisture anomalies) largely reflects the spatial distribution of soil types in WGEW. There does, however, appear to be a nearly vertical line in the North-central portion of the watershed, which corresponds to the boundary a rainfall grid pixel and would suggest that rainfall also influences near-surface soil moisture anomalies (figure 7-16(a)). The second EOF of near-surface moisture anomalies, which describes approximately an additional 10% of the variance, also reflects spatial heterogeneity in soil types, but suggests a significant topographic influence, as well (figure 7-16(b)). The third near-surface soil moisture anomaly EOF largely depicts soil type heterogeneity throughout WGEW, although interestingly the region in the North-central portion of the watershed that seems to indicate an influence of spatial heterogeneity in rainfall in the first EOF again depicts some behavior worth noting (figure 7-16(c)). In particular, there seems to be an area with a nearly East-West oriented southern boundary and North-South oriented western boundary which possesses a stronger topographic influence relative to surrounding areas of the same soil textural class (figure 7-16(c)). The fourth EOF of near-surface soil moisture anomalies is almost entirely reflective of the spatial distribution of soil types in WGEW (figure 7-16(d)).

The corresponding spatial distribution of the first four EOFs of profile-integrated soil moisture anomalies is shown in figure 7-17. The first EOF of profile-integrated soil moisture anomalies represents a fusion of soil type heterogeneity and topography (figure 7-17(a)). For a given soil type, topography appears to be the dominant control on the first EOF of profile moisture anomalies (figure 7-17(a)). The second EOF of profile moisture anomalies provides a similar depiction of a mix between topography and soil type heterogeneity (figure 7-17(b)). As noted in the first and third EOFs of near-surface moisture anomalies, the second profile moisture anomaly EOF seems to exhibit a sharp North-South oriented boundary in the North-cental portion of the watershed at the approximate location of the rainfall grid footprint (figure 7-17(b)). To the west of this boundary, topography seems to exert significant influence in the second profile moisture anomaly EOF relative to the region

to the east of the boundary, despite a similarity in soil type (figure 7-17(b)). This is suggestive that rainfall in this vicinity of the watershed may exert an important control on both near-surface and profile-integrated soil moisture anomalies. The third EOF of profile-integrated moisture anomalies again reflects both soil type and topographic heterogeneity, with topographic influence on profile moisture anomalies being more significant in the eastern portion of the watershed (figure 7-17(c)). Topography plays a significant role in describing profile-moisture anomalies captured by the fourth EOF, except in the area encompassed by the soil type in the South-central portion of the watershed.

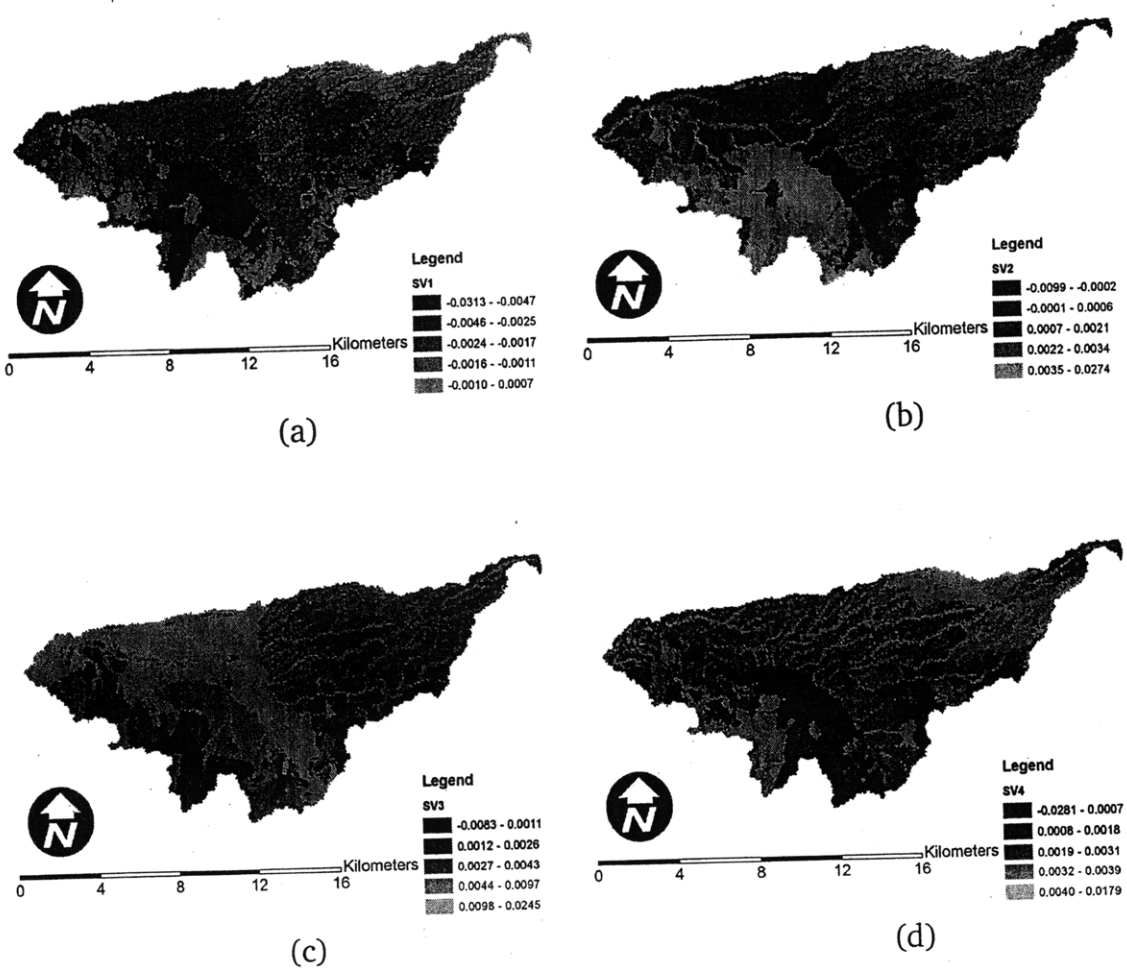


Figure 7-17: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile-integrated soil moisture anomalies at 144 hr into the simulation. They explain 14.2, 12, 10.2, and 8.5 percent of the variance in soil moisture, respectively.

As a caveat to the EOF analysis presented here, the marked influence of several distinct factors (e.g., soil type and topography) in a single EOF suggests that the ensemble of moisture anomalies is not perfectly separable through a linear decomposition such as SVD. This reflects, in part, the fact that the ensemble of moisture anomalies captured in this open loop experiment arises from a nonlinear set of processes. Spatial distributions of the first four EOFs of near-surface and profile-integrated soil moisture anomalies for all other time 72 hr time intervals in the open loop simulation are mostly similar to those presented here for hour 144 of the experiment are presented in **Appendix C.**

CHAPTER 8

ASSIMILATION EXPERIMENTS AND RESULTS

Experimental setup and results of three attempts to assimilate synthetic L-band microwave observations into tRIBS-VEGGIE are presented and discussed in this chapter. The three experiments reported here span a range of assumptions, and are designed to explore the performance and limitations of a data assimilation system, the core of which is the ensemble Kalman Filter (EnKF), to estimate the soil moisture state at hillslope scales modeled by tRIBS-VEGGIE. The first and most simple experiment explores assimilation of synthetic measurements of L-band microwave brightness temperature for soil moisture estimation at a single, flat computational element. This experiment seeks to demonstrate that assimilation of L-band measurements into the tRIBS-VEGGIE model is possible. The second experiment considers assimilation of coarse-scale brightness temperatures, assumed to represent a watershed-averaged measurement, for estimation of moisture at hillslope scales in a moderately rough topographic setting of uniform soil texture. This experiment is a demonstration that disaggregation of coarse-scale microwave brightness temperature measurements to hillslope scales, by conditioning on the model physics through the EnKF approach, is feasible. The final and most computationally demanding experiment investigates the assimilation of synthetic microwave radar measurements, at 3 km resolution, for estimation of soil moisture at hillslope scales in the Walnut Gulch Experimental Watershed (WGEW). This final case represents the most

plausible combination of factors that would be encountered in an operational or applied setting. Specifically, the WGEW is associated with spatial heterogeneity in soil type, rainfall, and topography, while the synthetic observations derived for the experiment are similar to anticipated products from NASA's forthcoming Soil Moisture Active-Passive mission. In this final experiment multiple notionally true soil moisture states (and therefore sequences of observations) are considered. The use of multiple plausible true realized moisture conditions (and corresponding observations) allows for a range of behavior in the soil parameters and hydrometeorological forcings used to derive the true soil moisture states and observations. Repeating the EnKF experiment with multiple sequences of observations derived from corresponding multiple potential true moisture realizations allows for assessment of the robustness of the EnKF machinery for hillslope-scale soil moisture estimation.

It is important to discuss an important experimental tool that is employed in each experiment. The use of Observing System Synthetic Experiments (OSSEs), in which the hydrological model is subjected to arbitrary realizations of the hydromterological forcings and parameterized by a single set of soil properties to generate a notionally true state and (through the measurement equations) observations, is motivated by two important factors. First, there are presently no spaceborne L-band microwave brightness temperature or backscatter observations, necessitating the use of the OSSE approach. Second, because the tRIBS-VEGGIE model is used in these OSSEs to generate the true state and observations, the performance of the EnKF can be assessed by the ability of the filter to retrieve the synthetic notionally true state and is independent of any structural model or observational errors or bias. The following sections of this chapter outline the experimental setup and assumptions that underly each data assimilation experiment, and present the associated results.

8.1 Experiment 1: Pixel-scale moisture estimation

8.1.1 Setup and assumptions

This first data assimilation experiment is designed to ascertain whether L-band microwave measurements would add value in estimation of the soil moisture state as simulated by tRIBS-VEGGIE under the most ideal of conditions. In this experiment, synthetic observations of L-band microwave brightness temperature are used to update the moisture state at a single, flat computational pixel that is comprised of 9 finite element layers.

A 27-day notionally true soil moisture state was simulated using tRIBS-VEGGIE by driving the model with the hydrometeorological forcings simulated for **Scenario 1** of the point-scale sensitivity analysis described in **Chapter 7** and providing as input a set of soil hydraulic and thermal properties (SHTPs) for an assumed loam soil drawn at random from an ensemble of soil parameter sets generated with the Restricted Pairing (RP)-based sampling scheme. Bare soil conditions are assumed in the true soil moisture, and throughout the pixel-scale estimation exercise. As a reminder to the reader, the time series of thermodynamic variables was simulated using the stochastic weather generator of Ivanov *et al.* [2007], conditioned on the input time series of hourly rainfall volues simulated with the Modified Bartlett-Lewis stochastic model. The seasonal and geographic context of these simulated hydrometeorological forcings is the month of July at the Tucson International Airport, Arizona, USA. Simulation begins at 0900 local time on July 1, and ceases at 0900 local time July 31.

Consistent with an approximate revisit time of NASA's forthcoming SMAP satellite, the microwave brightness temperature observing system outlined in **Chapter 4** was used to simulate brightness temperature observations in both the vertical and horizontal polarization every 72 hrs at 0900 local time based on the instantaneous true near-surface soil moisture simulated by tRIBS-VEGGIE. The assumed satellite sky position, consistent with the planned SMAP radiometer observing the

computational domain during the ascending limb of an orbit, is an azimuth angle (ζ_S) of 150° and an off-nadir look angle δ_S of 40°. Since the single pixel is flat, in this experiment the sensor off-nadir look angle is equivalent to the local angle of incidence, θ_L. The results are nine observations of the horizontally- and vertically-polarized L-band brightness temperature, the first occurring on July 4 at 0900 local time and the final observation occurring on July 31 at 0900 local time.

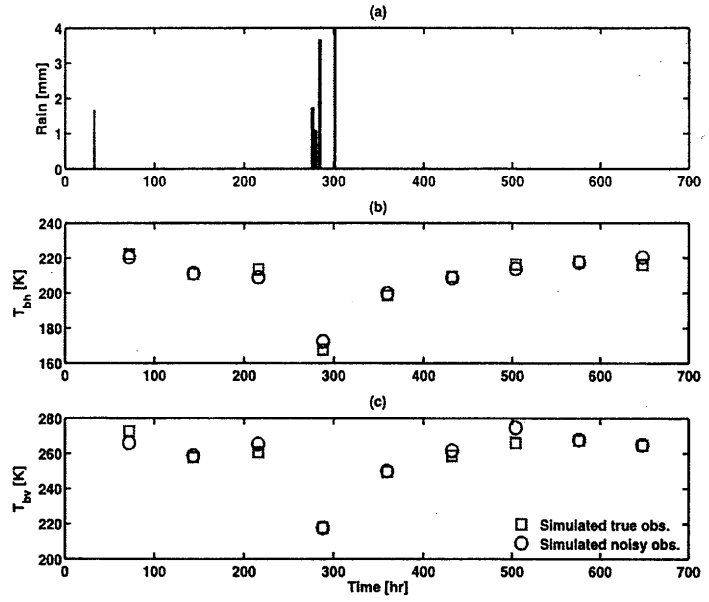


Figure 8-1: (a) True rain during experiment, (b) horizontally, and (c) vertically polarized microwave brightness temperature measurements simulated with the observing system developed in **Chapter 4**. Simulated observations based on the true state are shown as red squares, while the perturbed true observations are shown as blue circles.

It is assumed the instrument noise is additive and Gaussian, and that the errors in observation in the horizontal and vertical polarization are uncorrelated. The noise level in each polarization is zero mean and has a variance of 16 K^2, the required sensitivity of the SMAP radiometer. These assumptions define the structure of the observational error covariance matrix (R) in the EnKF algorithm. The simulated true brightness temperature observations are perturbed once with noise consistent with these assumptions to yield a set of simulated perturbed observations (figure 8-1).

In the data assimilation experiment, it is assumed uncertainty arises from both inadequate knowledge of the SHTPs and uncertainty in the hydrometeorological forcings required as input to tRIBS-VEGGIE. The 64-member ensemble of hydrometeorological forcings used in **Scenario 3** of the sensitivity analysis described in **Chapter 7** are used to drive the tRIBS-VEGGIE model for the nine EnKF analysis cycles. Similarly, a 64-member ensemble of soil parameters generated with the RP-based algorithm for an assumed loam soil in **Chapter 7** is used to parameterize the soil during the assimilation experiment. In this ensemble data assimilation experiment, each realization of the hydrometeorological forcings is paired to a corresponding realization of the soil parameters.

To recapitulate, the state vector in this first data assimilation experiment consists of the 10 soil moisture values at the boundaries of the 9 finite element layers of the soil column at the single computational pixel. Initial conditions for the EnKF experiment correspond to the moisture state at the end of a 27-day ensemble simulation.

An open loop (OL) ensemble simulation was also performed, assuming the same pairing between realizations of the hydrometeorological forcings and soil parameter sets as the data assimilation experiment, as well as the corresponding initial conditions. This OL experiment provides an ensemble soil moisture simulation during the same 27-day period that is independent of any observational data, since no data assimilation is performed in the OL experiment. Therefore, the OL ensemble is effectively a "worst case" scenario, against which the ability of the EnKF procedure to retrieve the true soil moisture state through assimilation of brightness temperatures can be assessed.

8.1.2 Pixel-scale results

Results of the OSSE for the single-pixel computational domain demonstrate the usefulness of the EnKF approach for estimating soil moisture in the near surface

(i.e., top 10 cm of the soil column) (figure 8-2) and in the entire profile (figure 8-3). Sequential assimilation of the brightness temperature observations reduces the ensemble standard deviation in the estimate of near-surface soil moisture, relative to the open loop simulation (figure 8-2(b) and (c)). In contrast to the filtered estimate, which benefits from ingesting observations, the ensemble standard deviation of the open loop estimate remains relatively unchanged during the duration of the experiment (figure 8-2(c)). Furthermore, assimilation of successive brightness temperature observations reduces the actual error in the filtered estimate of soil moisture, as measured by root-mean-square error (RMSE) relative to the true soil moisture time series (figure 8-2(d)). More specifically, sequential assimilation of the observations leads to increasingly improved near-surface moisture estimates, as evidenced by the gradual reduction in RMSE throughout the course of the experiment (figure 8-2(d)). The OL ensemble, on the other hand, largely exhibits the same RMSE in near-surface moisture throughout the simulation (figure 8-2(d)). At the end of the experiment, the RMSE in near-surface moisture of the OL ensemble is approximately twice that of the filtered estimate (figure 8-2(d)). Both the EnKF and OL estimates of near-surface soil moisture have relatively small bias, however, the occurrence of rainfall acts to increase bias in the near-surface soil moisture estimate (figure 8-2(e)). Although difficult to see, assimilation of observations leads to an immediate, although relatively small, decrease in soil moisture bias (figure 8-2(e)).

Turning to estimation of moisture throughout the soil column, although assimilation of the synthetic observations leads to a significant improvement in the estimate of soil moisture in the profile relative to the OL estimate, substantial improvement in the profile moisture estimate only occurs through sequential assimilation over several analysis cycles (figure 8-3). Initially, the true profile soil moisture is greater than the ensemble mean plus one ensemble standard deviation for both the OL and EnKF estimates (figures 8-3(b) and (c)). However, sequential assimilation of brightness temperature observations gradually improves the RMSE of the EnKF estimate of profile moisture, while the RMSE of the OL estimate effectively remains

the same throughout the simulation (figure 8-3(d)). The improvement in the EnKF estimate can also be seen through the gradual reduction in the bias of the EnKF estimate, which contrasts with a relatively unchanging bias in the OL estimate (figure 8-3(e)).

It it interesting to note that while the RMSE in the EnKF estimate decreases through sequential assimilation of brightness temperatures, the ensemble standard deviation of the EnKF estimate remains relatively constant throughout the experiment (figure 8-3(d)). At the beginning of the simulation, the fact that the EnKF standard deviation in profile moisture is substantially smaller than the EnKF RMSE in moisture estimate is a reflection of the fact that the estimate is initially biased. Hence, when the estimate of soil moisture is biased, then the ensemble standard deviation is a poor measure of error in the estimate, while the ensemble standard deviation is reasonable measure of estimator error when the underlying estimate is unbiased. In addition to the uncertainty associated with rainfall input, the difficulty in constraining bias in actual application of a soil moisture data assimilation system lies in the inability to adequately measure all of the outgoing moisture fluxes in a watershed. This will be discussed in more detail in the conclusion, with special emphasis paid to assimilation of discharge as a mass balance constraint on the assimilation system.

While this relatively simple pixel-scale experiment relies on a fairly contrived set of assumptions, it is nevertheless a valuable heuristic exercise to explore the impact of assimilation of measurements related to only near-surface variables on estimation of moisture throughout the entire profile. These results seem to suggest that, in the context of a hydrological process model, near-surface observations are of potential benefit to efforts to retrieve moisture in the entire profile. It should be pointed out, however, that the results of the pixel-scale experiment discussed above should not be construed as a general statement on assimilation of near-surface measurements for profile moisture estimation. Rather, for the particular combination of soil parameters drawn at random from an ensemble of possible parameter sets to

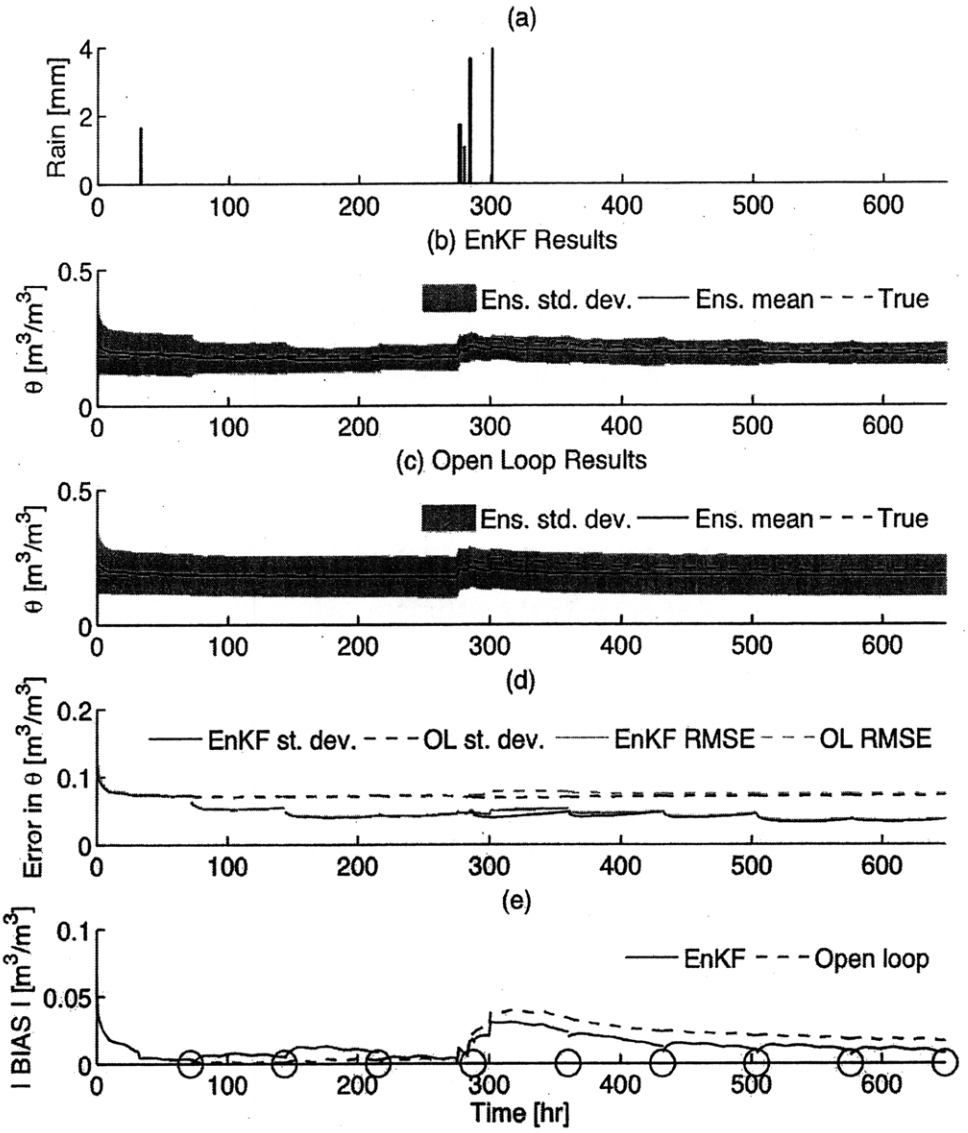


Figure 8-2: (a) True rain during experiment together with ensemble mean (solid black line), area encompassed by one standard deviation above and below the mean (gray area) and true (dashed black line) soil moistures for the (b) EnKF and (c) OL experiments. (d) Ensemble standard deviation of the EnKF (solid black line) and OL (dashed black line) ensembles and the RMSE of the EnKF (solid gray line) and OL (dashed gray line) estimates; and (e) the absolute value of the bias in the EnKF (solid black line) and OL (dashed black line) estimates. Observation times are indicated as solid black open circles on the x-axis of (e).

serve as the true soil parameters, the profile moisture content was reasonably well-retrieved by assimilation of near-surface brightness temperatures. Hence, a more

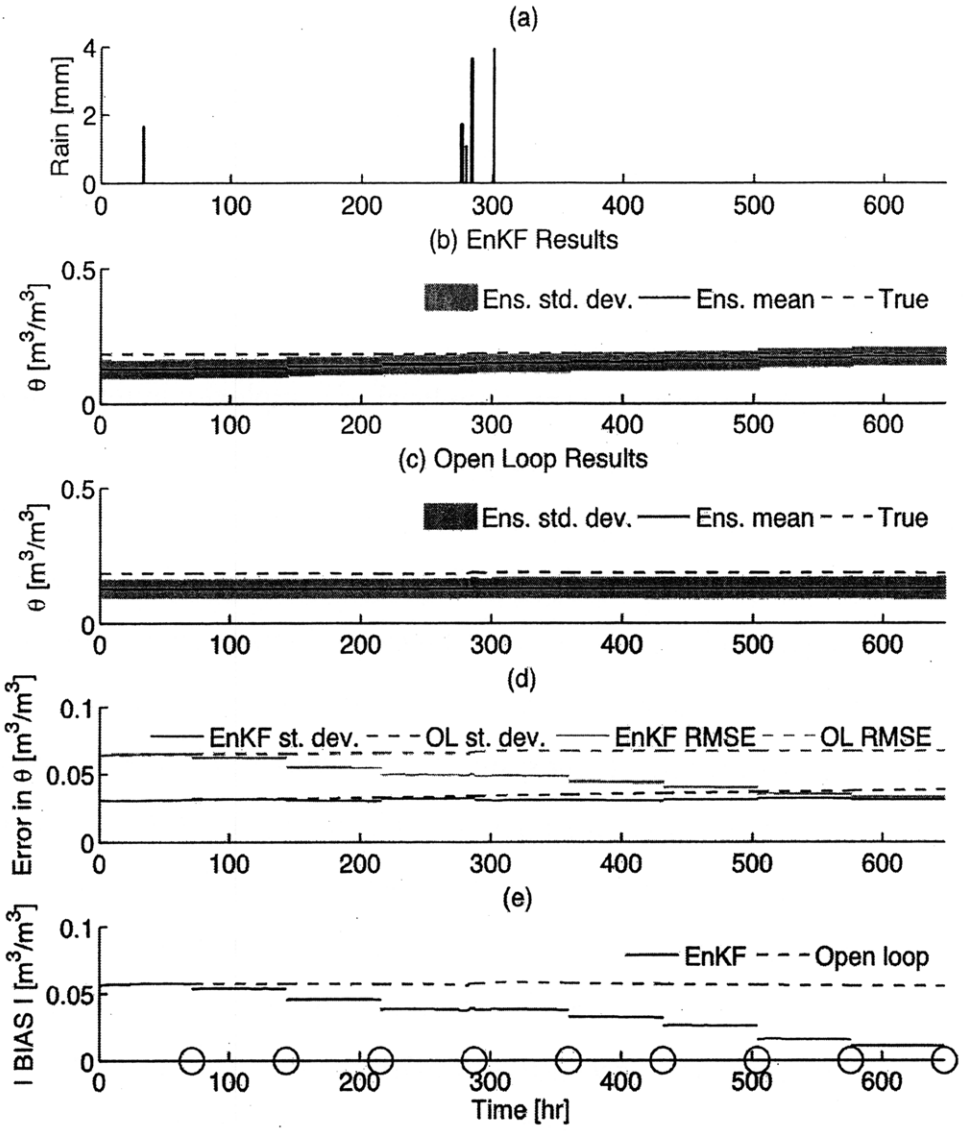


Figure 8-3: (a) True rain during experiment together with ensemble mean (solid black line), area encompassed by one standard deviation above and below the mean (gray area) and true (dashed black line) soil moistures for the (b) EnKF and (c) OL experiments. (d) Ensemble standard deviation of the EnKF (solid black line) and OL (dashed black line) ensembles and the RMSE of the EnKF (solid gray line) and OL (dashed gray line) estimates; and (e) the absolute value of the bias in the EnKF (solid black line) and OL (dashed black line) estimates. Observation times are indicated as solid black open circles on the x-axis of (e).

concise conclusion to this exercise would assert that, depending on the true soil parameters, assimilation of microwave brightness data can potentially be beneficial

in estimating the profile-integrated soil moisture.

8.2 Experiment 2: moisture estimation in a synthetic watershed

8.2.1 Setup and assumptions

This experiment represents an intermediate level of complexity in terms of spatial heterogeneity. The computational domain used in this experiment is a synthetic topographic field generated by a physically based landscape evolution model [Tucker and Bras, 1998; Tucker *et al.*, 2001b,a], and corresponding to a slope-dependent diffusive erosion mechanism. It is the same computational domain (referred to as the diffusive domain) used to demonstrate the properties of both the active and passive microwave observing systems in **Chapter 4**. The domain is characterized as having relatively shallow slopes, long hillslopes, and topographic relief that is significant but not extreme. The square domain is 2×2 km and contains 2401 active and approximately equally spaced computational nodes. At each of these computational nodes, 10 volumetric soil moisture states are modeled at the boundary of 9 discrete layers in the tRIBS-VEGGIE model. The watershed is assumed to be associated with a single soil textural class. While this computational domain is a simplified test bed, it offers greater complexity than the element domain considered above due to: (1) the additional process complexity associated with the potential for lateral sub-surface moisture redistribution, and (2) the increase in scale of the watershed and dimensionality of the associated state vector.

In similar fashion to the pixel-scale domain considered in the first experiment, a single 27-day notionally true soil moisture state was simulated using tRIBS-VEGGIE by driving the model with the hydrometeorological forcings simulated for **Scenario 1** of the point-scale sensitivity analysis of **Chapter 7**. Values of the SHTPs for this true simulation for an assumed sandy loam soil were drawn at random from an ensemble of soil parameter sets generated with the RP sampling scheme. Bare soil

conditions are again assumed in the simulation to develop the synthetic true soil moisture state, as well as in the ensemble data assimilation exercise. The simulation begins at 0900 local time on July 1, and ceases at 0900 local time July 31.

Synthetic microwave brightness temperature observations were again simulated based on the synthetic true soil moisture state using the passive observing system developed in **Chapter 4**. Simulated brightness temperatures in both the vertical and horizontal polarization were derived based on the instantaneous hillslope-scale distribution of soil moisture and temperature every 72 hrs at 0900 local time. Continuing with the assumption from the pixel-scale experiment, the satellite sky position is characterized by assumed values of azimuth angle (ζ_S) equal to 150° and an off-nadir look angle δ_S equal to 40°. Based on equation 4.34 and the local topographic conditions, the hillslope-scale distribution of simulated brightness temperatures is aggregated to the watershed scale.

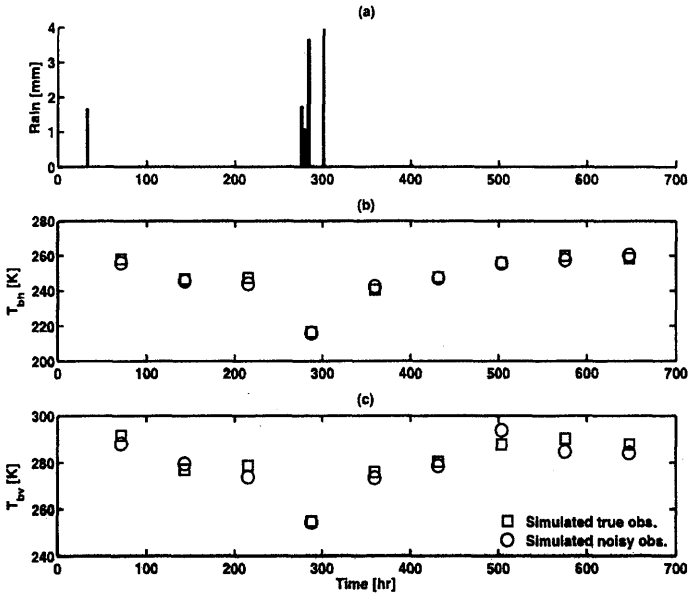


Figure 8-4: (a) True rain during experiment, (b) horizontally, and (c) vertically polarized microwave brightness temperature measurements simulated with the observing system developed in **Chapter 4**. Simulated observations based on the true state are shown as red squares, while the perturbed true observations are shown as blue circles.

Instrument noise is assumed again to be additive and Gaussian, and observa-

tional errors in the horizontal and vertical polarization are assumed uncorrelated. Assuming the observational errors are zero mean and have a variance of 16 K^2, the simulated true brightness temperature observations are perturbed once with noise consistent with these assumptions. This yields a set of watershed-scale simulated perturbed observations, shown together with the simulated true brightness temperature observations in figure 8-4.

Uncertainty in the hillslope-scale distribution of soil moisture is again assumed to arise from both inadequate knowledge of the SHTPs and uncertainty in the hydrometeorological forcings. The 64-member ensemble of hydrometeorological forcings used in **Scenario 3** described in **Chapter 7** and in the pixel-scale experiment above are used to drive the tRIBS-VEGGIE model for the nine EnKF analysis cycles. Spatial variability in rainfall over the synthetic watershed is assumed to be negligible, and hourly rainfall pulses are uniformly distributed in space. A 64-member ensemble of the set of soil parameters required by tRIBS-VEGGIE is generated using the RP-based algorithm for an assumed sandy loam soil. Again, each realization of the hydrometeorological forcings is paired to a corresponding realization of the soil parameters.

The state vector in this data assimilation experiment consists of the 10 soil moisture values characterizing the surface-normal soil moisture profile at each of the 2401 computational pixels in the diffusive erosion domain. Thus, the state vector for each ensemble replicate is 24010 by 1. The initial state vector for each replicate during the assimilation exercise was determined in a similar fashion to the initial state in the pixel-scale assimilation experiment. Because the size of the state vector is much larger than the number of observations at each analysis, this experiment tests the suitability of the EnKF to disaggregate the watershed-scale observations of brightness temperature by conditioning on the model physics, as embodied in a matrix containing an ensemble characterization of the soil moisture state.

Like the pixel-scale experiment, an open loop (OL) ensemble simulation was to derive a "worst case" scenario, against which the performance of the EnKF estimates

of hillslope-scale soil moisture can be compared.

8.2.2 Synthetic domain results

As seen in Figure 8-4, the fourth analysis cycle, which ends at 288 hr into the simulation is associated with a series of rainfall pulses. The rainfall event depresses the synthetic brightness temperatures approximately 25 K in the horizontal polarization and approximately 20 K in the vertical polarization, relative to the synthetic observations at hour 216. On the other hand, the final analysis cycle concludes at 648 hr into the simulation during a significantly long drying event. Results from these two analyses are discussed in detail here. Full graphical depictions of every analysis in the assimilation experiment are omitted here for brevity, but shown unabridged in **Appendix D.**

The EnKF estimate of the hillslope-scale distribution of profile-integrated soil moisture for the fourth analysis cycle (288 hr) shows a significant topographic signature, with channel and valley bottoms exhibiting higher moisture and topographically divergent areas (ridge lines) exhibiting the lowest moisture content (figure 8-5(a)). It should be noted, however, that there is not a large range of variability in ensemble mean profile moisture content throughout the watershed. The spatial variability of the EnKF mean near-surface moisture content, on the other hand, shows a wider range of variability in moisture content (figure 8-5(b)). Volumetric moisture content in the near surface, because of the relatively recent occurrence of rainfall, is substantially higher in the near surface compared with the profile average. Spatial patterns in EnKF mean near-surface soil moisture content are also largely organized around topographic gradients, with channels and valley bottoms exhibiting higher wetness (figure 8-5(b)). In contrast to the spatial distribution of EnKF mean profile-integrated moisture, however, topographically divergent areas are not the locations with the lowest EnKF mean near-surface soil moisture. Rather, the regions associated with the lowest ensemble mean near surface soil moisture are the locations with the largest slope within the domain (figure 8-5(b)). This

observation may underscore the role of lateral subsurface moisture redistribution as a constraint on the spatial distribution of ensemble mean near-surface moisture content. The spatial distribution of the standard deviation in profile-integrated moisture content bears a striking resemblance to the spatial distribution of ensemble mean profile moisture content (figure 8-5(c)). Areas of topographic convergence are associated with higher local variance in profile-integrated soil moisture, while areas of topographic divergence are associated with the lowest local variance in profile-integrated moisture content (figure 8-5(c)). The range of variability in the local ensemble standard deviation in profile-integrated moisture is relatively narrow, highlighting an additional similarity with the spatial pattern of ensemble mean profile moisture. While the spatial patterns of ensemble mean and variance in profile-integrated moisture content appear quite similar, the spatial distribution of variance in near-surface moisture content is markedly different from the corresponding pattern of ensemble mean near-surface moisture (figure 8-5(d)). Furthermore, although the variance of in near-surface moisture content is generally greater than the corresponding value of variance in profile-integrated moisture, the spatial structure is similar. Topographically divergent areas tend to exhibit lower values of ensemble variance in near-surface moisture content, while the channel and valley bottoms exhibit the highest values of near-surface moisture ensemble variance in the landscape (figure 8-5(d)). The steepest portions of the domain seem to be associated with moderate to high values of variance in near-surface moisture (figure 8-5(d)).

At the conclusion of the final analysis cycle (648 hr), after a significant amount of time has passed without rainfall, the spatial distribution of EnKF mean profile-integrated moisture is largely organized around topographic variability (figure 8-6(a)). Again, channel and valley bottoms are associated with higher EnKF mean profile moisture, while topographically divergent areas are associated with lower moisture (figure 8-6(a)). The profile-integrated moisture content at this analysis is also higher, and exhibits a greater range of variability than the spatial distribution of EnKF mean profile moisture at the fourth analysis cycle (figure 8-5(a)).

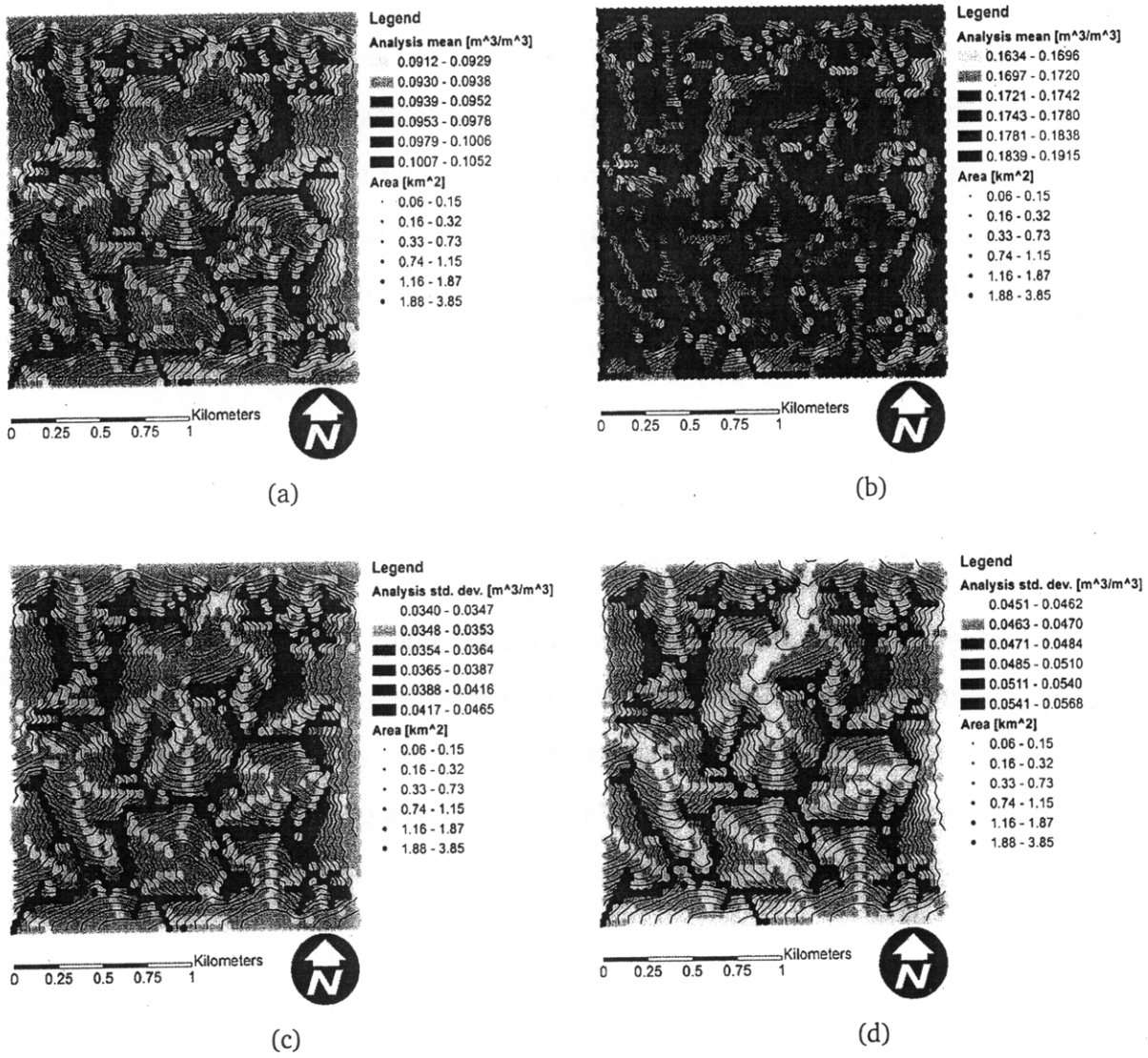


Figure 8-5: Analysis output, cycle 4 (288 hr). (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture. To maximize contrast, intervals are shown as quantiles.

The range of variability in the spatial distribution of EnKF mean near-surface moisture is, by contrast with the results from the fourth analysis, more similar to the profile-integrated moisture values (figure 8-6(b)). This indicates that the near-surface, because of a significantly long period without rain, is not substantially dissimilar from the remainder of the profile. The spatial variation of EnKF mean near-surface moisture is similar in spatial structure to the corresponding distribution in near-surface moisture at the fourth analysis. Specifically, topographically

convergent areas exhibit higher EnKF mean near-surface moisture, while areas with steeper slopes tend to be associated with lower near-surface moisture content (figure 8-6(b)). Topographic ridge-lines as areas of the lowest values of EnKF mean near-surface moisture seem to be more prominent during the final analysis, when compared to the fourth analysis cycle (figure 8-6(b)). In a pattern that is similar to the fourth analysis cycle, the spatial pattern of EnKF standard deviation in profile-integrated moisture content is similar to the distribution of EnKF mean profile moisture content (figure 8-6(c)). The spatial pattern of EnKF variance in near-surface moisture, while organized along topographic gradients, exhibits an interesting spatial structure. Similar to the spatial pattern in near-surface variance at the fourth analysis cycle, the channel and valley bottoms exhibit higher variance in near-surface moisture (figure 8-6(d)). However, steep and West-facing slopes exhibit moderate variance in near-surface moisture content while South- and East-facing slopes exhibit much lower values of variance in near-surface moisture (figure 8-6(d)). This pattern bears some resemblance to patterns of incoming solar radiation, particularly in the morning when the Sun is low in the Eastern sky, and could potentially be related to the spatial pattern of evaporation.

The notion that, in semiarid locations in particular, the greatest ensemble variance in soil moisture could be associated with the topographically convergent areas within the landscape was discussed in the context of the spatially distributed sensitivity analysis outlined in **Chapter 7**. It was argued there that this could be a result of the uncertainty in soil hydraulic properties, particularly residual and saturation moisture content and saturated hydraulic conductivity. The rationale posits that at topographically divergent areas (i.e., ridgelines), uncertain rainfall is partitioned into infiltration and runoff at uncertain rates that depend on the soil textural class and the uncertain antecedent moisture. Further, the lateral redistribution and runon mechanisms in tRIBS-VEGGIE lead to an uncertain amount of runoff and infiltrated water being redistributed downslope, to a topographically lower pixel. This pixel, in turn, is also exposed to uncertain incoming rainfall and uncertain soil moisture conditions. This topographic amplification of uncertainty in precipitation,

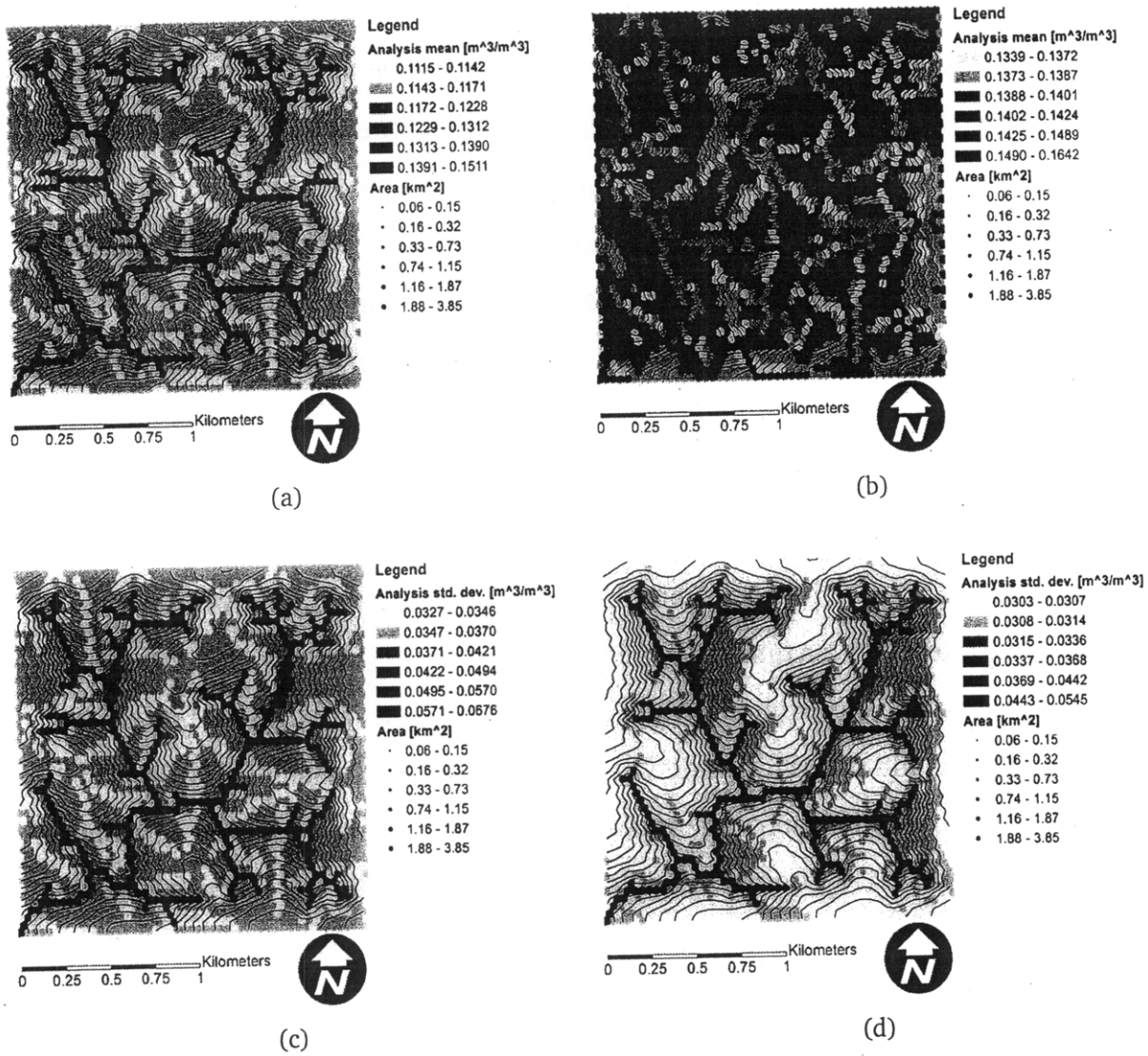


Figure 8-6: Analysis output, cycle 9 (648 hr). (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture. To maximize contrast, intervals are shown as quantiles.

runoff, and lateral redistribution in a downstream direction results in higher ensemble variance in the topographic lows of the domain. This reasoning is inextricably bound to the fact that moisture redistribution in the subsurface and runon on the surface are parameterized as slope-dependent processes in tRIBS-VEGGIE, which results in areas of topographic convergence exhibiting a higher ensemble mean soil moisture content. Hence, while the topographic variation of a watershed may provide some qualitative insight into the spatial distribution of hillslope-scale moisture

content, the amplification of uncertainty in moisture by topography correspondingly decreases the predictability in areas of topographic convergence. Again, this rationale may hold for semiarid regions in which the water table is at significant depth beneath the surface and channels are mostly ephemeral, in more humid regions groundwater-surface water interaction would tend to reduce amplification of uncertainty by topography. In humid areas, downslope boundary conditions would correspond to saturation, thereby limiting the amount of upslope runoff that could re-infiltrate downslope and suppressing subsurface lateral redistribution in the unsaturated zone.

At the fourth analysis cycle, the EnKF estimate of profile-integrated and near-surface soil moisture is less error prone than the OL estimate, relative to the synthetic true state. The maximum RMSE in the EnKF estimate of profile-integrated moisture content is approximately 0.085 [m^3/m^3] (figure 8-7(a)), whereas the corresponding maximum RMSE in the OL estimate is approximately 0.102 [m^3/m^3] (figure 8-7(c)). Furthermore, the minimum RMSE in the OL estimate of profile-integrated moisture is approximately 0.084 [m^3/m^3] (figure 8-7(c)), which is equivalent to the greatest RMSE in the EnKF estimate of profile moisture content. In both the EnKF and OL cases, the highest RMSEs occur in the channel and valley bottoms (figures 8-7(a) and (c)). Similar conclusions can be reached for the estimate of soil moisture in the near-surface. The highest RMSE in the EnKF estimate of near-surface moisture is approximately 0.056 [m^3/m^3] (figure 8-7(b)), while the lowest RMSE in the OL estimate is approximately 0.057 [m^3/m^3] (figure 8-7(d)). Again topographically convergent areas exhibit the highest values of RMSE, while topographically divergent areas exhibit the smallest RMSE values (figures 8-7(b) and (d)).

At the ninth analysis, which corresponds to the end of the experiment, it remains apparent that the EnKF estimate of profile-integrated and near-surface soil moisture is less error prone than the OL estimate, as measured by the RMSE relative to the synthetic true state. The maximum RMSE in the EnKF estimate of

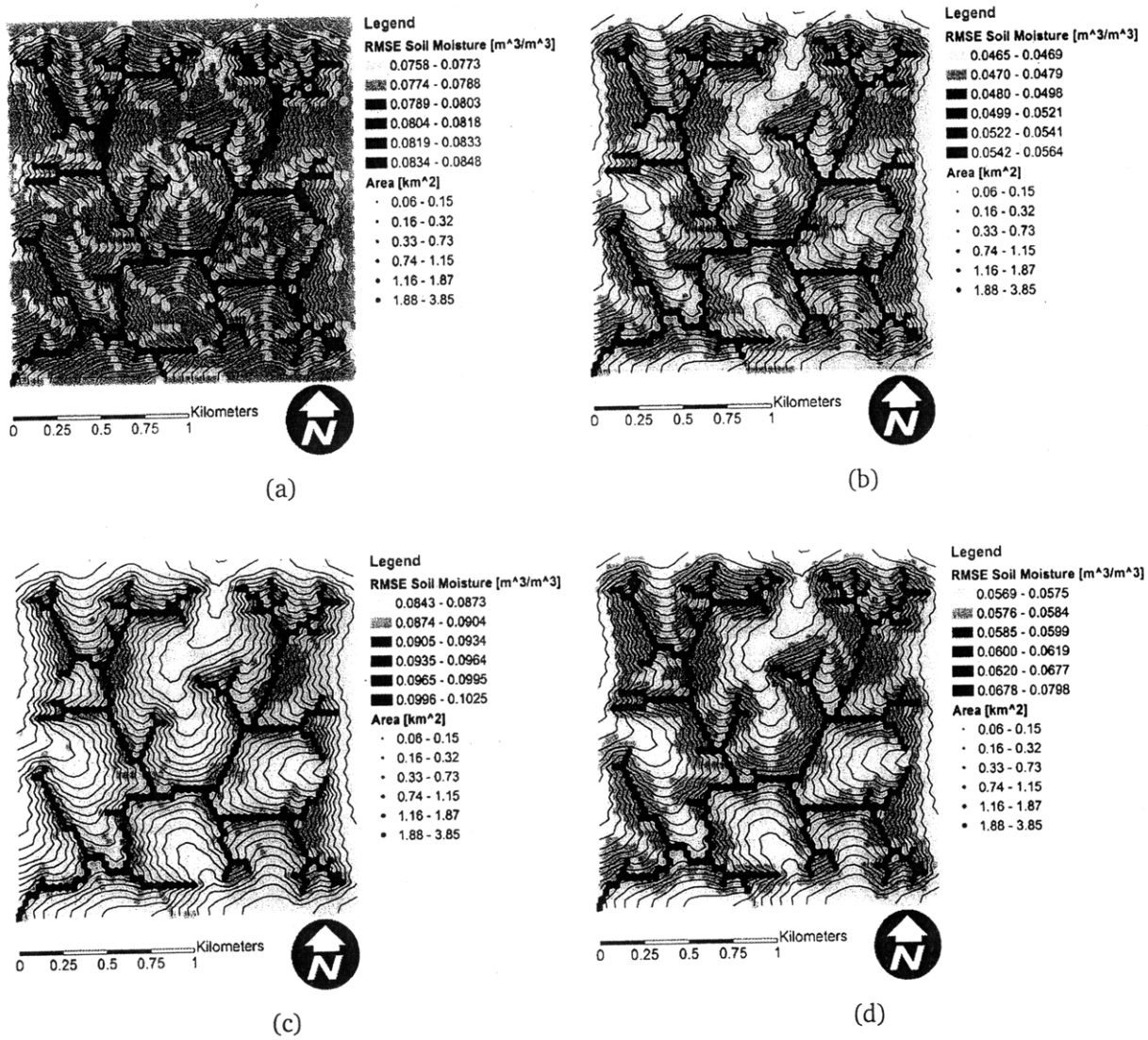


Figure 8-7: RMSE relative to truth, analysis cycle 4 (288 hr). (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture. To maximize contrast, intervals are shown as quantiles.

profile-integrated moisture content is approximately 0.072 [m^3/m^3] (figure 8-8(a)), whereas the minimum RMSE in the OL estimate is approximately 0.080 [m^3/m^3] (figure 8-8(c)). The spatial structure of RMSE errors appears nearly identical for both the EnKF and OL simulations. The channel and valley bottoms exhibit the greatest RMSE, while the surrounding hillslopes exhibit the smallest (figures 8-7(a) and (c)). Interestingly, in terms of the profile-integrated estimate both the EnKF and OL estimates show a strong contrast in RMSE between channels and hillslopes. The highest RMSE in the EnKF estimate of near-surface moisture at the end of the

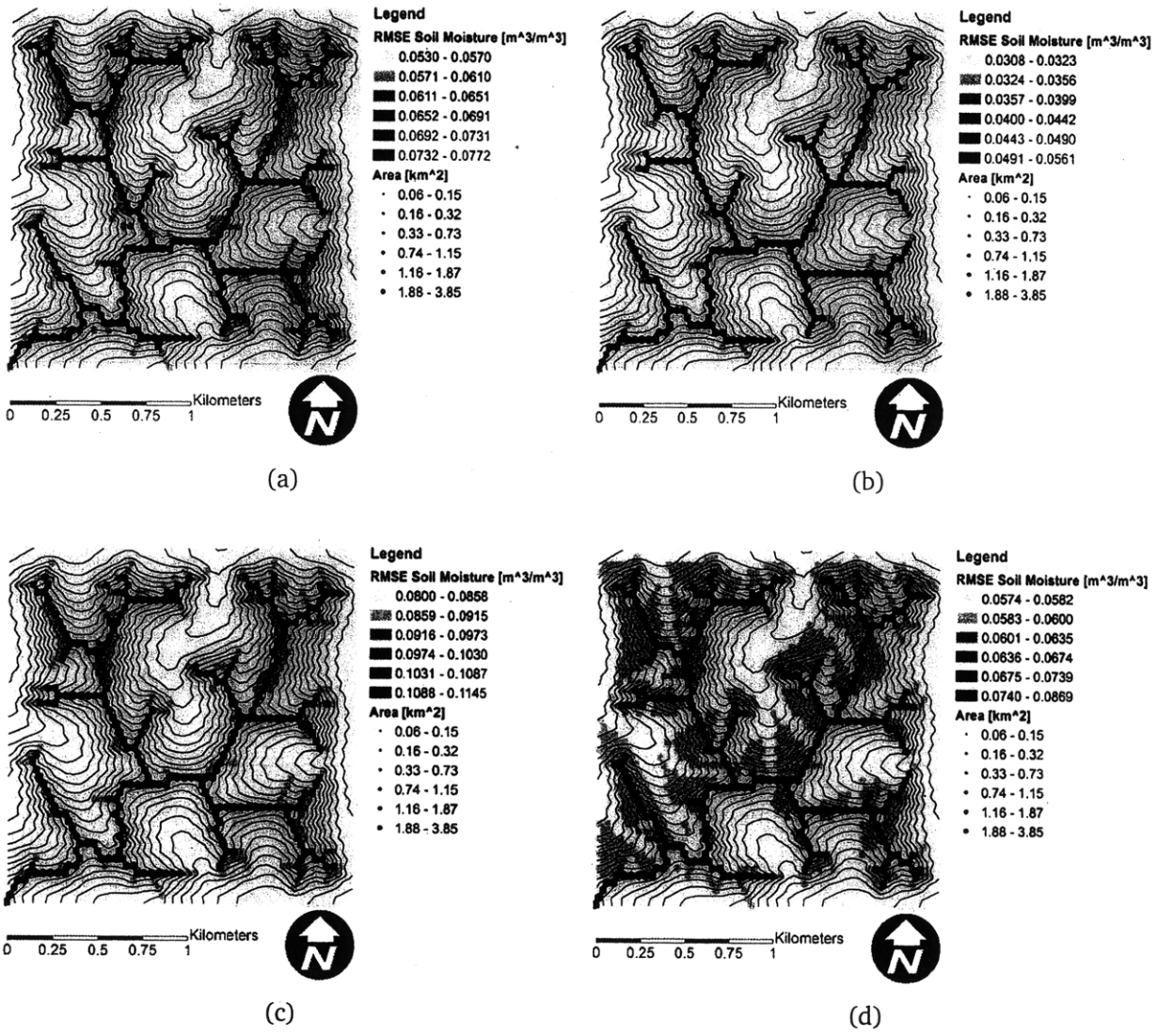


Figure 8-8: RMSE relative to truth, analysis cycle 9 (648 hr). (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture. To maximize contrast, intervals are shown as quantiles.

experiment is approximately 0.056 [m^3/m^3] (figure 8-7(b)). By comparison, the lowest RMSE in the OL estimate is approximately 0.057 [m^3/m^3] (figure 8-7(d)). Again topographically convergent areas exhibit the highest values of RMSE, while topographically divergent areas exhibit the smallest RMSE values (figures 8-7(b) and (d)).

Of particular interest in this spatially distributed data assimilation exercise is how the new information contained in the observations is distributed across the

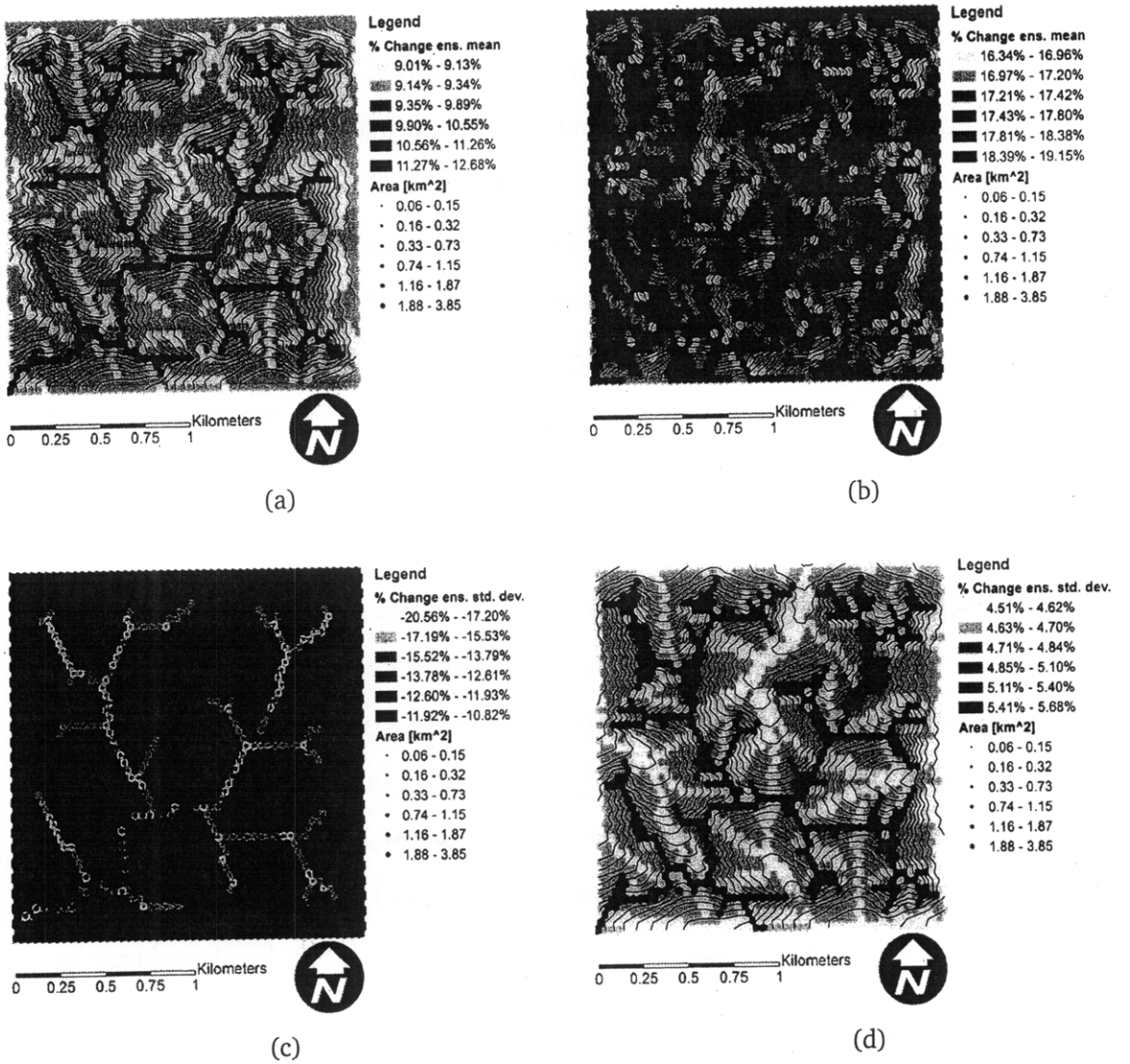


Figure 8-9: Analysis cycle 4 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation. To maximize contrast, intervals are shown as quantiles.

landscape through the EnKF update process, and the sensitivity of the analysis to sharp changes in the observations (such as the depression in the brightness temperature at fourth analysis cycle). The EnKF update increased the ensemble mean in the estimate of profile-integrated soil moisture by anywhere from 9 to 12.7 percent at the fourth analysis (figure 8-9(a)). The change in the ensemble mean near-surface moisture estimate as a result of the EnKF update was substantially greater,

ranging from 16.3 to 19.2 percent (figure 8-9(b)). In both cases, the largest analysis increments are found in the channel and valley bottoms (figures 8-9(a) and (b)). Because the EnKF update increased the mean near-surface and profile-integrated moisture everywhere in the watershed, the observations act to add moisture to the ensemble. This addition of moisture through assimilation of the watershed-scale brightness temperature may reflect some combination of the nature of the true soil properties and hydrometeorological forcings. For instance, a true precipitation that is large relative to the ensemble behavior of the uncertain rainfall forcings may lead to greater soil moisture in the true state. Additionally, the true soil properties may be such that infiltration capacity is greater than the ensemble mean behavior.

Investigating the reduction in the local ensemble variance due to the EnKF update reveals a substantial decrease in the EnKF estimate of variance in profile-integrated moisture content, reflecting improved confidence in the profile-integrated moisture estimate (figure 8-9(c)). The largest decreases in local variance conform to the organization of valleys and channels in the watershed, while the smallest decreases are found largely in areas with the largest topographic slope or areas of topographic divergence (figure 8-9(c)). On the other hand, the local variance in near-surface moisture actually increases between 4.5 and 5.7 percent due to the EnKF update (figure 8-9(d)). The increase in the local ensemble variance in near surface moisture underscores the fact that the EnKF, unlike the Kalman Filter from which it originated, may not always result in a reduction of uncertainty in all states because the assumptions of linearity and Gaussianity that underlie the Kalman update may not be met in an ensemble framework. While difficult to reconcile with the corresponding decrease in variance in the EnKF profile-integrated moisture estimate, it is important to recognize that the near-surface moisture is computed using the top 3 soil moisture states of the profile. Hence this modest increase in local variance is confined to roughly 30% of the state space.

At the end of the simulation, during a prolonged drying period, the EnKF is nevertheless useful in correcting errors in both near-surface and profile-integrated

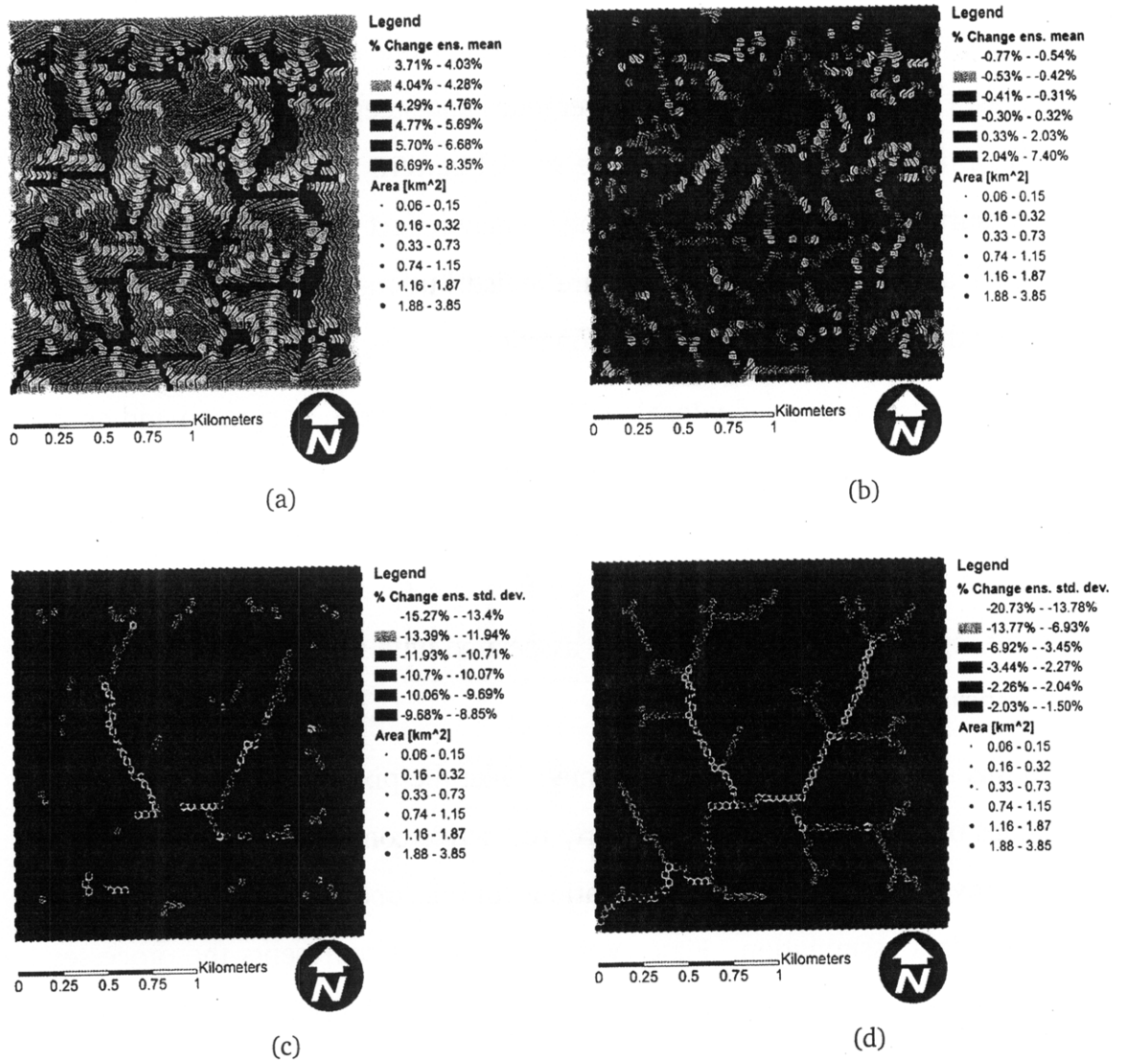


Figure 8-10: Analysis cycle 9 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation. To maximize contrast, intervals are shown as quantiles.

estimates of soil moisture. The update is associated with a modest increase in the ensemble mean estimate of profile moisture (figure 8-10(a)), while being associated with both increases and decreases in the ensemble mean estimate of near-surface moisture (figure 8-10(b)). Increases in mean profile-integrated moisture associated with the update range from approximately 3.7 to 8.3 percent, with channel and valley bottoms being the locations with the largest increases in the ensemble mean

profile moisture (figure 8-10(a)). Ensemble mean near-surface soil moisture increases in the valley bottoms across the update step, while mean near-surface moisture decreases in areas of topographic divergence and steep slope (figure 8-10(b)). During this prolonged drying episode the spatial pattern of changes in ensemble mean near-surface moisture due to the update may reflect the differences between the ensemble and true behavior of moisture redistribution, potentially revealing to some degree the nature of the true soil properties.

The EnKF update reduces the local variance in both profile-integrated and near-surface soil moisture estimates (figures 8-10(c) and (d)). The reduction of local variance due the update ranges from 8.9 to 15.3 percent for the profile moisture estimate and 1.5 to 20.7 percent for the near-surface moisture estimate. In both cases, the areas of topographic convergence are associated with the greatest decrease in local soil moisture variance.

While this spatially distributed experiment, like the pixel-scale experiment before it, relies on assumptions that are fairly restrictive compared with actual conditions, it conveys some important implications for hillslope-scale moisture estimation through data assimilation. First, with a model that represents the processes responsible for moisture redistribution in the landscape, identification of plausible sources of uncertainty in the modeled soil moisture state, and an adequate representation of that uncertainty, the EnKF approach seems promising as a mechanism for translating geophysically observed quantities from one scale to modeled-soil moisture states at another. Second, the local variance in soil moisture, particularly the profile-integrated estimate, is not always decreased. Because the EnKF approach employs an ensemble-based estimate of the covariance structure in the update step, it can be sensitive to sampling errors associated with small ensemble sizes. In particular, spurious correlations (particularly at depth within the soil column) that occur because of a relatively small ensemble sizes could potentially result in analysis increments that, while small, are in the opposite direction of the true moisture. When considering the profile-integrated moisture state, erroneous

analysis increments are amplified because deeper soil layers are associated with greater thickness in the tRIBS-VEGGIE computational mesh. Hence, while errors in the volumetric moisture state at depth may be relatively small, they may represent a substantial error when integrated over their respective soil layer thickness. Although this may be a shortcoming of a Kalman-like update step, it is tempered by the ease with which the EnKF can be implemented and the improvement in the estimate that is more frequently observed. Furthermore, this shortcoming could conceivably be overcome with no changes to the current framework by using larger ensembles.

As with the pixel-scale experiment, it should be noted that these results reflect the particular combination of soil parameters drawn at random to serve as the true soil parameters. For the set of parameters and forcings chosen to represent truth in this experiment, the EnKF approach appears beneficial in reducing the error in the estimate of both profile-integrated and near-surface moisture content. Furthermore, the following experiment makes an effort to account for potential variation in the true parameters and forcings.

8.3 Experiment 3: moisture estimation in WGEW

8.3.1 Setup and assumptions

This final experiment seeks to establish the most realistic set of assumptions for an effort to estimate soil moisture at hillslope-scales through assimilation of microwave observations. The setting considered is the Walnut Gulch Experimental Watershed, maintained by the US Department of Agriculture, and employed for the spatially-distributed sensitivity analysis performed in **Chapter 7**. This experiment extends hillslope-scale data assimilation to an environment exhibiting: (1) topographic variability, (2) soil textural class heterogeneity, and a spatial extent large enough to exhibit spatial variability in (3) rainfall and (4) radar backscatter obser-

vations.

Because the conclusions of the first two experiments are somewhat specific to a single realization of the true soil moisture state, as simulated by tRIBS-VEGGIE under the influence of a particular combination of SHTPs and hydrometeorological forcings, this experiment considers four potential realizations of the soil moisture state. Therefore, four 27-day realizations of the true soil moisture state were simulated using tRIBS-VEGGIE for four different combinations of hydrometeorological forcings and SHTPs. A description of a reclassification of soil textural categories to facilitate the use of the RP procedure to generate stochastic realization of the soil parameters require by tRIBS-VEGGIE was described in **Chapter 7**. The spatial boundaries of soil units (i.e., areas of relatively homogeneous soil texture) were obtained from the Soil Survey Geographic (SSURGO) database maintained and published by the USDA and the distribution of reclassified soil textures shown in the previous chapter (figure 7-6). Values of the SHTPs for each of the four true simulations, constrained to the spatial organization of reclassified soil textures shown in figure 7-6, were drawn at random from an ensemble of soil parameter sets generated with the RP sampling scheme. Bare soil conditions are again assumed in this final data assimilation experiment. Hydrometeorological forcings for the synthetic true simulations were obtained by simulating an additional four realizations based on the 1024 realizations of hydrometeorological forcings used for the WGEW sensitivity analysis in the previous chapter. Specifically, the stochastic multiplicative cascade is used both to perturb the four simulated true hourly rainfall time series and to disaggregate hourly rainfall pulses to a 4×4 km grid overlaid upon the watershed. The four time series of hourly simulated rainfall are supplied as input to the stochastic weather generator to derive four corresponding time series of thermodynamic forcings. The ensemble average behavior of the hydrometeorological forcings was shown in figure 7-7. The true simulations begin at 0900 local time on August 1, and cease at 0900 local time August 31.

Four time sequences of L-band (1.26 GHz) microwave radar backscatter obser-

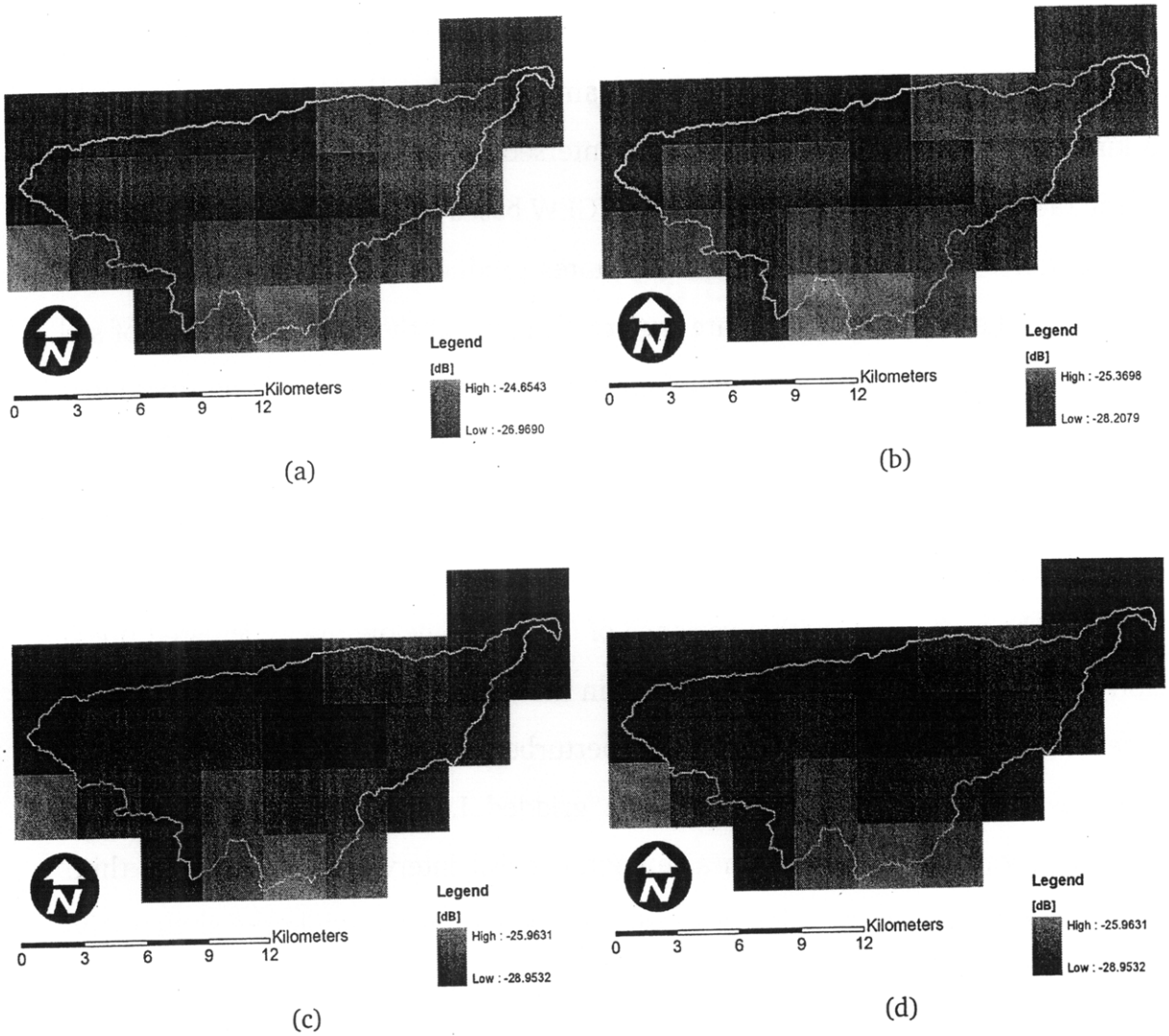


Figure 8-11: Average observations across the four candidate observations in the (a) horizontally-copolarized state at the third analysis (216 hr), (b) vertically-copolarized state at the third analysis (216 hr), (b) horizontally-copolarized state at the sixth analysis (432 hr), and (d) vertically-copolarized state at the third analysis (432 hr).

vations are simulated at 72 hr intervals based on the synthetic true soil moisture states at 0900 local time using the active observing system developed in **Chapter 4**. Simulated radar backscatter in both the horizontally- and vertically-copolarized states was simulated at each computational pixel in WGEW and resampled to a 3 $\times$ 3 km grid, consistent with the anticipated resolution of the SMAP level 1 radar product. Continuing with the assumptions from the previous two experiments, t

it is again assumed that the satellite azimuth angle (ζ_S) equals 150° and the off-nadir look angle δ_S equals 40°. Allowing for the presence of radar pixels not falling completely within WGEW, the spatial resampling of synthetic observations to the 3 km square grid reveals that WGEW is intersected by 30 pixels. Hence, the synthetic radar backscatter observations of WGEW based on the true states in both the horizontally- and vertically-copolarized states produces 60 observations every 72 hrs. Since there are four candidate true realizations of the true distribution of soil moisture, there are correspondingly four time sequences of synthetic observations.

Following the planned specifications of the SMAP radar instrument, the noise is assumed to be additive and Gaussian, and observational errors in the horizontally- and vertically-copolarized are assumed uncorrelated. Observational errors are assumed to be zero mean, with a variance of 0.25^2 dB^2, below the anticipated precision of the SMAP radar. Each observation in the collection of synthetic observations based on the four true moisture states is perturbed once with noise consistent with these assumptions. This yields a synthetic, gridded, L-band radar product at a spatial resolution of 3 km and with a temporal revisit interval of 72 hrs. The third analysis (216 hr) occurs amidst a substantial rainfall event, while the sixth analysis (432 hr) occurs during an extended drying episode, providing a two different wetness regimes that will provide a useful contextual contrast throughout discussion of this data assimilation experiment. The average observation across the four candidate observations in both the horizontally- and vertically-copolarized states is shown in figure 8-11 for both the third and sixth analysis. Average observations across the four candidate sets of observations are shown for the other 7 analysis cycles in **Appendix D**.

For each of the four candidate sequences of observations, the EnKF estimation procedure was applied with an ensemble size of 256 replicates. The SHTPs for the ensemble are the first 256 of the 1024 soil parameter sets for WGEW generated for the sensitivity analysis in the previous chapter. Similarly, the time series of hydrometeorological forcings for the ensemble are the first 256 of the 1024 realizations

of hydrometeorological forcings simulated with the stochastic rainfall and weather models for the spatially distributed sensitivity analysis in the previous chapter. Finally, the initial conditions for the 256 replicates of the EnKF experiments are the same initial conditions for the first 256 of 1024 initial conditions for the sensitivity analysis in the previous chapter. Across the four EnKF experiments, the paired realizations of hydrometeorological forcings and soil parameters are maintained. Thus, between the four experiments only the observations differ.

Similar to the previous experiment, the state vector in this synthetic experiment is much larger than the number of observations: 194,470 by 1 for the state vector compared with 60 by 1 for the vector of observations. Hence, this experiment also tests the suitability of the EnKF to disaggregate the coarse observations to hillslope scales by conditioning on the model physics. This experiment, however, encompasses a much broader range of spatial variability due to the heterogeneity in soil textural classes, rainfall, and radar backscatter observations than the previous experiment which considered only spatial variability in topography.

In this final data assimilation experiment, the OL ensemble used to characterize the "worst case" hillslope-scale soil moisture estimation scenario is the 1024-replicate ensemble simulation used in the sensitivity analysis in the previous chapter.

8.3.2 Results: spatially distributed

As discussed above, the third and sixth analyses (216 and 432 hrs into the simulation, respectively) provide a contrast in overall wetness regimes, and are the focus of the discussion here. The unabridged results from all 9 analyses is presented in Appendix D.

At the third analysis (216 hr), which is temporally concurrent with a significant rainfall event, the EnKF mean profile-integrated soil moisture reflects an interplay between soil type heterogeneity and topography (figure 8-12(a)). The average

EnKF mean profile moisture estimate varies between 0.09 and 0.33 [m^3/m^3], and while soil type heterogeneity is the most powerful descriptor of the profile moisture content (see figure 7-6 for the distribution of soil types), topography seems to be the most important factor in describing spatial variability in the EnKF profile moisture estimate within a particular soil type (figure 8-12(a)). By comparison, average near-surface ensemble mean soil moisture ranges from 0.15 to 0.40 [m^3/m^3], reflecting the coincident occurrence rainfall throughout the watershed (figure 8-12(b)). Spatial heterogeneity in soil types again explains the most marked contrasts in the near-surface moisture average EnKF estimate, but spatial variability in rainfall plays a significant role in the variability of EnKF mean near-surface moisture within soil types (figure 8-12(b)). This is evidenced in figure 8-12(b) as the square features in the Northwest and North-central portions of WGEW. Topography also plays a minor role in explaining the spatial heterogeneity in EnKF mean near-surface moisture, particularly in the Western-most and Southwestern portions of the watershed (figure 8-12(b)).

Throughout the WGEW, the EnKF estimate of profile-integrated moisture is relatively accurate, with more than half of the watershed exhibiting an average RMSE[1] in profile moisture of 0.06 [m^3/m^3] or less (figure 8-12(c)). Consistent with the second EnKF experiment, the channel and valley network tends to be associated with higher average RMSE, although the spatial pattern of average RMSE in profile moisture estimates also exhibits a dependence on soil type heterogeneity (figure 8-12(c)). The average accuracy of the EnKF estimate in near-surface soil moisture, on the other hand, is comparatively better than the profile-integrated estimate. The majority of the watershed exhibits an average RMSE of 0.04 [m^3/m^3] or less, and all but a few pixels exhibit average RMSE in the near-surface moisture estimate of 0.07 [m^3/m^3] (figure 8-12(d)). Again, topographically low portions of the watershed seem to be associated with highest RMSE in the near-surface estimate (figure 8-12(d)).

[1]The term average RMSE is meant as the spatial distribution of local RMS errors averaged across each of the four EnKF experiments.

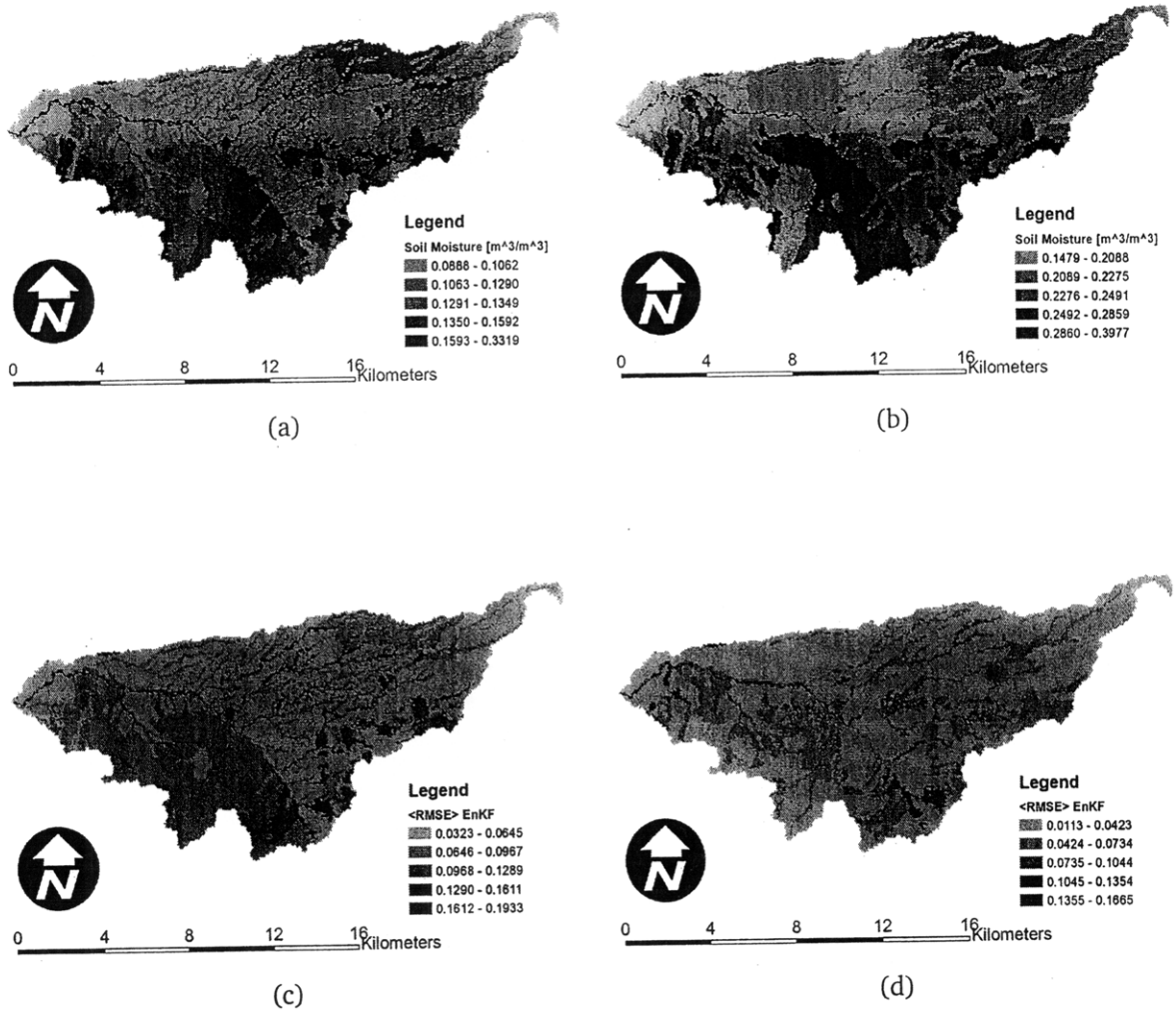
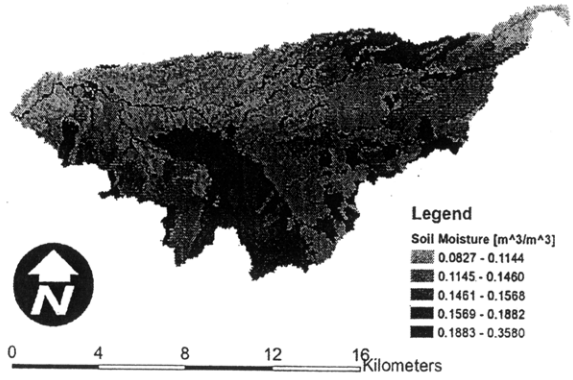
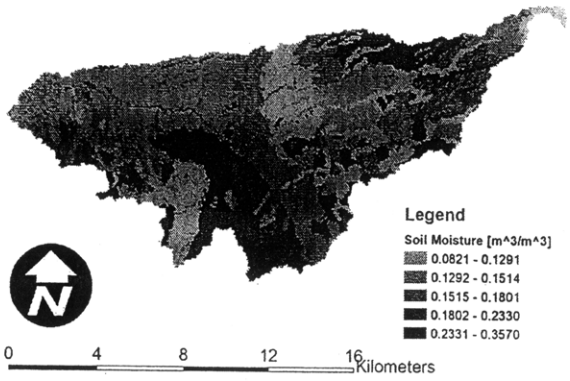


Figure 8-12: At the third analysis (216 hr), the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.

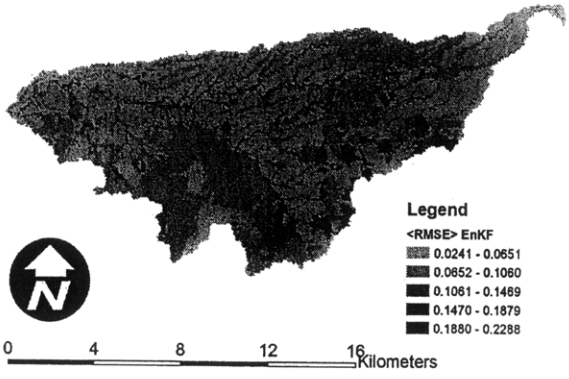
During the sixth analysis (432 hrs into the simulation), which is occurs during an extended drying episode, the average EnKF estimate of profile-integrated moisture content ranges from approximately 0.08 to 0.36 [m^3/m^3] throughout WGEW (figure 8-13(a)). Spatial heterogeneity in soil textural classes is the most important factor influencing the spatial distribution of the EnKF ensemble mean profile moisture. However, the channel network, particularly in the lower part of WGEW



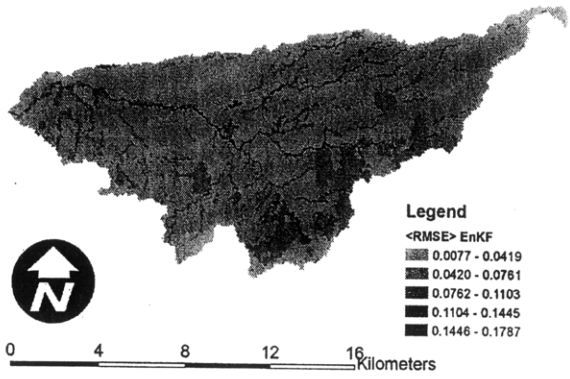
(a)



(b)



(c)



(d)

Figure 8-13: At the sixth analysis (432 hr), the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.

is associated with higher average ensemble mean profile moisture contents. In the sandy loam soil that encompasses a majority of the area in the Northwestern quadrant of WGEW in particular, the channel network is associated with substantially higher moisture contents than the surrounding terrain (figure 8-13(a)). In the near-surface, the EnKF mean estimate averaged across the four analyses is strongly organized around spatial heterogeneity in soil types, and exhibits the same range

of values as the profile moisture estimate (figure 8-13(b)). It is interesting to note that despite the extended drying period, there remains some evidence of the impact of spatial heterogeneity in rainfall, as evidenced by the vertical line in the North-central portion of the watershed (figure 8-13(b)).

The EnKF estimate of profile-integrated moisture is again relatively accurate after the sixth analysis, with a substantial portion of the watershed exhibiting an average RMSE in profile moisture of 0.07 [m^3/m^3] or less (figure 8-13(c)). Within certain soil types, such as sandy loam soil that encompasses much of the Northwest part of the watershed, topography plays a secondary role in explaining the spatial distribution of averaged RMSE in the EnKF profile moisture esitmate. Comparing the average RMSEs in the EnKF profile moisture estimate, there are some localized soil types that seem to exhibit large RMS error at both analysis cycles, exhibiting RMSE values of up to 0.23 [m^3/m^3] (figure 8-13(c)). Investigating spatial pattern in average RMSE in the near-surface EnKF moisture estimate reveals a reasonably high degree of accuracy throughout much of the watershed (figure 8-13(d)). The majority of the watershed is associated with average RMS errors in the near-surface moisture estimate of 0.04 [m^3/m^3] or less, with topography being the most obvious factor that describes spatial patterns in average RMSE (figure 8-13(d)). Channels are again associated with the highest average RMSE values.

Critically important assessing the performance of the EnKF as a means of hillslope-scale moisture estimation in this experiment is the accuracy of the EnKF estimates of relative to the corresponding estimates from the OL simulation. One useful way to visualize the relative performance of the EnKF and OL estimates of moisture is to plot the spatial distribution of the ratio of average EnKF RMSE to average OL RMSE. Values of this metric that are less than unity indicate locations of the watershed where the EnKF provides a more accurate estimate of the true moisture state than the OL estimate. During the third analysis, the EnKF estimate of profile integrated moisture is modestly outperforming the OL estimate (figure 8-14(a)). Variation in topography and soil textural class seem to be equally important in

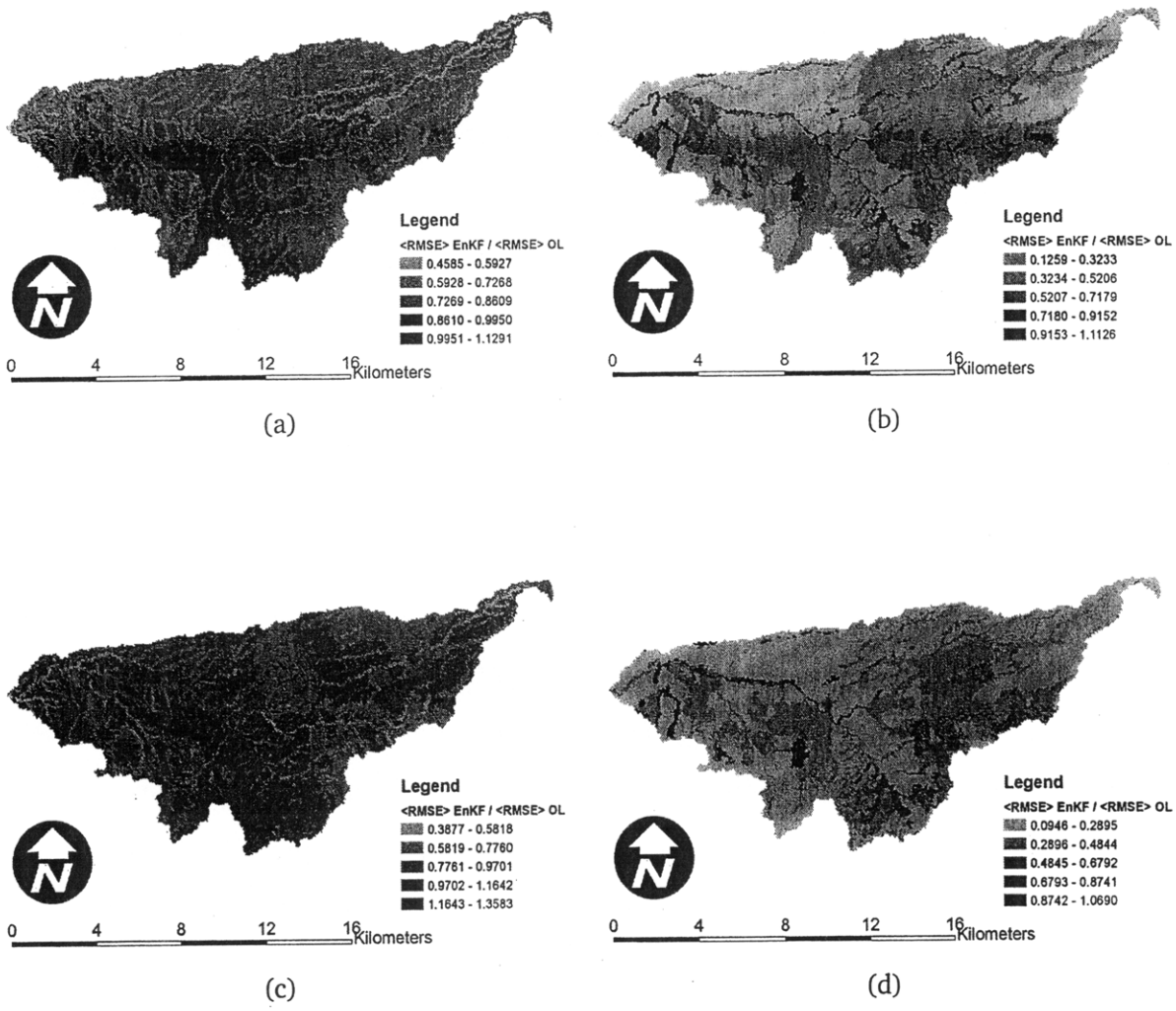


Figure 8-14: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture estimate at the third analysis (216 hr), (b) near-surface soil moisture estimate at the third analysis (216 hr), (c) profile-integrated soil moisture estimate at the sixth analysis (432 hr), and (d) near-surface soil moisture estimate at the sixth analysis (432 hr).

describing the spatial distribution of relative performance, with channels and valley bottoms seemingly being locations in which the accuracy of the EnKF estimate of profile-integrated moisture is substantially better than the OL estimate. In the near-surface, by contrast, at the third analysis the EnKF estimate is substantially outperforming the OL estimate in much of the watershed (figure 8-14(b)). In much of the watershed, the EnKF RMSE in near surface moisture is less than 32 per-

cent of the corresponding OL RMSE. However, contrary to the pattern in profile moisture performance, in terms of the near-surface moisture the channel network seems to be associated with only minor outperformance or slight underperformance of the EnKF estimate (figure 8-14(b)). Moreover, as seen in figure 8-14(b), some relatively localized soil types tend to be associated with EnKF RMS errors in near-surface moisture that are not substantially better than the corresponding OL RMS errors. Moreover, a slight imprint of the rainfall grid can be seen in the relative accuracy of the EnKF and OL estimates of near-surface soil moisture (figure 8-14(b)). This reflects the ability of the EnKF update scheme to effectively correct errors in precipitation by assimilation of observations related to near-surface moisture. At the sixth analysis, which is associated with relative dry conditions, the performance of the EnKF estimate of profile-integrated moisture, inferred from the ratio of the average EnKF RMSE to the average OL RMSE, only marginally outperforms the OL estimate in much of the watershed (figure 8-14(c)). While soil type heterogeneity accounts for some of the contrast in relative performance, in this analysis the EnKF profile-integrated moisture estimate interestingly seems to be more accurate in the channel and valley network, in a similar fashion to the third analysis cycle (figure 8-14(c)). In the near-surface, the EnKF is again substantially more accurate at predicting the true distribution of soil moisture at the sixth analysis cycle (figure 8-14(d)). In much of the watershed, the average EnKF RMSE of the near-surface moisture estimate is less than or equal to 29 percent of the average OL RMSE (figure 8-14(d)). Like the corresponding results at the third analysis, the channels and valley bottoms along with some relatively localized soil types seem to be the only locations where the EnKF is only marginally more accurate or slightly less accurate in predicting the true near-surface moisture state at the sixth analysis (figure 8-14(d)). Results from both the third and sixth analyses, along with other analysis results displayed in **Appendix D**, seem to indicate that the EnKF approach produces substantially better estimates of near-surface soil moisture than the OL approach. Furthermore, using the EnKF to assimilate observations of microwave backscatter often leads to moderately more accurate estimates of profile-integrated soil mois-

ture. While channels and valley bottoms are areas of the watershed in which the EnKF performs best relative to an OL estimate of profile soil moisture, they also coincide with areas of the watershed in which the EnKF performs worst in estimation of near-surface moisture, relative to the OL estimate.

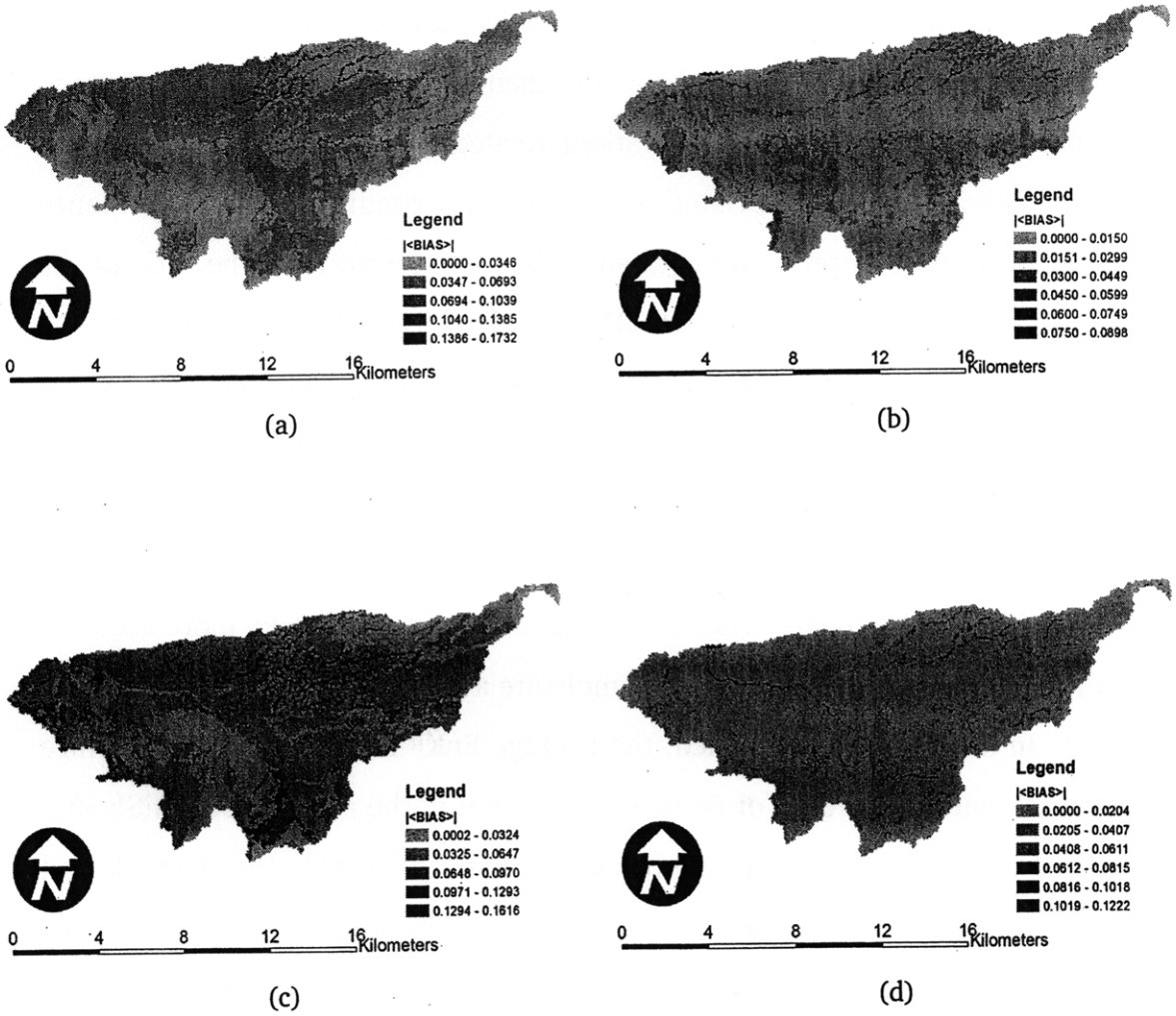


Figure 8-15: The local average bias in the EnKF estimate for the: (a) profile-integrated soil moisture estimate at the third analysis (216 hr), (b) near-surface soil moisture estimate at the third analysis (216 hr), (c) profile-integrated soil moisture estimate at the sixth analysis (432 hr), and (d) near-surface soil moisture estimate at the sixth analysis (432 hr).

The bias in the EnKF estimate also provides an important measure of the performance of the data assimilation system. As mentioned in the interpretation of

results from the first experiment, the ensemble variance in the EnKF estimate of soil moisture may be a poor reflection of actual estimation error if the estimate is biased.

During the third analysis, the absolute value of bias in the estimate of the profile-integrated moisture content, averaged across the four EnKF experiments, is less than 0.04 [m^3/m^3] in approximately half of WGEW and less than 0.07 [m^3/m^3] in much of the remainder of the watershed (figure 8-15(a)). It should be pointed out that in figure 8-15(a) the sandy loam soil occupying much of the watershed in the Northwestern part of the watershed associated with values of average bias between 0.04 and 0.07 [m^3/m^3], which is potentially problematic since the average ensemble mean soil moisture at the same analysis tends to be low. On the other hand, the average bias in the near-surface EnKF moisture estimate at the third analysis is less than 0.02 [m^3/m^3] in the vast majority of WGEW, and less than 0.03 [m^3/m^3] in much of the rest of the watershed (figure 8-15(b)). It is interesting to note that there a faint topographic signature in the spatial organization of the average bias, as well as a faint signature of the rainfall grid (figure 8-15(b)). The average bias in the estimate of the profile-integrated moisture at the sixth analysis is less than 0.07 [m^3/m^3] throughout much of the WGEW, while the maximum average value of bias is approximately 0.16 [m^3/m^3] (figure 8-15(c)). In the North-central and Northeast portion of the watershed, there is a moderate topographic signature in the spatial organization of average bias in the profile moisture estimate. Channel and valley bottoms seem to be associated with some of the more moderately-high observed values of average bias in the estimate of profile moisture (figure 8-15(c)). In this sixth analysis, as in the third analysis, the sandy loam soil occupying much of the Northwestern portion of the watershed exhibits a moderate level of average bias. Turning to the average bias in the EnKF estimate of near-surface soil moisture at the sixth analysis cycle, the vast majority of the watershed is associated with values less than or equal to 0.02 [m^3/m^3] (figure 8-15(d)). The few areas of the watershed that are associated values of average bias that are on the moderate to high end of the observed range fall mostly along the channel network (figure 8-15(d)).

The spatial patterns of average bias in both profile-integrated and near-surface EnKF estimates reveal underscore a few important conclusions. First, consistent with previously discussed results related to the spatial patterns in the average EnKF ensemble mean and average estimator error, the EnKF estimation procedure is more reliable in estimation of the true near-surface rather than profile-integrated soil moisture. Second, the average value of bias in the estimate of profile-integrated soil moisture largely follows the spatial pattern of soil textural classes, with some secondary influence associated with topographic variability.

8.3.3 Results: pixel-scale soil moisture dynamics

Beyond the EnKF estimate of the spatial distribution of soil moisture and the estimator error at times of observations, it is important to consider the soil moisture dynamics during the forecast step between observation times. To investigate these effects, two pixels within WGEW were selected for monitoring during the EnKF experiment (figure 8-16). These pixels correspond roughly to the spatial location of two intensively monitored sites within the WGEW, the Lucky Hills site (pixel 3378) and the Kendall site (pixel 4496). It should be underscored here that the pixel-scale results presented here are only diagnostics, and no additional assimilation or processing was performed at these sites.

The soil moisture dynamics during the entire 27-day EnKF experiment are presented, as an example, for the first set of synthetic radar backscatter observations[2]. At pixel number 3378, the true soil moisture is anomalously high during most of the simulation period, but somewhat low at the first analysis cycle. As can be seen in figure 8-17(b), the EnKF update step tends to both reduce the ensemble variance in the near-surface soil moisture estimate, as well as move the ensemble mean behavior toward the true near-surface moisture content. The EnKF update is capable of producing fairly large analysis increments at this location, in response to the assimilated observations. During the forecast step, as expected, the ensemble variance

[2]Results for the remaining three observations are shown in **Appendix D**

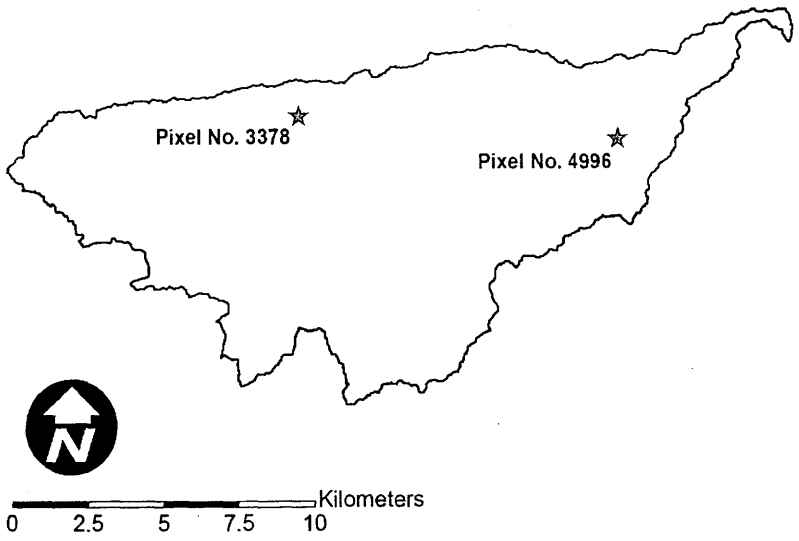


Figure 8-16: Locations of the two observation pixels within WGEW in the EnKF simulations. Pixel 3378 corresponds roughly to the location of the Lucky Hills experimental site, and pixel 4996 corresponds approximately in location to the Kendall experimental site.

tends to increase and the estimate becomes more biased as time elapses beyond the analysis (figure 8-17(b)). This is due to a combination of factors, among which are a combination of true parameters that lead to anomalously high true near-surface moisture contents during the experiment. Nevertheless, the EnKF update relatively well forces the ensemble estimate of near surface moisture toward the observation. On the other hand, the EnKF update does little to cause the ensemble estimate of the profile-integrated moisture content toward the true profile moisture (figure 8-17(c)). To evidence, at the fourth analysis (360 hr), the EnKF update leads to an increase in the ensemble mean estimate of profile moisture, when the true profile-integrated moisture content is actually substantially below the prior estimate. This seems to indicate that while the EnKF update of soil moisture in the near-surface tends to push the ensemble estimate toward the true near-surface moisture (on which the observations are based), this can occur at the expense of accuracy in the profile-integrated estimate of moisture content.

Turning to pixel 4996 (approximately the location of the Kendall experimental

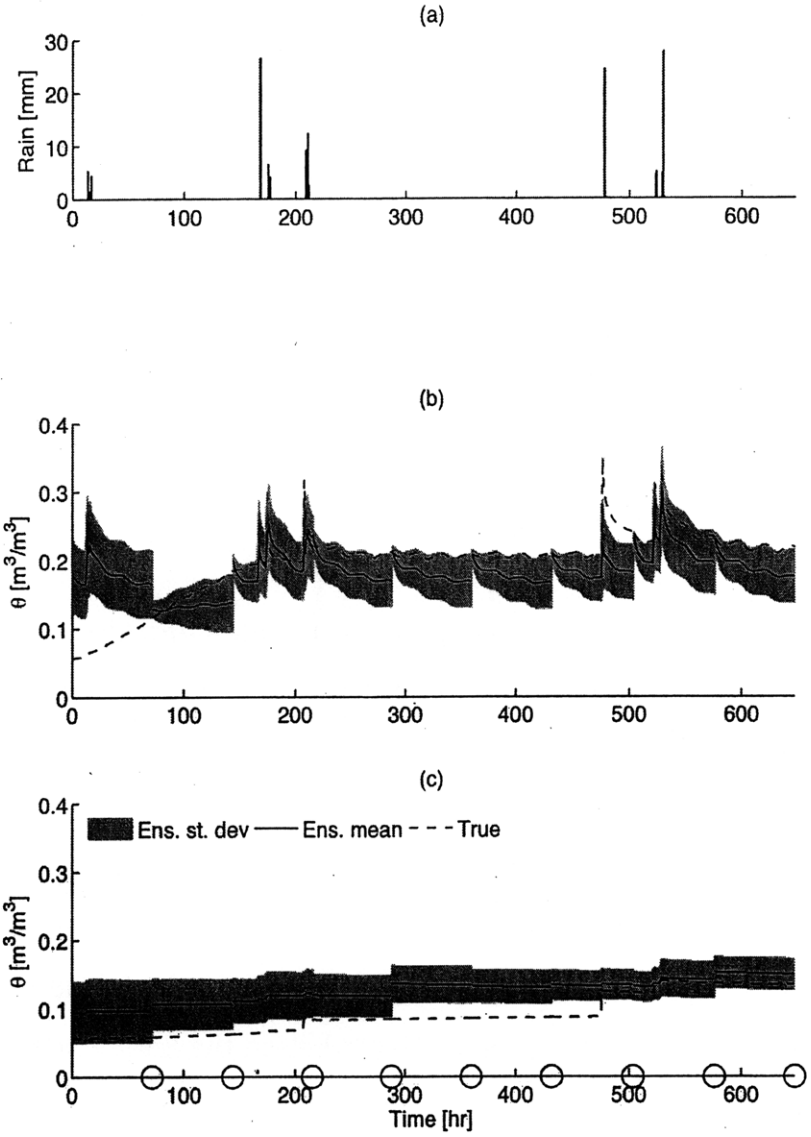


Figure 8-17: Example of the pixel-scale behavior near the Lucky Hills experimental site during one of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

site), again the true soil moisture evolution is anomalously high (figure 8-18(b)). Consistent with the results at pixel 3378, the EnKF update significantly reduces the ensemble variance in the near-surface soil moisture estimate, and moves the

ensemble mean toward the true near-surface content. Large analysis increments at this location, in response to the assimilated observations, are again seen during the update, and the ensemble variance and estimator bias tend both to increase as the time since the last analysis increases (figure 8-18(b)). At this pixel, the quality of the ensemble estimate of the profile-integrated moisture content seems unresponsive to assimilation of the radar backscatter observations based on the true near-surface moisture content (figure 8-18(c)). In particular, the update at analyses 2 through 7 actually increases the profile-integrated moisture content, despite the fact that the true soil moisture content is lower than the ensemble mean estimate during that same period of time (figure 8-18(c)). In keeping with the results from pixel 3378, this seems to indicate that while the EnKF update moves the near-surface estimate toward the truth, this does not generically hold true for the profile-integrated moisture content.

Observations of the pixel-scale moisture dynamics reveal some interesting implications of the data assimilation system posed here. Specifically, while assimilation of quantities related to the near-surface soil moisture may dramatically increase the accuracy in the estimate of the near-surface soil moisture, it can also induce or increase the bias in the profile integrated moisture content. As alluded to in the second data assimilation experiment and above, bias in the estimate of water stored in the unsaturated zone is problematic in applying data assimilation concepts such as the one employed here for applications where the total mass of water in the system is critical, such as irrigation scheduling and landslide prediction. For example, in scheduling application of irrigation water based on an estimate of soil water deficit retrieved from a soil moisture data assimilation system similar to this one, using the ensemble mean estimate of profile-integrated moisture content from figure 8-18(c) would lead to a potentially severe under-prediction of soil water deficit. Similarly, using that same ensemble estimate of profile moisture content in a slope stability analysis would potentially lead to erroneously high pore pressures and correspondingly low safety factors. Such a scenario could conceivable lead to a false-alarm landslide prediction, which would be associated with societal costs ranging from

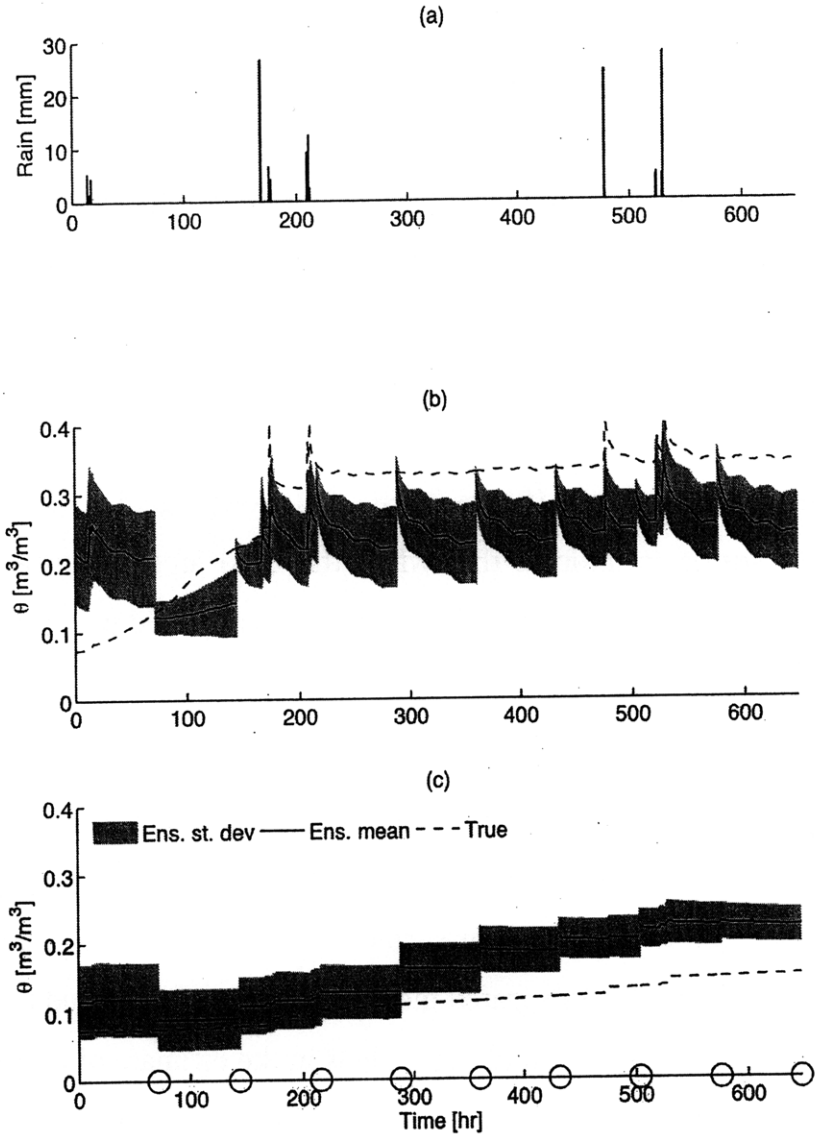


Figure 8-18: Example of the pixel-scale behavior near the Kendall experimental site during one of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

lost productivity to "disaster fatigue." Without, additional constraints on the soil moisture data assimilation system, such as assimilation of discharge observations that would impose a strong constraint on the ensemble mass balance, or better es-

timating the soil hydraulic and thermal properties, it is reasonable to suspect that the estimate of moisture in the unsaturated zone retrieved from an assimilation system like that posed here may be deficient for purposes of applications. Additional constraints such as these would potentially also lead to more accurate (i.e., less ensemble variance and bias) near-surface moisture estimates during the forecast step of the EnKF procedure. Assimilation of discharge and parameter estimation will be highlighted as areas of potentially important future research in the concluding chapter.

CHAPTER 9

DISCUSSION AND CONCLUSIONS

This final chapter summarizes the primary original contributions of the thesis work to soil moisture remote sensing and data assimilation. This if followed by an outline of a few potential areas for future research that could are immediately related to the reasearch findings of this thesis. The chapter concludes with a brief argument for the importance of applications as a potential verification tool for hillslope-scale soil moisture data assimilation systems such as the one proof-of-concept version formulated in this thesis.

9.1 Contribution

Stated simply, the primary original contribution of this thesis work is the construction and demonstration through a series of synthetic experiments of a data assimilation system to estimate the hillslope-scale distribution of soil moisture using a process hydrology model and L-band microwave measurements from space. The findings presented in the previous chapter demonstrate that the data assimilation approach followed here can lead to an improvement in the predicted distribution of soil moisture, particularly in the near-surface environment, that is manifest through a reduction in the error relative to a series of synthetic true states, and a reduction in the bias of the ensemble mean estimate. Moreover, it was shown that

the EnKF update resulted in soil moisture information gain at the hillslope scale, despite the fact that the observations were associated with a much coarser spatial scale. In one of the most realistic synthetic experiments to date in hillslope-scale soil moisture estimation, it was demonstrated that L-band microwave radar backscatter observations at a spatial resolution of 3 km substantially improve the estimate of near-surface soil moisture at hillslope scales in the Walnut Gulch Experimental Watershed when assimilated using the EnKF. Average values of RMSE in near-surface soil moisture across four observations representing four potentially true soil moisture distributions are substantially lower when observations are assimilated using the EnKF, compared with an open loop (OL) estimate. Furthermore, assimilation of observations leads to a reduction in the bias, averaged across the four sets of observations, at the time of the analysis. The spatial structure of estimator error reveals varying influence due to soil type heterogeneity, topography, and (when it occurs) rainfall. These results suggest that, at the time of the analysis, assimilation of observations can assist in identifying errors in the hydrometeorological forcings (particularly the spatial distribution of rainfall) and soil parameters. The success of the data assimilation system was the result of three critical preconditions being met:

1. Identification of the factors contributing to uncertainty in soil moisture prediction at hillslope scales, which allowed for development of effective means for explicitly modeling the uncertainty in the soil properties and hydrometeorological forcings required as input to the hydrological model. In particular, the application of the Restricted Pairing sampling algorithm was shown to reduce estimator variance in first- and second-order ensemble soil moisture statistics at small ensemble sizes.

2. The tools developed for representing the uncertainty in the parameters and forcings facilitated a thorough investigation of the of the spatiotemporal sensitivity in soil moisture in the near-surface and throughout the profile arising from uncertainty in the hydrometeorological forcings and inadequate knowl-

edge of the soil parameters.

3. Formulation of an L-band microwave observing system that captured the heterogeneity in factors affecting emission/reflection of microwave energy simulated by the model as well as the role of topography on the observing geometry was a critical component of the data assimilation approach.

The contributions associated with and implications of each of these preconditions are elaborated in greater detail below.

9.1.1 Modeling uncertainties in forcings and parameters

Chapters 5 and 6 of the thesis were devoted to outlining the methods by which uncertainty in the hydrometeorlogical forcings (**Chapter 5**) and inadequate knowledge of the hydraulic and thermal properties of the soil within a watershed were explicitly modeled in the ensemble data assimilation framework (**Chapter 6**).

To summarize **Chapter 5**, the hydrometeorological forcings were subdivided into rainfall and thermodynamic drivers. The temporal evolution of hourly rainfall was modeled using the Modified Bartlett-Lewis (MBL) approach, which stochastically generates storm arrivals and the associated intrastorm cellular structure of the storm. A stochastic multiplicative cascade was used both to perturb the hourly rainfall volumes and disaggregate them in space to spatial resolutions consistent with operational weather radars (i.e., NEXRAD). The stochastic weather generator of Ivanov *et al.* [2007] was used to derive the hourly time series of thermodynamic drivers (air and dew temperatures, wind speed, sky cover fraction, and incoming solar radiation) based on a time series of hourly rainfall simulated by the MBL model. This approach, which combined three computationally inexpensive stochastic models, is used to generate ensembles of the hydrometeorological inputs to the tRIBS-VEGGIE model that were completely internally consistent.

In **Chapter 6**, a stochastic approach was outlined to generate ensembles of the soil hydraulic and thermal properties (SHTPs) that characterize the hydrologic be-

havior of the soil in the tRIBS-VEGGIE model. The methodological construct is based on a Latin Hypercube (LH)-based sampling strategy. The LH approach ensures that, given the marginal behavior of a parameter, values of the parameter associated with low probability of occurrence but potentially high consequence to the model outputs are represented, even at small sample sizes. The so-called Restricted Pairing (RP) approach extends the LH approach to allow for correlation control among the parameters. The RP algorithm was applied to stochastically generate the soil hydraulic and thermal properties to the tRIBS-VEGGIE model. The soil database of Schaap and Leij [1998] was partitioned into categorical soil textural classes, based on the USDA classification scheme. Within each textural class, soil parameters were subjected to a statistical analysis to estimate the rank correlation among the parameters and assign marginal distributions to each parameter based on its empirical histogram. A series of pixel-scale experiments was performed to investigate the behavior of soil moisture ensembles in which the soil parameters were generated using the RP approach and compared to corresponding ensembles in which the soil parameters were randomly sampled from their respective marginal distributions. Ensemble experiments were repeated a number of times for varying ensemble sizes to determine the stability of the ensemble estimates of mean and variance in near-surface soil moisture. It was found that the soil moisture ensembles modeled with soil parameters obtained using the RP algorithm converged to stable estimates of the ensemble first- and second-order statistics at much smaller ensemble sizes than the correspond soil moisture ensembles associated with the simple random sampling approach. Moreover, because soil moisture ensembles simulated with parameters derived from the simple random sampling approach contained some ensemble members that exhibited hydrologically unlikely behavior, the ensemble variance was generally higher compared with the ensembles derived with soil parameters generated by RP. This behavior was verified over a range of soil textural classes. The principal benefit of the RP-based sampling approach for soil moisture data assimilation using the EnKF is the ability to achieve low variance in the estimates of first- and second-order ensemble statistics at relatively small sam-

ple sizes. Stated another way, careful attention to the way in which uncertain soil parameters are sampled can result in substantial computational saving associated with reducing the ensemble size without a corresponding increase in sampling error. This implication is of immense importance when formulating data assimilation systems with models possessing high state dimension and/or extensive numerically complexity, such as tRIBS-VEGGIE. As such, application of the RP algorithm for sampling soil properties is a relatively novel, but potentially powerful, way of improving the economics of a hillslope-scale soil moisture data assimilation system as it relates to computational demands.

9.1.2 Spatiotemporal factors affecting uncertainty in soil moisture

The methods for representing uncertainty in the hydrometeorological forcings and soil hydraulic and thermal properties outlined in **Chapter 5 and 6** facilitated the detailed uncertainty analysis in **Chapter 7**. The sensitivity analysis was subdivided into investigation of the temporal evolution of uncertainty at the point-scale under varying assumptions about factors contributing to uncertainty in soil moisture and investigation of the spatial features underlying the spatial distribution of soil moisture and its associated uncertainty. Both of these investigations were conducted through open loop ensemble simulation.

Important implications for hillslope-scale soil moisture data assimilation were identified as a result of the point-scale sensitivity analysis. In particular, the point-scale sensitivity analysis suggests that representing uncertainty in the SHTPs required as input to tRIBS-VEGGIE is critical to avoid problems associated with unreasonably low variance in the ensemble estimate of the soil moisture state. Further, while imperfect knowledge of the soil parameters was the dominant source of uncertainty in the ensemble soil moisture response to forcing, representing uncertainty in hydrometeorlogicalal forcings is associated with little additional computational burden because of the relatively simple nature of the stochastic models used to generate the forcings. However, computational demands may be substantial in

situations in which the hydrometeorological forcings are derived from a physically-based numerical weather prediction model. Results of the point-scale sensitivity analysis seem to indicate that in such scenarios, adequate attention should be paid to characterizing uncertainty in SHTPs. In the data assimilation experiments summarized in **Chapter 8**, uncertainty in SHTPs, rainfall, and thermodynamic forcings are explicitly represented because of the relatively small burden of modeling uncertainty in the hydrometeorological forcings, and to maintain a degree of realism in acknowledging the uncertainty in available hydrometeorological data.

The spatially distributed sensitivity analysis revealed that, given uncertainty in both the hydrometeorological forcings and SHTPs, the mean near-surface soil moisture largely reflects a combination of the heterogeneity in soil types, as well as the spatial organization of the channel and valley network. As might be expected, the channel network is mostly associated with higher values of ensemble mean near-surface soil moisture. Soil type heterogeneity and, to a lesser extent, topography, also account for much of the spatial distribution of ensemble standard deviation in near-surface soil moisture. Locations of higher ensemble mean near-surface soil moisture tend to be associated with higher ensemble variance and when visualizing the corresponding distribution of ensemble coefficient of variation leads, the role of topography in the spatial distribution of the standard deviation relative to the ensemble mean is much more prominent. The highest values of the local coefficient of variation in the near-surface moisture are seen in the channel and valley bottoms. In both the near-surface and profile-integrated soil moisture distributions, channel and valley bottoms are typically associated with higher ensemble variance in soil moisture. This is likely a result of the uncertainty in soil hydraulic properties. As discussed in **Chapter 7**, topography tends to amplify uncertainty in precipitation, runoff, and lateral redistribution in a downstream direction, resulting in higher ensemble variance in the topographic lows of the domain. This reasoning may be valid for semiarid regions where the water table is at significant depth and channels are mostly ephemeral. Although the current version of tRIBS-VEGGIE is not capable of capturing such behavior, groundwater-surface water interactions would

potentially curtail this amplification in more humid regions because a saturated downslope subsurface boundary condition, for instance, would limit the amount of upslope runoff that could re-infiltrate downslope and suppress subsurface lateral redistribution.

An analysis of empirical orthogonal functions (EOFs) in near-surface and profile-integrated moisture content revealed that much of the variance explained by the ensemble of moisture anomalies can be captured with fewer than 50. The spatial patterns associated with several EOFs were found to represent a complex interplay between spatial variability in soil types, topography, and rainfall. The marked influence of several distinct factors (e.g., soil type and topography) seen in several EOFs suggests that the ensemble of moisture anomalies are not perfectly separable using a linear decomposition. This is a result of the fact that the ensemble of moisture anomalies captured in the open loop sensitivity analyses arise from a nonlinear set of processes governing the redistribution of moisture across the watershed.

9.1.3 Formulation of the L-band microwave observing systems

Chapter 4 outlined the formulation of passive and active L-band microwave observing systems. The passive microwave system takes the soil moisture and temperature outputs of the tRIBS-VEGGIE model and, given a satellite sky position in the form of an azimuth and zenith angle, predicts the microwave brightness temperature at the observing sensor. The second takes as input the spatial distribution of soil moisture simulated by the tRIBS-VEGGIE model and (again depending on satellite sky position) predicts the microwave backscatter measured at the receiving sensor.

One innovation in the development of these systems is the explicit representation of topography, as captured by the topographic slope and aspect, on the local incidence and polarization rotation angles. This work in particular carries some important implications for soil moisture remote sensing that are summarized here. On

an individual hillslope topography affects the amount of solar radiation received, therefore impacting local soil moisture conditions, temperature conditions, and dominant vegetation. Moreover, the topographic patterns in soil moisture and incoming solar radiation induce correlated patterns in vegetation biomass, water content and height Ivanov *et al.* [2008a,b, 2007] which enhances topographic contrasts in the surface states that control emission and reflection of microwave radiation. Hillslope-scale topography also controls the incidence angle and polarization rotation required to capture topographic effects in modeling the microwave brightness temperature and radar backscatter at the sensor. Significant variation in modeled hillslope-scale brightness temperatures and radar backscatter are generally a result of variation in: (1) soil moisture and temperature, (2) vegetation water content, height, and temperature, and (3) local incidence angle and polarization rotation associated with topography. At the sensor, modeled microwave observations are sensitive to the satellite azimuth angle and zenith angle because the histogram of hillslope-scale incidence angles reflects both the distribution of slopes and aspects within the observation area and the location of the satellite with respect to the observation area at the time of observation. The magnitude of this sensitivity is a function primarily of the degree of topographic ruggedness (or smoothness).

The work outlined in **Chapter 4** in formulation of the observing systems is of importance for applications that require modeling surface microwave emission in areas of non-negligible topographic variation, particularly development of soil moisture retrieval algorithms, soil moisture data assimilation, and geostatistical techniques for disaggregation of brightness temperature observations to hillslope scales.

9.2 Future research

While this work represents an important and original contribution to the field of hydrologic estimation, the findings also serve to identify some significant challenges that still need to be addressed. The most obvious challenge that this thesis high-

lights as deserving additional research effort is in reducing the bias in the estimate of the profile-integrated estimate of soil moisture.

One explanation for the observation of bias in the profile-integrated moisture estimate in WGEW during the entire experiment is that the experiment is not long enough to assimilate a sufficient number of microwave observations and simulate sufficiently many wetting-drying cycles to allow the profile moisture EnKF mean to converge to the true profile moisture value. Because integrating moisture content over the entire soil profile tends to dampen hydrologic fluctuations in the near-surface, there may be little correlation between moisture contents in the deeper portion of the profile and the predicted observations at the analysis. A lack of correlation between deeper moisture contents and the predicted observations would lead to smaller analysis increments, or increments in the opposite direction of the true profile moisture. Therefore, performing the sequential estimation procedure over a longer period of time may result in a profile-moisture estimate that eventually moves toward the true profile moisture value.

Another way to better constrain soil moisture estimates throughout the entire profile would be to develop methods for assimilating point-scale observations. Water content reflectometers and tensiometers are just examples of technologies that can accurately and continuously provide information about soil moisture throughout the profile. These observations are highly localized, however, and due to fine-scale heteorgeneity in soil properties they may not be representative of an entire tRIBS-VEGGIE pixel. Therefore, assuming that the point-scale observations constitute direct observations of the tRIBS-VEGGIE state at the pixel may be misleading. The problem, therefore, is to find an effective observing system that can translate pixel-scale quantities to point-scale predicted observations. One way of approaching this problem would be to imbed a hydrologic model of much higher spatial resolution (centimeters) within the tRIBS-VEGGIE pixel. tRIBS-VEGGIE would supply the boundary conditions to the high-resolution model, which would then evolve states and fluxes at spatial scales more similar to that of the point-scale observa-

tions. Although such a multi-scale modeling approach is an attractive way to integrate models and data from across a wide range of spatial scales, the fine-scale model embedded within the tRIBS-VEGGIE model pixel would likely have a computational burden on the order and possibly significantly greater than the hillslope-scale EnKF effort pursued in this thesis. This is particularly true if the fine-scale model would have to be run in an ensemble model Therefore, while potentially promising, the computational burdens associated with this approach are currently prohibitively expensive.

Bias in the profile-integrated moisture content is largely a result of the model error associated with the uncertainty in the soil hydraulic and thermal properties that persists throughout the simulation. As discussed in **Chapter 2**, there is a historical recognition that an important source of error in data assimilation frameworks arises from the error in the model structure and in the parameters required as input to the model. Neglect of model error can lead to unrealistically low ensemble approximations of the state error covariance and/or lead to bias in the estimate. The Restricted Pairing approach to stochastic generation of soil parameters outlined in **Chapter 6** was implemented to address the first of these pitfalls. However, as was seen explicitly in the previous chapter in the WGEW experiment, assimilation of radar observations led to substantial improvement in prediction of near-surface moisture while at the same time giving biased estimates of the profile-integrated soil moisture. As was seen in the temporal soil moisture dynamics at the pixel-scale, the ensemble mean near-surface soil moisture diverged from the true near-surface soil moisture as time elapsed after the model was re-initialized with the analyzed soil moisture. The profile moisture content was often seen moving in the opposite direction of the true profile-integrated moisture content. A significant reason for these findings are that the uncertainty in the SHTPs persists during the experiment and information contained in the observations does not propagate to the parameters, as well as the state.

Augmenting the parameters to the state during the analysis would allow the ob-

servations to better constrain the uncertainty in soil parameters, while only marginally increasing the dimension of the state vector. The increase in the size of the state vector associated with parameter augmentation would be equal to the product of the number of soil types in the domain and the number of parameters to be updated. Supposing, saturation and residual moisture content, saturated hydraulic conductivity, and the Brooks-Corey parameters are assumed to be the variables most responsible for uncertainty in soil moisture, the increase in the size of the state vector for WGEW due to soil hydraulic properties would only be $41 \times 5 = 205$. In assimilation of radar backscatter observations, it should be noted that the predicted observations are particularly sensitive to the parameters of the surface roughness correlation function, the rms height and correlation length. It may also be beneficial, therefore, to augment the two roughness parameters for each soil type, which in WGEW would add an additional 82 elements to the state vector, for a cumulative addition of 287 variables. Given that the soil moisture state vector is approximately 90 times this value, the added expense in the update step is rather paltry. State augmentation has been shown in the literature to be beneficial in estimation of both the states and parameters with lumped hydrological models, through assimilation of discharge observations (e.g., Hino [1973] and Kitanidis and Bras [1979]). More recently Vrugt *et al.* [2005] incorporated the strengths of both global optimization techniques and data assimilation into a single framework to simultaneously estimate the states and parameters of a lumped hydrological model.

However, a significant drawback of augmenting the states with the SHTPs in assimilation of microwave observations of quantities reflecting soil moisture near the surface is the physical reality that often surface soil properties can differ substantially from parameters in the rest of the soil column. The current implementation of tRIBS-VEGGIE does not generally handle vertically heterogeneous soils, and therefore any soil parameters updated based on assimilation of observations of near-surface phenomena may not reduce the bias problem seen here. Another potentially powerful way of addressing the issue of parameter uncertainty as it affects the estimate of moisture in the unsaturated zone is to assimilate discharge, which

imposes a strong constraint on the system. In particular, the Ensemble Kalman Smoother (EnKS) could potentially be applied to update historical states and parameters based on current observations and predictions of discharge. Unlike most smoothing algorithms the EnKS does not require a backward model, requiring only the EnKF estimates during the smoothing window. A physical argument can be made that the length of the smoothing window should be related to measures of travel time, such as time of concentration, in the watershed under study. The EnKS has already been applied successfully in soil moisture estimation by Dunne and Entekhabi [2006], who used the EnKS approach to improve estimates of moisture content and energy flux. In a similar fashion to Dunne and Entekhabi [2006], the EnKS could also be used to update historical estimates of the soil moisture state based on current observations of microwave brightness temperature and backscatter. Because the application of the EnKS is a relatively straightforward extension of the EnKF estimation procedure pursued here, it is a reasonable near-future outgrowth of this research.

The data assimilation experiments considered only bare soil conditions. In WGEW, and many semiarid areas more generally, vegetation plays an immensely important role in the dynamics of soil moisture in the near-surface and throughout the profile, and may even be a dynamic link between groundwater and the atmosphere. However, vegetation also complicates the modeling of radar backscatter based on the surface moisture conditions. As discussed in **Chapter 4**, previous work [Narayan and Lakshmi, 2008; Narayan *et al.*, 2006; Njoku *et al.*, 2002] has shown a monotonic relationship between the change in successive backscatter observations and the change in soil moisture over the same interval in areas with significant vegetation cover. Assimilation of temporal L-band radar backscatter change represents one potentially simple way of extending the hillslope-scale data assimilation system formulated here to areas with significant vegetation cover. At the first analysis, the predicted observation of radar backscatter at the current analysis step and at the initialization are computed based on the radar observing system based on the moisture state at the current analysis and initialization, and the temporal change

in predicted backscatter computed. With the temporal change in backscatter observations, the EnKF estimation machinery is used to update the temporal moisture change. The analysis increment of temporal moisture changes are all assigned to the analysis moisture state at the current analysis step. Using the nonlinear ecohydrology model, the analyzed soil moisture state is propagated forward to the next analysis cycle. The predicted observation of radar backscatter at the current analysis step on the forecast moisture state and at the previous analysis based on the analyzed moisture state are computed based on the radar observing system. Again the EnKF procedure is used to update the temporal moisture change based on the temporal change in backscatter observations, and the increments assigned to the analysis moisture state at the current analysis. This procedure to assimilate values of temporal change in radar observations is a relatively straightforward extension of the current data assimilation system, requiring very little additional code development.

9.2.1 The importance of applications

At present, there are no L-band measurements to allow for application of the data assimilation estimation procedure outlined in this thesis in any real context. Furthermore, the fact that there are no current data or foreseeable data sources that could serve to quantitatively validate the hillslope-scale soil moisture estimates derived from the EnKF procedure pursued in this thesis necessitates other approaches to verification. Applications are one potentially important and viable way to provide some degree of forecast verification. The applications discussed in the introductory chapter require an estimate of the soil moisture state as input to some geoprocessing routine to yield a forecast with some actionable information. For instance, a landslide prediction scheme would take the EnKF soil moisture estimate as input and provide a spatial mapping of landslide risk over some immediate time horizon that could be acted upon. Alternatively, an irrigation scheduling system may take the EnKF moisture estimate to produce a map of required irrigation application depths

based on a spatially distributed assessment of rooting depth-integrated soil water deficit. And while the additional geoprocessing of the soil moisture estimate required for the particular application can itself introduce uncertainties, an improved estimate of soil moisture state arising from assimilation of remotely sensed data nevertheless improves the quality of the actionable data. The verification of the hillslope-scale soil moisture state estimate obtained from assimilation of available microwave data, therefore, lies in the quality of the decisions made based on the actionable data derived from the forecast. A reasonable objective of hillslope-scale data assimilation problems like the one addressed in this thesis is to *enhance the utility of the data derived from the estimate of the moisture state in the context of the decision support systems associated with particular applications.* This objective leads to an iterative process of calibration in which decision-makers provide feedback, often in the form of soft data, to modelers and forecasters in an effort to improve the quality of the forecasts of both the state (e.g., reducing RMSE and bias) and of the decision variables of interest derived from the state estimate (e.g., 95% chance of landslide in location 'A', 75% slow-go conditions on path 'C' to target).

APPENDIX A

DETERMINING OBSERVATION TIME

This appendix provides an expansion of the observing system that determines when a location on the Earth's surface is under observation based on the orbital characteristics of the satellite carrying the sensor. As demonstrated in "**Chapter 4: Formulation of the observing system**," the local incidence angle of a particular location can be calculated directly through the spherical law of cosines provided that the following are known: (1) the local topographic slope, (2) the local topographic aspect, (3) the azimuth angle to the subsatellite point, and (4) the zenith angle from the locally vertically up direction to the satellite. The first two parameters must be calculated from a map of the terrain. However, the second two depend on the motion of the satellite with respect to the area being observed. The satellite azimuth (ζ_S) and zenith (δ_S) angles can be determined directly by knowing the latitude and longitude of the site at which microwave energy emission is being modeled (l_R, ϕ_R), and the latitude and longitude of the sub-satellite point (l_{SS}, ϕ_{SS}). The sub-satellite point is the point on the surface of the Earth below the satellite at nadir. Knowing (l_R, ϕ_R) and (l_{SS}, ϕ_{SS}), the azimuth angle to the satellite can be computed directly from the spherical law of sines as follows:

$$\sin(\zeta_S) = \frac{\sin(l_{SS} - l_R)}{\sin(\gamma)}, \text{ when } l_{SS} > l_R \text{ and } \phi_{SS} > \phi_R \qquad \text{(A.1)}$$

$$\sin\left(\zeta_S - \frac{\pi}{2}\right) = \frac{\sin(\phi_{SS} - \phi_R)}{\sin(\gamma)}, \text{ when } l_{SS} > l_R \text{ and } \phi_{SS} \leq \phi_R \quad \text{(A.2)}$$

$$\sin\left(\zeta_S - \pi\right) = \frac{\sin(l_{SS} - l_R)}{\sin(\gamma)}, \text{ when } l_{SS} \leq l_R \text{ and } \phi_{SS} < \phi_R \quad \text{(A.3)}$$

$$\sin\left(\zeta_S - \frac{3\pi}{2}\right) = \frac{\sin(\phi_{SS} - \phi_R)}{\sin(\gamma)}, \text{ when } l_{SS} \leq l_R \text{ and } \phi_{SS} > \phi_R \quad \text{(A.4)}$$

$$\begin{aligned}\text{haversin}\,(\gamma) = {} & \text{haversin}\,(\phi_{SS} - \phi_R) \\ & + \cos(\phi_{SS})\cos(\phi_R)\,\text{haversin}\,(l_{SS} - l_R)\end{aligned} \quad \text{(A.5)}$$

where the haversine formula is given as,

$$\text{haversin}\gamma = \sin^2\left(\frac{\gamma}{2}\right). \quad \text{(A.6)}$$

Applying the law of sines to the triangle connecting the satellite sky position (S), the site being observed (R), and the center of the Earth (E), the zenith angle of the satellite, δ_S can be computed,

$$\frac{\sin(\pi - \delta_S)}{A_\rho + z} = \frac{\sin\gamma}{\sqrt{A_\rho^2 + (A_\rho + z)^2 - 2A_\rho(A_\rho + z)\cos\gamma}}, \quad \text{(A.7)}$$

where A_ρ is the Authalic radius of the Earth (approximately), and z is the nominal satellite altitude.

Satellite sky position can be determined by solving equations 4.9-4.15 when the proximity of S and R is such that the R is within the viewing footprint of the satellite. What remains in determining the satellite sky position coordinates (ζ_S and δ_S) is calculating the time evolution of the subsatellite point (l_{SS}, ϕ_{SS}) relative to some space-time datum , referred to as the nodal crossing. The datum is specified by the time at which the subsatellite point intersects a given latitude

and longitude (often the intersection of the equator and Prime Meridian) during an ascending (South to North) pass. The time evolution of the subsatellite point can be modeled using a simple model of satellite orbital kinematics. Prior to describing the orbital kinematics model, the following are assumptions made to simplify the orbital model and make it conform to characteristics of planned and conceived soil moisture observing satellites:

1. The satellite is Sun-synchronous, Low Earth Orbiting, and in a nearly circular orbit

2. The sensor is conically scanning

3. Each time step of the orbital simulation corresponds to an instantaneous sensor scan, which can be approximated as a circular band on the surface of the Earth, centered at the subsatellite point. The width of the scanning band is determined by the half-power beam width (HPBW) of the sensor

4. The watershed being modeled is sufficiently small that if the point is within the scanning band, the entire watershed is observable

The inclination angle and orbital period of the satellite determine its orbital kinematics. The angle of inclination, defined as positive counter-clockwise from the equator, of the satellite is computed by solving,

$$\Omega_S = -\frac{3J_2\sqrt{GM}a_e^2 a^{-7/2}\cos i}{2}, \tag{A.8}$$

for i. In the above Ω_S is the mean angular speed at which the Earth orbits the Sun, J_2 is the dynamical form factor of the Earth to account for its non-sphericity, GM is the gravitational constant-Earth mass product, a_e is the equatorial radius of the Earth, and a is the semi-major axis of the satellite. As a consequence of the Sun-synchronous orbit assumption, i will be between $\pi/2$ and π (i.e., retrograde) and the motion of the satellite will be from East to West relative to the surface of the Earth. The orbital period of the satellite, P_n is calculated as,

$$P_n = 2\pi\sqrt{\frac{a^3}{GM}}\left(1 + \frac{3J_2 a_e^2}{4a^2}\left\{1 - 3\cos^2 i + \frac{1 - 5\cos^2 i}{(1+e^2)^2}\right\}\right), \tag{A.9}$$

where the only additional parameter introduced is the satellite eccentricity, e, which is zero as per the circular orbit assumption. For near-circular orbit the longitude of the subsatellite point is,

$$l_{SS} = -\left(\frac{2\pi t}{P_n} + \Omega_E t - 2\pi n_O\right) \tag{A.10}$$

where t is the time since nodal crossing, Ω_E is the mean angular speed of the rotation of the Earth about its axis, and n_O is the number of orbits made since nodal crossing.

The latitude of the subsatellite point at time t, is then calculated as,

$$\sin\phi_{SS} = \sin\varphi \sin i. \tag{A.11}$$

From the law of sines the angle φ can be determined as follows:

$$\frac{\sin\varphi}{\sin(\pi/2)} = \frac{\sin(2\pi t/P_n)}{-\cos(\pi - i)\cos(\pi/2) + \sin(\pi - i)\sin(\pi/2)\cos(2\pi t/P_n)}. \tag{A.12}$$

Equations 4.16 through 4.20 describe the time evolution of the subsatellite point (l_{SS}, ϕ_{SS}) at time t after the satellite nodal crossing. Figure 4-2 demonstrates an example of the evolution of the subsatellite point for more than 3 orbits. The characteristics of the satellite modeled correspond to those of the Hydrosphere State Mission satellite (HYDROS).

A site R is observed when,

$$R_{inner} \leq \gamma A_\rho, \text{ and,}$$

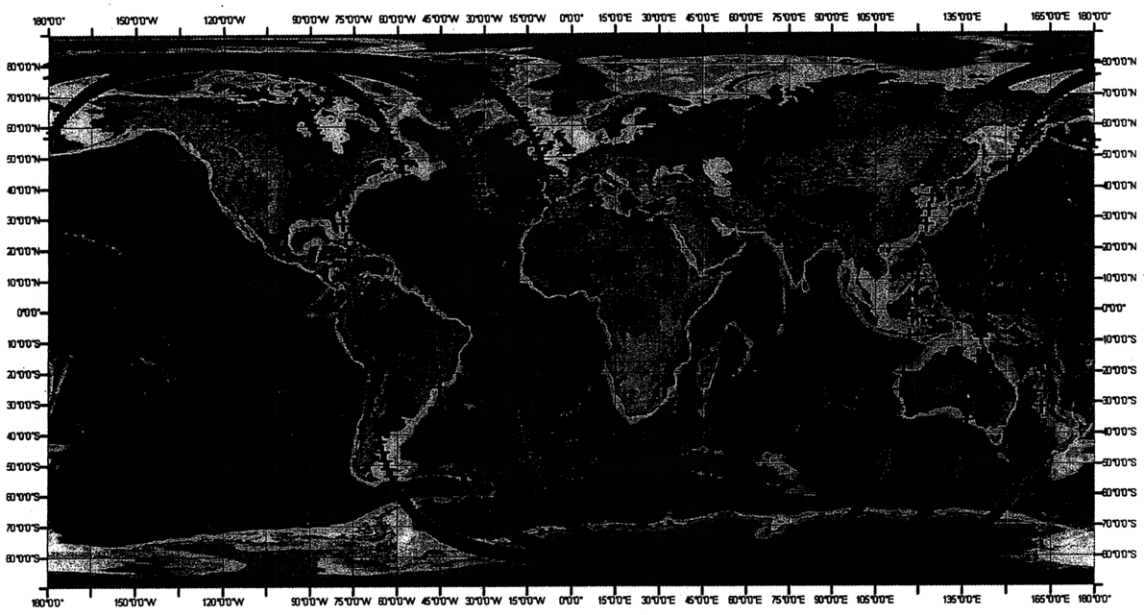


Figure A-1: An example of the time evolution of the subsatellite point after a nodal crossing at the intersection of the equator and Prime Meridian. The time interval between points is 60 s, and the characteristics of the satellite were chosen based on the HYDROS satellite.

$$R_{outer} \geq \gamma A_{\rho}, \tag{A.13}$$

where R_{inner} is the great circle distance along the surface of the Earth connecting the subsatellite point with the inner concentric ring of the scanning band, and R_{outer} is the great circle distance from the subsatellite point to the outer concentric ring of the scanning band. The great circle distances R_{inner} and R_{outer} are related to their respective arc angles (ρ_{inner} and ρ_{outer}) made between the subsatellite point, the Earth center, and the inner and outer edges of the scanning band through,

$$\begin{aligned} R_{inner} &= A_{\rho}\rho_{inner}, \text{ and} \\ R_{outer} &= A_{\rho}\rho_{outer}. \end{aligned} \tag{A.14}$$

In turn, ρ_{inner} and ρ_{outer} are calculated as,

$$\rho_{inner} = \pi - \left(\frac{A_\rho + z}{A_\rho} \sin^{-1} \left(\omega - \frac{\beta}{2} \right) + \left(\omega - \frac{\beta}{2} \right) \right), \text{ and}$$
$$\rho_{outer} = \pi - \left(\frac{A_\rho + z}{A_\rho} \sin^{-1} \left(\omega + \frac{\beta}{2} \right) + \left(\omega + \frac{\beta}{2} \right) \right). \quad \text{(A.15)}$$

The additional parameters are properties of the sensor: β is the half power beam width (HPBW), and ω is the off-nadir look angle.

The following summarizes the algorithm by which equations 4.16 through 4.23 are used to determine the satellite sky position given by ζ_S and δ_S:

1. Calculate the evolution of the subsatellite (l_{SS}, ϕ_{SS}) point after a nodal crossing (equations 4.16-4.20)

2. Compute the great circle distance between R and the subsatellite point and compare with R_{inner} and R_{outer} (equations 4.21-4.23)

3. If the point R is within the scanning swath of the sensor, calculate ζ_S and δ_S (equations 4.9-4.15)

APPENDIX B | ANALYSIS OF SOIL HYDRAULIC AND THERMAL PROPERTIES

In this appendix, results of the statistical analysis of the soil property meta-database aggregated by Schaap and Leij [1998]. The first section presents the descriptive statistics and Spearman correlation matrices of each variable used in the restricted pairing sampling algorithm presented in Chapter 6. The second section displays the histogram of each soil property by categorical soil type along, with candidate parametric probability distribution functions (pdfs) fit to the data for each variable and soil class. The parameters of pdf used in the restricted pairing stochastic sampling algorithm are also presented in tabular format.

B.1 Descriptive statistics and Spearman correlation matrices

The following two tables show the sample means (Table B.1) and standard deviations (Table B.2) for each of the 8 variables in the analysis, displayed according to their associated soil textural class. Tables B.3-B.11 present the Spearman correlation matrices for the 8 variables in the analysis by soil textural class.

Table B.1: Sample means of soil hydraulic and thermal parameters by soil texture

Soil texture	ρ_b [kg/m^3]	K_S [mm/hr]	θ_R [m^3/m^3]	θ_S [m^3/m^3]	h_b [mm]	λ_p [-]	$C_{s,\text{solids}}$ [J/(m^{3}K)]	$k_{s,\text{dry}}$ [J/(msK)]
Clay	1393.301	52.120	0.099	0.467	-1533.641	0.206	2301672	4.826
Sandy clay loam	1592.740	22.565	0.058	0.382	-1222.309	0.304	2205654	7.023
Silty clay loam	1356.167	12.178	0.088	0.476	-2640.397	0.591	2323871	4.318
Clay loam	1456.428	19.419	0.065	0.440	-1803.798	0.337	2258334	5.818
Sandy loam	1541.871	47.176	0.039	0.386	-683.103	0.458	2166725	7.913
Loam	1413.496	28.872	0.058	0.439	-1461.720	0.500	2212161	6.874
Silt loam	1418.978	29.504	0.060	0.436	-3309.559	0.852	2256133	5.868
Sand	1535.267	491.091	0.047	0.375	-335.457	2.286	2136393	8.607
Loamy sand	1535.528	120.768	0.045	0.388	-544.689	0.720	2151145	8.270

Table B.2: Sample standard deviation of soil hydraulic and thermal parameters by soil texture

Soil texture	ρ_b [kg/m^3]	K_S [mm/hr]	θ_R [m^3/m^3]	θ_S [m^3/m^3]	h_b [mm]	λ_p [-]	$C_{s,\text{solids}}$ [J/(m^{3}K)]	$k_{s,\text{dry}}$ [J/(msK)]
Clay	230.102	217.340	0.118	0.078	2155.694	0.184	33338	0.763
Sandy clay loam	177.396	50.032	0.072	0.065	1886.965	0.392	13668	0.313
Silty clay loam	119.964	16.224	0.093	0.080	3109.640	0.797	25533	0.584
Clay loam	200.095	36.869	0.079	0.069	3838.874	0.335	15498	0.355
Sandy loam	177.939	97.492	0.045	0.066	920.202	0.429	13225	0.303
Loam	228.629	73.177	0.058	0.091	1976.410	0.472	20068	0.459
Silt loam	139.294	61.671	0.053	0.057	3172.784	0.812	49803	1.139
Sand	129.888	865.189	0.023	0.058	221.717	1.590	4880	0.112
Loamy sand	198.677	160.830	0.034	0.072	748.416	0.628	6311	0.144

Table B.3: Spearman rank correlation matrix: Clay

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.3535	0.0576	-0.9283	-0.0980	0.1161	-0.6286	0.6286
K_S	-0.3535	1.0000	0.2605	0.2441	0.2792	0.0860	0.0864	-0.0864
θ_R	0.0576	0.2605	1.0000	-0.0569	0.0462	0.6854	-0.1560	0.1560
θ_S	-0.9283	0.2441	-0.0569	1.0000	0.1385	-0.1687	0.6091	-0.6091
h_b	-0.0980	0.2792	0.0462	0.1385	1.0000	-0.4253	-0.1561	0.1561
λ_p	0.1161	0.0860	0.6854	-0.1687	-0.4253	1.0000	-0.1159	0.1159
$C_{s,\text{solids}}$	-0.6286	0.0864	-0.1560	0.6091	-0.1561	-0.1159	1.0000	-1.0000
$k_{s,\text{dry}}$	0.6286	-0.0864	0.1560	-0.6091	0.1561	0.1159	-1.0000	1.0000

Table B.4: Spearman rank correlation matrix: Sandy clay loam

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.3606	0.0037	-0.7250	-0.0624	-0.1838	-0.2078	0.2078
K_S	-0.3606	1.0000	0.0603	0.1487	0.3299	0.1840	-0.0492	0.0492
θ_R	0.0037	0.0603	1.0000	0.1262	-0.0709	0.6979	0.0134	-0.0134
θ_S	-0.7250	0.1487	0.1262	1.0000	0.1139	0.2661	0.2297	-0.2297
h_b	-0.0624	0.3299	-0.0709	0.1139	1.0000	-0.3177	0.0405	-0.0405
λ_p	-0.1838	0.1840	0.6979	0.2661	-0.3177	1.0000	-0.2365	0.2365
$C_{s,\text{solids}}$	-0.2078	-0.0492	0.0134	0.2297	0.0405	-0.2365	1.0000	-1.0000
$k_{s,\text{dry}}$	0.2078	0.0492	-0.0134	-0.2297	-0.0405	0.2365	-1.0000	1.0000

Table B.5: Spearman rank correlation matrix: Silt clay loam

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.5562	-0.0753	-0.8222	0.0300	-0.2506	-0.4207	0.4207
K_S	-0.5562	1.0000	0.1651	0.6593	0.3810	0.1309	0.3025	-0.3025
θ_R	-0.0753	0.1651	1.0000	0.0678	-0.1218	0.7703	-0.2841	0.2841
θ_S	-0.8222	0.6593	0.0678	1.0000	0.2345	0.1783	0.3585	-0.3585
h_b	0.0300	0.3810	-0.1218	0.2345	1.0000	-0.4609	-0.0541	0.0541
λ_p	-0.2506	0.1309	0.7703	0.1783	-0.4609	1.0000	-0.2247	0.2247
$C_{s,\text{solids}}$	-0.4207	0.3025	-0.2841	0.3585	-0.0541	-0.2247	1.0000	-1.0000
$k_{s,\text{dry}}$	0.4207	-0.3025	0.2841	-0.3585	0.0541	0.2247	-1.0000	1.0000

Table B.6: Spearman rank correlation matrix: Clay loam

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.5097	0.1570	-0.5277	-0.0714	0.0510	-0.0760	0.0760
K_S	-0.5097	1.0000	-0.0890	0.5644	0.2041	-0.0710	0.0135	-0.0135
θ_R	0.1570	-0.0890	1.0000	0.1217	-0.1388	0.7402	0.0963	-0.0963
θ_S	-0.5277	0.5644	0.1217	1.0000	0.1216	0.0968	0.5062	-0.5062
h_b	-0.0714	0.2041	-0.1388	0.1216	1.0000	-0.6002	0.0835	-0.0835
λ_p	0.0510	-0.0710	0.7402	0.0968	-0.6002	1.0000	0.1270	-0.1270
$C_{s,\text{solids}}$	-0.0760	0.0135	0.0963	0.5062	0.0835	0.1270	1.0000	-1.0000
$k_{s,\text{dry}}$	0.0760	-0.0135	-0.0963	-0.5062	-0.0835	-0.1270	-1.0000	1.0000

Table B.7: Spearman rank correlation matrix: Sandy loam

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.3823	0.0427	-0.8055	0.0367	-0.1542	0.1081	-0.1082
K_S	-0.3823	1.0000	0.0998	0.3671	0.2823	0.1829	-0.2155	0.2156
θ_R	0.0427	0.0998	1.0000	0.1391	-0.1929	0.7434	0.2061	-0.2061
θ_S	-0.8055	0.3671	0.1391	1.0000	0.0288	0.2429	0.0298	-0.0297
h_b	0.0367	0.2823	-0.1929	0.0288	1.0000	-0.4145	-0.2801	0.2801
λ_p	-0.1542	0.1829	0.7434	0.2429	-0.4145	1.0000	-0.0631	0.0631
$C_{s,\text{solids}}$	0.1081	-0.2155	0.2061	0.0298	-0.2801	-0.0631	1.0000	-1.0000
$k_{s,\text{dry}}$	-0.1082	0.2156	-0.2061	-0.0297	0.2801	0.0631	-1.0000	1.0000

Table B.8: Spearman rank correlation matrix: Loam

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.3739	-0.0284	-0.8266	-0.1940	-0.0922	-0.3052	0.3052
K_S	-0.3739	1.0000	-0.1342	0.4363	0.3554	-0.1555	0.0269	-0.0269
θ_R	-0.0284	-0.1342	1.0000	0.0550	-0.3097	0.7255	0.1132	-0.1132
θ_S	-0.8266	0.4363	0.0550	1.0000	0.2055	0.1090	0.3377	-0.3377
h_b	-0.1940	0.3554	-0.3097	0.2055	1.0000	-0.6283	-0.0688	0.0688
λ_p	-0.0922	-0.1555	0.7255	0.1090	-0.6283	1.0000	-0.0344	0.0344
$C_{s,\text{solids}}$	-0.3052	0.0269	0.1132	0.3377	-0.0688	-0.0344	1.0000	-1.0000
$k_{s,\text{dry}}$	0.3052	-0.0269	-0.1132	-0.3377	0.0688	0.0344	-1.0000	1.0000

Table B.9: Spearman rank correlation matrix: Silt loam

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.4467	0.0866	-0.6324	-0.3667	0.0471	0.1121	-0.1119
K_S	-0.4467	1.0000	-0.2958	0.1741	0.4105	-0.2229	-0.2709	0.2708
θ_R	0.0866	-0.2958	1.0000	0.2227	-0.3049	0.6924	0.2680	-0.2680
θ_S	-0.6324	0.1741	0.2227	1.0000	0.3832	0.0701	0.1453	-0.1455
h_b	-0.3667	0.4105	-0.3049	0.3832	1.0000	-0.5970	-0.0650	0.0649
λ_p	0.0471	-0.2229	0.6924	0.0701	-0.5970	1.0000	-0.0736	0.0734
$C_{s,\text{solids}}$	0.1121	-0.2709	0.2680	0.1453	-0.0650	-0.0736	1.0000	-1.0000
$k_{s,\text{dry}}$	-0.1119	0.2708	-0.2680	-0.1455	0.0649	0.0734	-1.0000	1.0000

Table B.10: Spearman rank correlation matrix: Sand

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.1894	-0.0230	-0.7050	0.0389	0.0375	0.0935	-0.0935
K_S	-0.1894	1.0000	0.0335	-0.0413	0.5915	0.4088	-0.1936	0.1936
θ_R	-0.0230	0.0335	1.0000	-0.0080	0.0695	0.2114	0.1502	-0.1501
θ_S	-0.7050	-0.0413	-0.0080	1.0000	-0.1480	-0.2530	-0.1108	0.1107
h_b	0.0389	0.5915	0.0695	-0.1480	1.0000	-0.0757	0.1719	-0.1720
λ_p	0.0375	0.4088	0.2114	-0.2530	-0.0757	1.0000	-0.2957	0.2959
$C_{s,\text{solids}}$	0.0935	-0.1936	0.1502	-0.1108	0.1719	-0.2957	1.0000	-1.0000
$k_{s,\text{dry}}$	-0.0935	0.1936	-0.1501	0.1107	-0.1720	0.2959	-1.0000	1.0000

Table B.11: Spearman rank correlation matrix: Loamy sand

	ρ_b	K_S	θ_R	θ_S	h_b	λ_p	$C_{s,\text{solids}}$	$k_{s,\text{dry}}$
ρ_b	1.0000	-0.3344	0.1674	-0.8058	0.0629	-0.0251	0.1087	-0.1087
K_S	-0.3344	1.0000	0.2080	0.3008	0.5620	0.3927	-0.0721	0.0721
θ_R	0.1674	0.2080	1.0000	-0.1078	0.0164	0.6740	0.2264	-0.2264
θ_S	-0.8058	0.3008	-0.1078	1.0000	0.0313	-0.0511	0.1094	-0.1094
h_b	0.0629	0.5620	0.0164	0.0313	1.0000	-0.1449	0.0209	-0.0209
λ_p	-0.0251	0.3927	0.6740	-0.0511	-0.1449	1.0000	-0.1518	0.1518
$C_{s,\text{solids}}$	0.1087	-0.0721	0.2264	0.1094	0.0209	-0.1518	1.0000	-1.0000
$k_{s,\text{dry}}$	-0.1087	0.0721	-0.2264	-0.1094	-0.0209	0.1518	-1.0000	1.0000

B.2 Histograms and distribution parameters

Comparing the empirical histograms with candidate parameterized pdfs, ρ_b, θ_R, θ_S, h_b, λ_p, $C_{s,\text{solids}}$, and $k_{s,\text{dry}}$ were found to behave most closely to the two-parameter beta distribution while $\log K_S$ was found to be well approximated by a normal distribution. Additionally, there are a substantial number of records with $\theta_R = 0$, and therefore the marginal behavior of θ_R was modeled as a mixed discrete-continuous distribution. This mixed distribution places an atom of probability at $\theta_R = 0$ with a mass estimated empirically from the fraction of records with $\theta_R = 0$, while the distribution of θ_R conditioned on $\theta_R > 0$ is modeled as a two parameter beta distribution.

B.2.1 The beta distribution and parameter estimation

The probability density function for a random variable x that follows a two-parameter beta distribution can be analytically written as,

$$f_x(x) = \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)}, \tag{B.1}$$

where the beta function $\mathrm{B}(a,b)$ is a normalization that insures that the beta distribution pdf integrates to unity, and is given by,

$$\mathrm{B}(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt. \tag{B.2}$$

The parameters of the beta distribution are estimated from the method of moments,

$$a = \bar{x}\left(\frac{\bar{x}(1-\bar{x})}{v} - 1\right), \tag{B.3}$$

and

$$b = (1 - \bar{x}) \left(\frac{\bar{x}(1 - \bar{x})}{v} - 1 \right). \tag{B.4}$$

In the above $\bar{x}$ and v are the sample mean and variance of the data, respectively, and can be estimated as,

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{B.5}$$

and

$$\bar{x} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2, \tag{B.6}$$

where N is the number of records.

B.2.2 Data and analysis results

Table B.12: Beta distribution parameters: Bulk density [kg/m^3]

	a	b	Minimum value	Maximum value
Clay	1.973	0.904	724	1700
Sandy clay loam	5.774	1.190	490	1820
Silty clay loam	3.312	1.347	710	1760
Clay loam	2.742	2.721	1050	1660
Sandy loam	7.591	3.604	640	1970
Loam	4.713	2.335	492	1870
Silt loam	5.565	2.360	780	1690
Sand	5.641	5.210	1070	1965
Loamy sand	3.869	1.364	700	1830

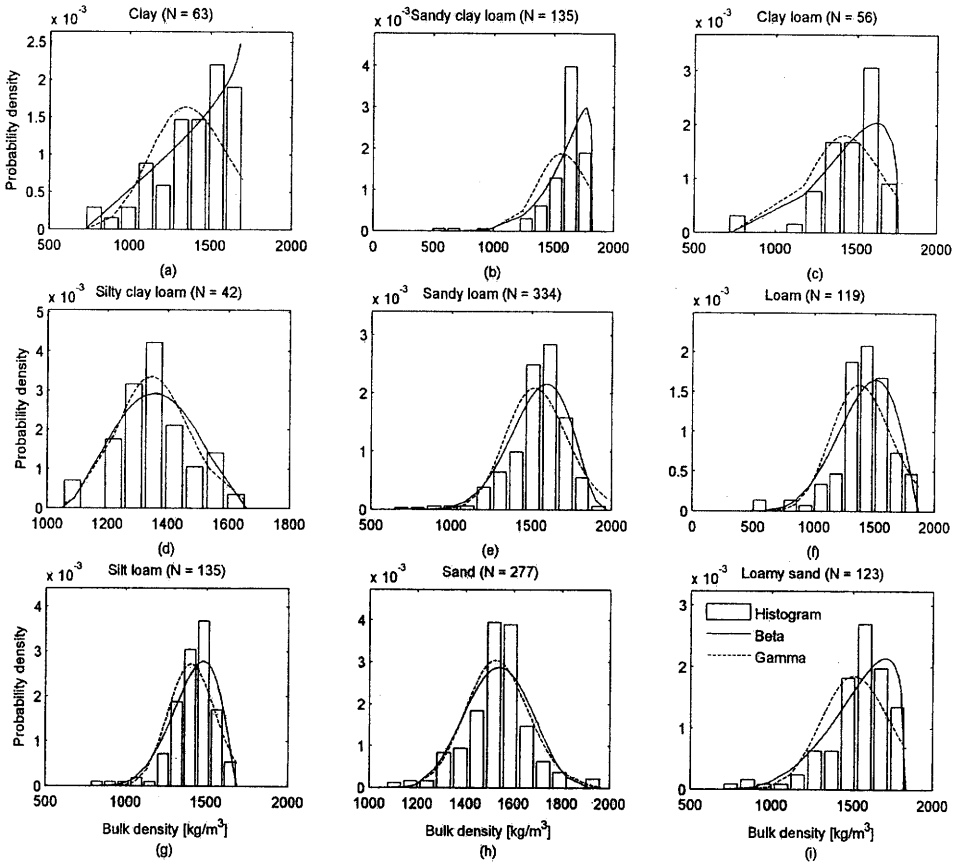


Figure B-1: Histograms of bulk density and fit trial distributions for (a) clay, (b) sandy clay loam, (c) clay loam, (d) silty clay loam, (e) sandy loam, (f) loam, (g) silt loam, (h) sand, and (i) loamy sand.

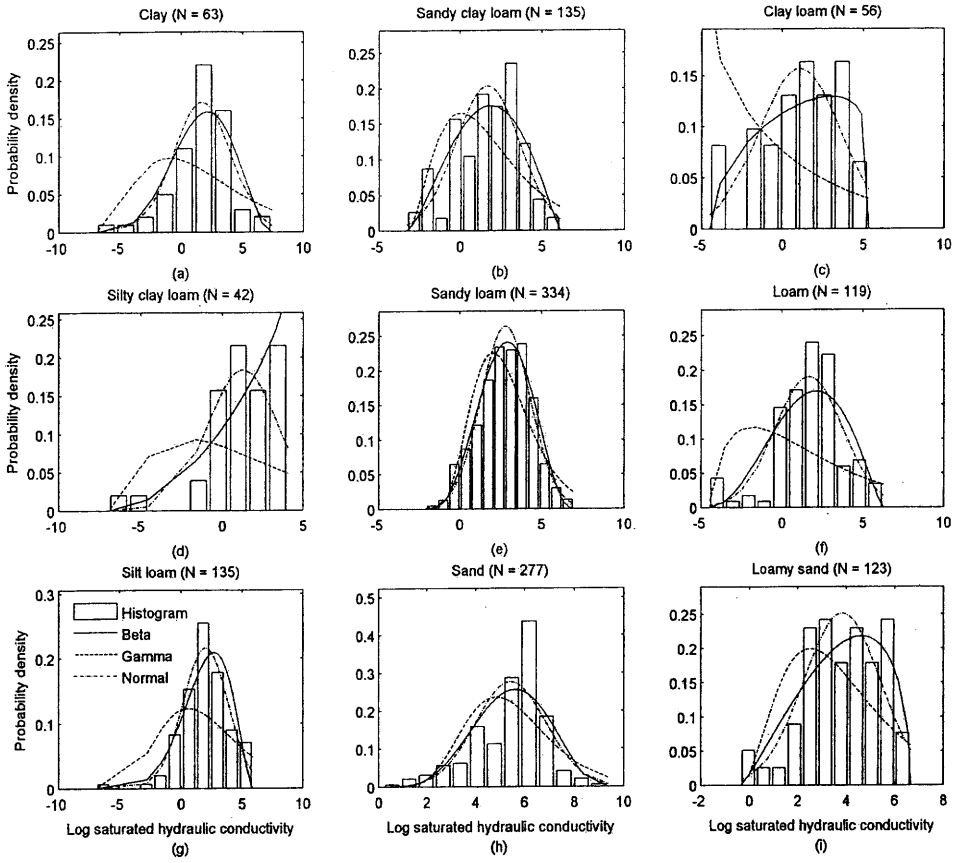


Figure B-2: Histograms of log of saturated hydraylic conductivity and fit trial distributions for (a) clay, (b) sandy clay loam, (c) clay loam, (d) silty clay loam, (e) sandy loam, (f) loam, (g) silt loam, (h) sand, and (i) loamy sand.

Table B.13: Gaussian distribution parameters: Log saturated hydraulic conductivity [log(mm/hr)]

	$\bar{x}$	s
Clay	1.667	2.333
Sandy clay loam	1.705	1.962
Silty clay loam	1.082	2.544
Clay loam	1.281	2.170
Sandy loam	2.813	1.514
Loam	1.669	2.093
Silt loam	1.984	1.847
Sand	5.431	1.440
Loamy sand	3.830	1.587

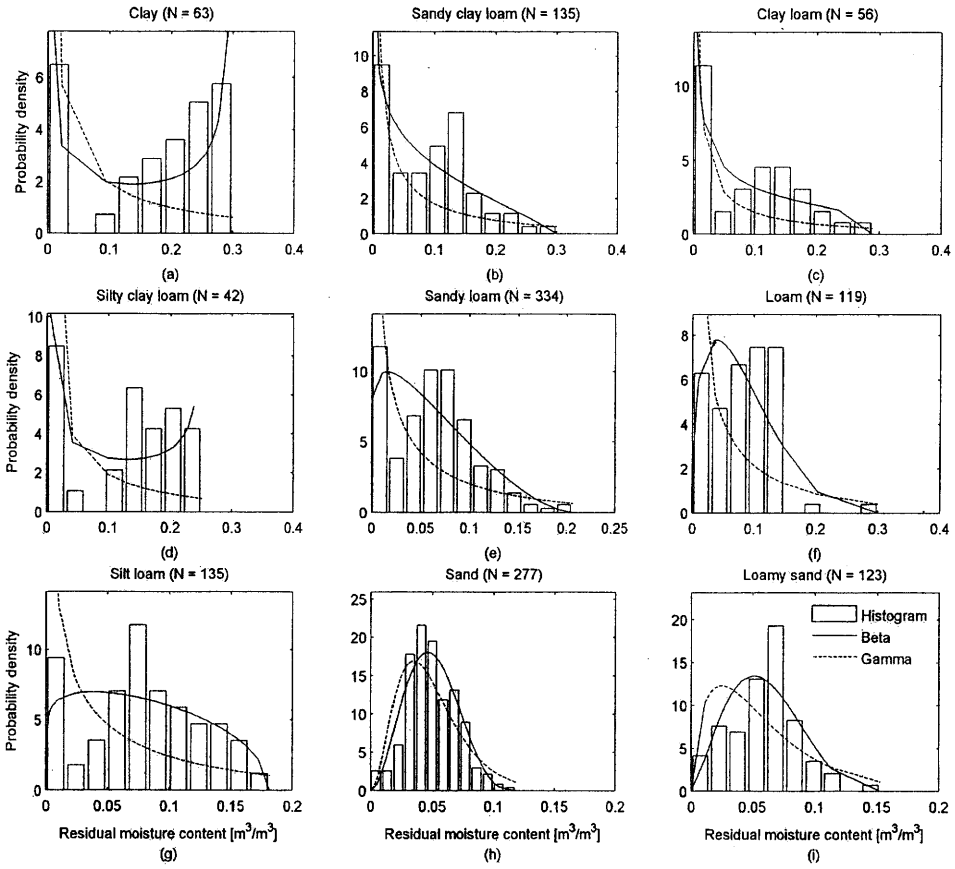


Figure B-3: Histograms of residual moisture content and fit trial distributions for (a) clay, (b) sandy clay loam, (c) clay loam, (d) silty clay loam, (e) sandy loam, (f) loam, (g) silt loam, (h) sand, and (i) loamy sand.

Table B.14: Beta distribution parameters: Residual moisture content [m^3/m^3]

	a	b	Minimum value	Maximum value	Probability $\theta_R = 0$
Clay	0.496	0.384	0.000001	0.300	0.413
Sandy clay loam	0.770	1.844	0.000001	0.300	0.348
Silty clay loam	0.546	1.222	0.000001	0.289	0.268
Clay loam	0.534	0.552	0.000001	0.252	0.286
Sandy loam	1.134	2.654	0.000001	0.206	0.362
Loam	1.413	3.815	0.000001	0.300	0.286
Silt loam	1.128	1.487	0.000001	0.182	0.237
Sand	3.074	4.228	0.000003	0.118	0.061
Loamy sand	2.496	3.972	0.000001	0.153	0.228

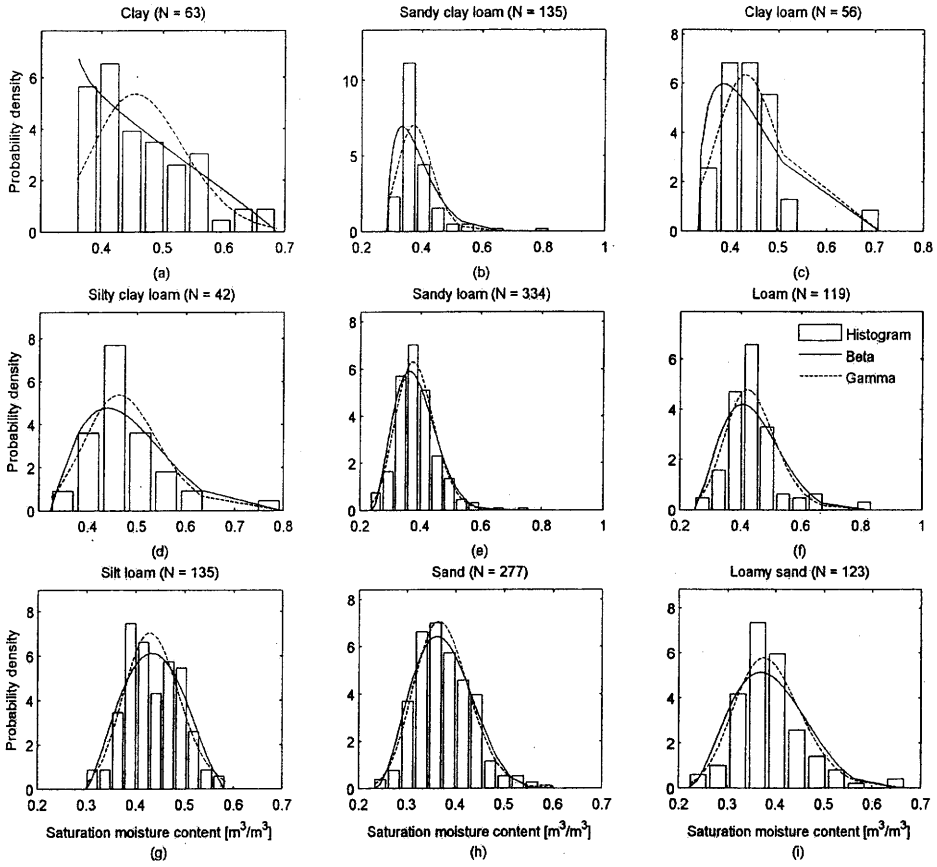


Figure B-4: Histograms of saturation moisture content and fit trial distributions for (a) clay, (b) sandy clay loam, (c) clay loam, (d) silty clay loam, (e) sandy loam, (f) loam, (g) silt loam, (h) sand, and (i) loamy sand.

Table B.15: Beta distribution parameters: Saturation moisture content [m^3/m^3]

	a	b	Minimum value	Maximum value
Clay	0.949	1.921	0.358	0.687
Sandy clay loam	1.656	7.467	0.285	0.820
Silty clay loam	1.400	3.621	0.335	0.712
Clay loam	2.169	4.576	0.323	0.797
Sandy loam	3.483	8.559	0.234	0.760
Loam	2.696	5.613	0.247	0.838
Silt loam	2.496	2.695	0.300	0.584
Sand	3.355	5.267	0.232	0.600
Loamy sand	3.017	5.049	0.220	0.670

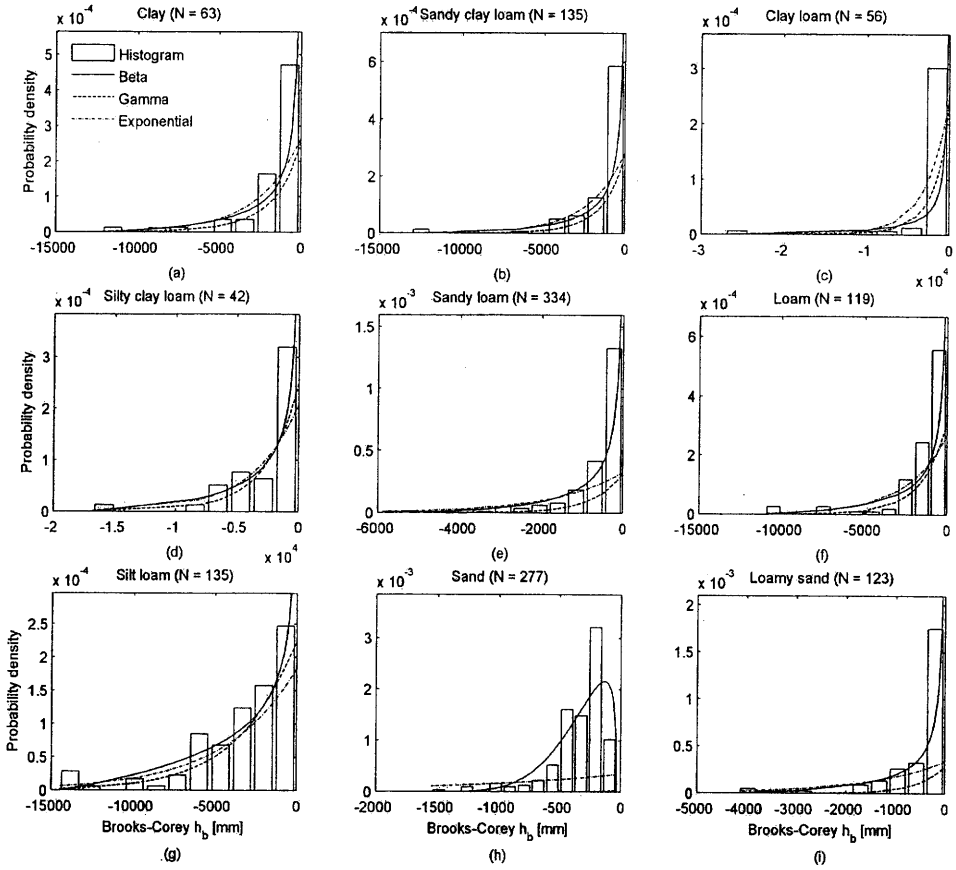


Figure B-5: Histograms of Brooks-Corey air entry pressure and fit trial distributions for (a) clay, (b) sandy clay loam, (c) clay loam, (d) silty clay loam, (e) sandy loam, (f) loam, (g) silt loam, (h) sand, and (i) loamy sand.

Table B.16: Beta distribution parameters: Brooks-Corey air entry pressure [mm]

	a	b	Minimum value	Maximum value
Clay	0.311	2.170	-10.00	-12150.67
Sandy clay loam	0.282	2.759	-10.00	-13089.01
Silty clay loam	0.138	1.941	-10.00	-27100.27
Clay loam	0.445	2.395	-14.50	-16778.52
Sandy loam	0.359	2.771	-10.10	-5878.89
Loam	0.330	2.175	-21.02	-10964.91
Silt loam	0.607	2.058	-10.01	-14492.75
Sand	1.293	5.183	-32.60	-1548.95
Loamy sand	0.315	2.126	-10.01	-4149.38

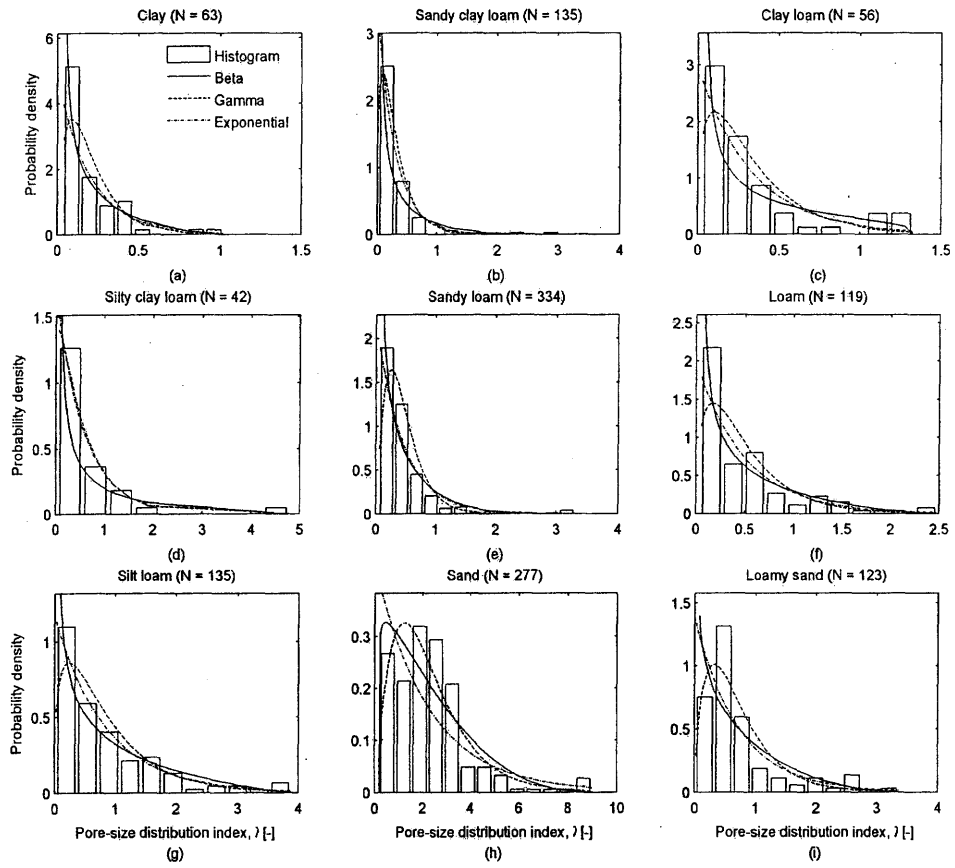


Figure B-6: Histograms of Brooks-Corey pore distribution index and fit trial distributions for (a) clay, (b) sandy clay loam, (c) clay loam, (d) silty clay loam, (e) sandy loam, (f) loam, (g) silt loam, (h) sand, and (i) loamy sand.

Table B.17: Beta distribution parameters: Brooks-Corey pore distribution index [dimensionless]

	a	b	Minimum value	Maximum value
Clay	0.505	2.463	0.040	1.018
Sandy clay loam	0.378	3.615	0.022	3.011
Silty clay loam	0.400	1.296	0.030	1.328
Clay loam	0.306	2.332	0.040	4.790
Sandy loam	0.622	4.468	0.063	3.295
Loam	0.540	2.407	0.055	2.480
Silt loam	0.596	2.190	0.029	3.881
Sand	1.066	3.448	0.211	9.000
Loamy sand	0.758	2.879	0.026	3.355

Table B.18: Beta distribution parameters: Volumetric heat capacity [J/(m^{3}K)]

	a	b	Minimum value	Maximum value
Clay	0.923	1.230	2250381.0	2369994.7
Sandy clay loam	1.247	1.451	2181290.1	2234012.5
Silty clay loam	0.990	1.371	2234183.5	2291770.5
Clay loam	1.294	1.736	2279619.2	2383245.7
Sandy loam	2.060	2.115	2137028.9	2197206.2
Loam	2.676	2.916	2162802.1	2265952.3
Silt loam	1.808	2.324	2156617.9	2384055.1
Sand	1.316	1.858	2128000.0	2148240.4
Loamy sand	2.889	2.020	2132794.8	2163974.8

Table B.19: Beta distribution parameters: Dry thermal conductivity [J/(msK)]

	a	b	Minimum value	Maximum value
Clay	1.230	0.923	3.263	6.000
Sandy clay loam	1.451	1.247	6.375	7.581
Silty clay loam	1.371	0.990	5.053	6.371
Clay loam	1.736	1.294	2.960	5.331
Sandy loam	2.115	2.060	7.217	8.593
Loam	2.916	2.676	5.644	8.004
Silt loam	2.324	1.808	2.942	8.145
Sand	1.858	1.316	8.337	8.800
Loamy sand	2.020	2.889	7.977	8.690

B.3 Time evolution of ensemble statistics for various ensemble sizes

Figures B-7 through B-13 depict the time evolution of 20 estimates of ensemble mean (figures B-7(a) through B-13(a)) and ensemble standard deviation (figures B-7(b) through B-13(b)) of soil moisture for varying ensemble sizes ranging from 16 ensemble replicates (figure B-7) through 1024 replicates (figure B-13). In these figures, solid black lines indicate ensemble simulations in which SHTP uncertainty was represented using the Restricted Pairing (RP) approach, while grey dashed lines indicate ensemble simulations in which SHTP uncertainty was treated with simple random sampling (SRS). Note that generally as the ensemble size increases, ensem-

ble mean and standard deviation in soil moisture are more consistently estimated. However, the ensemble statistics from the experiments that used the RP approach to sample the soil parameters are more consistent at lower ensemble sizes, relative to experiments that used SRS to generate realizations of the soil parameters.

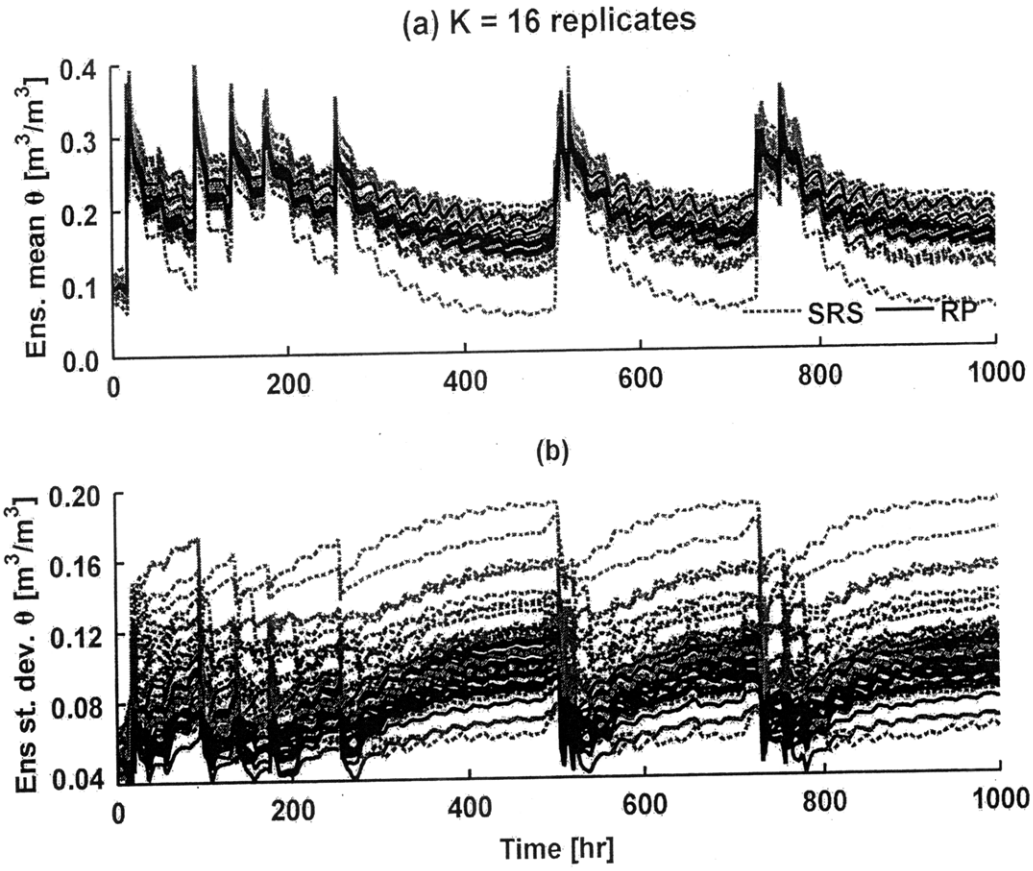


Figure B-7: For twenty 16 member ensembles, the time evolution of: (a) soil moisture in the top 10 cm [m^3/m^3], and (b) the standard deviation in soil moisture [m^3/m^3]. Gray dashed lines show ensembles in which SHTPs were generated using random sampling while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

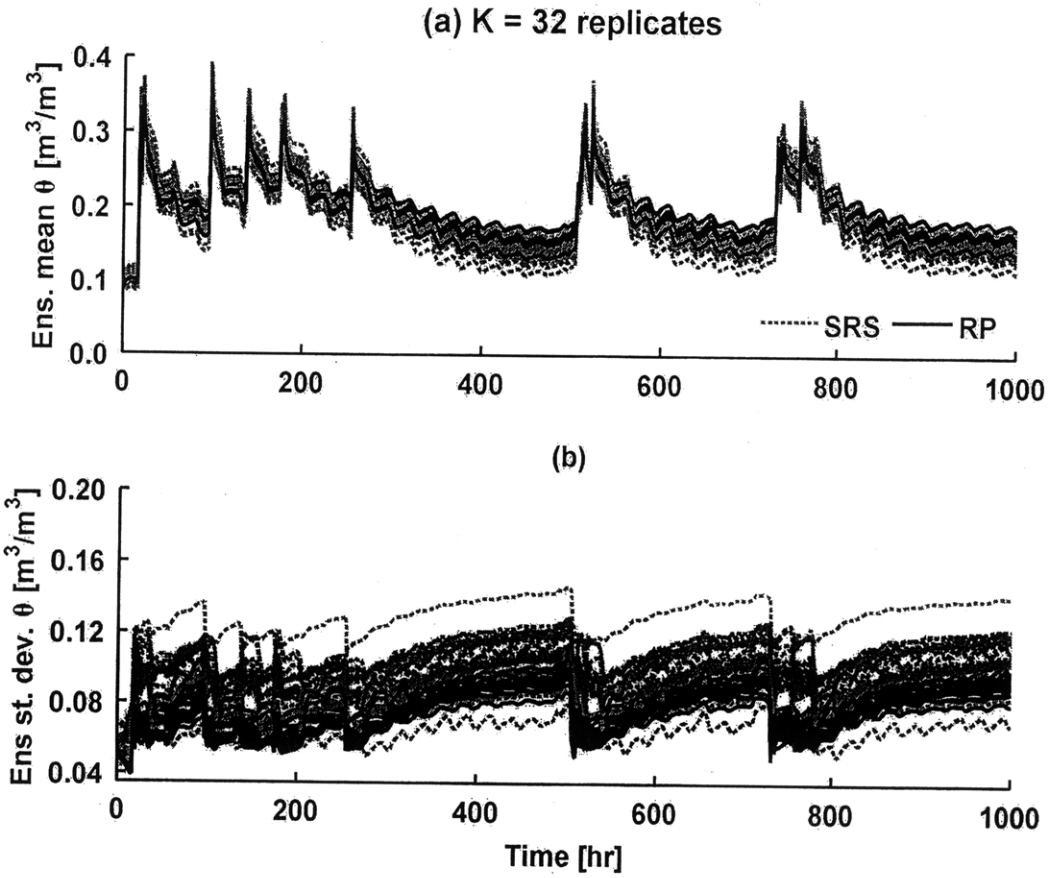


Figure B-8: For twenty 32 member ensembles, the time evolution of: (a) soil moisture in the top 10 cm [m^3/m^3], and (b) the standard deviation in soil moisture [m^3/m^3]. Gray dashed lines show ensembles in which SHTPs were generated using random sampling while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

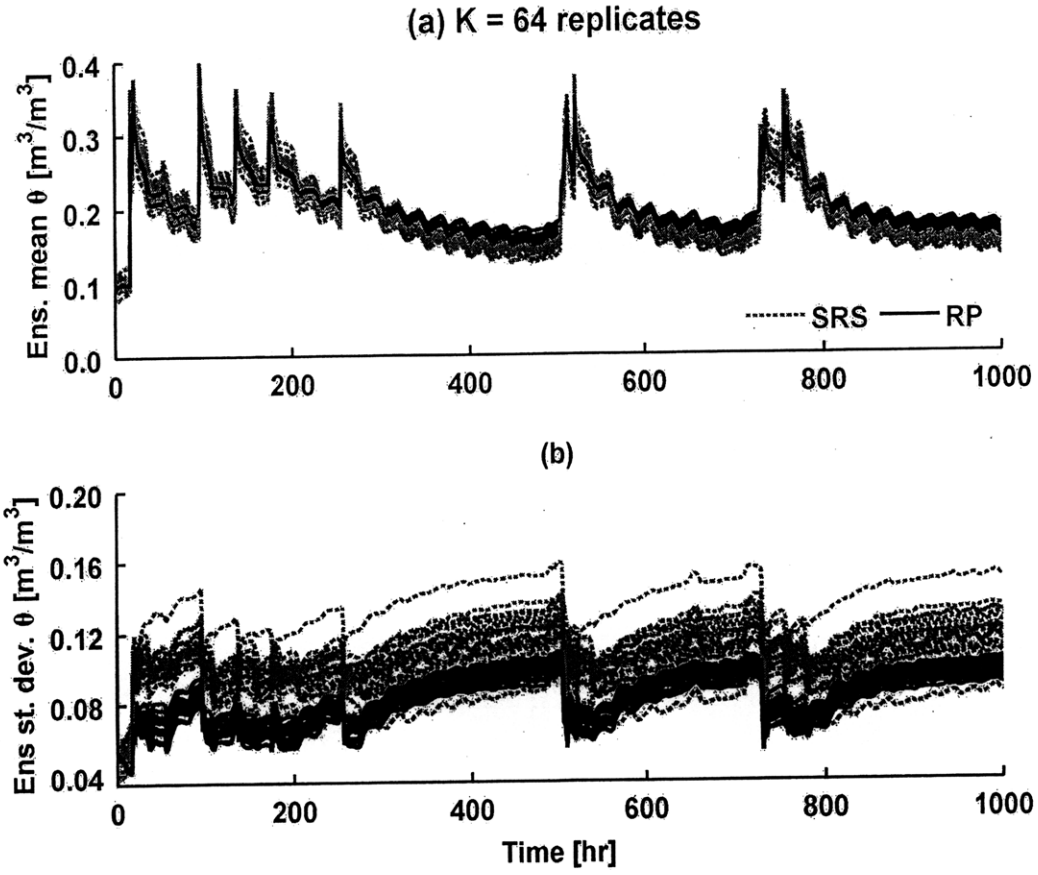


Figure B-9: For twenty 64 member ensembles, the time evolution of: (a) soil moisture in the top 10 cm [m^3/m^3], and (b) the standard deviation in soil moisture [m^3/m^3]. Gray dashed lines show ensembles in which SHTPs were generated using random sampling while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

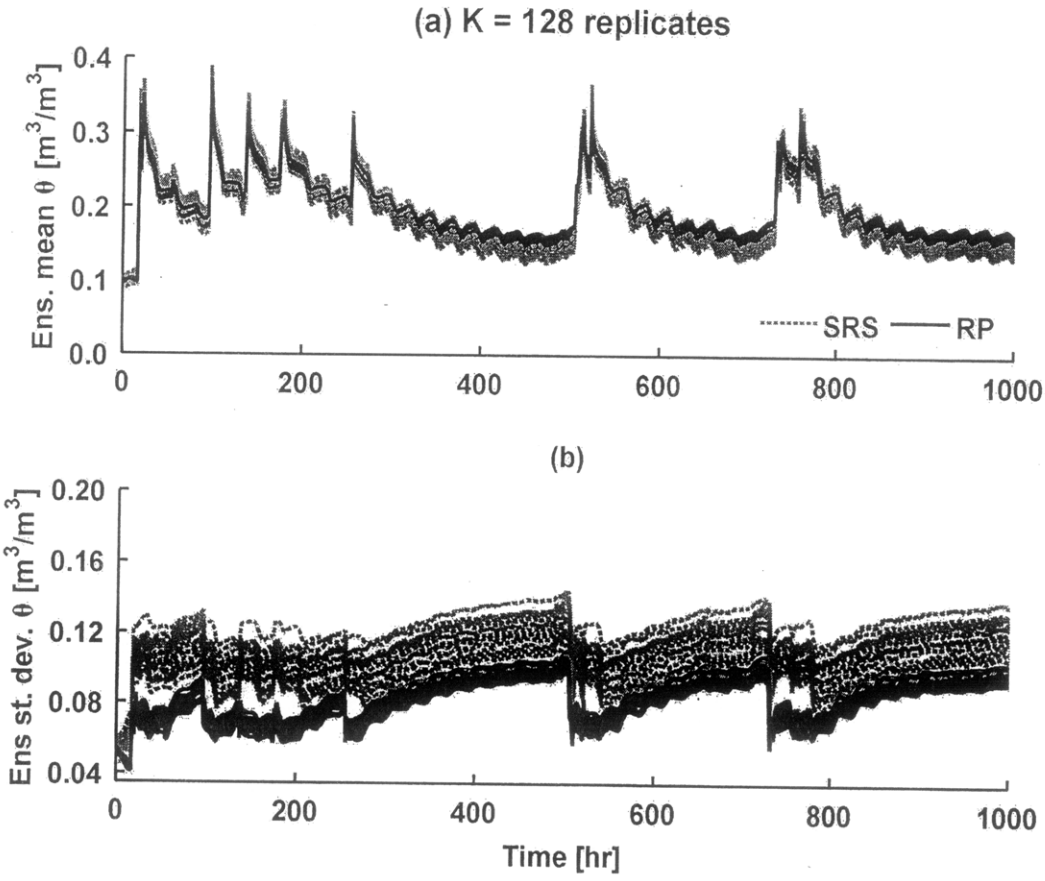


Figure B-10: For twenty 128 member ensembles, the time evolution of: (a) soil moisture in the top 10 cm [m^3/m^3], and (b) the standard deviation in soil moisture [m^3/m^3]. Gray dashed lines show ensembles in which SHTPs were generated using random sampling while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

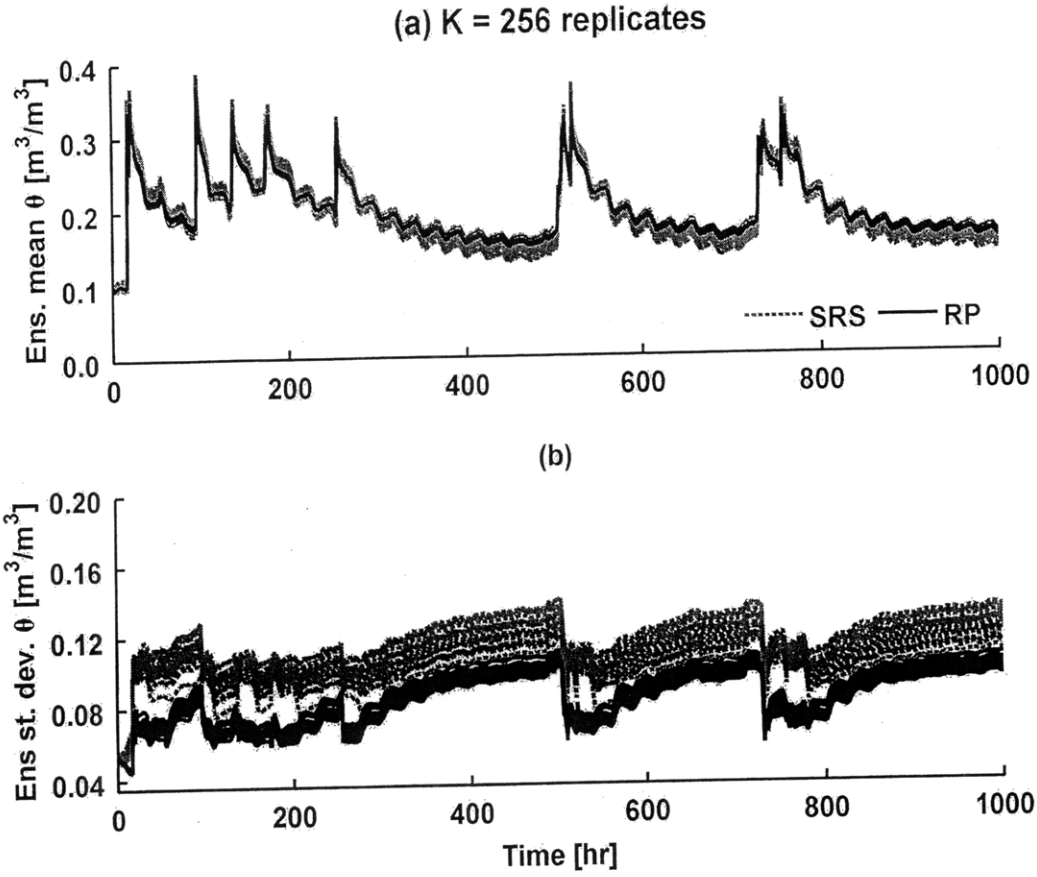


Figure B-11: For twenty 256 member ensembles, the time evolution of: (a) soil moisture in the top 10 cm [m^3/m^3], and (b) the standard deviation in soil moisture [m^3/m^3]. Gray dashed lines show ensembles in which SHTPs were generated using random sampling while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

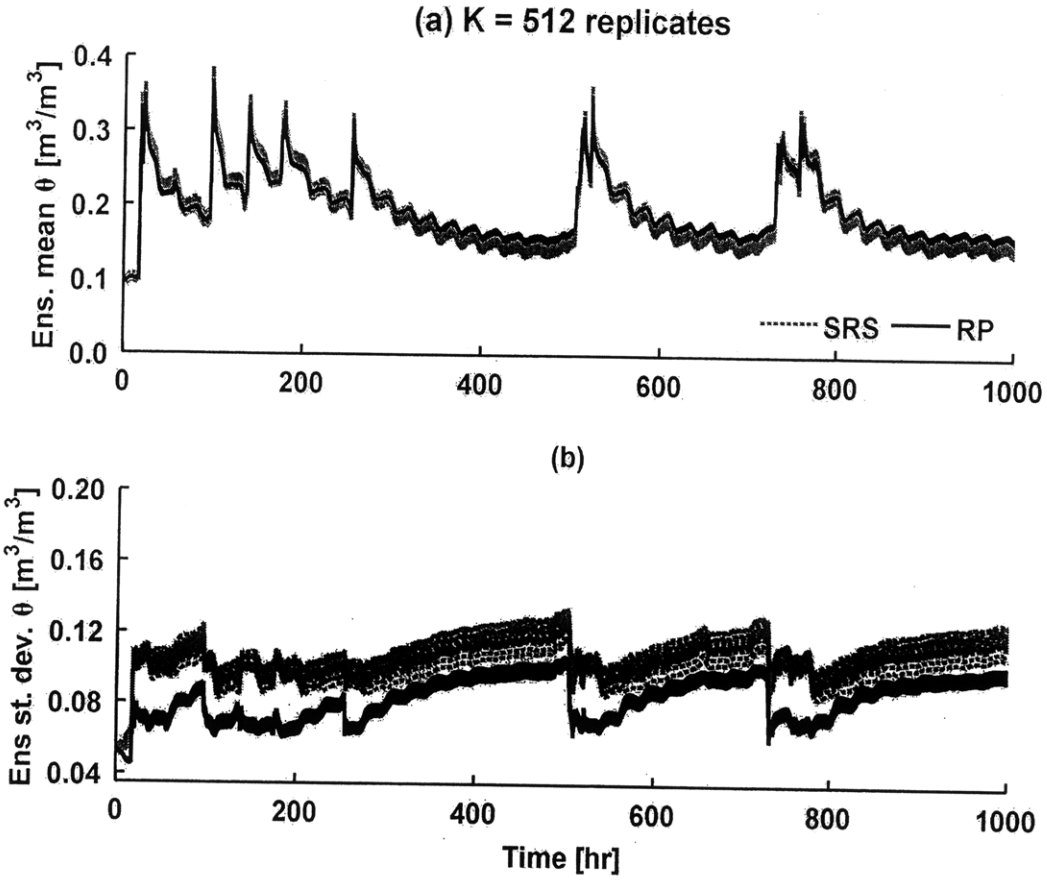


Figure B-12: For twenty 512 member ensembles, the time evolution of: (a) soil moisture in the top 10 cm [m^3/m^3], and (b) the standard deviation in soil moisture [m^3/m^3]. Gray dashed lines show ensembles in which SHTPs were generated using random sampling while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

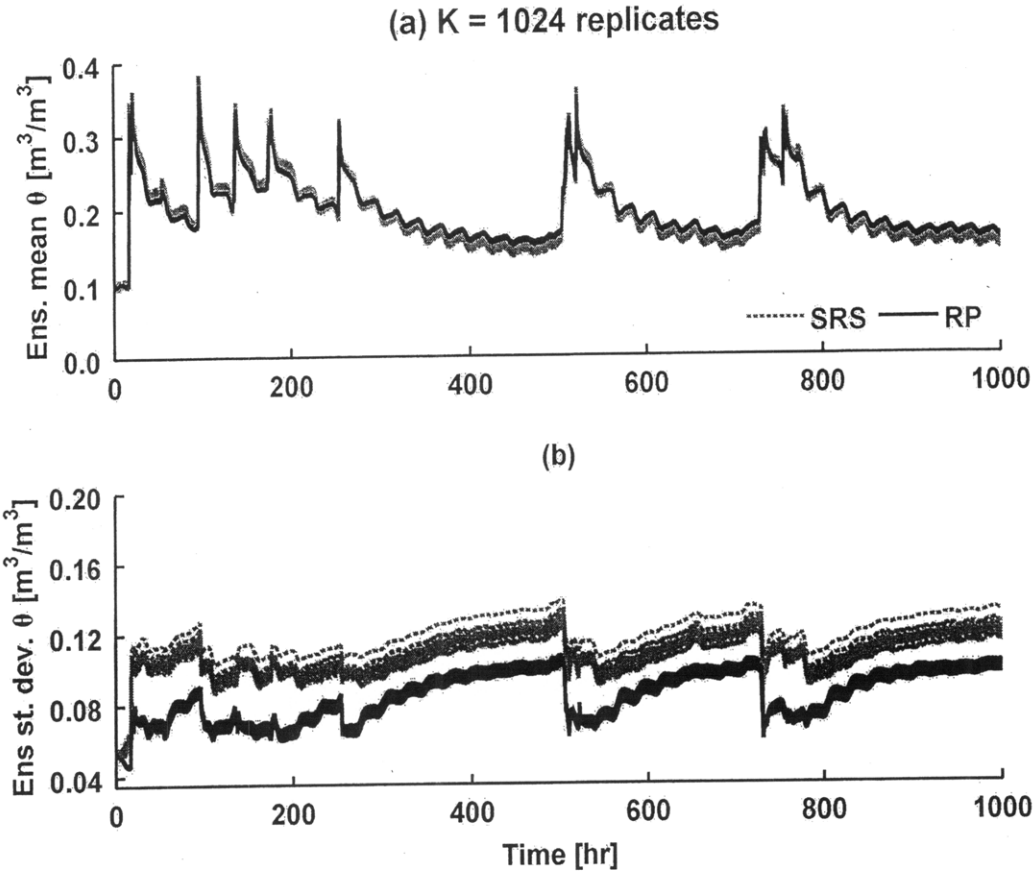


Figure B-13: For twenty 1024 member ensembles, the time evolution of: (a) soil moisture in the top 10 cm [m^3/m^3], and (b) the standard deviation in soil moisture [m^3/m^3]. Gray dashed lines show ensembles in which SHTPs were generated using random sampling while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

APPENDIX C

SENSITIVITY ANALYSIS OF SOIL MOISTURE (EXPANDED)

This appendix represents an unabridged presentation of the graphical results from spatially distributed soil moisture sensitivity analysis presented in **Chapter 7**. This appendix is split into three sections. The first section displays the results from a suite of point-scale sensitivity analyses. For each ensemble of hydrometeorological forcings developed for **Scenarios 2-5** in **Chapter 7** of the thesis, the near-surface soil moisture response is simulated under the assumption of deterministic soil hydraulic and thermal properties. These corresponding sets of sensitivity analyses are referred to as **Scenarios 2A-5A**. As outlined in the thesis, **Scenarios 2-5** are related to **Scenarios 2A-5A** in the following manner: **Scenario 2A** uses the same hydrometeorological forcings as **Scenario 2**, but assumes the soil properties are fixed through the ensemble, etc. Figures C-1 through C-4 compare the time evolution of near-surface soil moisture for each hydrometeorological forcings scenario.

The second and third sections depict results of the spatially distributed sensitivity analysis performed with Walnut Gulch Experimental Watershed. The second section shows results related to the near-surface (top 10 cm) soil moisture, while the third depicts results related to the profile-integrated soil moisture. Within each section dedicated to results of the spatially distribute sensitivity analysis, graphical results are subdivided into two subsections. The first subsection presents the first- and second-order statistics (ensemble mean, standard deviation, and coefficient of

variation) of the ensemble soil moisture distribution at each 72 hr interval of the simulation. The second subsection presents the first 4 empirical orthogonal functions (EOFs) at each 72 hour interval of the simulation, obtained through a singular through a singular value decomposition of a the matrix containing the ensemble of soil moisture anomalies at each time 72 hr interval. The percent variance explained by each of the four EOFs shown is indicated in the figure caption.

C.1 Point scale results

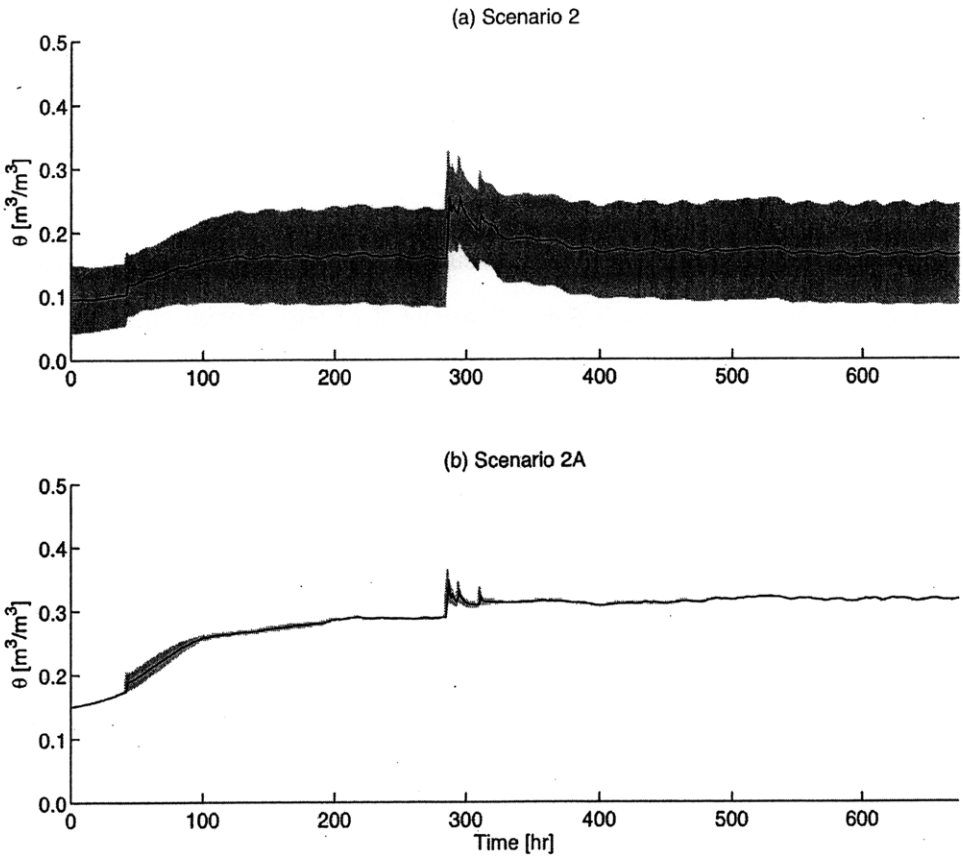


Figure C-1: Ensemble soil moisture response to the hydrometeorological forcings developed for **Scenario 2** for: (a) stochastic soil hydraulic and thermal properties generated with the Restricted Pairing-based approach, and (b) deterministic soil hydraulic and thermal properties. The solid black line in each plot represents the ensemble mean, while the grey shaded area encompasses one ensemble standard deviation above and below the ensemble mean soil moisture response.

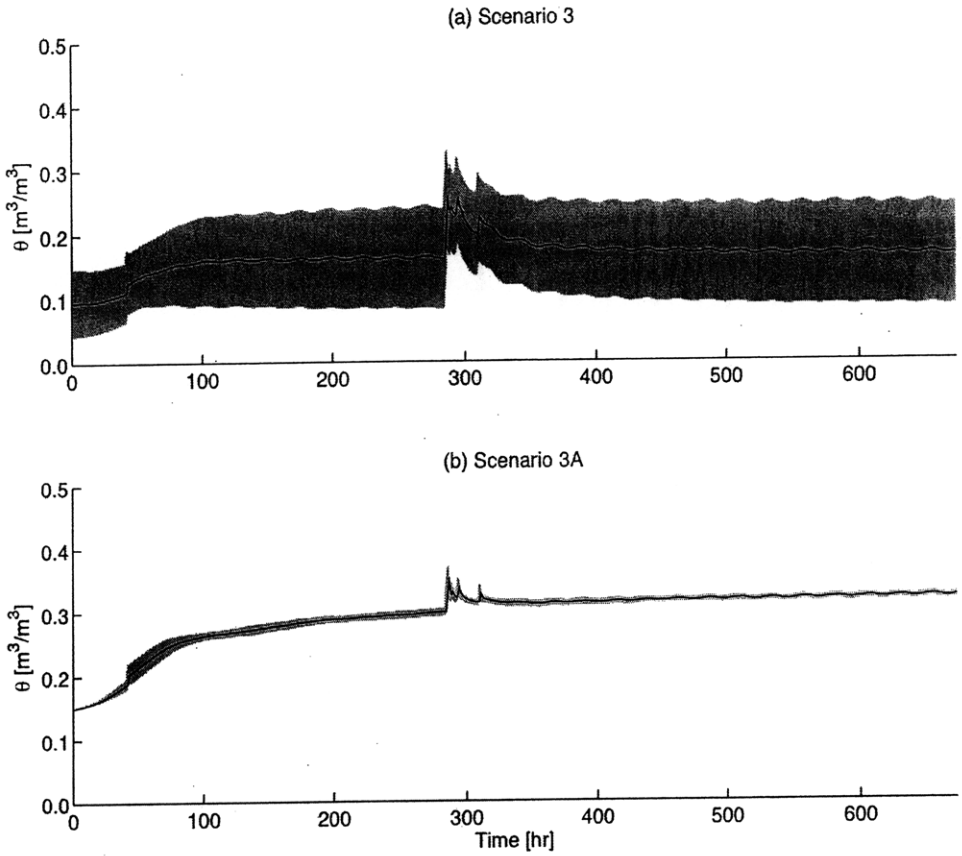


Figure C-2: Ensemble soil moisture response to the hydrometeorological forcings developed for **Scenario 3** for: (a) stochastic soil hydraulic and thermal properties generated with the Restricted Pairing-based approach, and (b) deterministic soil hydraulic and thermal properties. The solid black line in each plot represents the ensemble mean, while the grey shaded area encompasses one ensemble standard deviation above and below the ensemble mean soil moisture response.

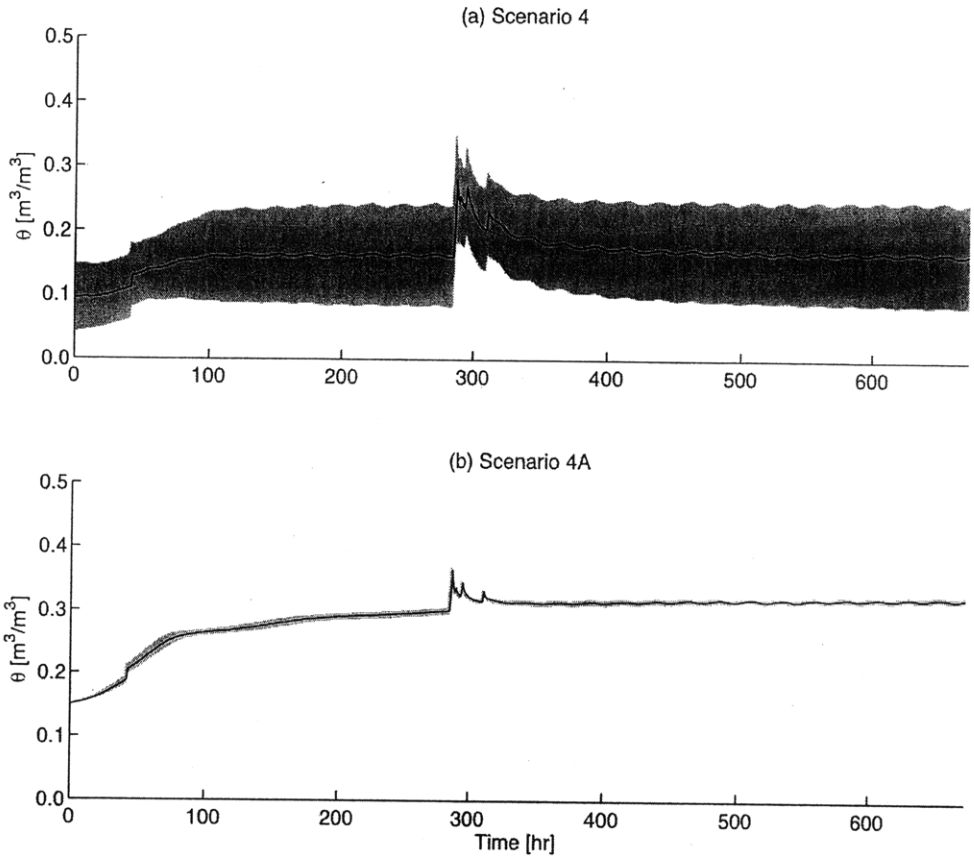


Figure C-3: Ensemble soil moisture response to the hydrometeorological forcings developed for **Scenario 4** for: (a) stochastic soil hydraulic and thermal properties generated with the Restricted Pairing-based approach, and (b) deterministic soil hydraulic and thermal properties. The solid black line in each plot represents the ensemble mean, while the grey shaded area encompasses one ensemble standard deviation above and below the ensemble mean soil moisture response.

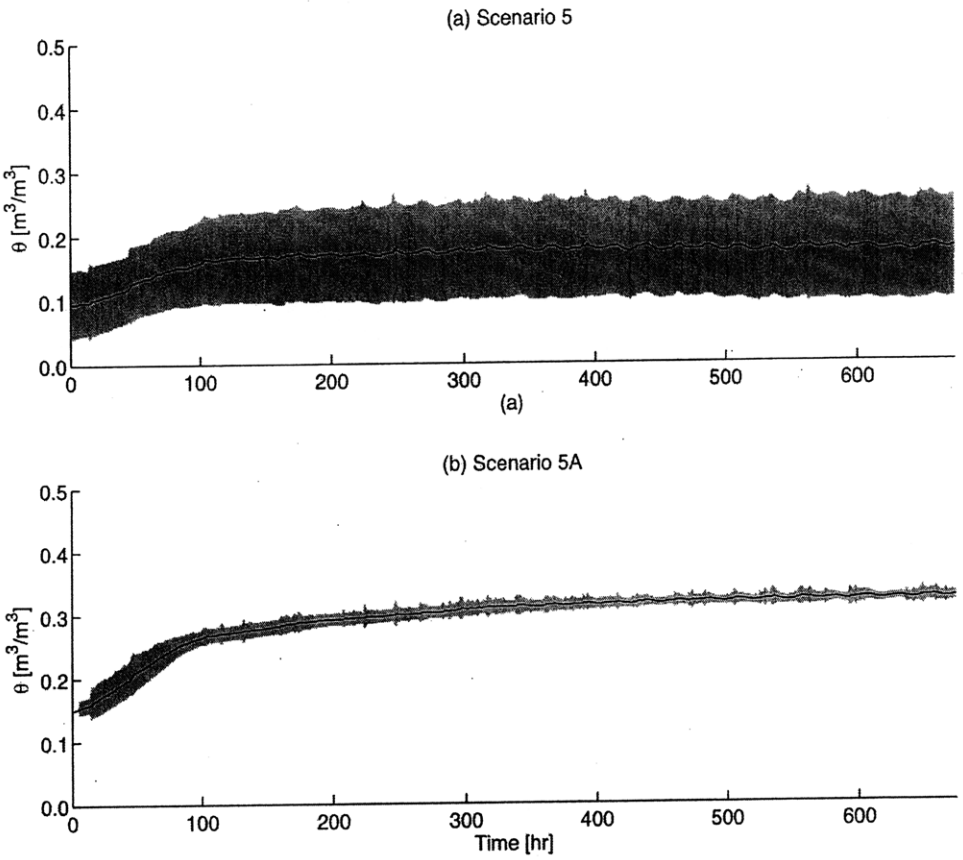
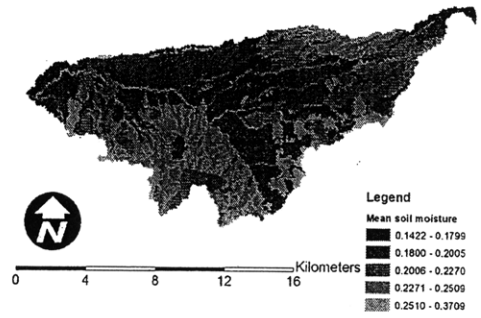


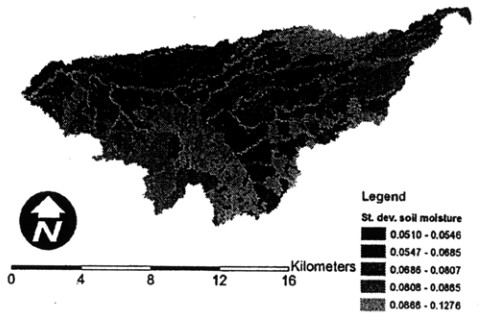
Figure C-4: Ensemble soil moisture response to the hydrometeorological forcings developed for **Scenario 5** for: (a) stochastic soil hydraulic and thermal properties generated with the Restricted Pairing-based approach, and (b) deterministic soil hydraulic and thermal properties. The solid black line in each plot represents the ensemble mean, while the grey shaded area encompasses one ensemble standard deviation above and below the ensemble mean soil moisture response.

C.2 Spatially distributed near-surface soil moisture results

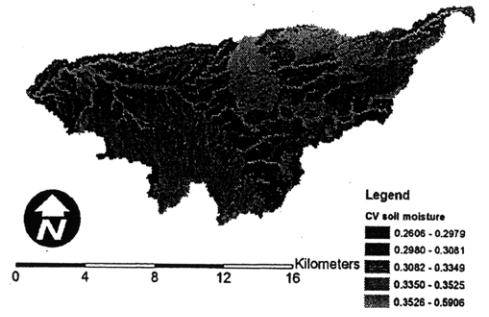
C.2.1 First- and second-order moments



(a)

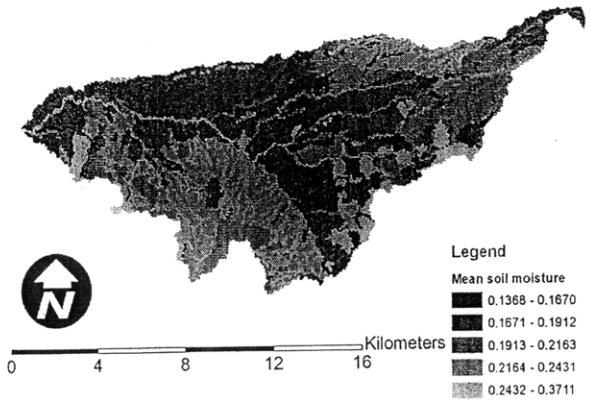


(b)

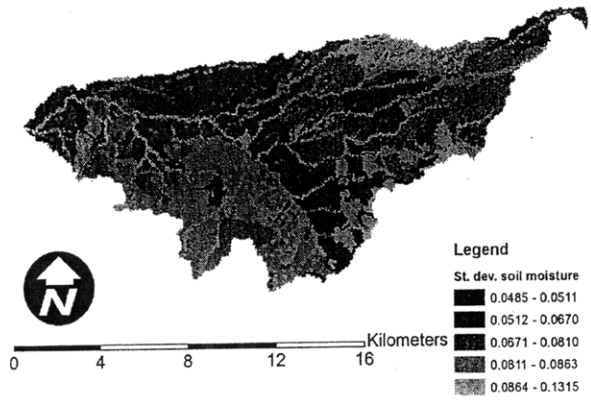


(c)

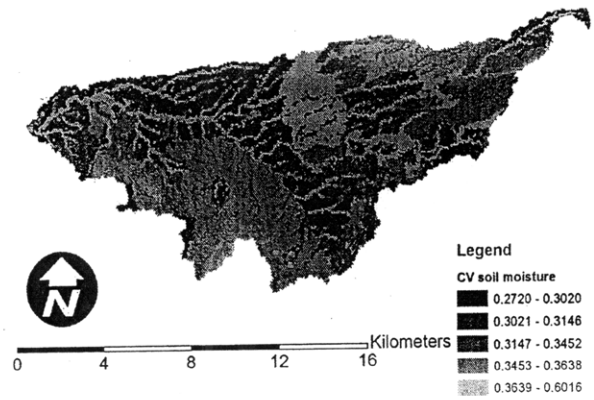
Figure C-5: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 0 hrs into the 1024 replicate open loop simulation.



(a)

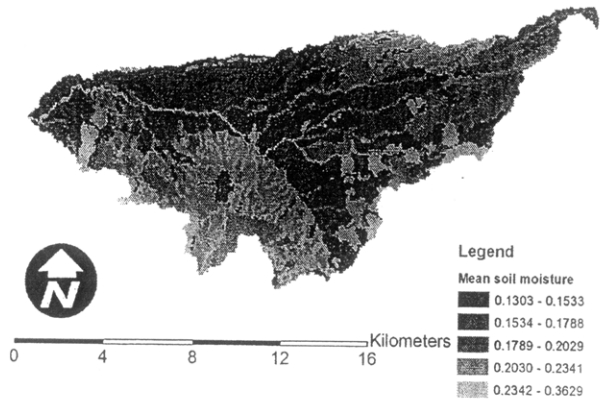


(b)

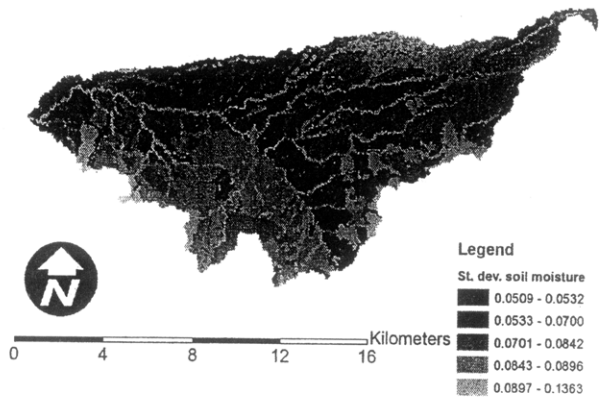


(c)

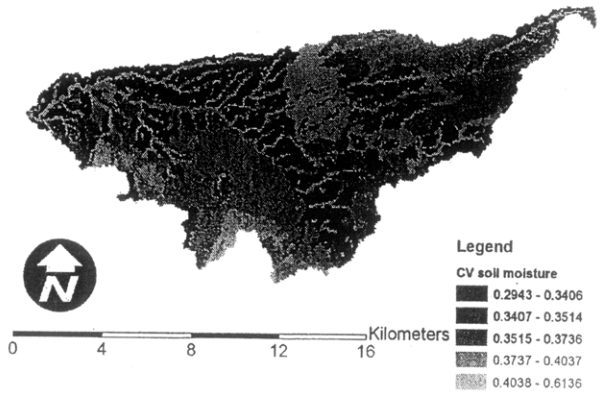
Figure C-6: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 72 hrs into the 1024 replicate open loop simulation.



(a)

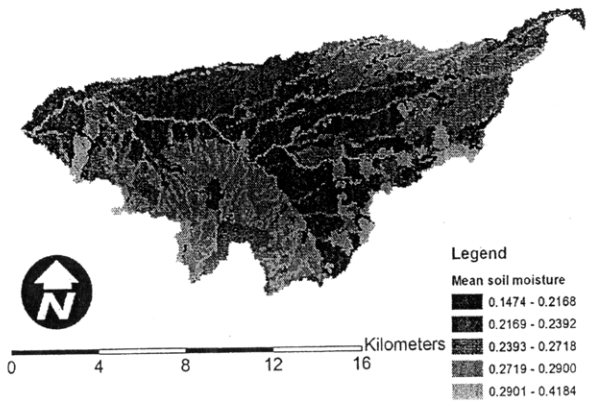


(b)

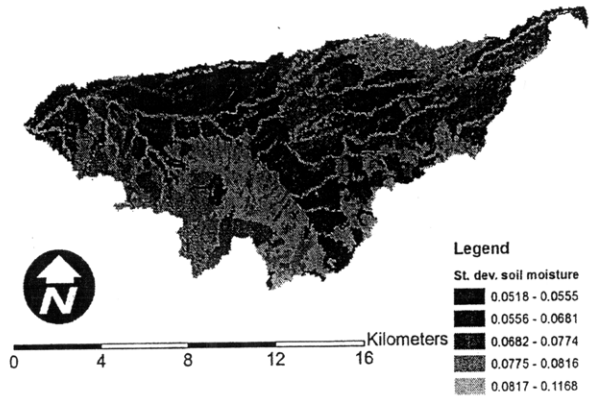


(c)

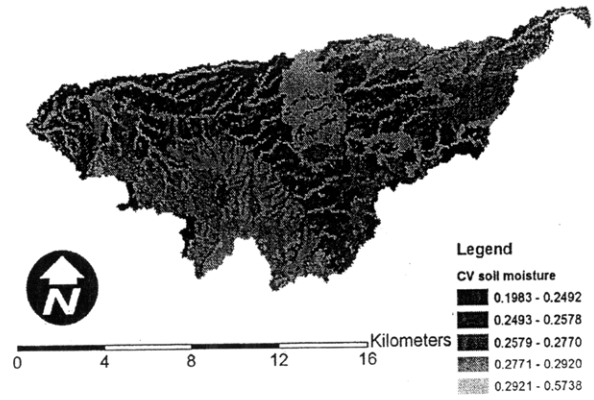
Figure C-7: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 144 hrs into the 1024 replicate open loop simulation.



(a)

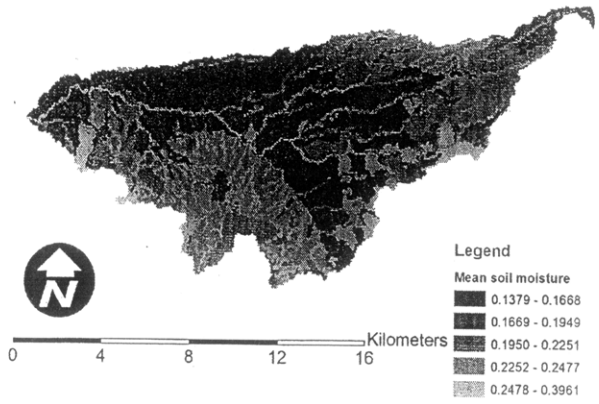


(b)

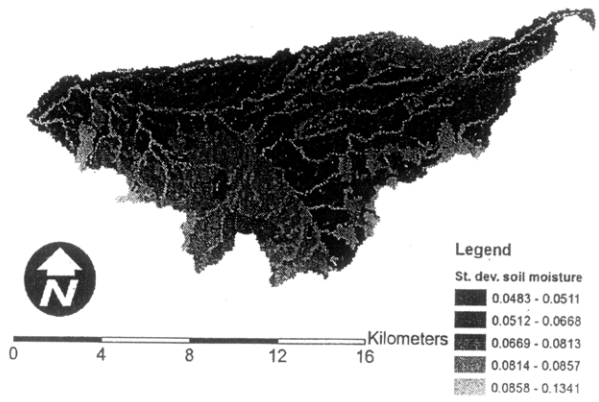


(c)

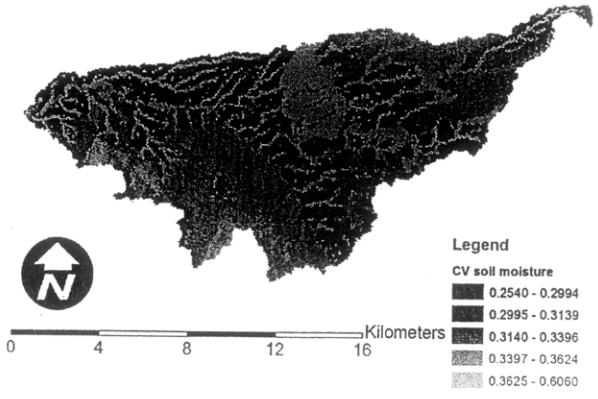
Figure C-8: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 216 hrs into the 1024 replicate open loop simulation.



(a)

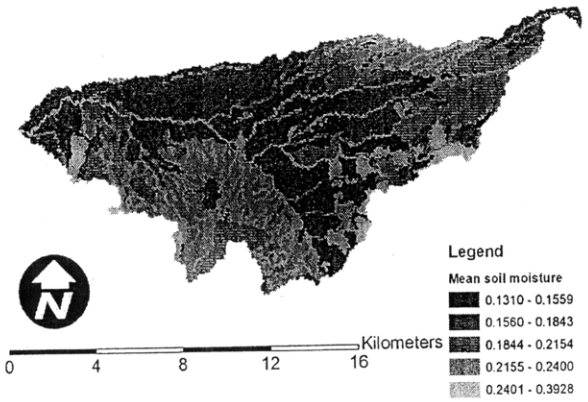


(b)

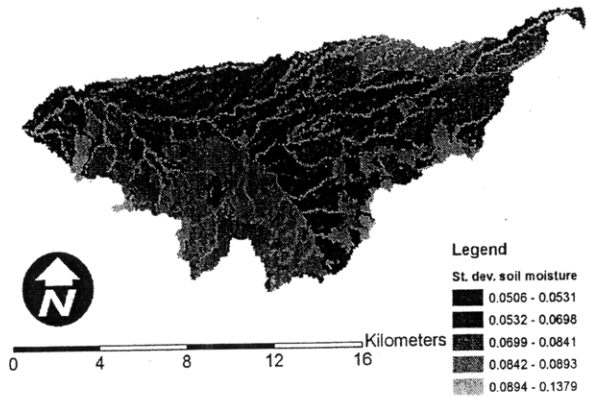


(c)

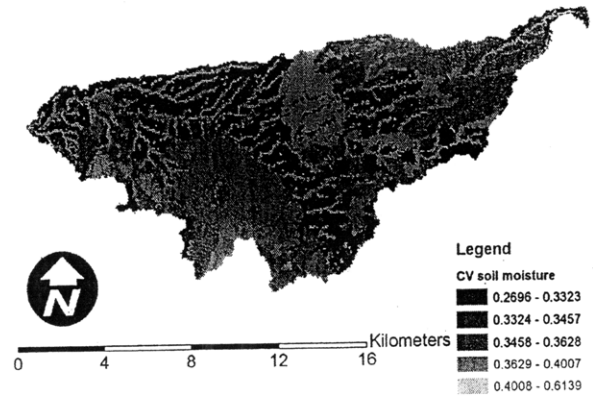
Figure C-9: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 288 hrs into the 1024 replicate open loop simulation.



(a)

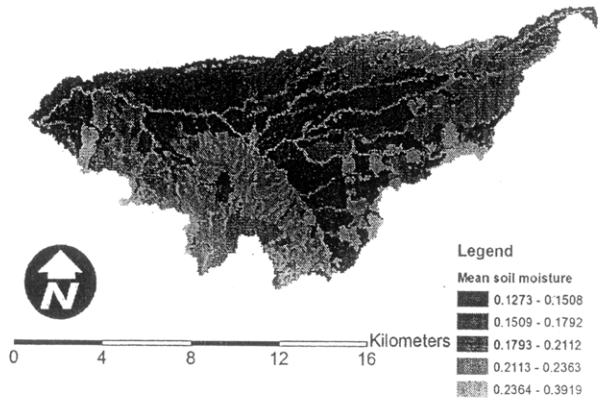


(b)

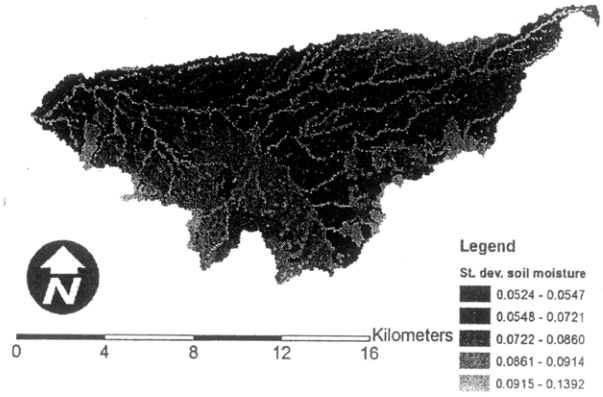


(c)

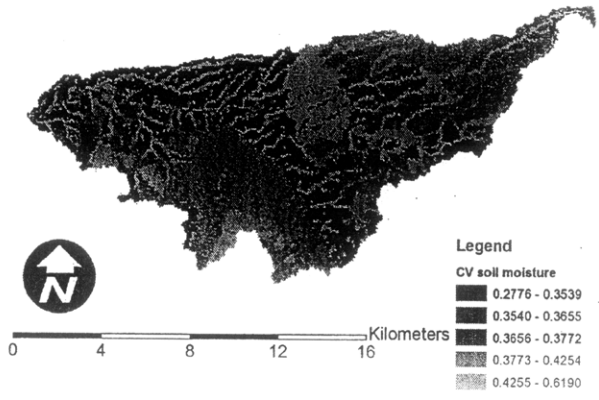
Figure C-10: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 360 hrs into the 1024 replicate open loop simulation.



(a)

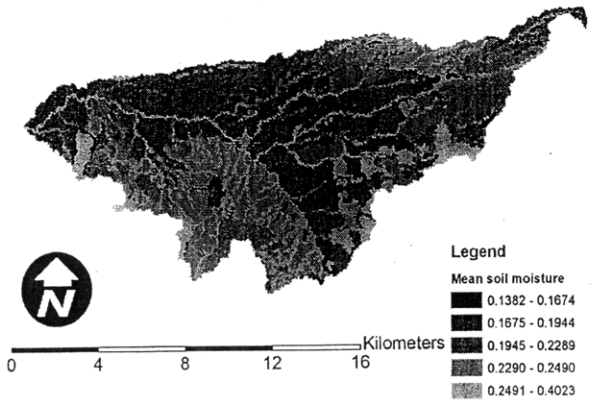


(b)

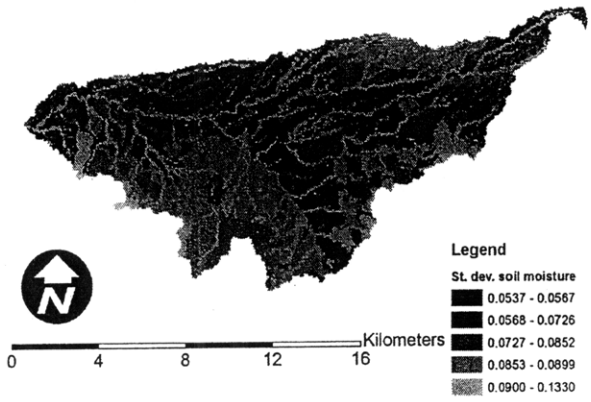


(c)

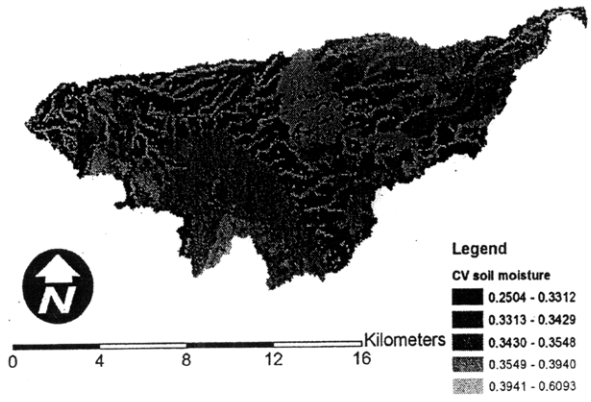
Figure C-11: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 432 hrs into the 1024 replicate open loop simulation.



(a)

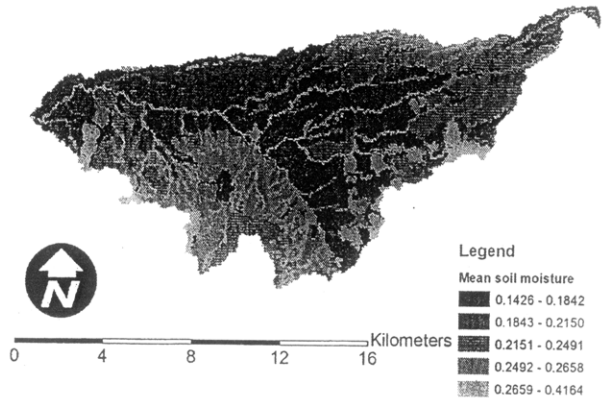


(b)

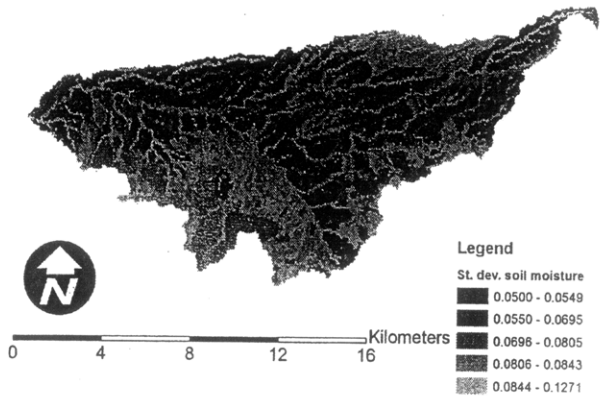


(c)

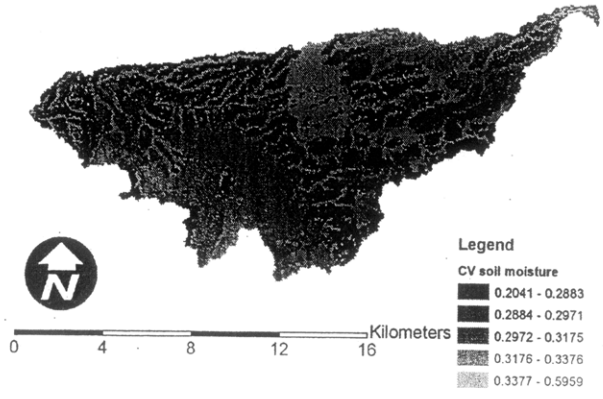
Figure C-12: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 504 hrs into the 1024 replicate open loop simulation.



(a)

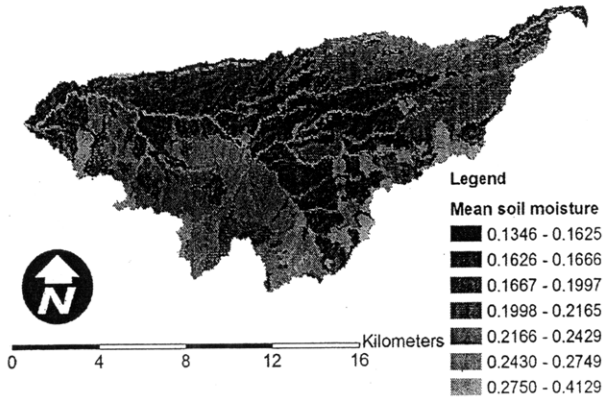


(b)

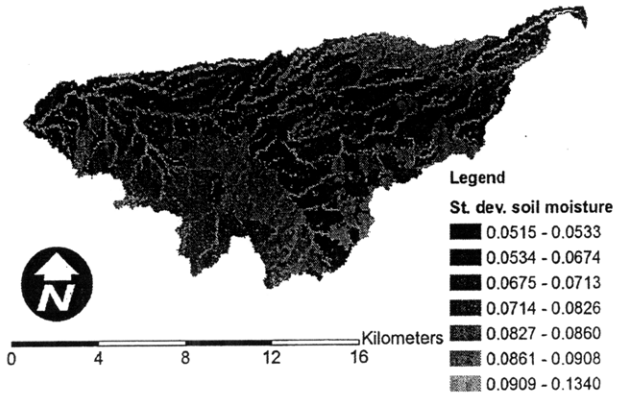


(c)

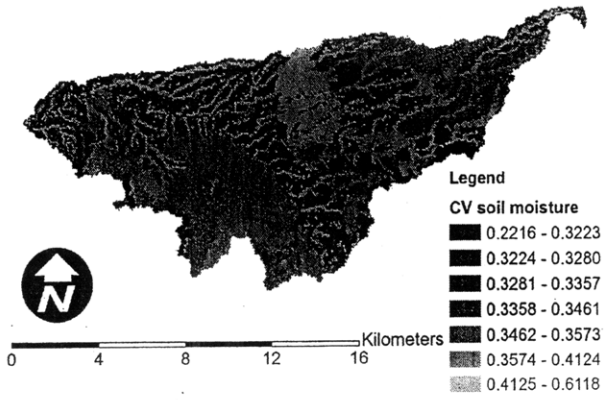
Figure C-13: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 576 hrs into the 1024 replicate open loop simulation.



(a)



(b)



(c)

Figure C-14: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in near-surface soil moisture at 648 hrs into the 1024 replicate open loop simulation.

C.2.2 Spatially distributed empirical Orthogonal Function (EOF) analysis

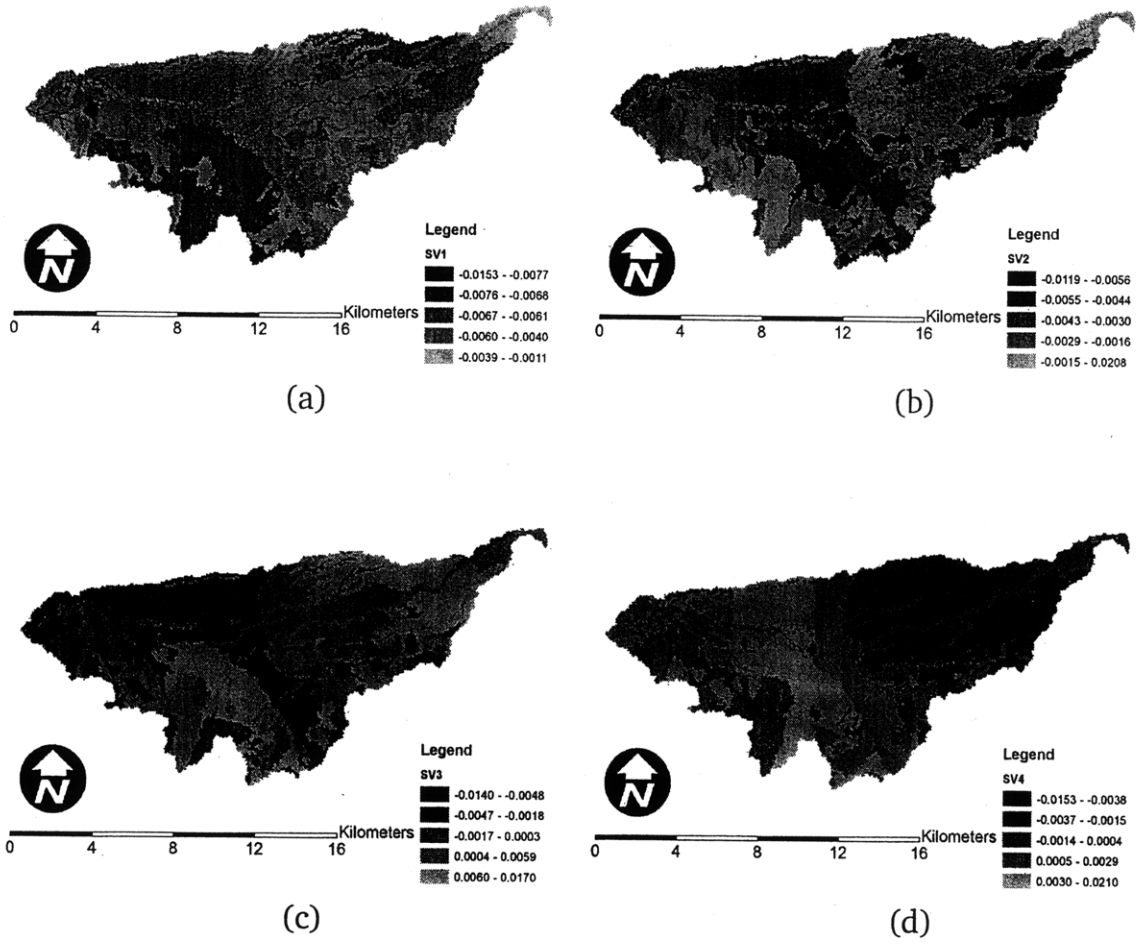


Figure C-15: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 0 hr into the simulation. They explain 13.5, 12, 8.6, and 8.2 percent of the variance in soil moisture, respectively.

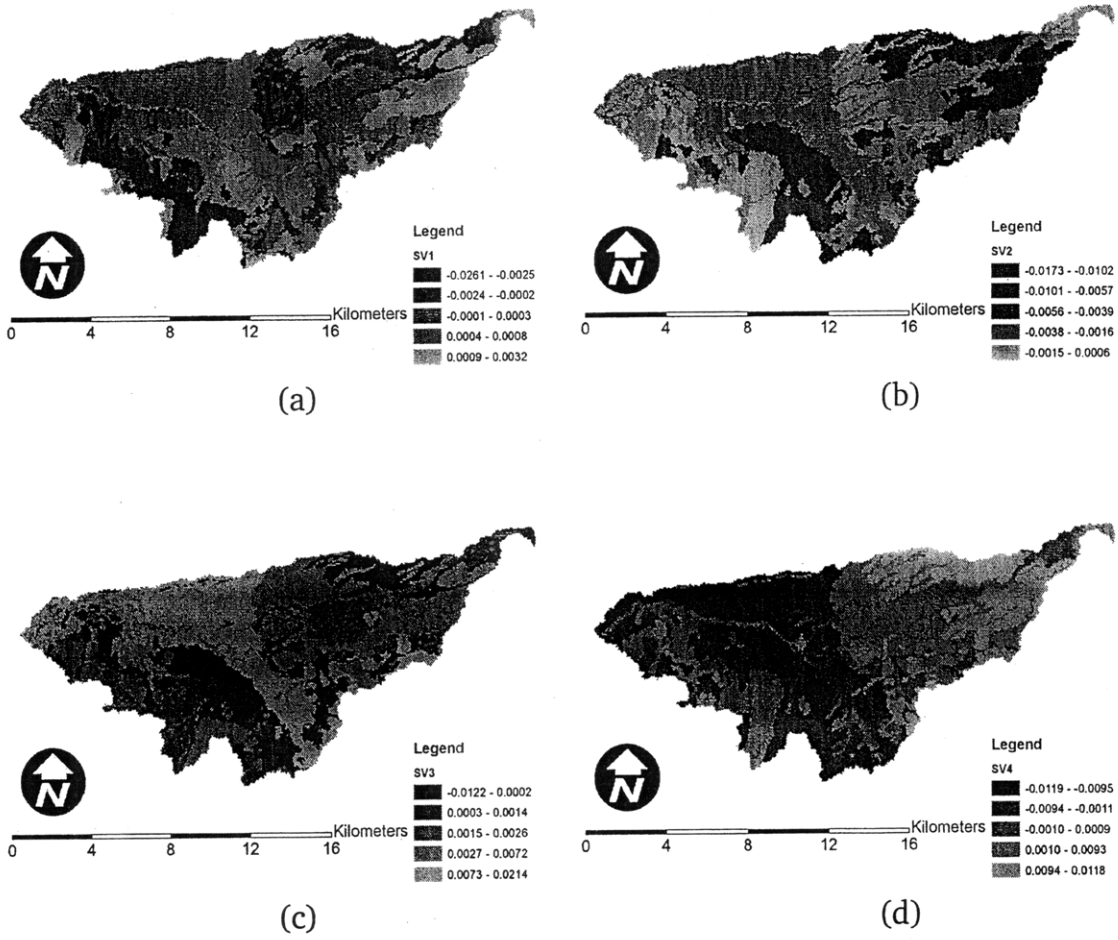


Figure C-16: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 72 hr into the simulation. They explain 12.3, 10.1, 8.6, and 8.3 percent of the variance in soil moisture, respectively.

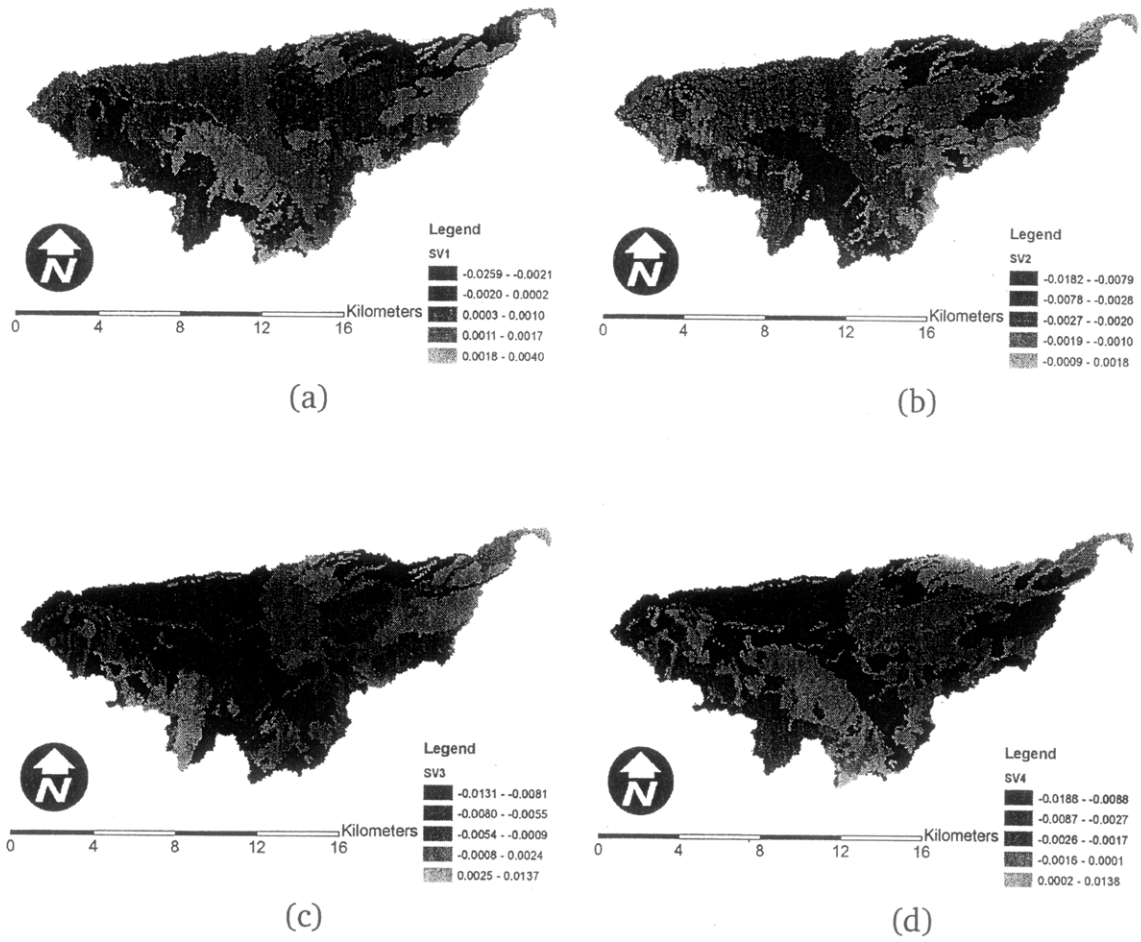


Figure C-17: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 144 hr into the simulation. They explain 12, 9.8, 8.6, and 8.4 percent of the variance in soil moisture, respectively.

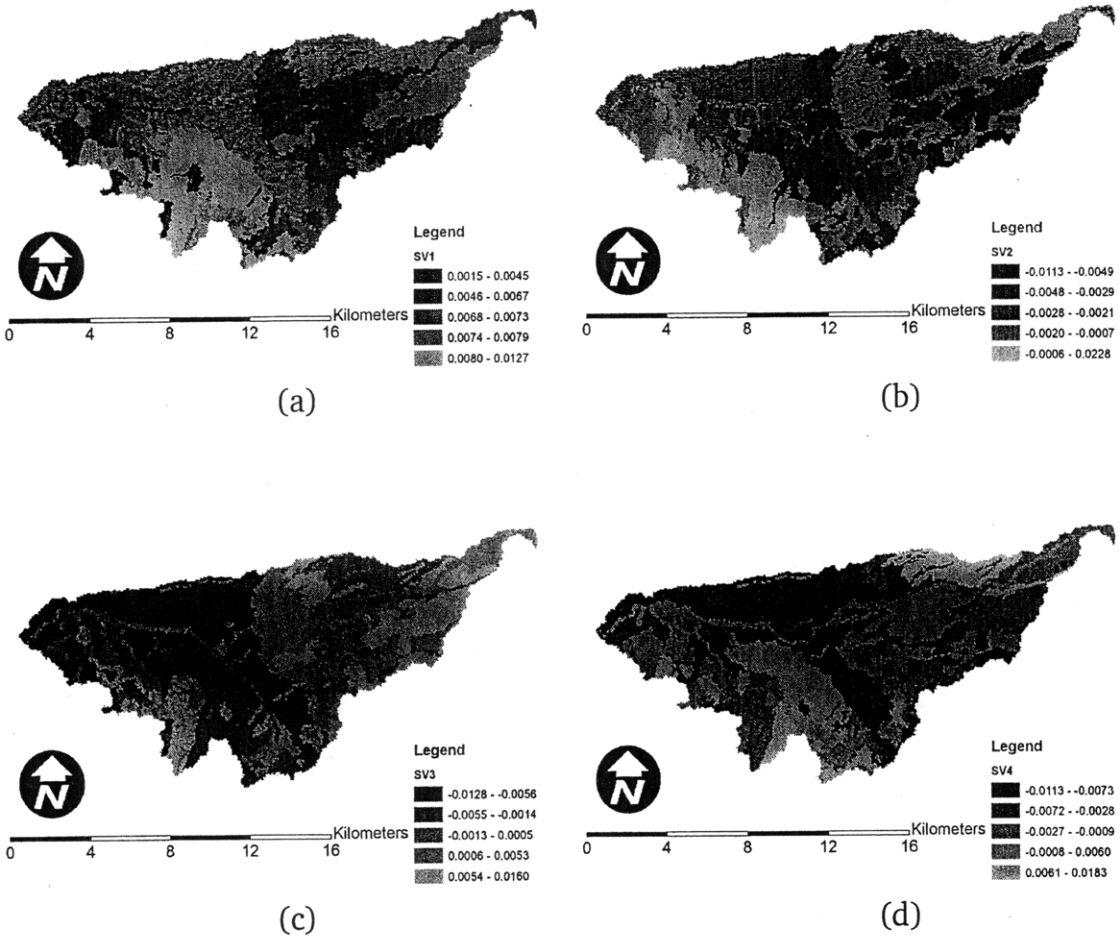


Figure C-18: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 216 hr into the simulation. They explain 14.1, 10.1, 8.6, and 7.9 percent of the variance in soil moisture, respectively.

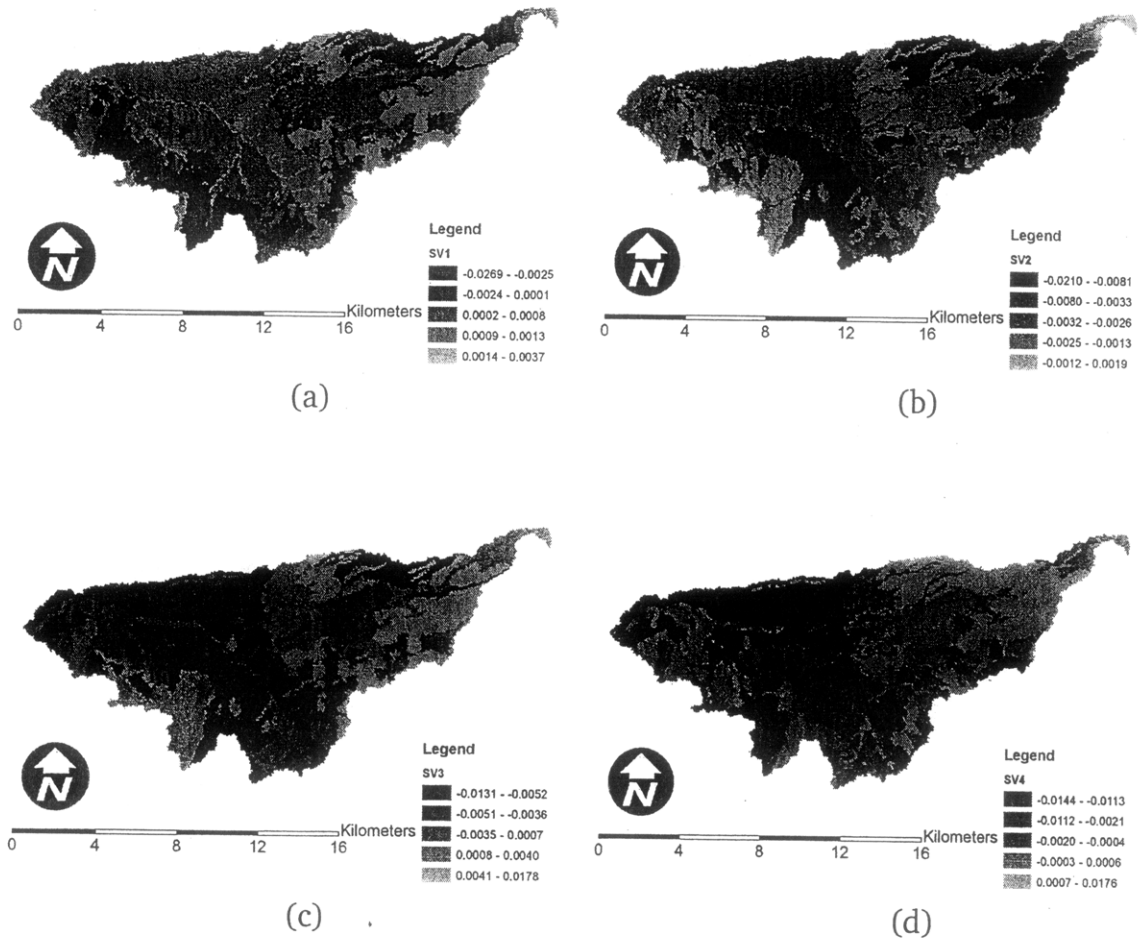


Figure C-19: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 288 hr into the simulation. They explain 12, 9.8, 8.7, and 8 percent of the variance in soil moisture, respectively.

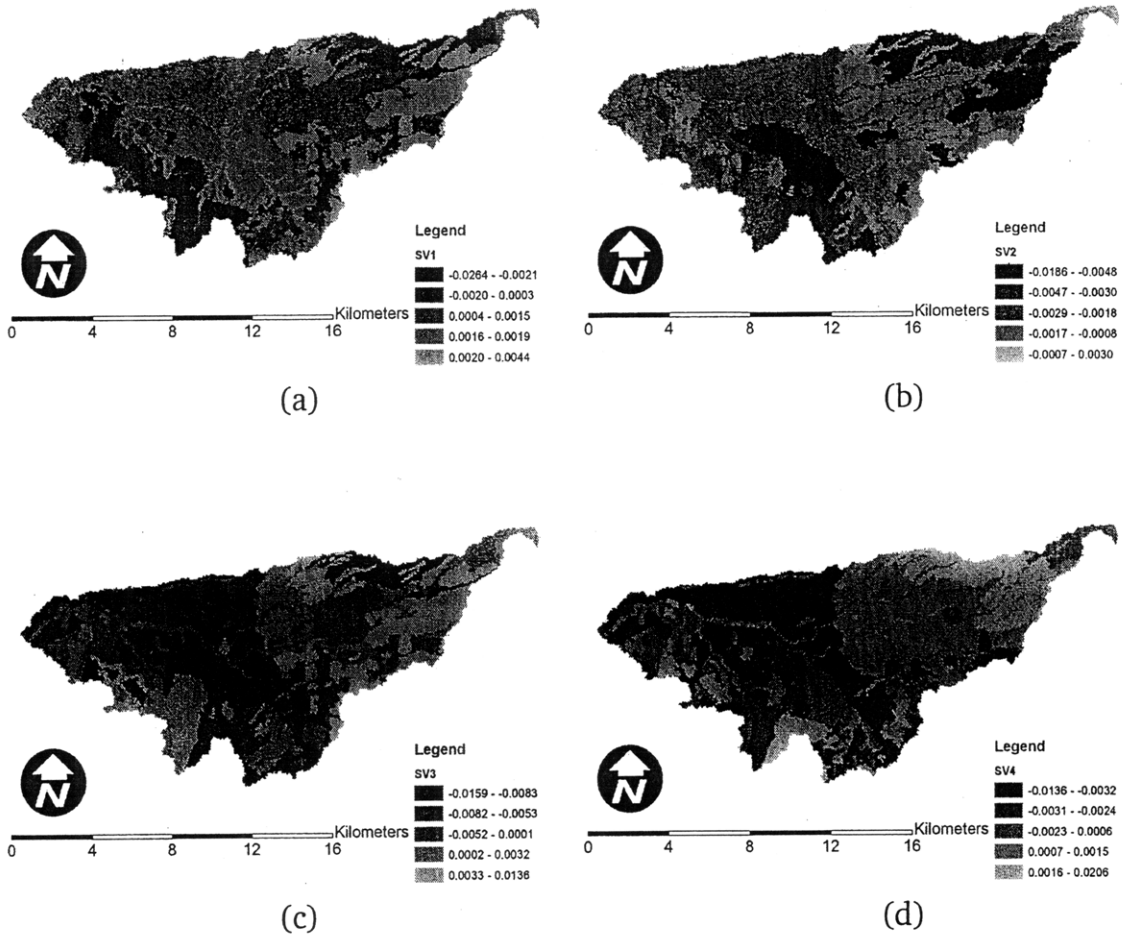


Figure C-20: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 360 hr into the simulation. They explain 12, 9.8, 8.7, and 8 percent of the variance in soil moisture, respectively.

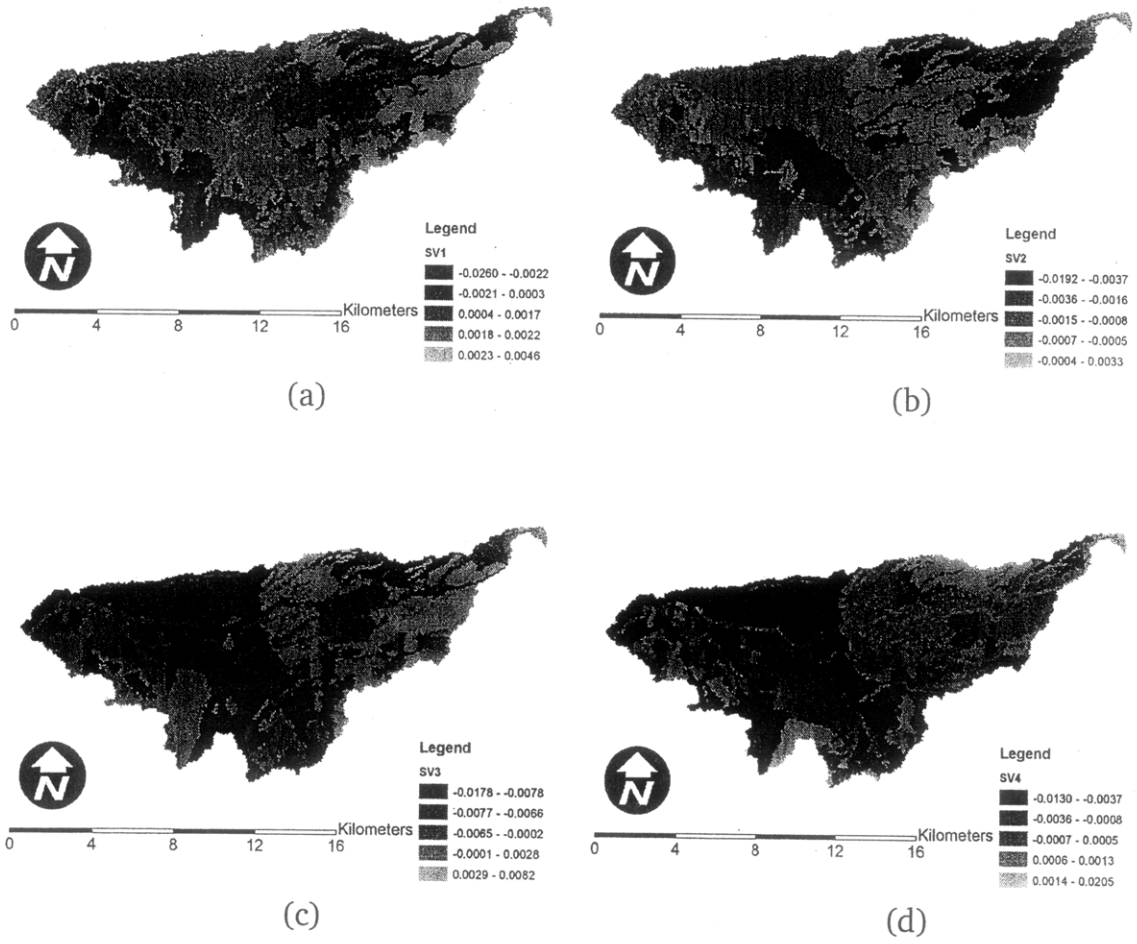


Figure C-21: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 432 hr into the simulation. They explain 12, 9.8, 8.7, and 8 percent of the variance in soil moisture, respectively.

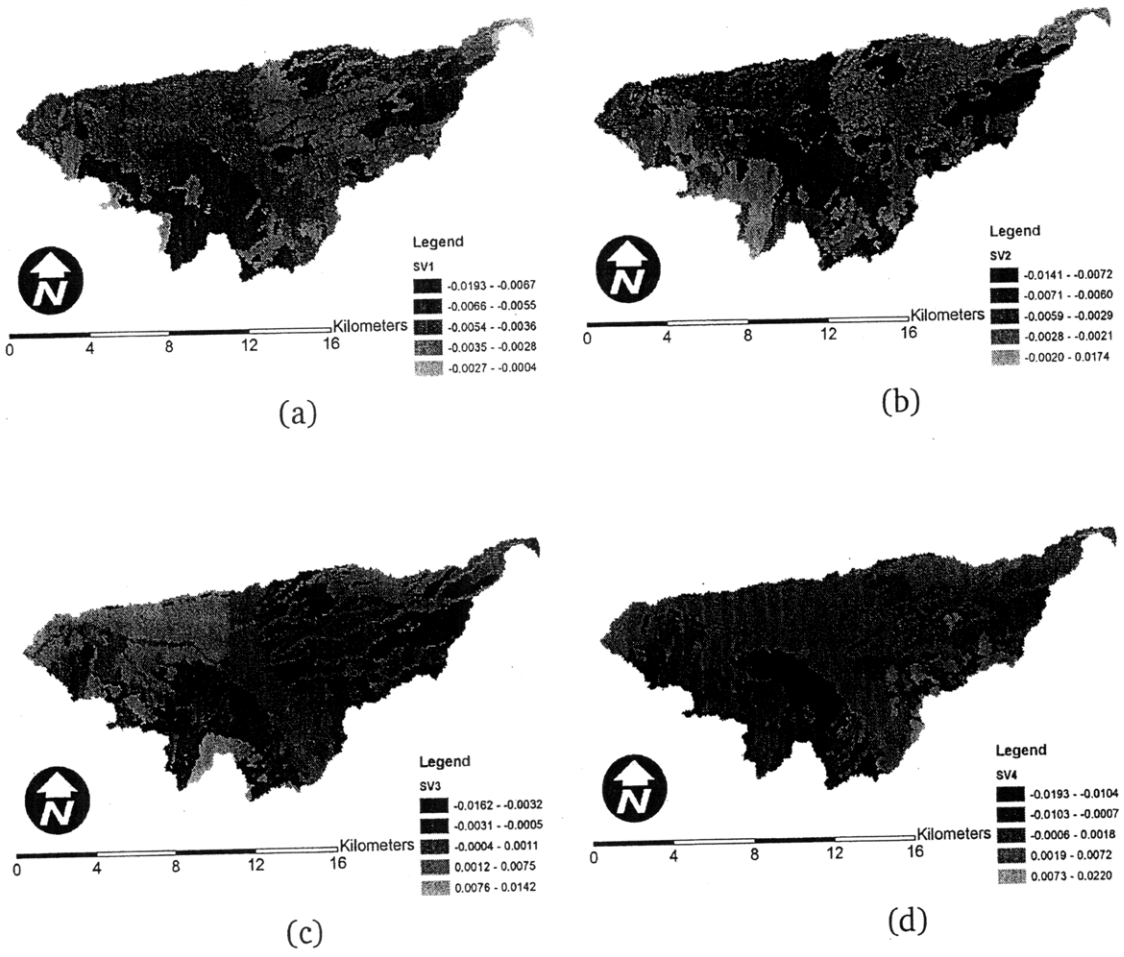


Figure C-22: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 504 hr into the simulation. They explain 12, 11, 8.7, and 7.8 percent of the variance in soil moisture, respectively.

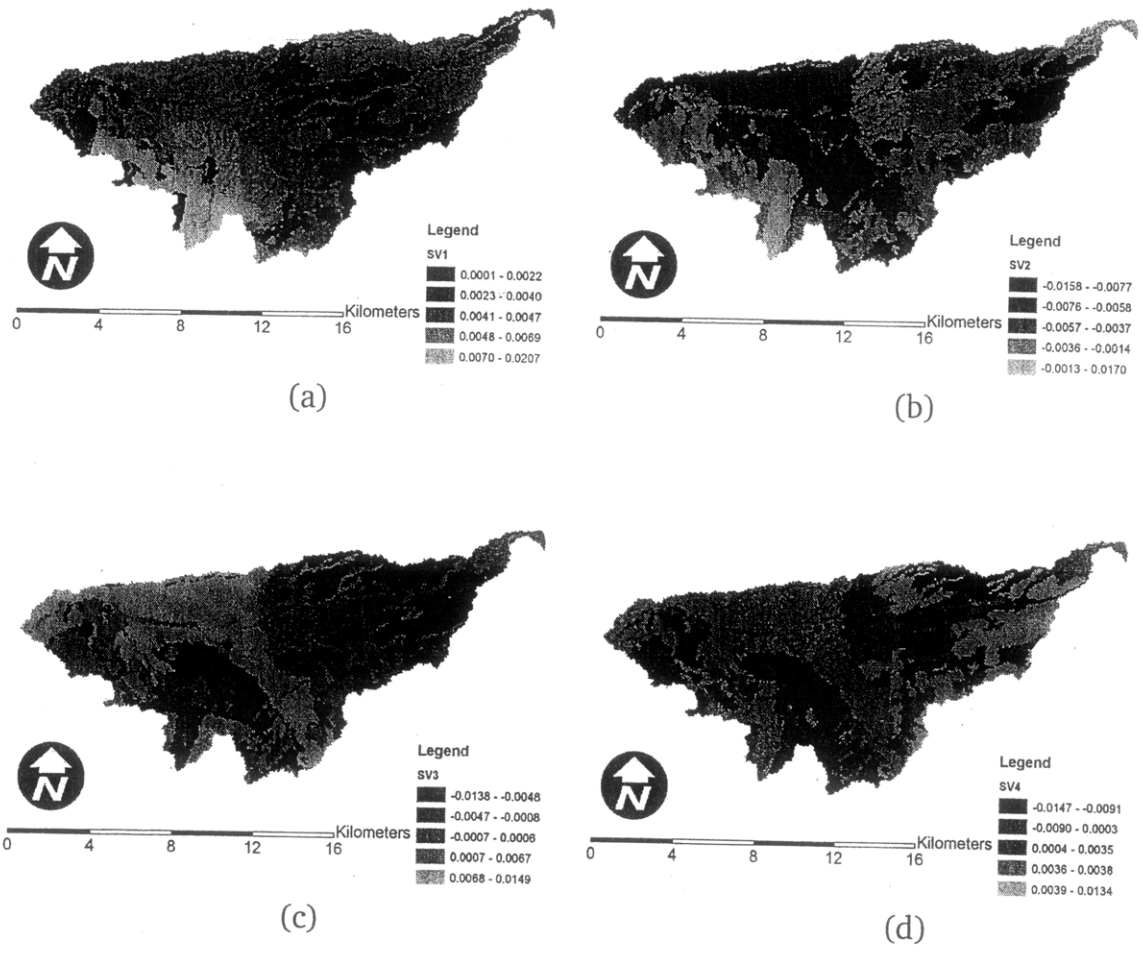


Figure C-23: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 576 hr into the simulation. They explain 11.5, 10.5, 8.7, and 7.8 percent of the variance in soil moisture, respectively.

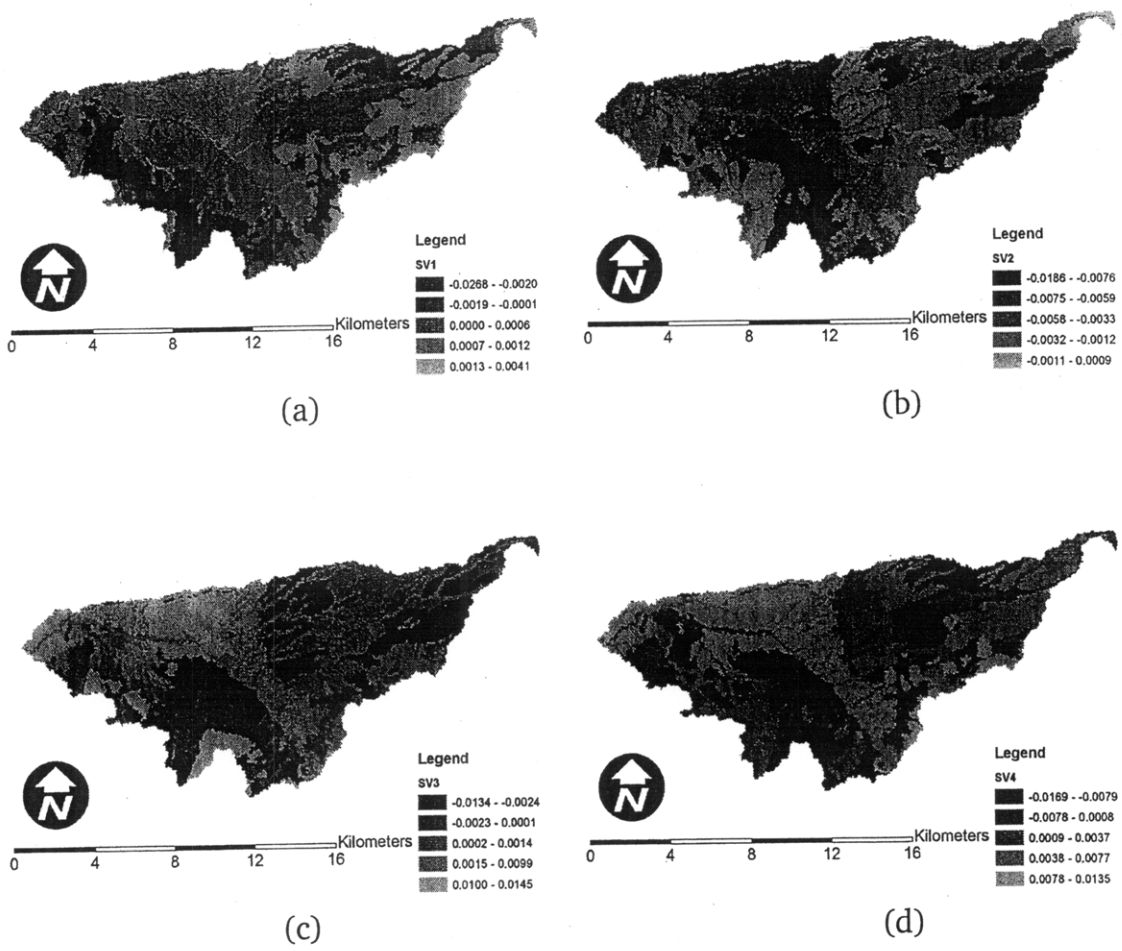
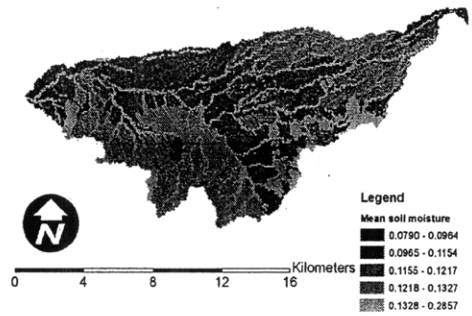


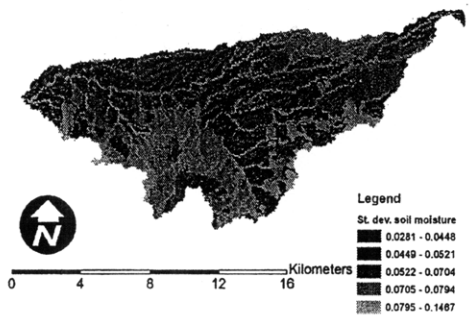
Figure C-24: The (a) first, (b) second, (c) third, and (d) fourth EOFs of near-surface soil moisture at 648 hr into the simulation. They explain 11, 10, 8.7, and 7.9 percent of the variance in soil moisture, respectively.

C.3 Profile soil moisture results

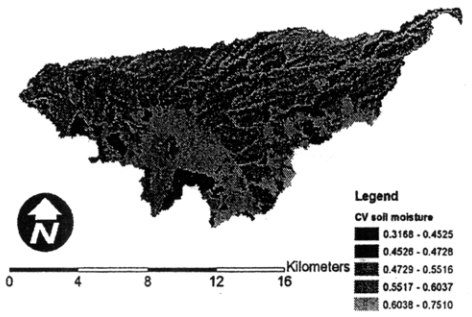
C.3.1 First- and second-order moments



(a)

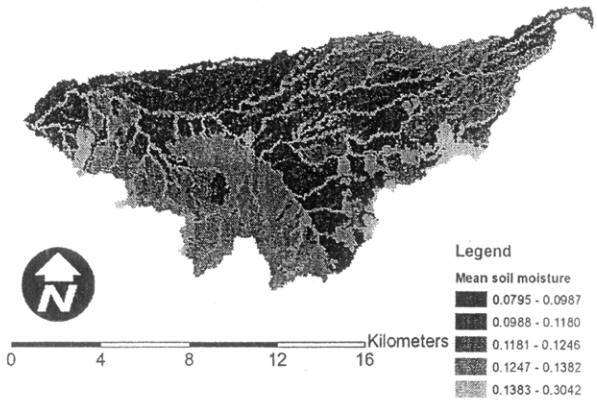


(b)

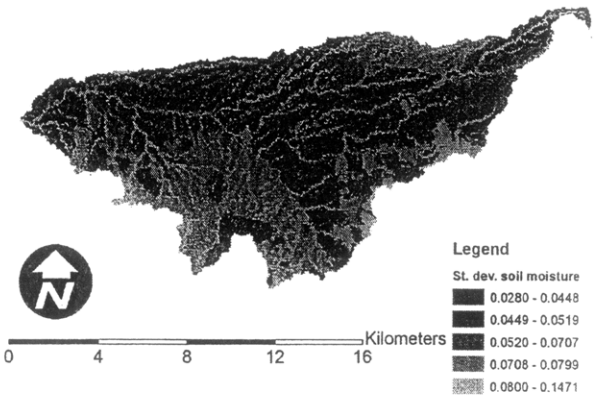


(c)

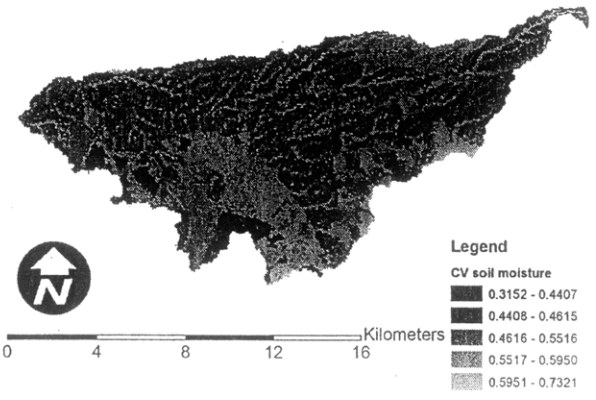
Figure C-25: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 0 hrs into the 1024 replicate open loop simulation.



(a)

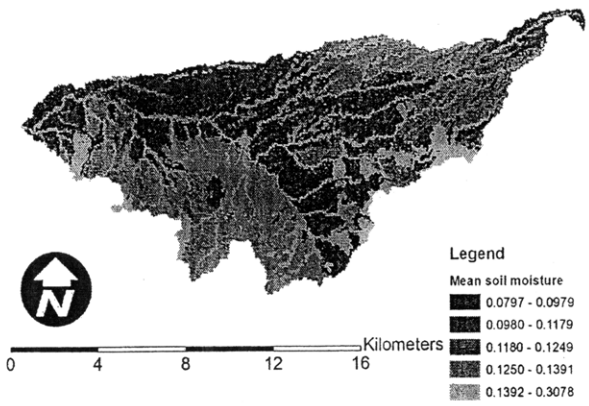


(b)

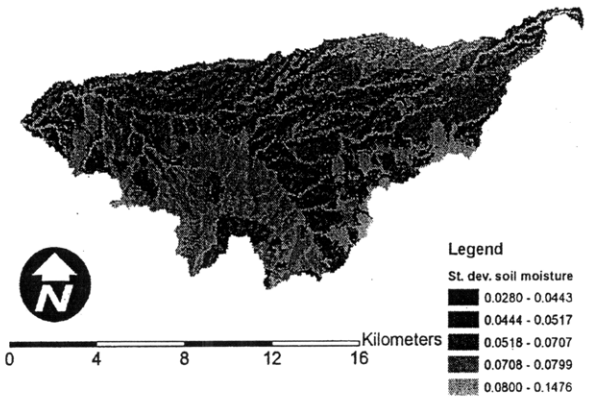


(c)

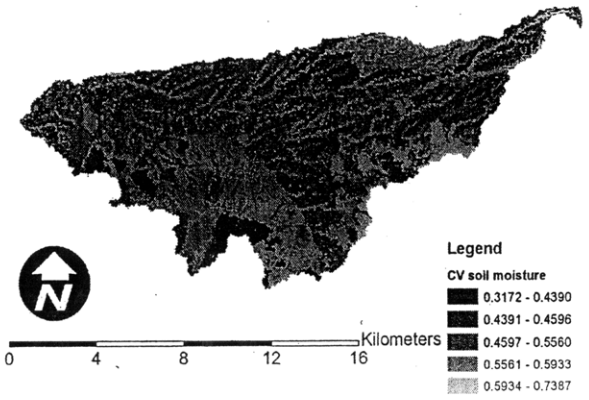
Figure C-26: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 72 hrs into the 1024 replicate open loop simulation.



(a)

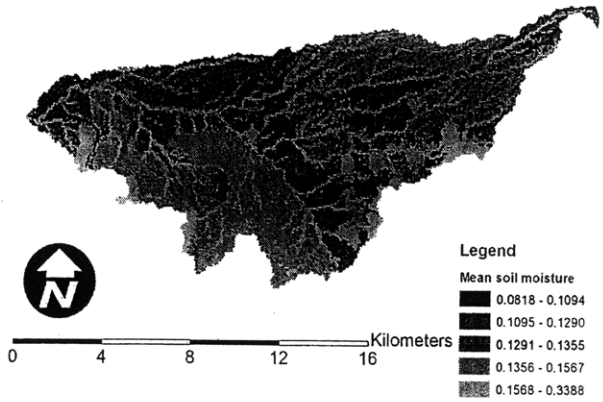


(b)

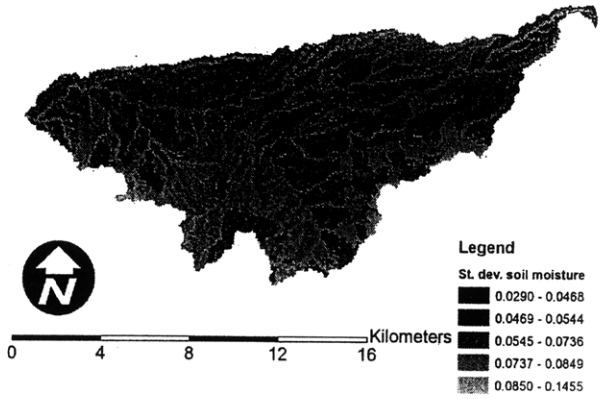


(c)

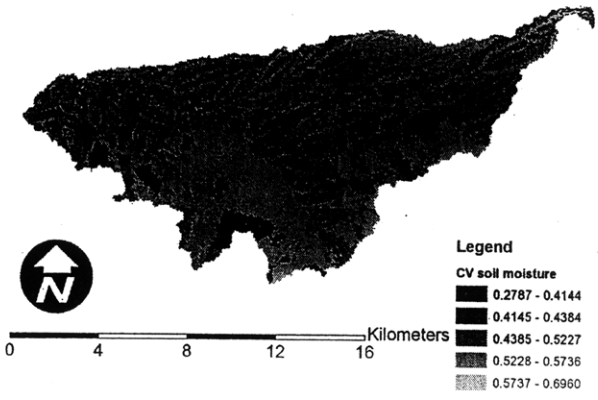
Figure C-27: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 144 hrs into the 1024 replicate open loop simulation.



(a)

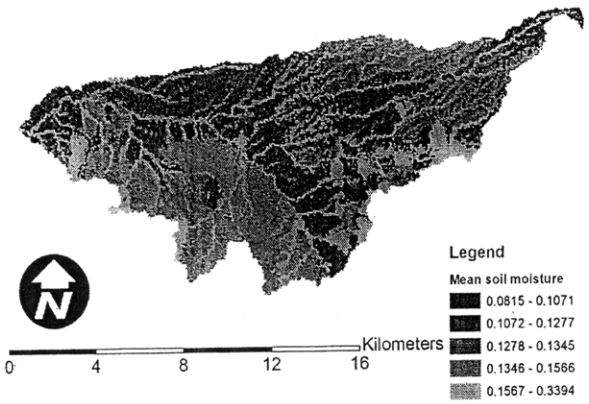


(b)

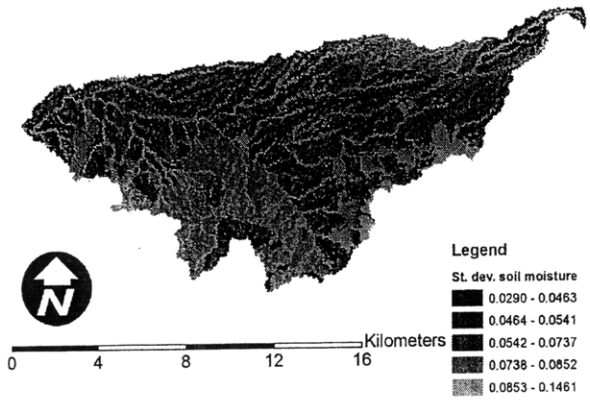


(c)

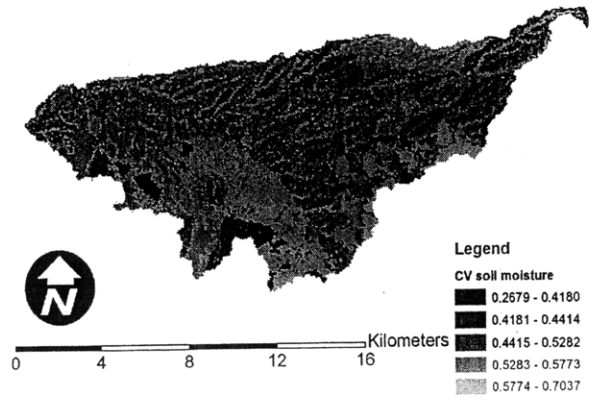
Figure C-28: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 216 hrs into the 1024 replicate open loop simulation.



(a)

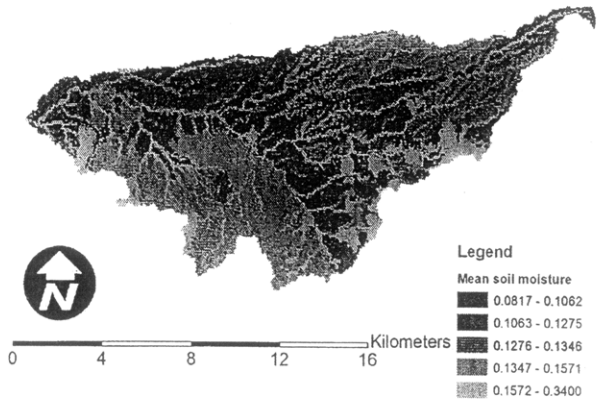


(b)

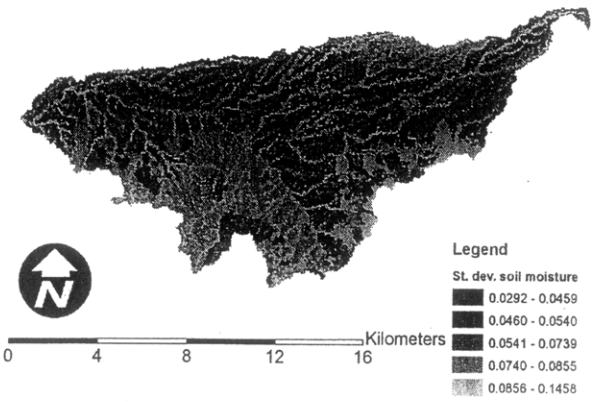


(c)

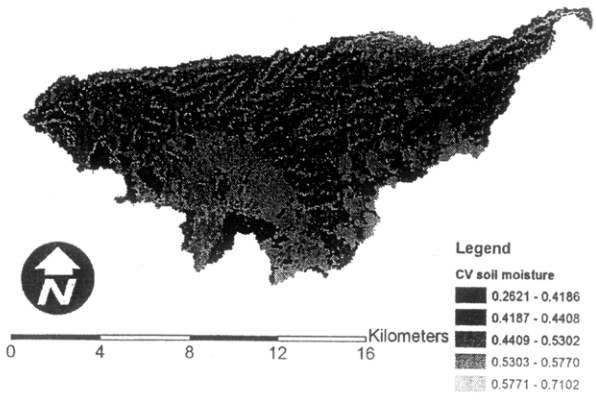
Figure C-29: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 288 hrs into the 1024 replicate open loop simulation.



(a)

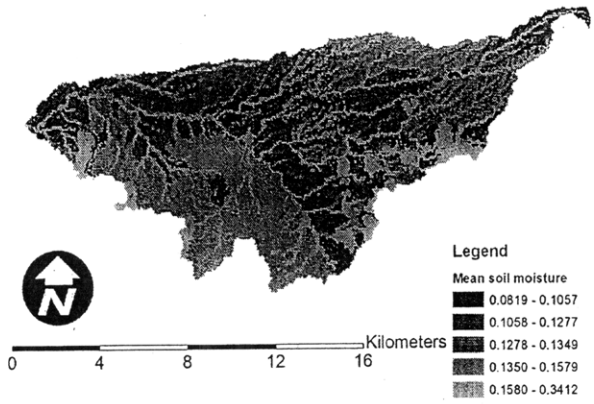


(b)

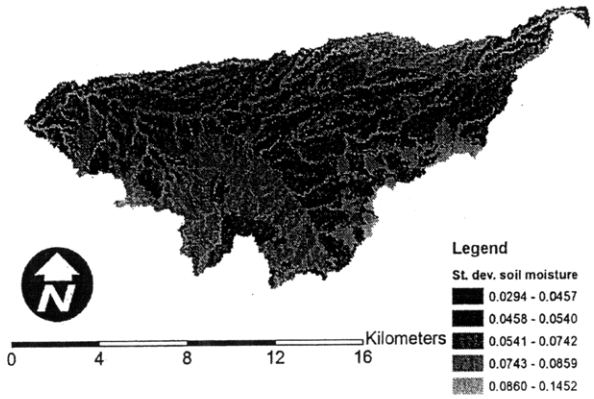


(c)

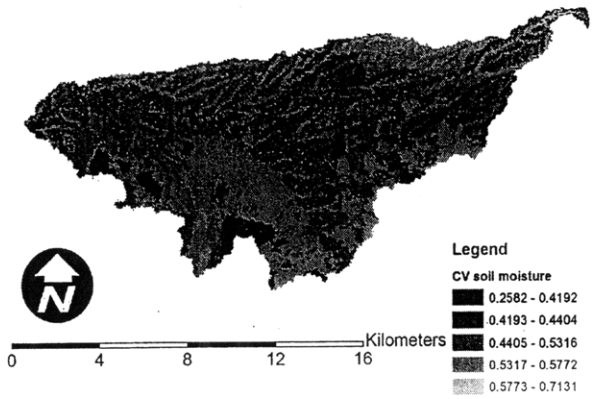
Figure C-30: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 360 hrs into the 1024 replicate open loop simulation.



(a)

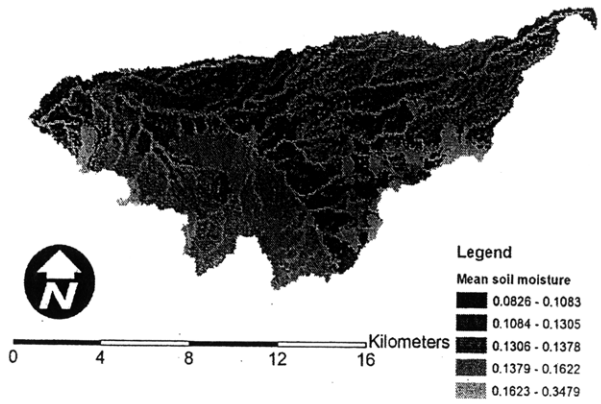


(b)

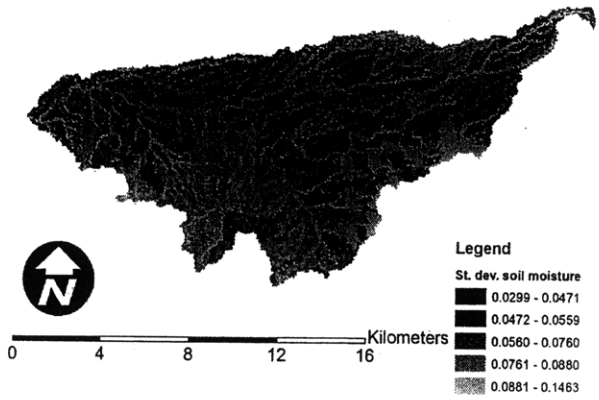


(c)

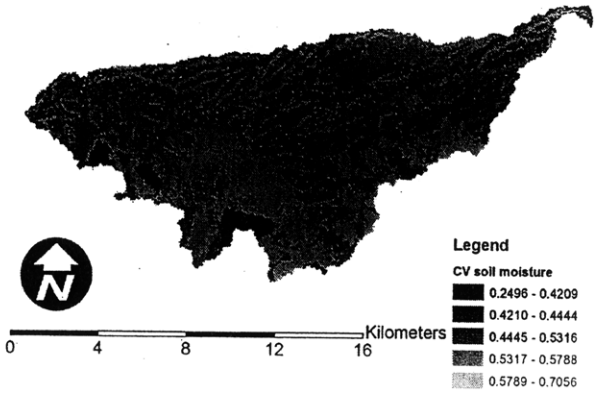
Figure C-31: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 432 hrs into the 1024 replicate open loop simulation.



(a)

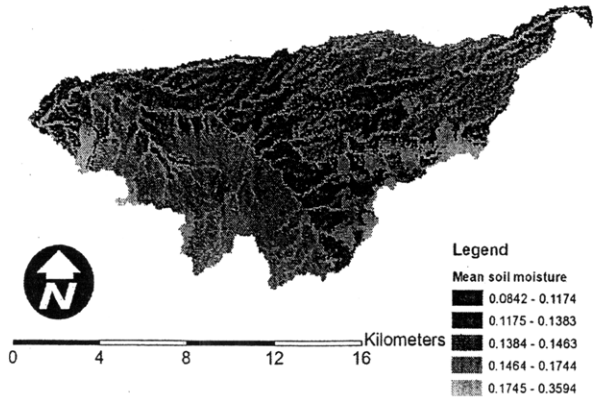


(b)

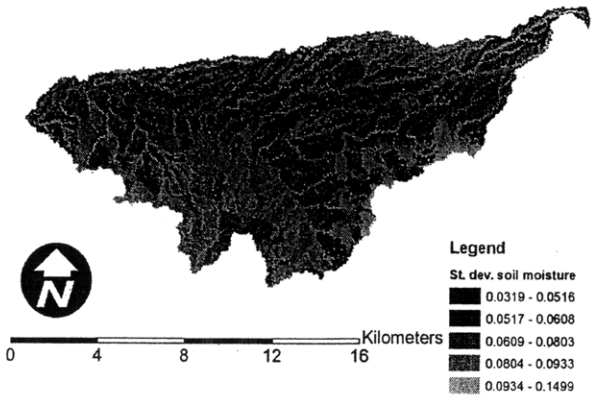


(c)

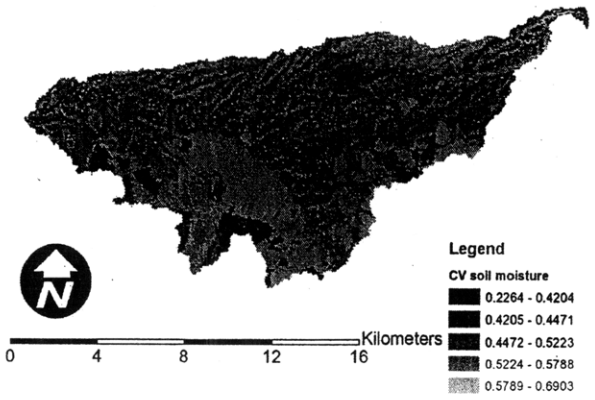
Figure C-32: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 504 hrs into the 1024 replicate open loop simulation.



(a)

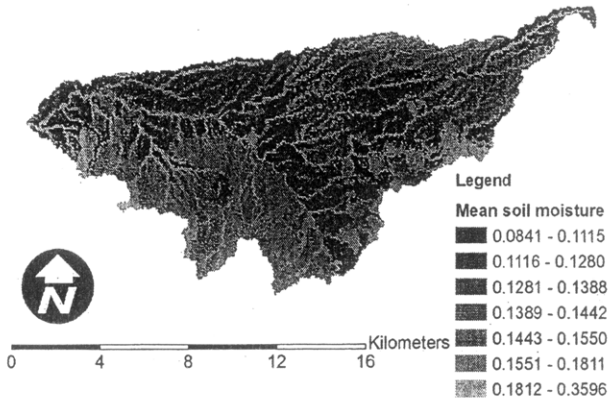


(b)

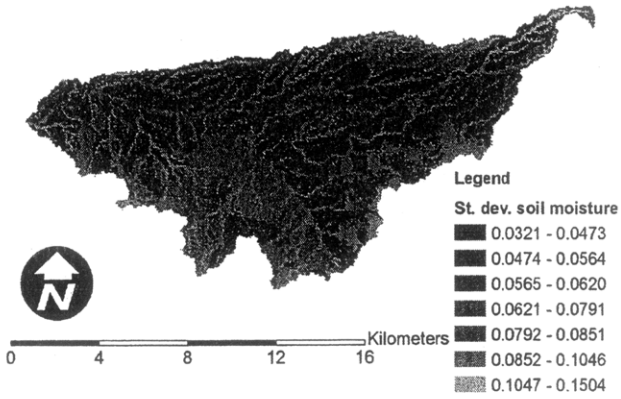


(c)

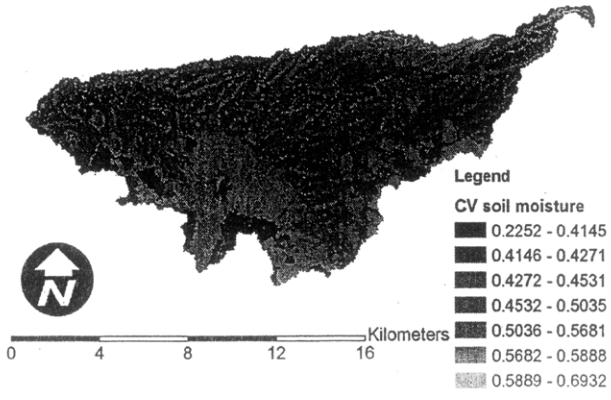
Figure C-33: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 576 hrs into the 1024 replicate open loop simulation.



(a)



(b)



(c)

Figure C-34: Ensemble (a) mean, (b) local standard deviation, and (c) local coefficient of variation in profile soil moisture at 648 hrs into the 1024 replicate open loop simulation.

C.3.2 Empirical Orthogonal Function (EOF) analysis

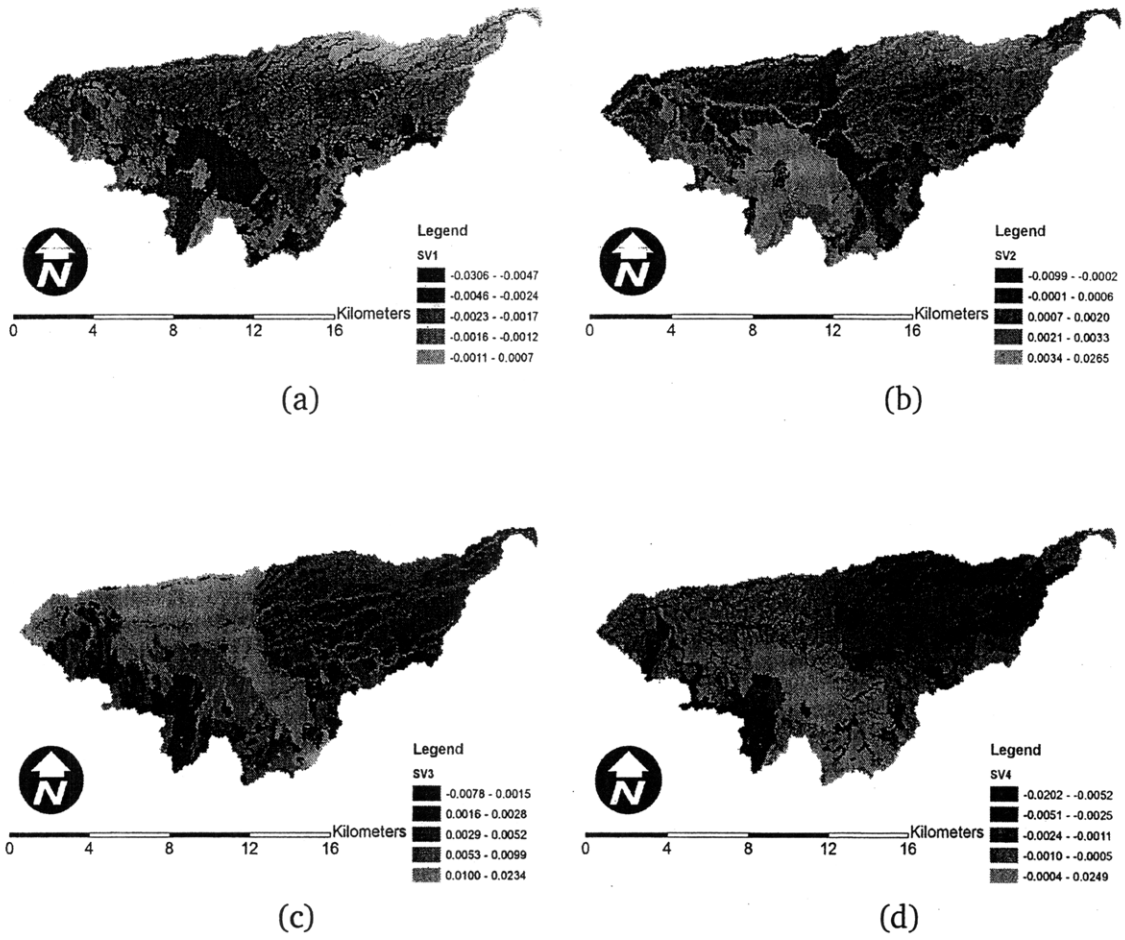


Figure C-35: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 0 hr into the simulation. They explain 14.5, 12, 10.2, and 8.5 percent of the variance in soil moisture, respectively.

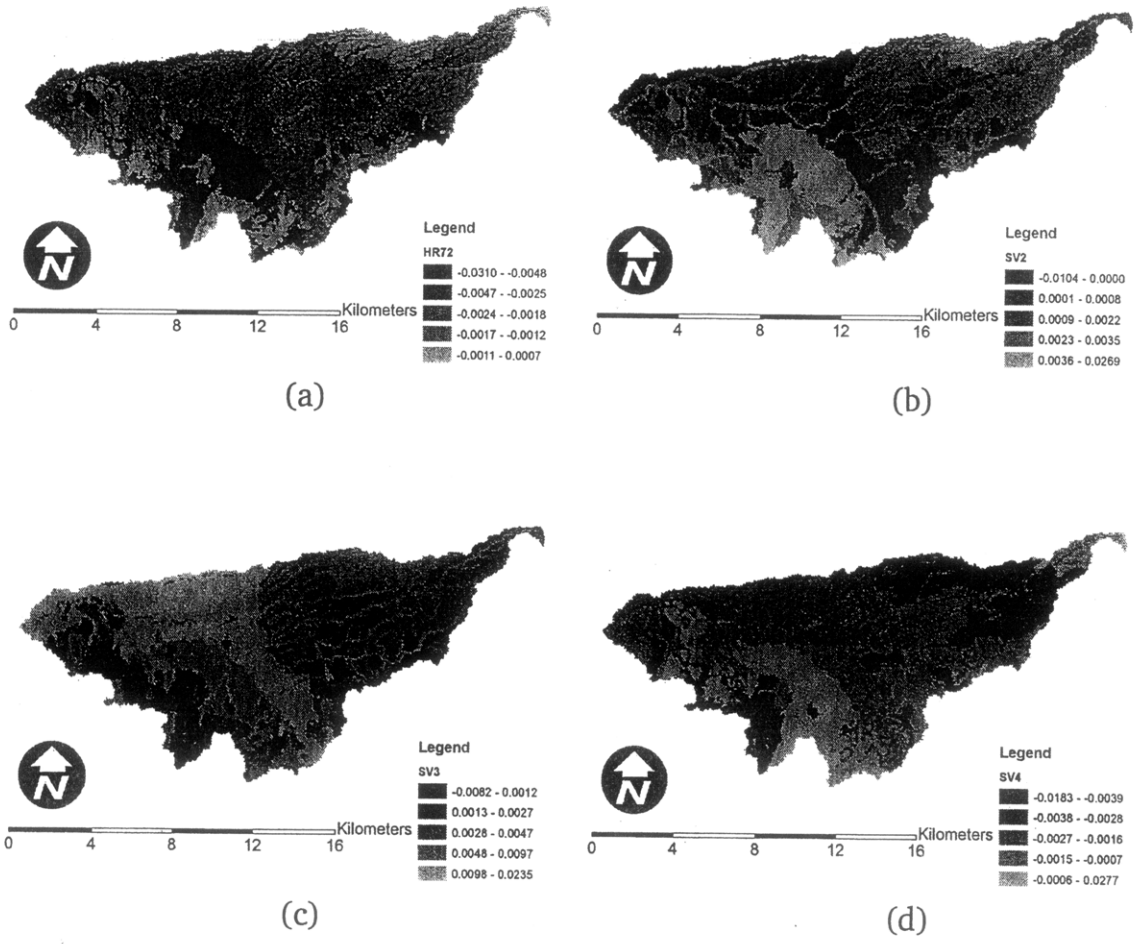


Figure C-36: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 72 hr into the simulation. They explain 14.3, 12, 10.1, and 8.3 percent of the variance in soil moisture, respectively.

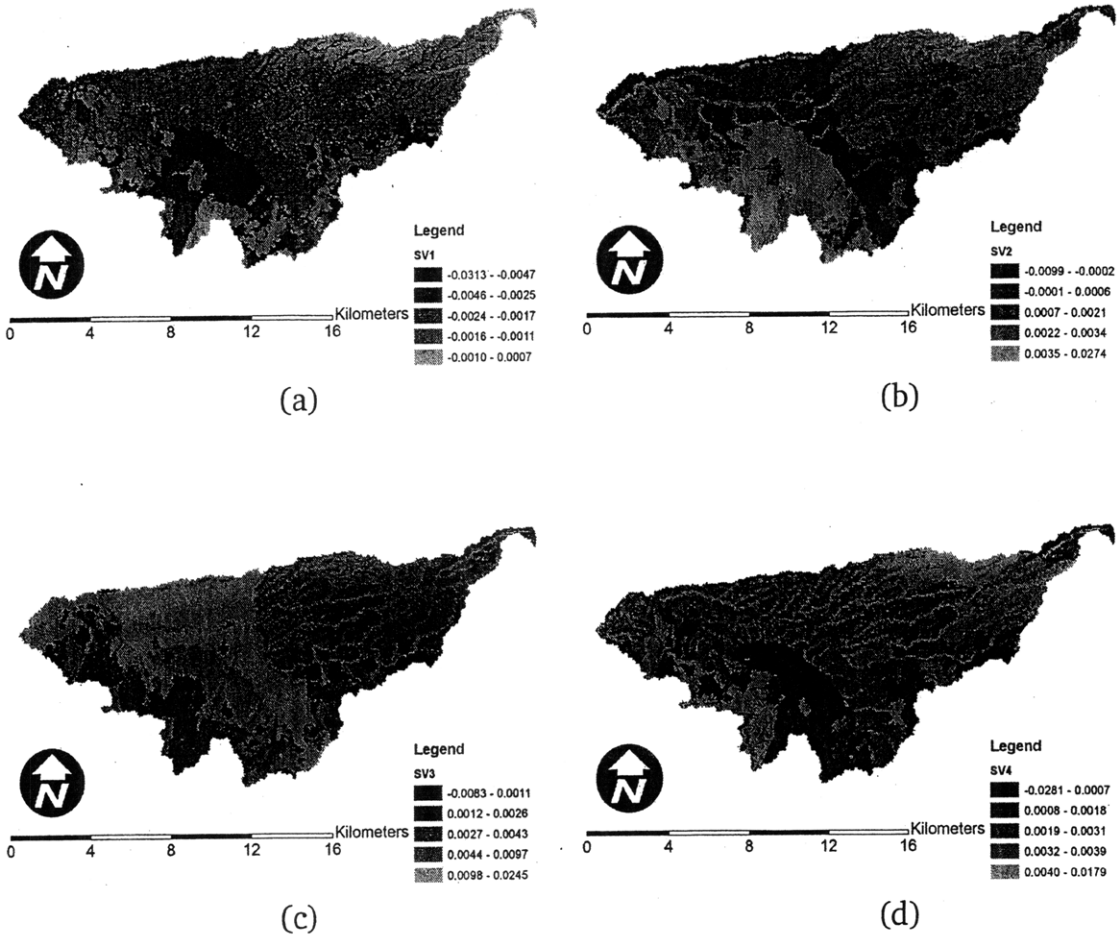


Figure C-37: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 144 hr into the simulation. They explain 14.1, 12, 10.1, and 8.3 percent of the variance in soil moisture, respectively.

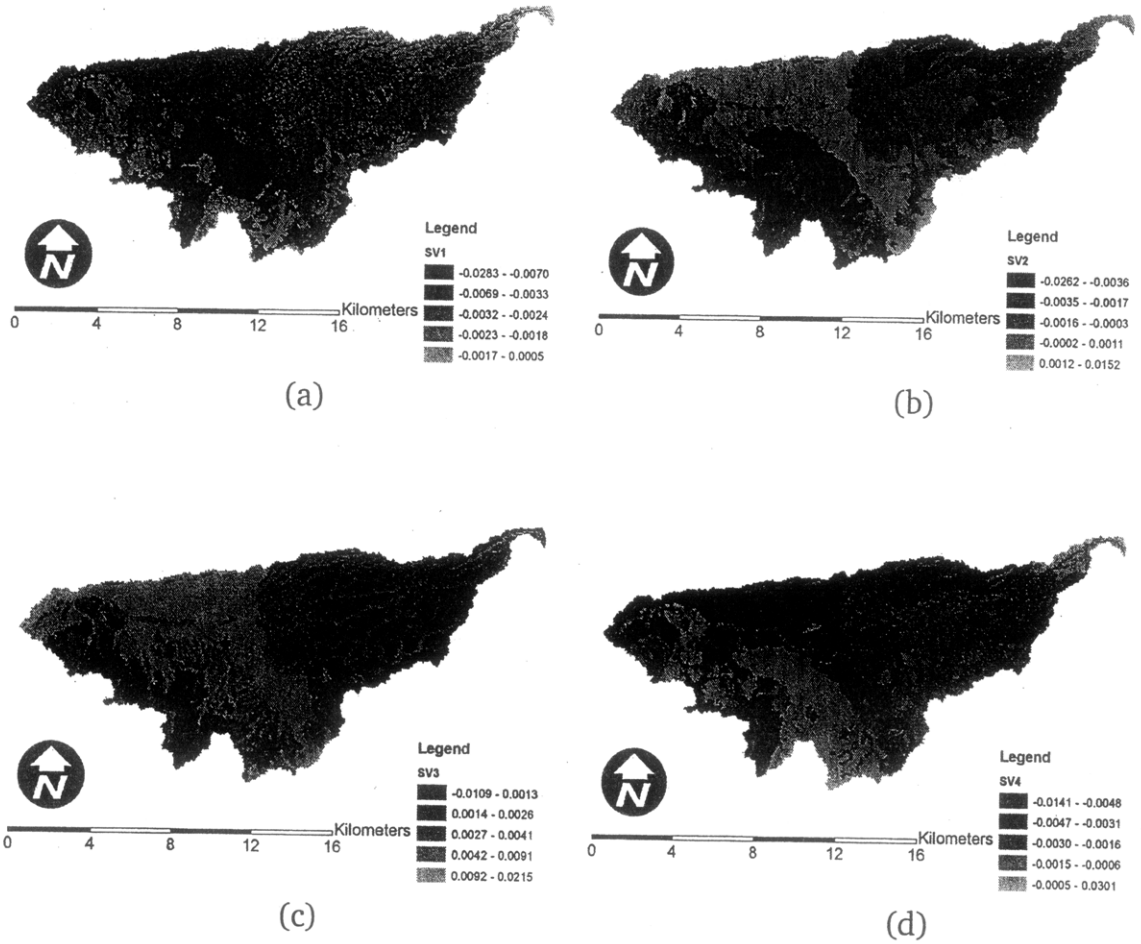


Figure C-38: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 216 hr into the simulation. They explain 14, 12, 10.1, and 8.3 percent of the variance in soil moisture, respectively.

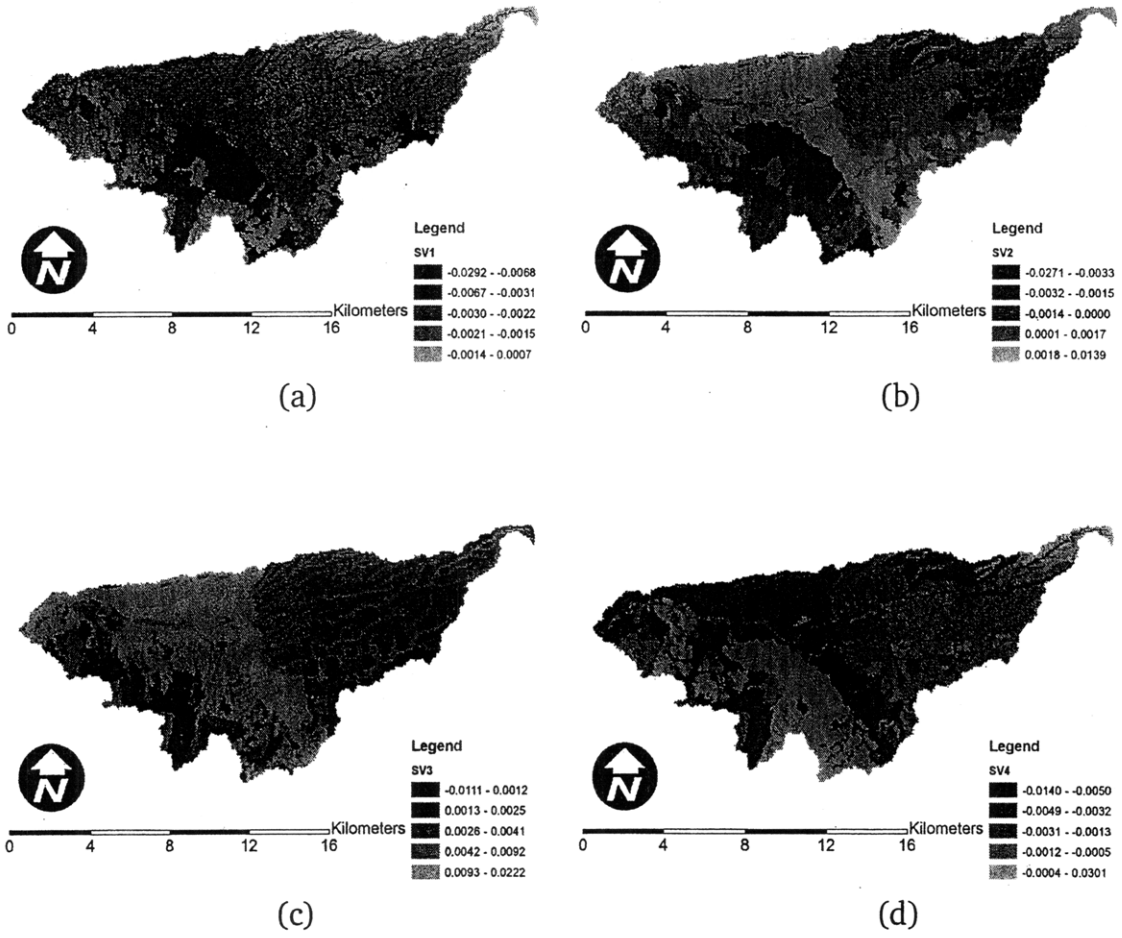


Figure C-39: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 288 hr into the simulation. They explain 14, 12, 10.1, and 8.3 percent of the variance in soil moisture, respectively.

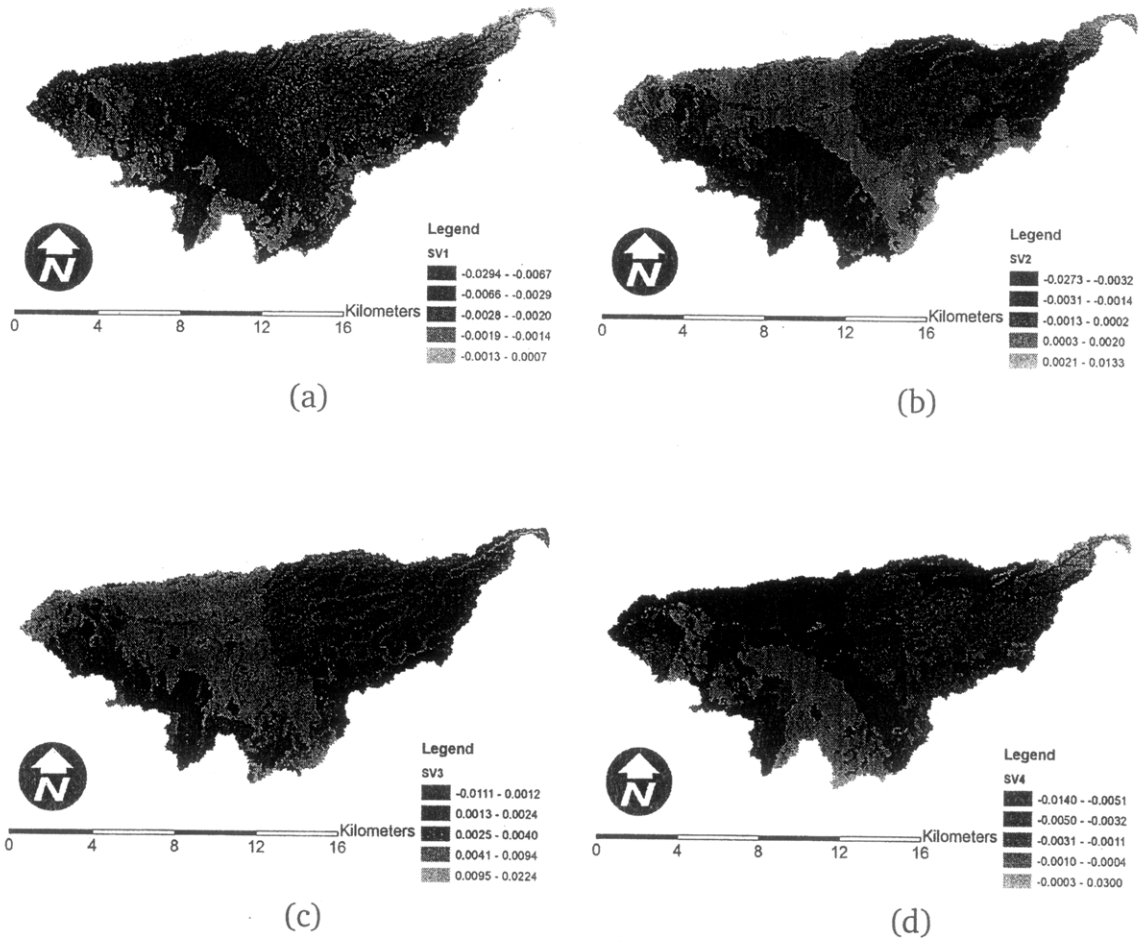


Figure C-40: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 360 hr into the simulation. They explain 14, 12, 10.1, and 8.3 percent of the variance in soil moisture, respectively.

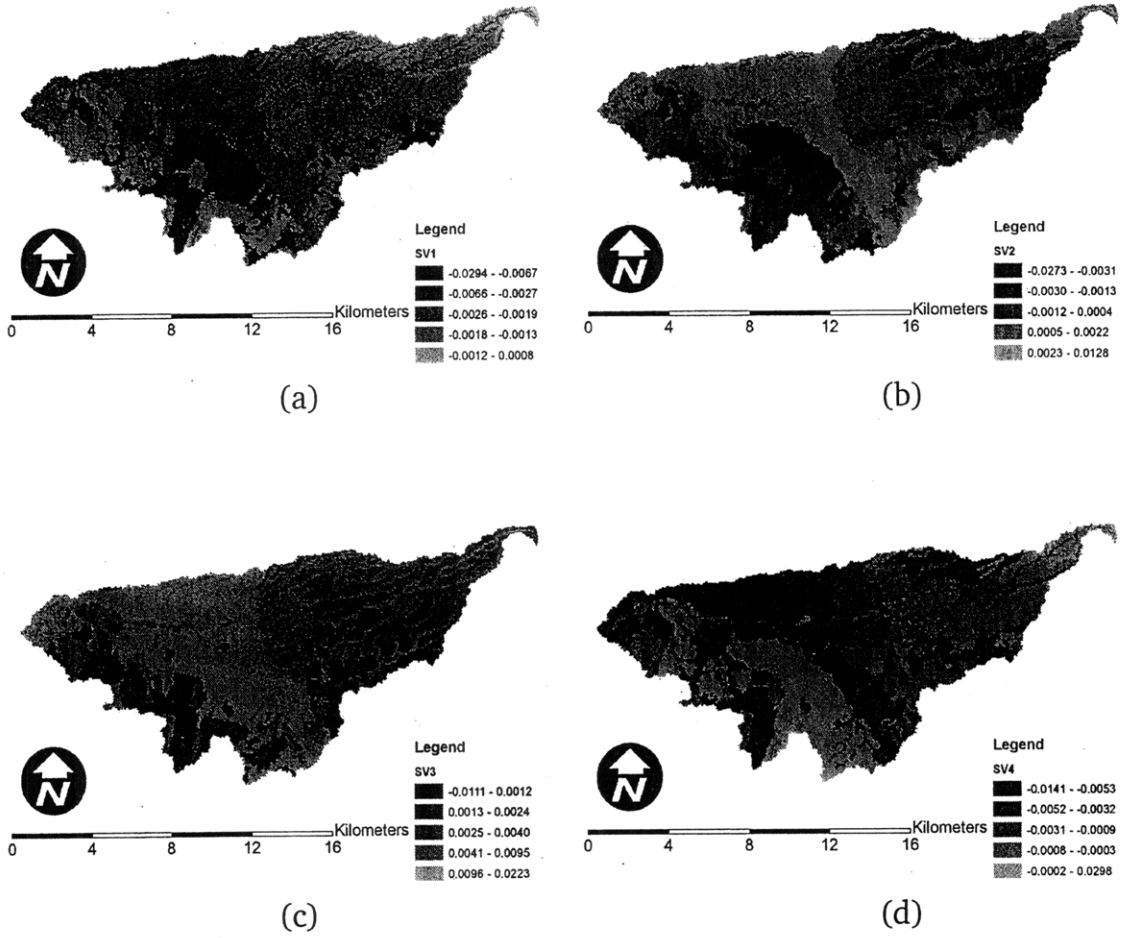


Figure C-41: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 432 hr into the simulation. They explain 14, 12, 10.2, and 8.3 percent of the variance in soil moisture, respectively.

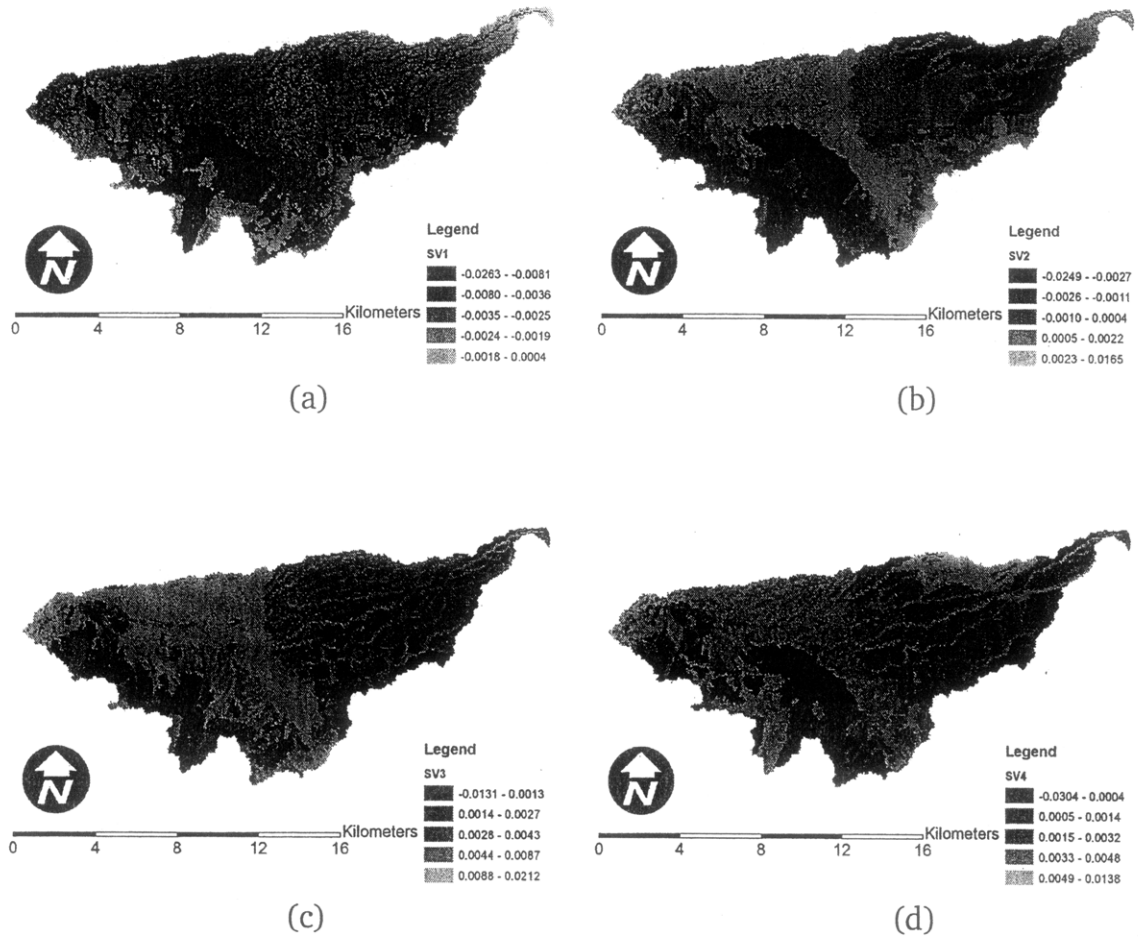


Figure C-42: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 504 hr into the simulation. They explain 13.9, 12, 10.2, and 8.5 percent of the variance in soil moisture, respectively.

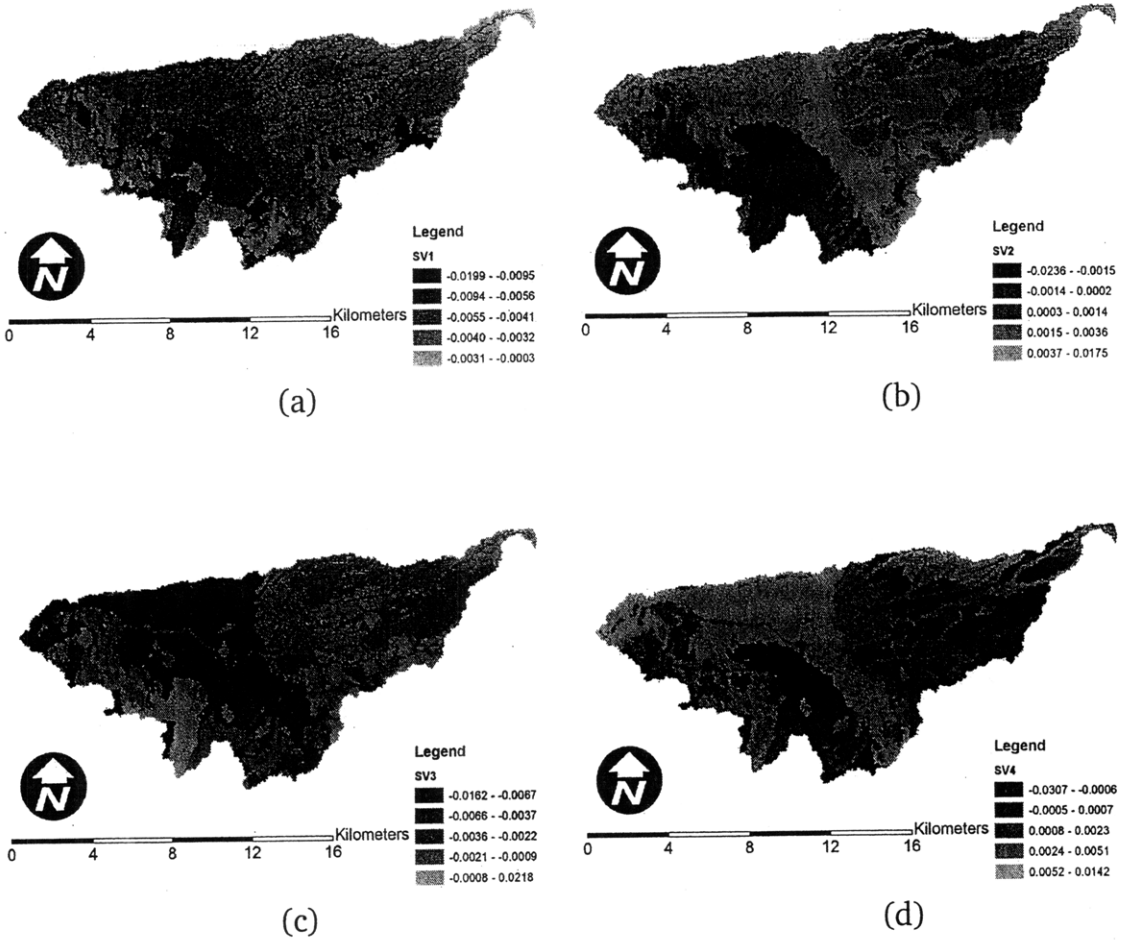


Figure C-43: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 576 hr into the simulation. They explain 14.1, 12, 10.1, and 8.5 percent of the variance in soil moisture, respectively.

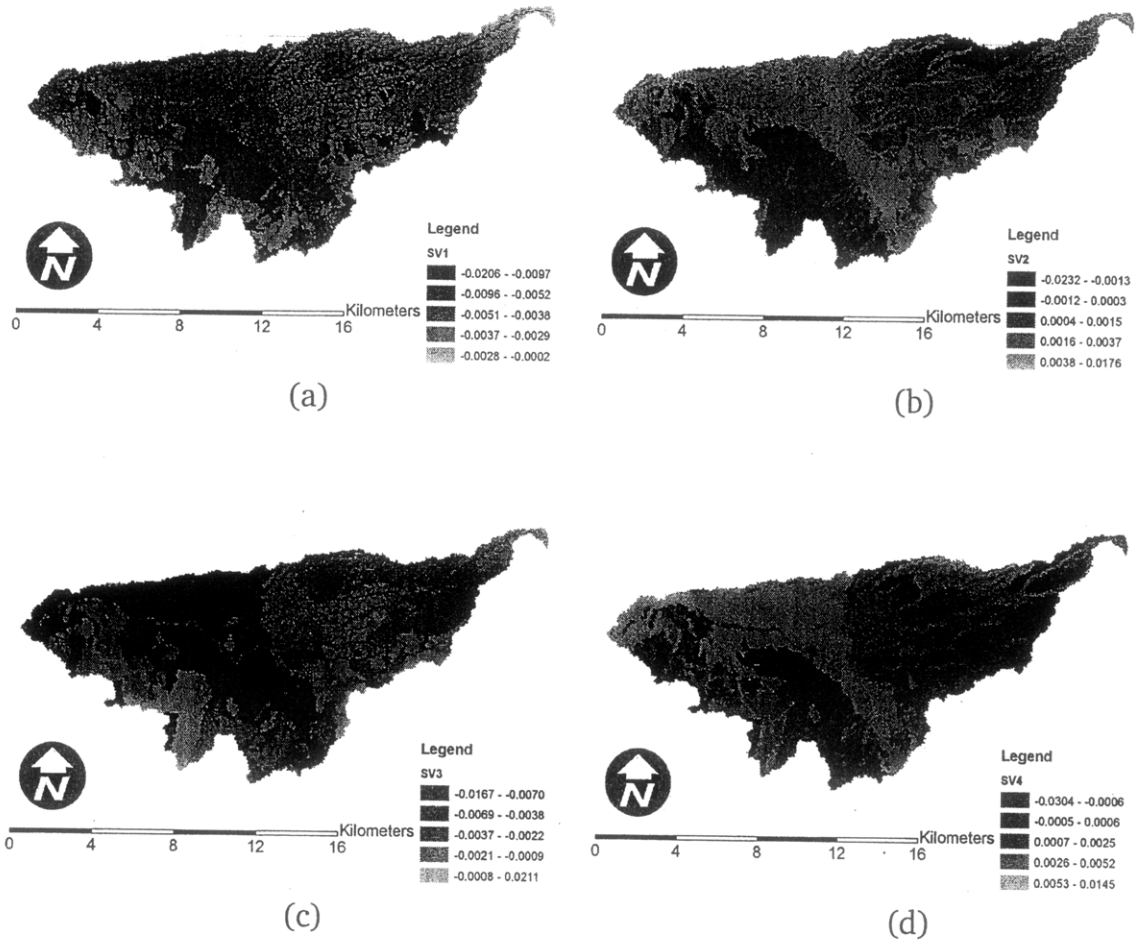


Figure C-44: The (a) first, (b) second, (c) third, and (d) fourth EOFs of profile soil moisture at 648 hr into the simulation. They explain 14, 12, 10.1, and 8.5 percent of the variance in soil moisture, respectively.

APPENDIX D

DATA ASSIMILATION EXPERIMENTS (EXPANDED)

This appendix provides expanded graphical results of the spatially-distributed data assimilation experiments described in **Chapter 8**.

D.1 Expanded results: assimilation of brightness temperature data

D.1.1 Estimator error

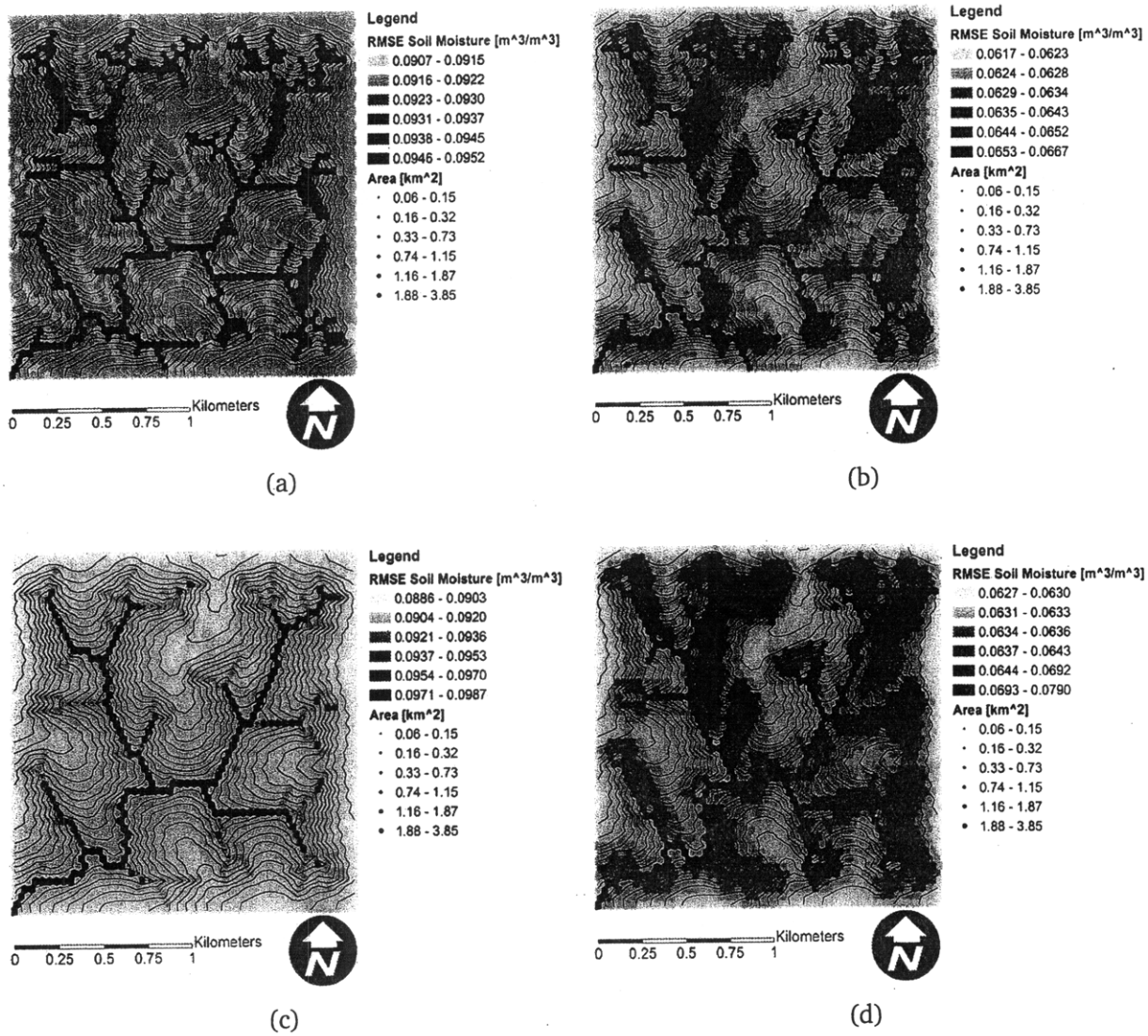


Figure D-1: RMSE relative to truth, filter-forecast cycle 1. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

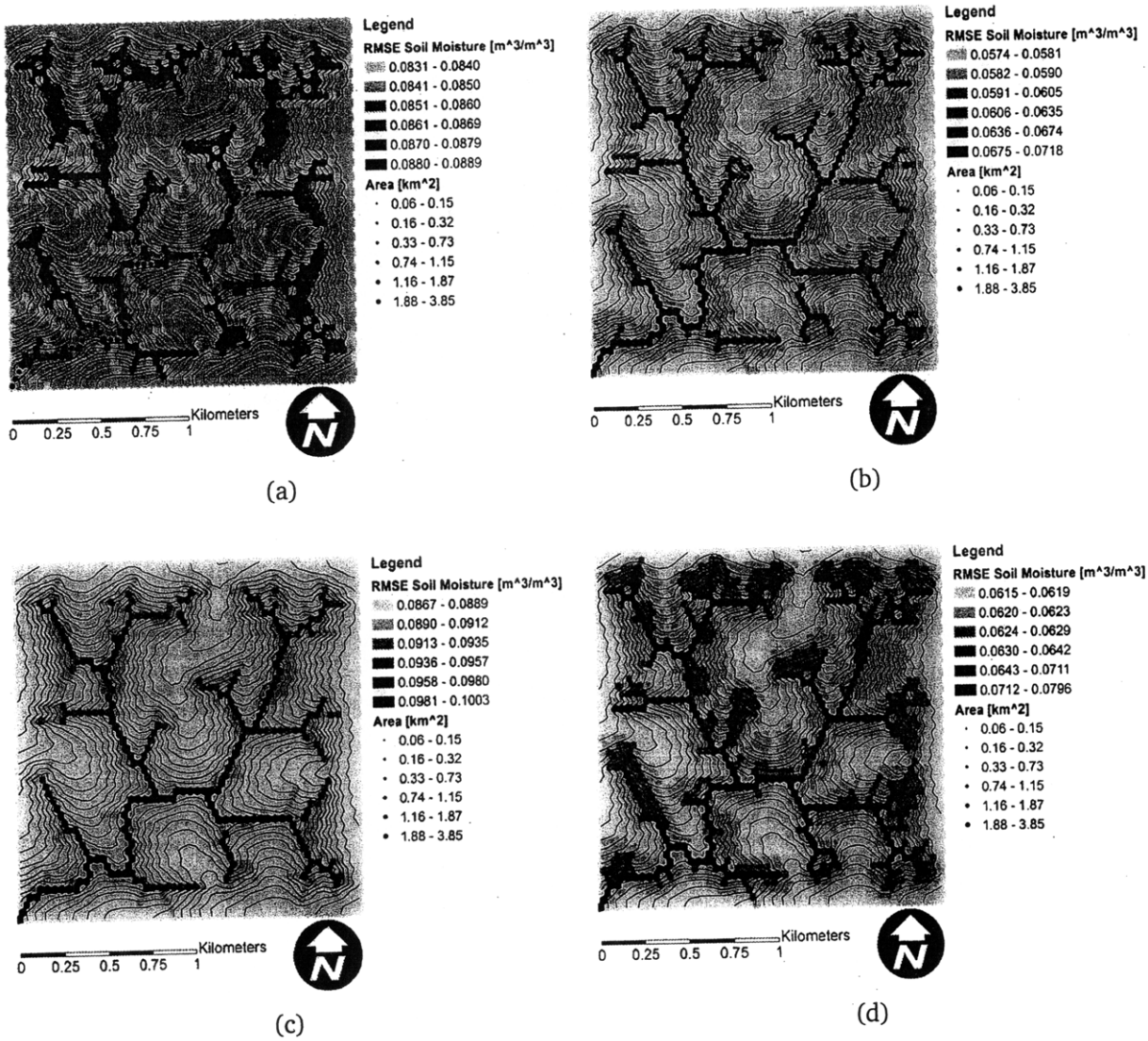


Figure D-2: RMSE relative to truth, filter-forecast cycle 2. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

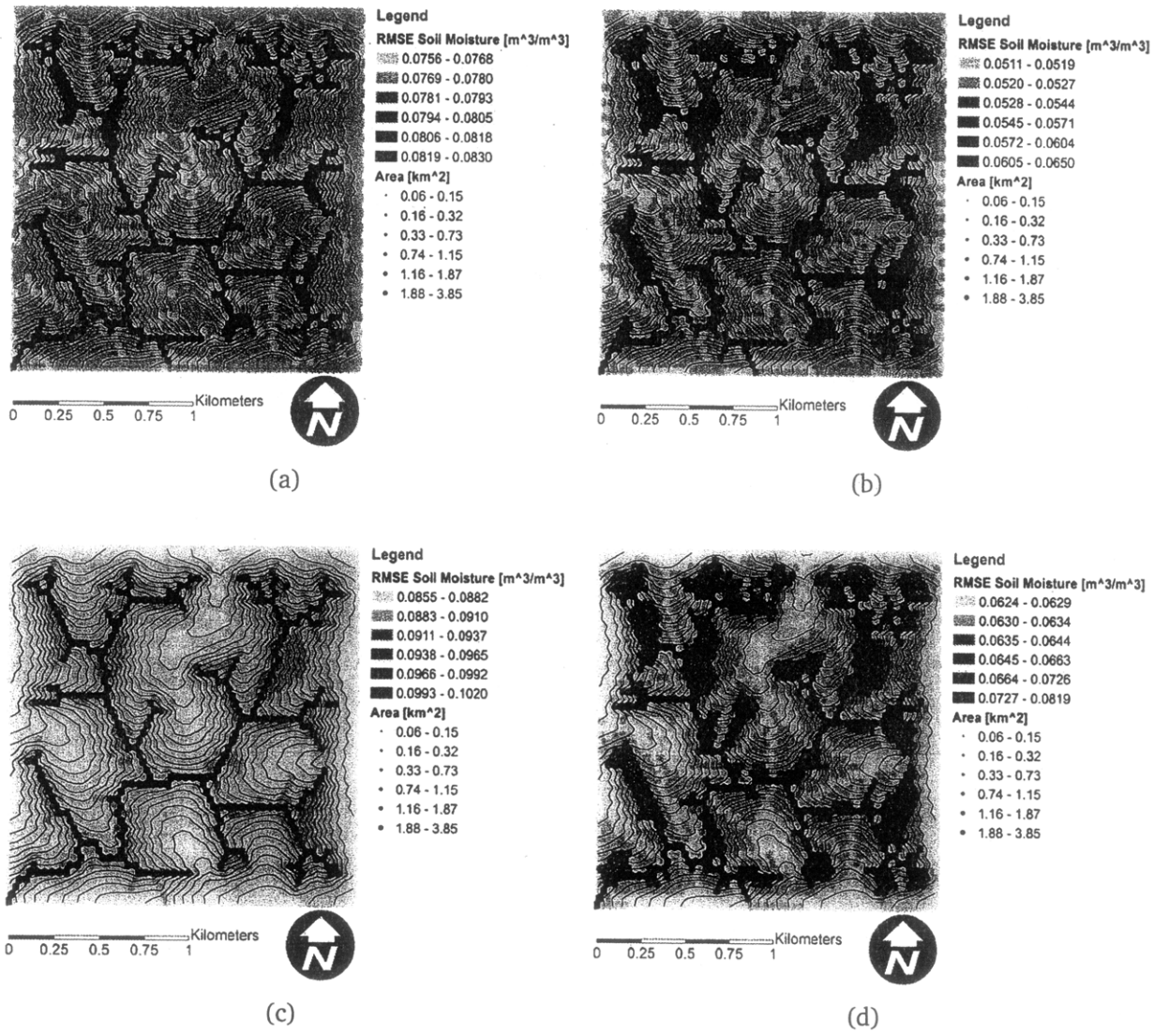


Figure D-3: RMSE relative to truth, filter-forecast cycle 3. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

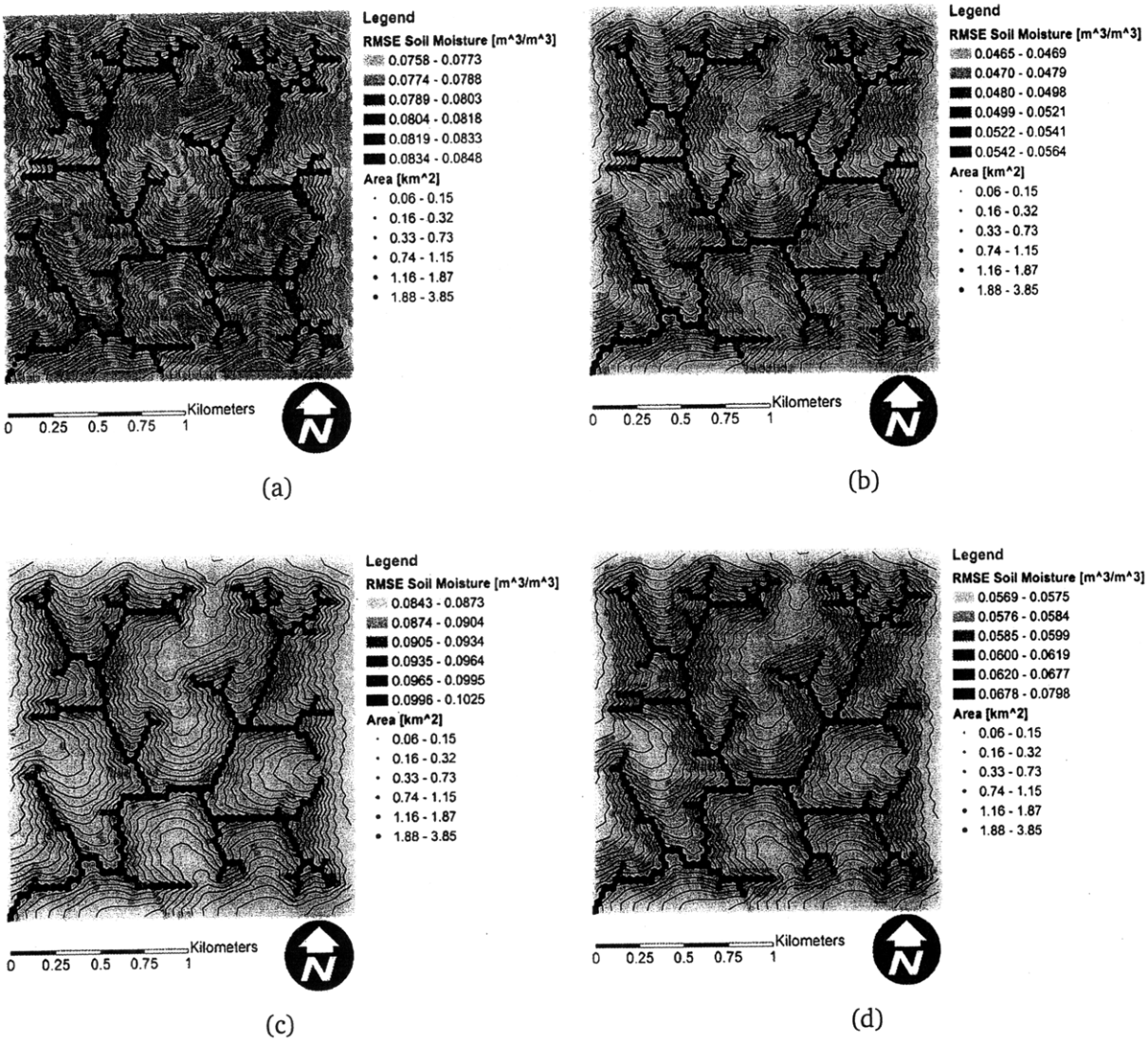


Figure D-4: RMSE relative to truth, filter-forecast cycle 4. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

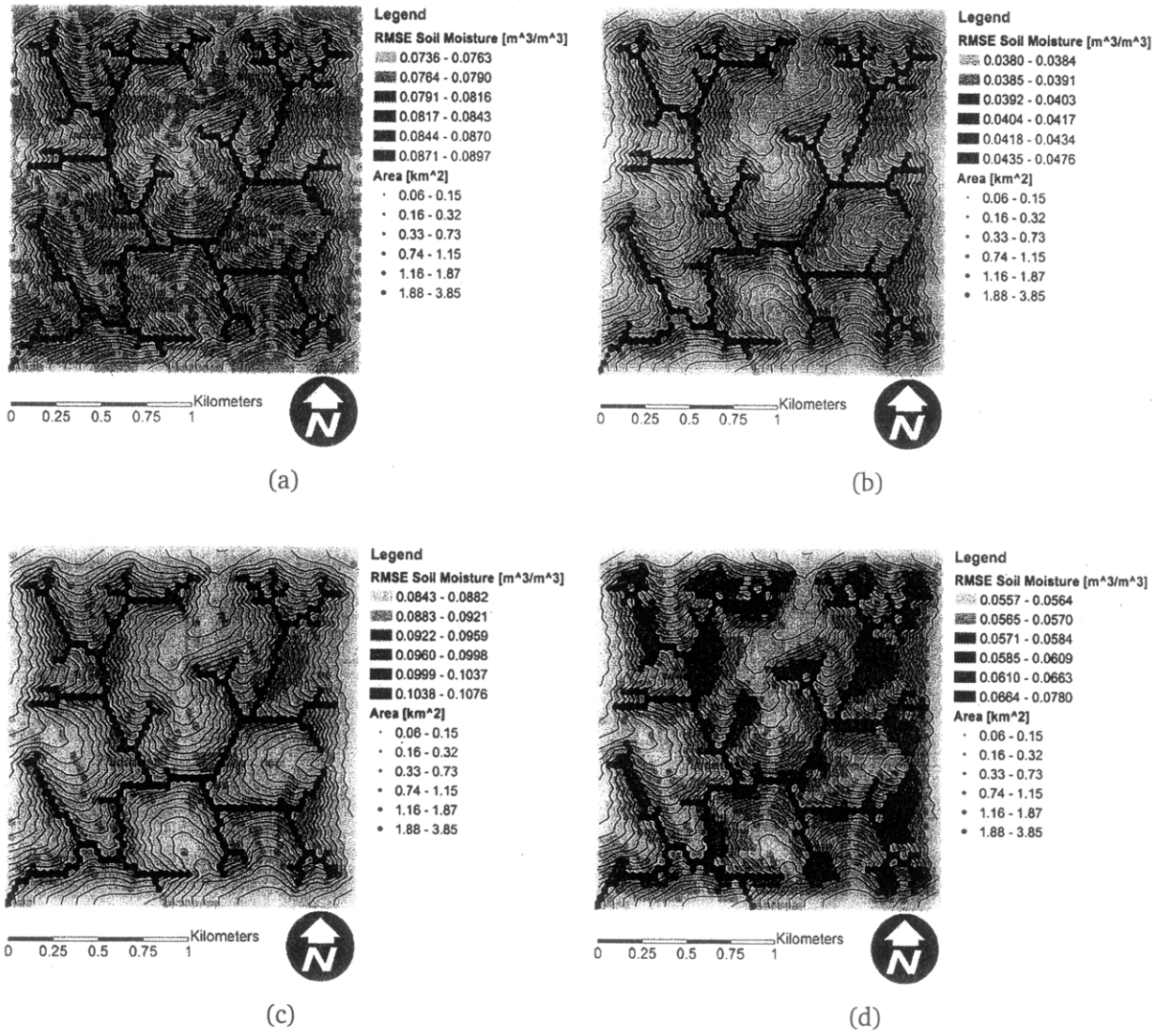


Figure D-5: RMSE relative to truth, filter-forecast cycle 5. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

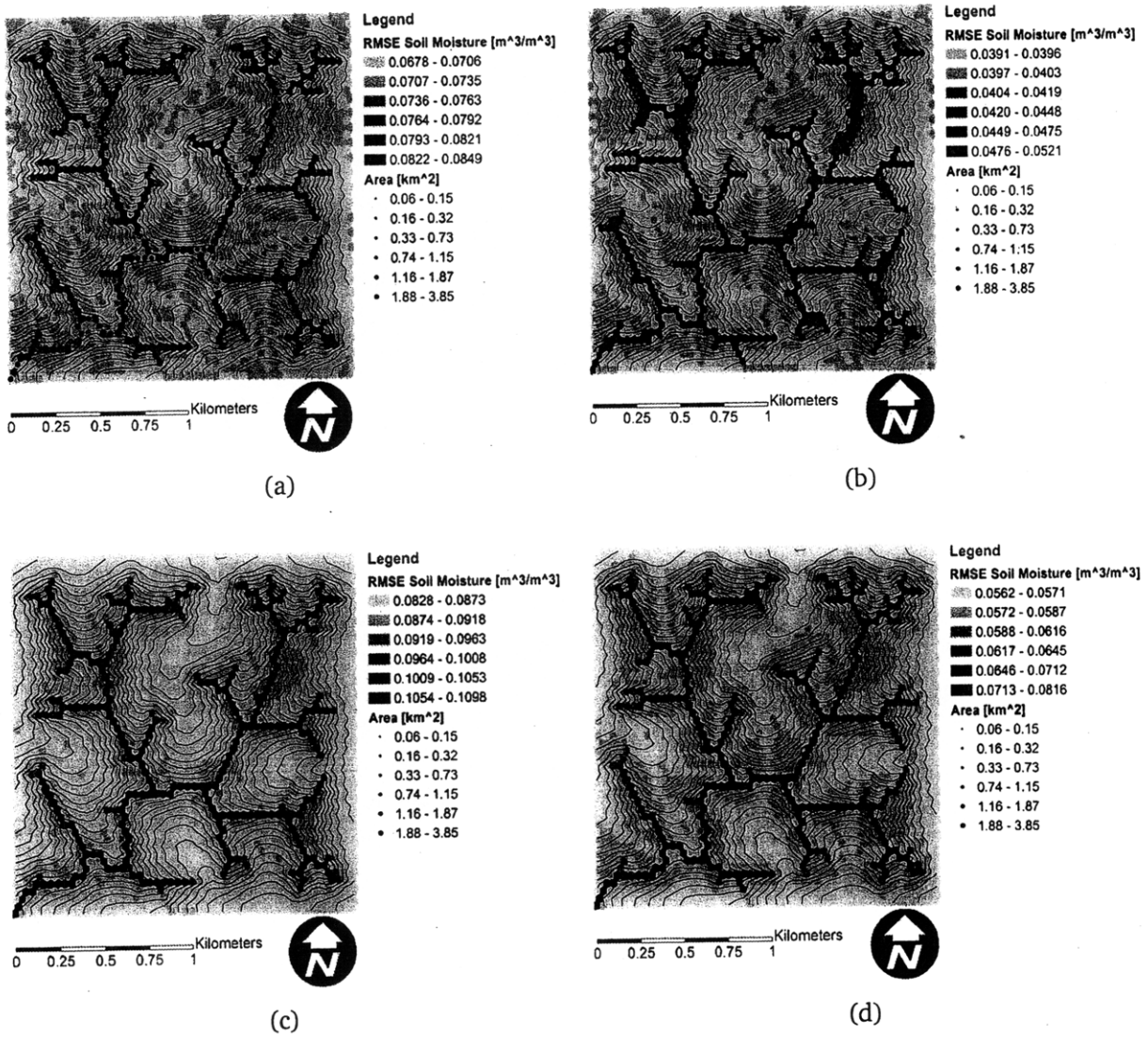


Figure D-6: RMSE relative to truth, filter-forecast cycle 6. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

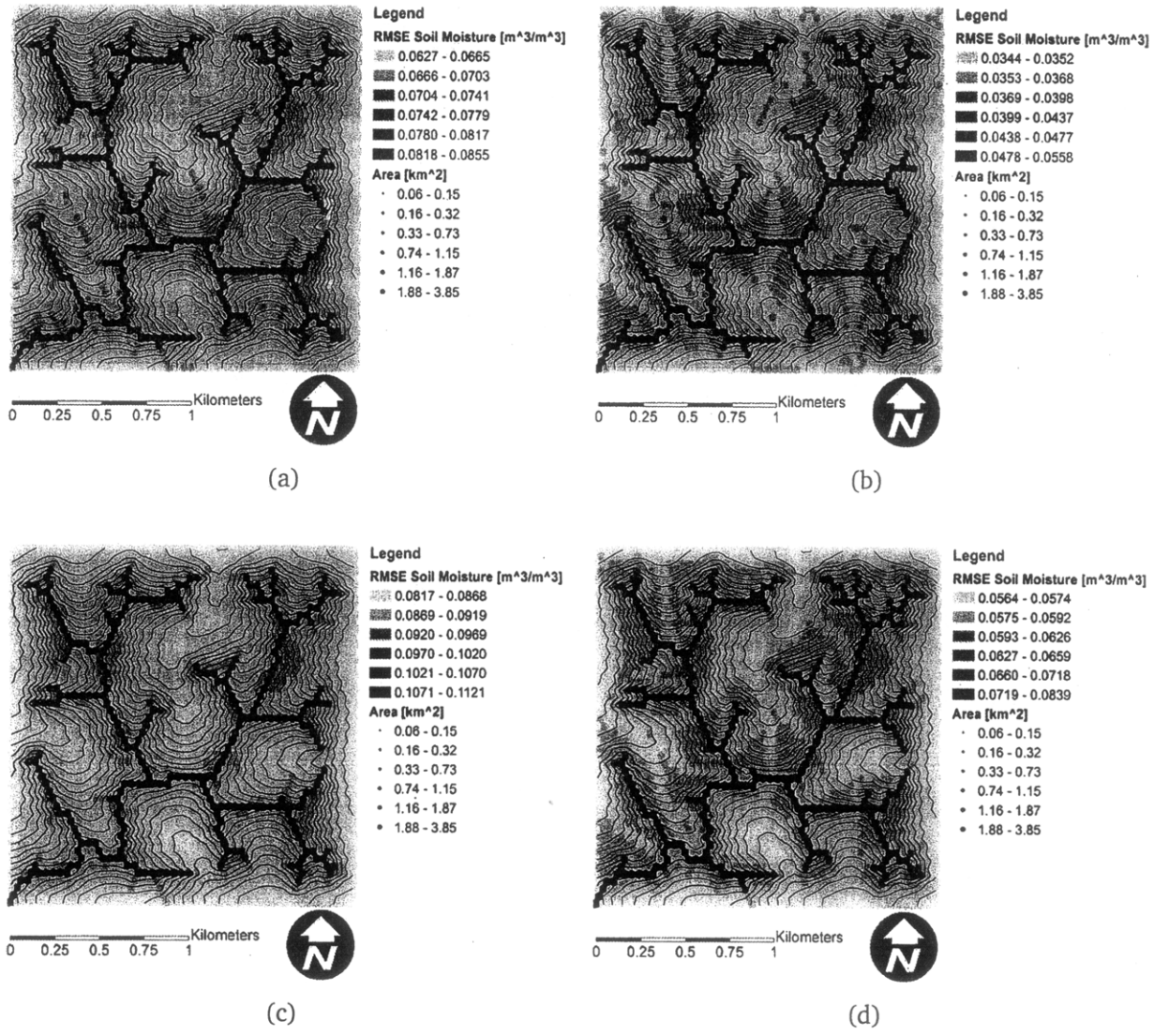


Figure D-7: RMSE relative to truth, filter-forecast cycle 7. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

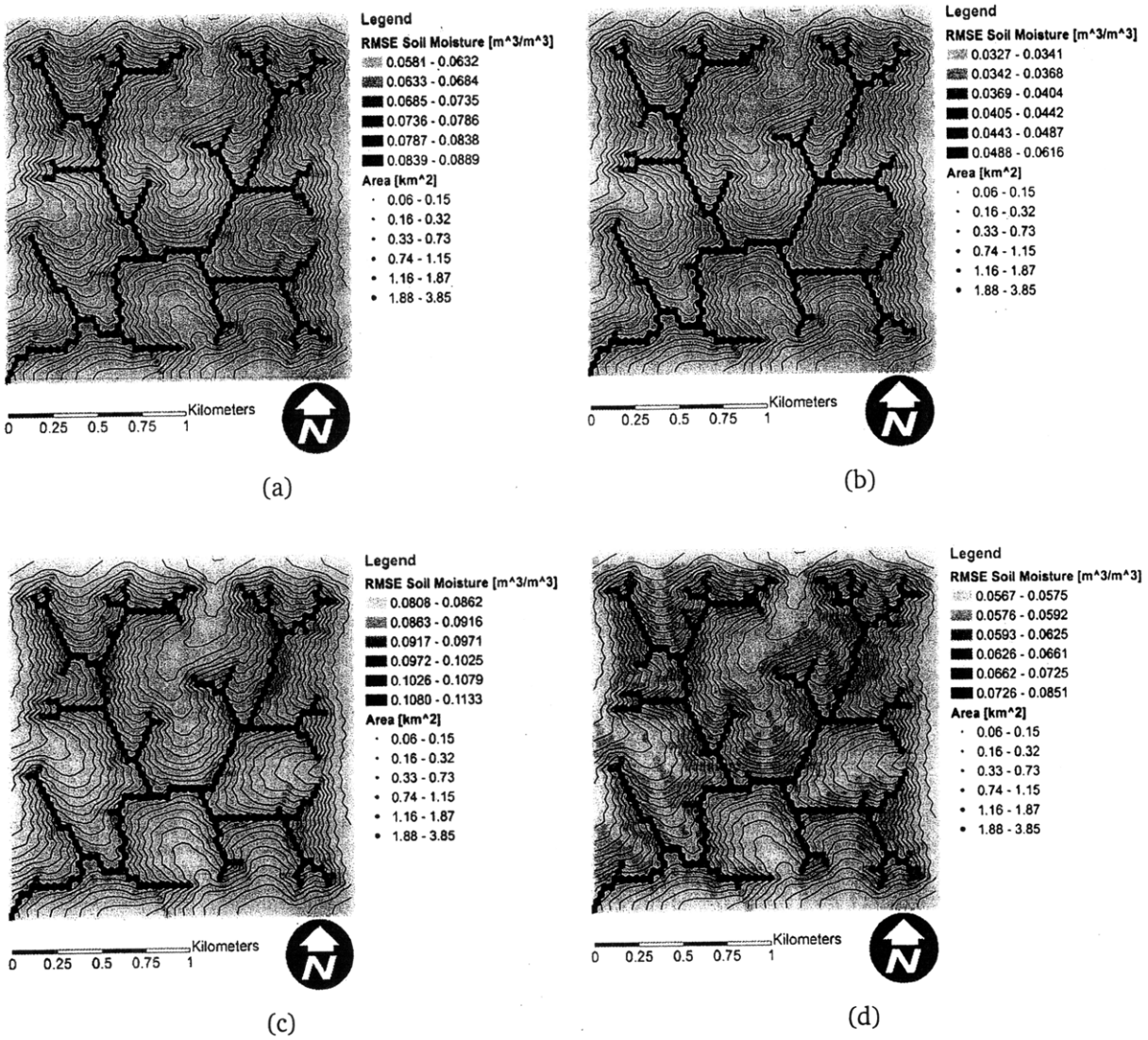


Figure D-8: RMSE relative to truth, filter-forecast cycle 8. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

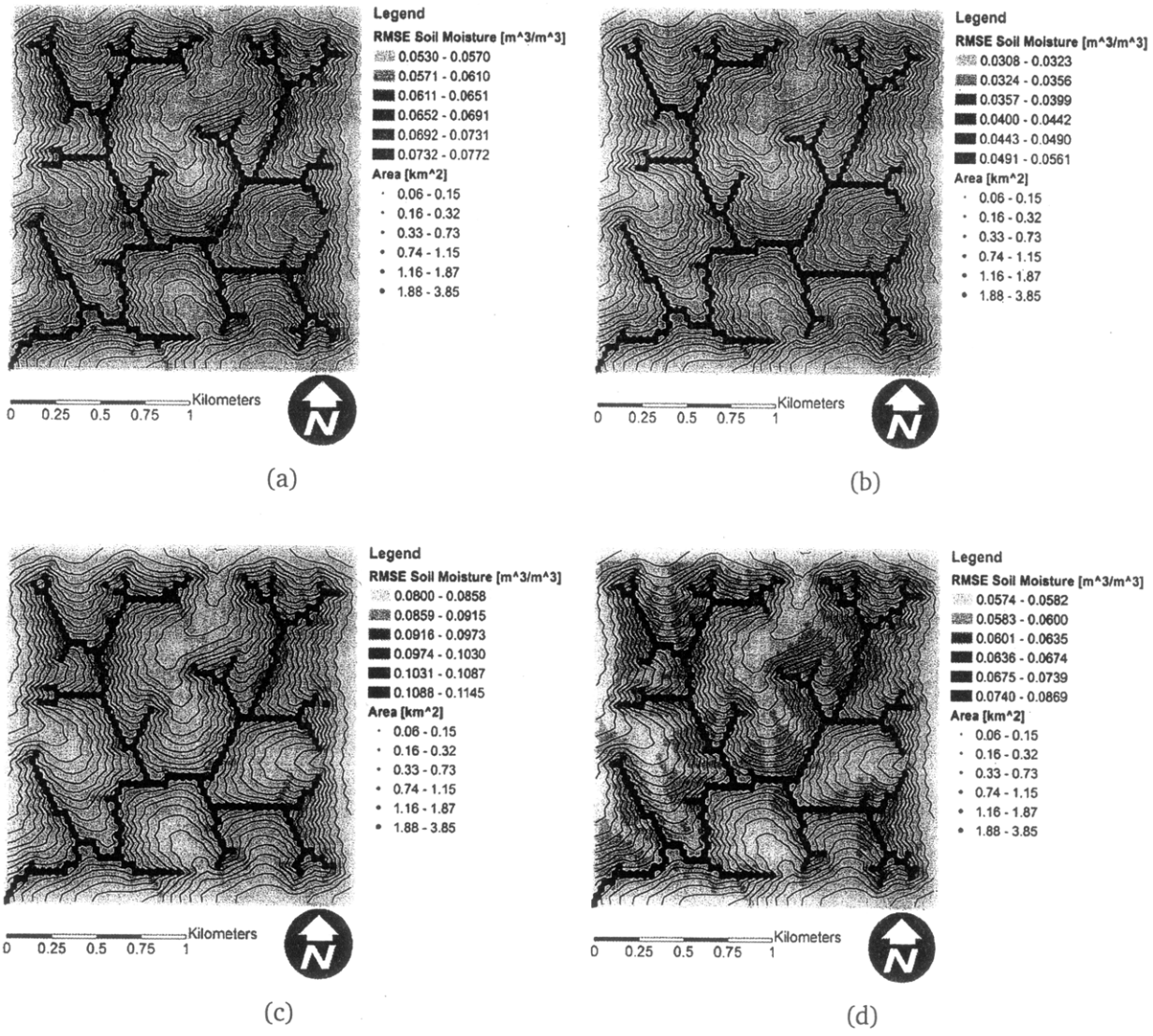


Figure D-9: RMSE relative to truth, filter-forecast cycle 9. (a) EnKF profile moisture, (b) EnKF near-surface moisture, (c) open-loop profile moisture, (d) open-loop near-surface moisture

D.1.2 Analysis outputs

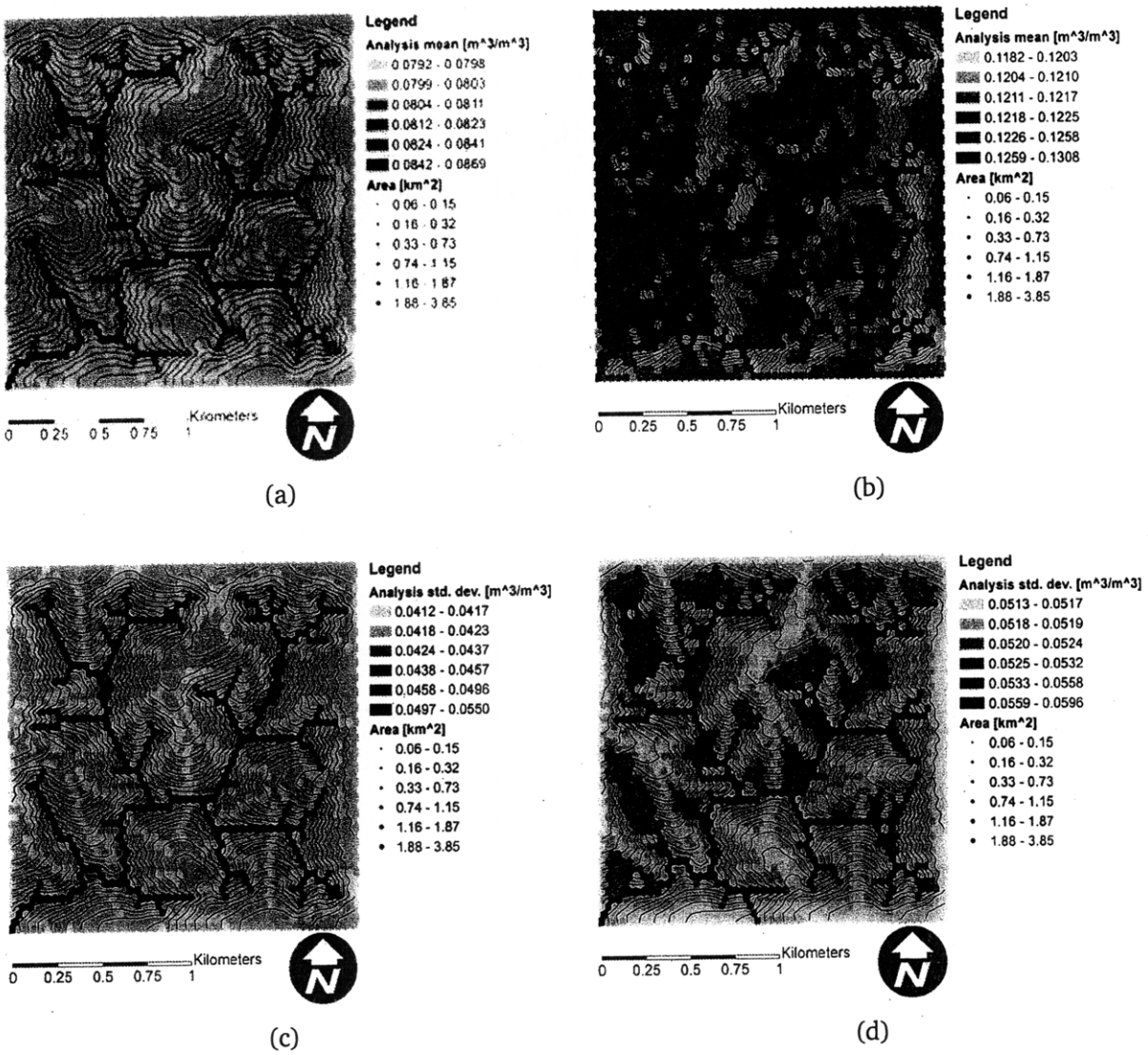
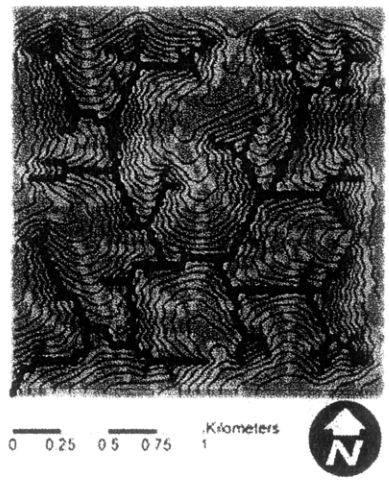
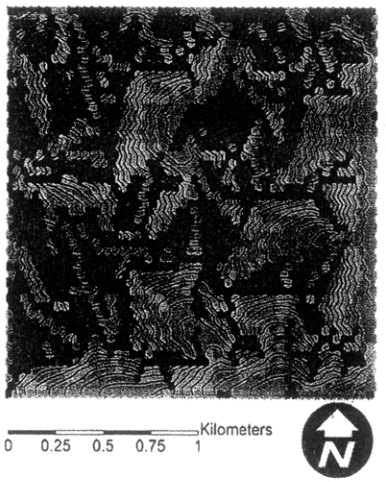


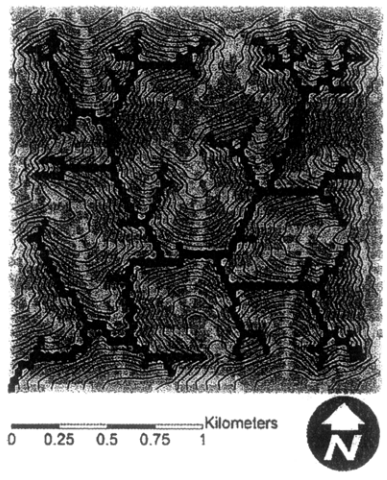
Figure D-10: Analysis output, filter-forecast cycle 1. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture



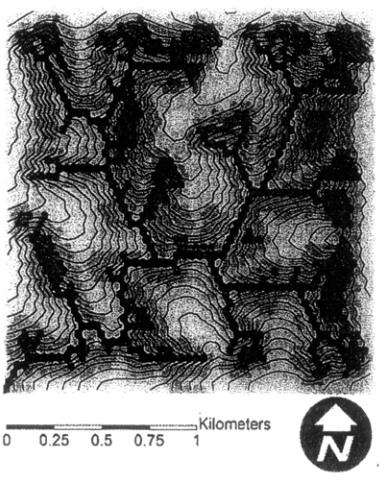
(a)



(b)



(c)



(d)

Figure D-11: Analysis output, filter-forecast cycle 2. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture

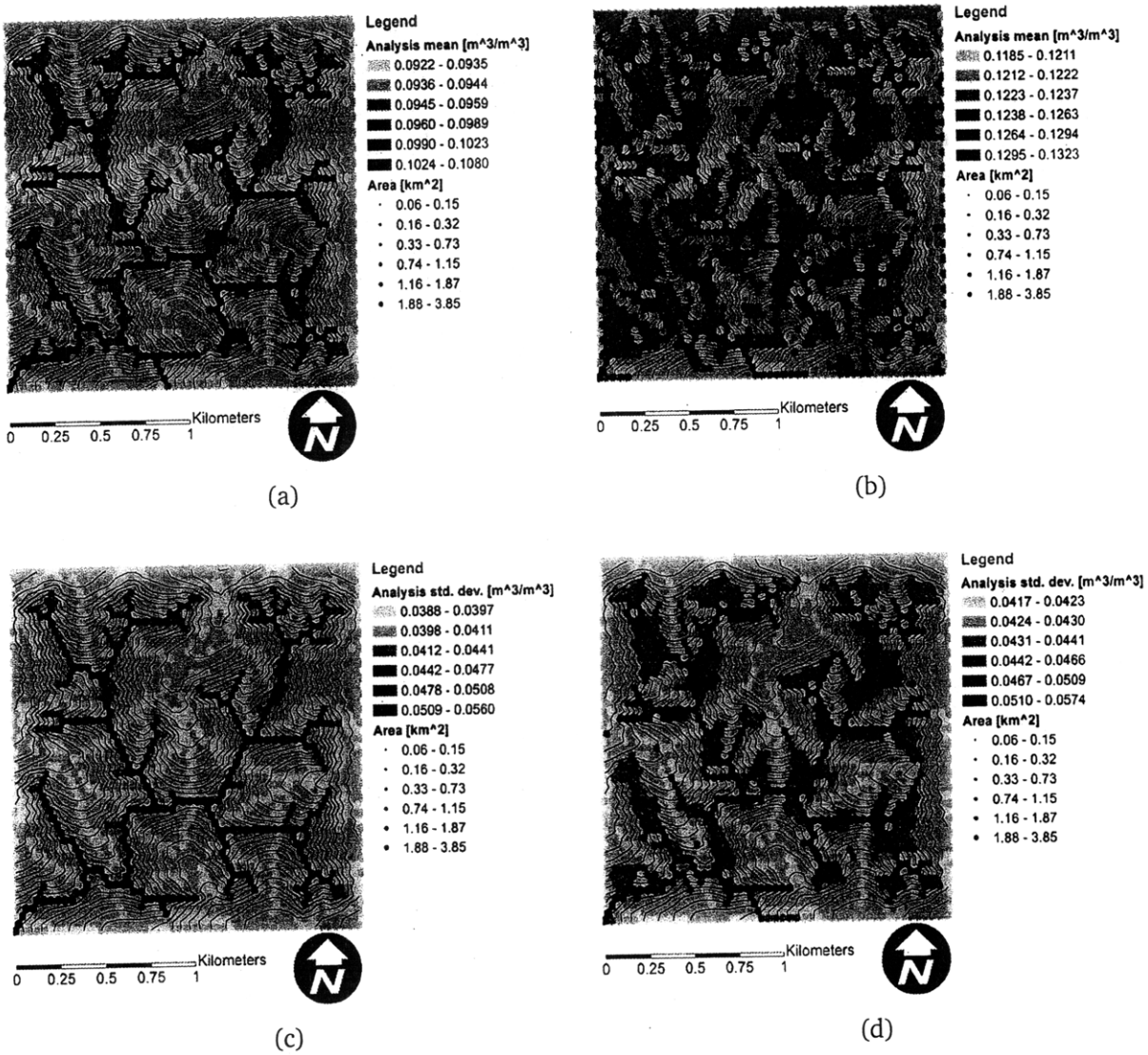
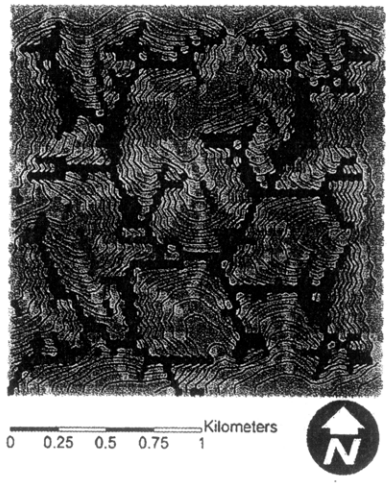
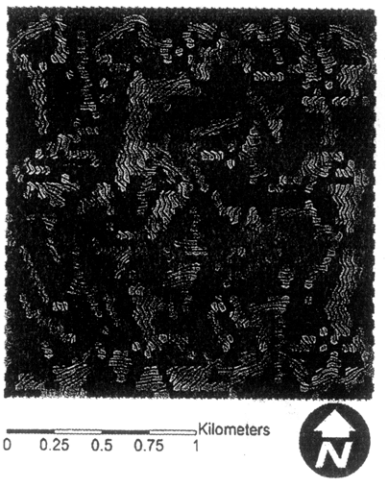


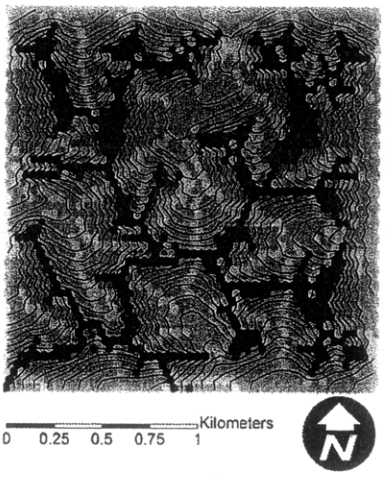
Figure D-12: Analysis output, filter-forecast cycle 3. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture



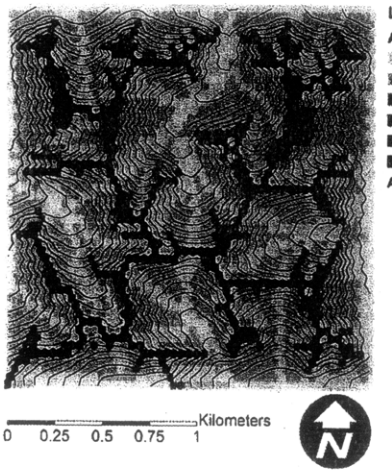
(a)



(b)

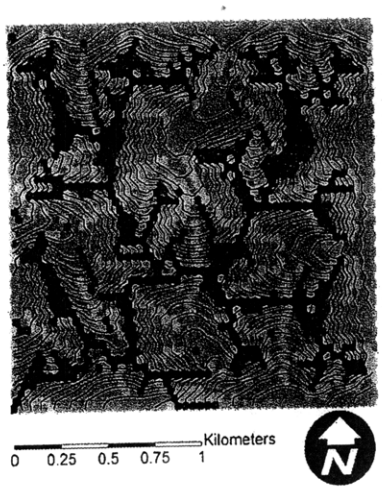


(c)

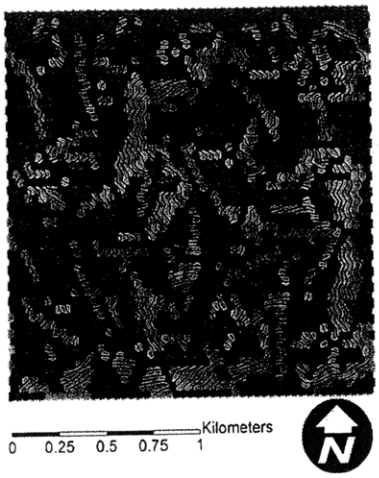


(d)

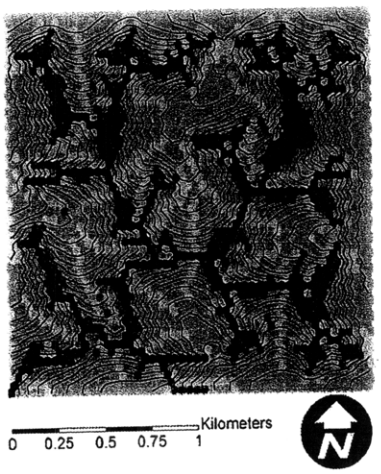
Figure D-13: Analysis output, filter-forecast cycle 4. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture



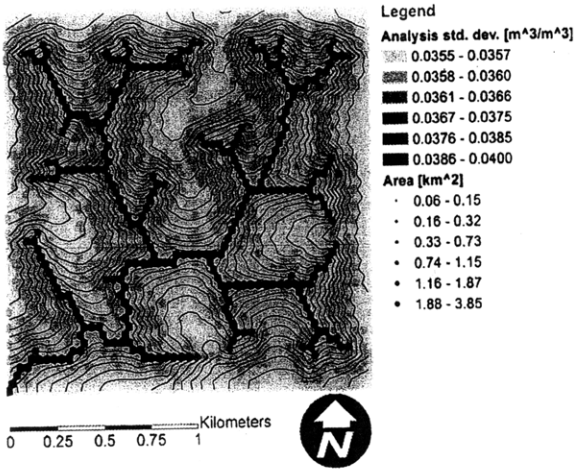
(a)



(b)



(c)



(d)

Figure D-14: Analysis output, filter-forecast cycle 5. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture

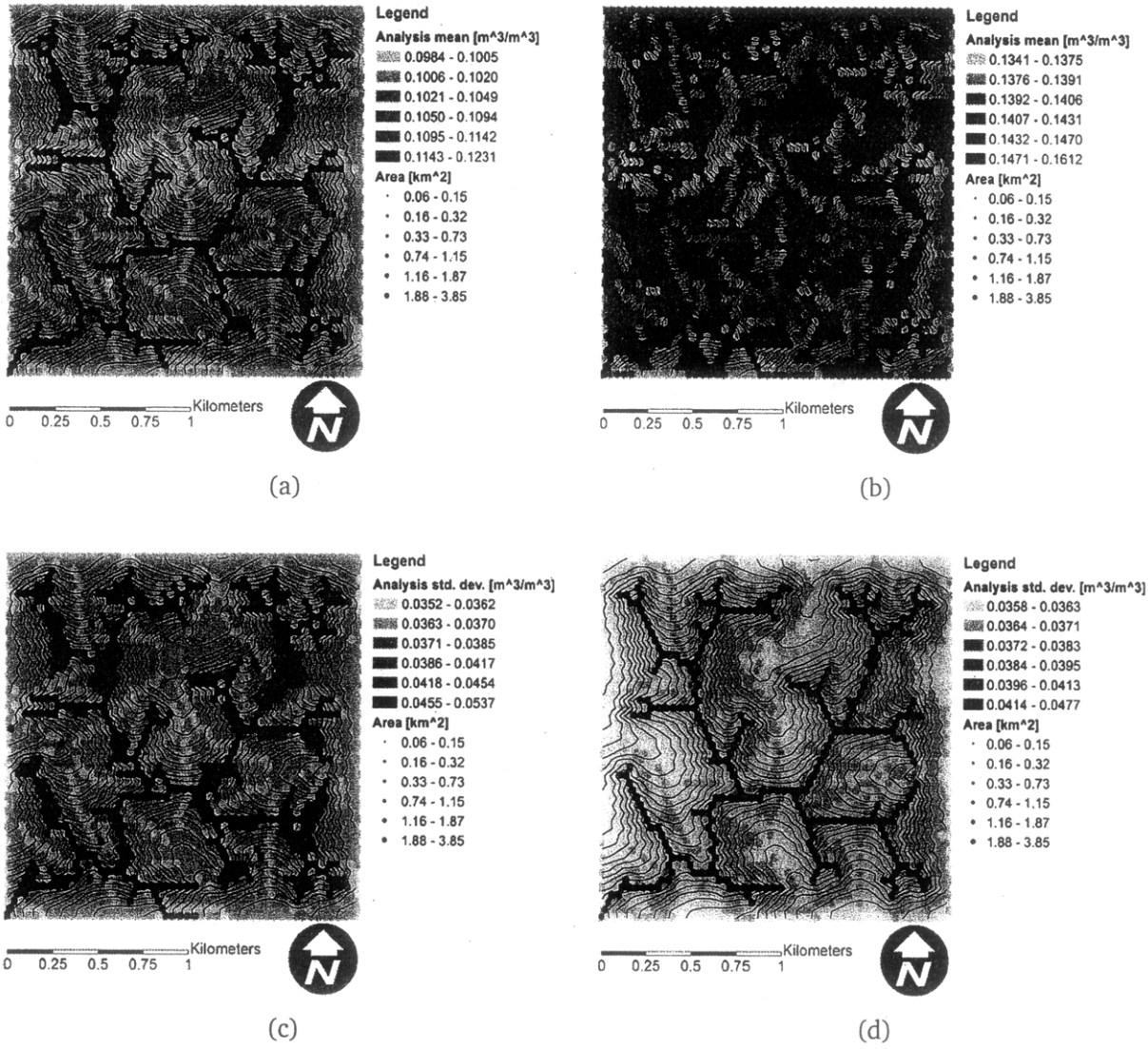


Figure D-15: Analysis output, filter-forecast cycle 6. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture

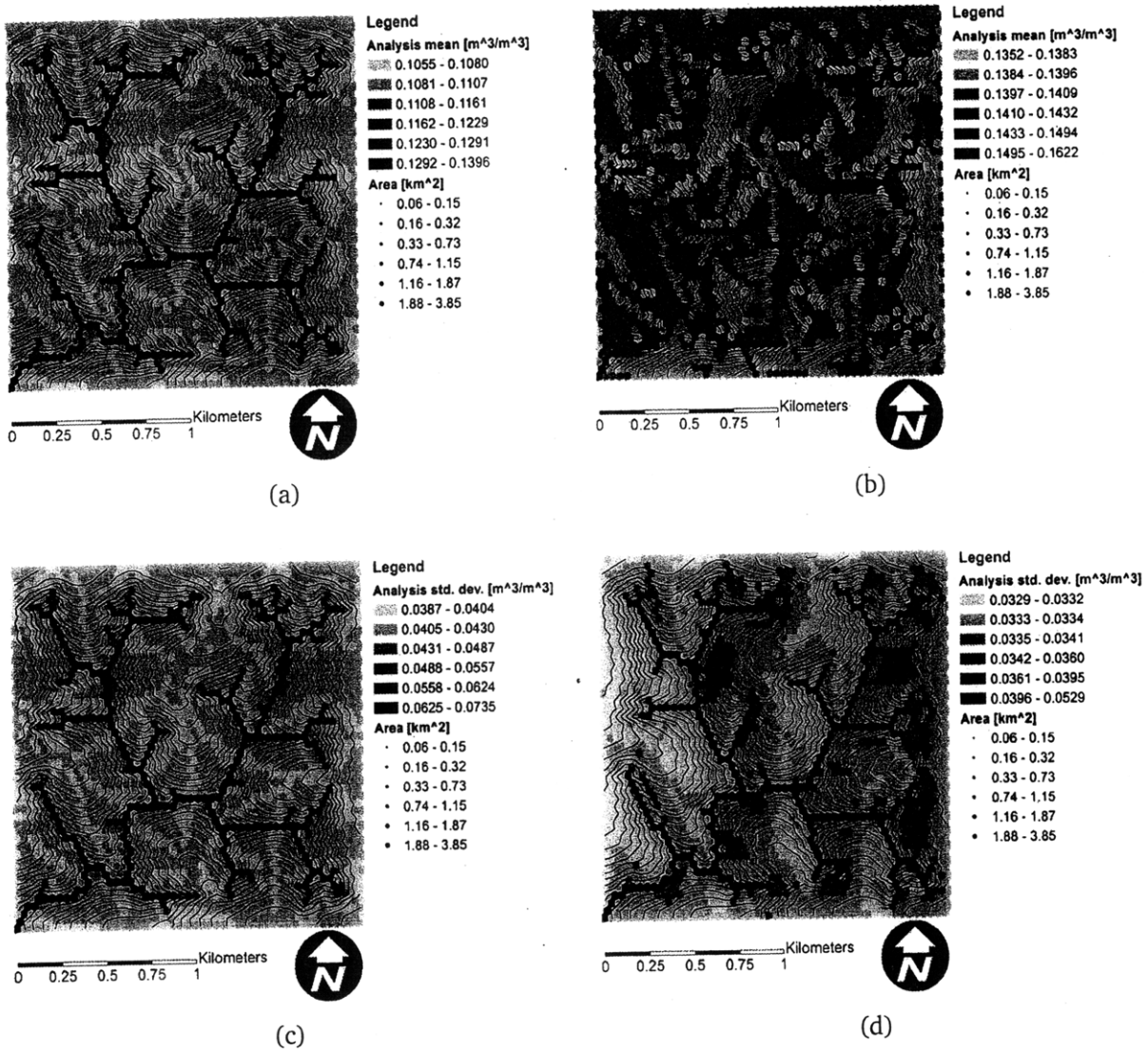


Figure D-16: Analysis output, filter-forecast cycle 7. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture

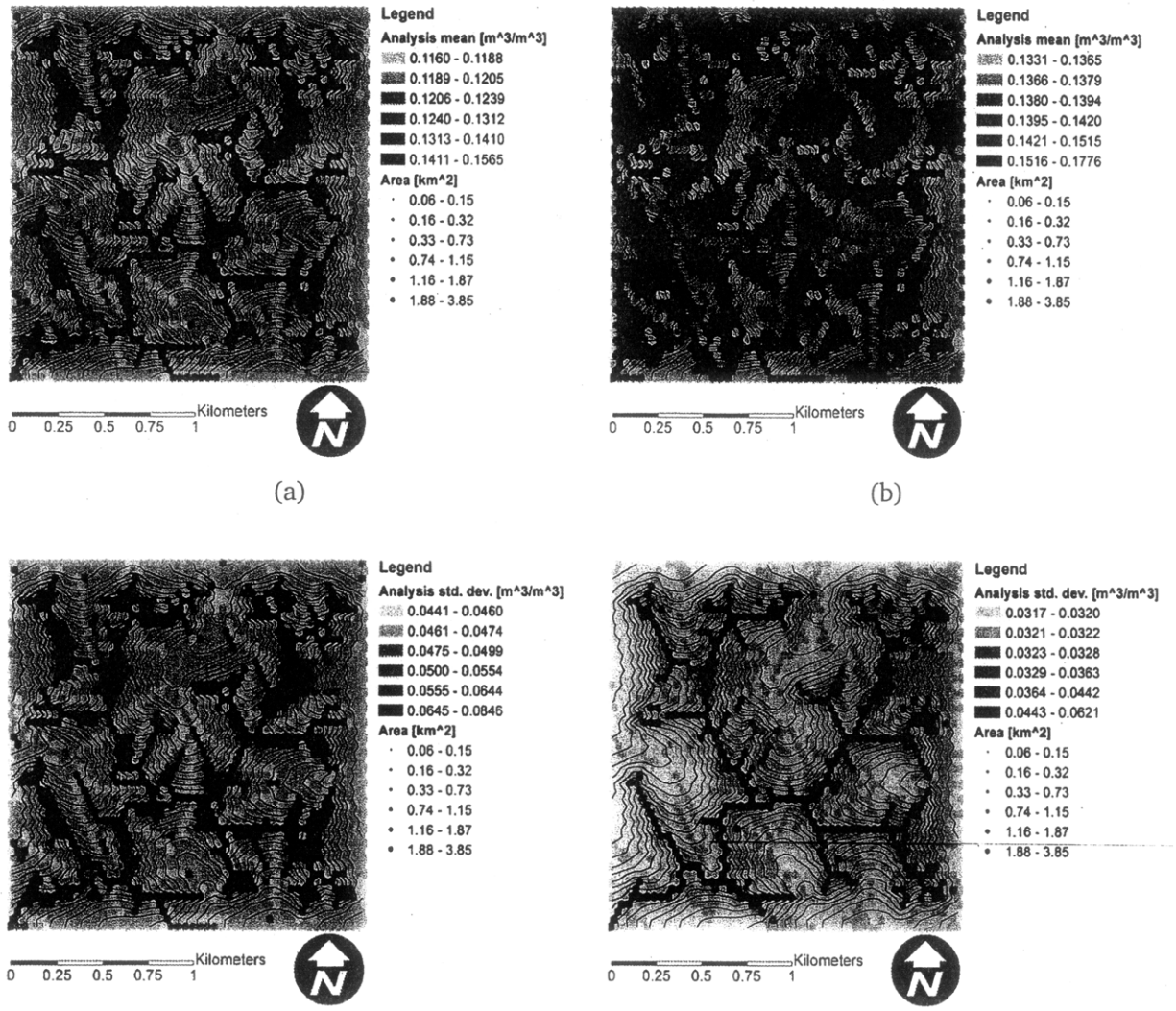


Figure D-17: Analysis output, filter-forecast cycle 8. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture

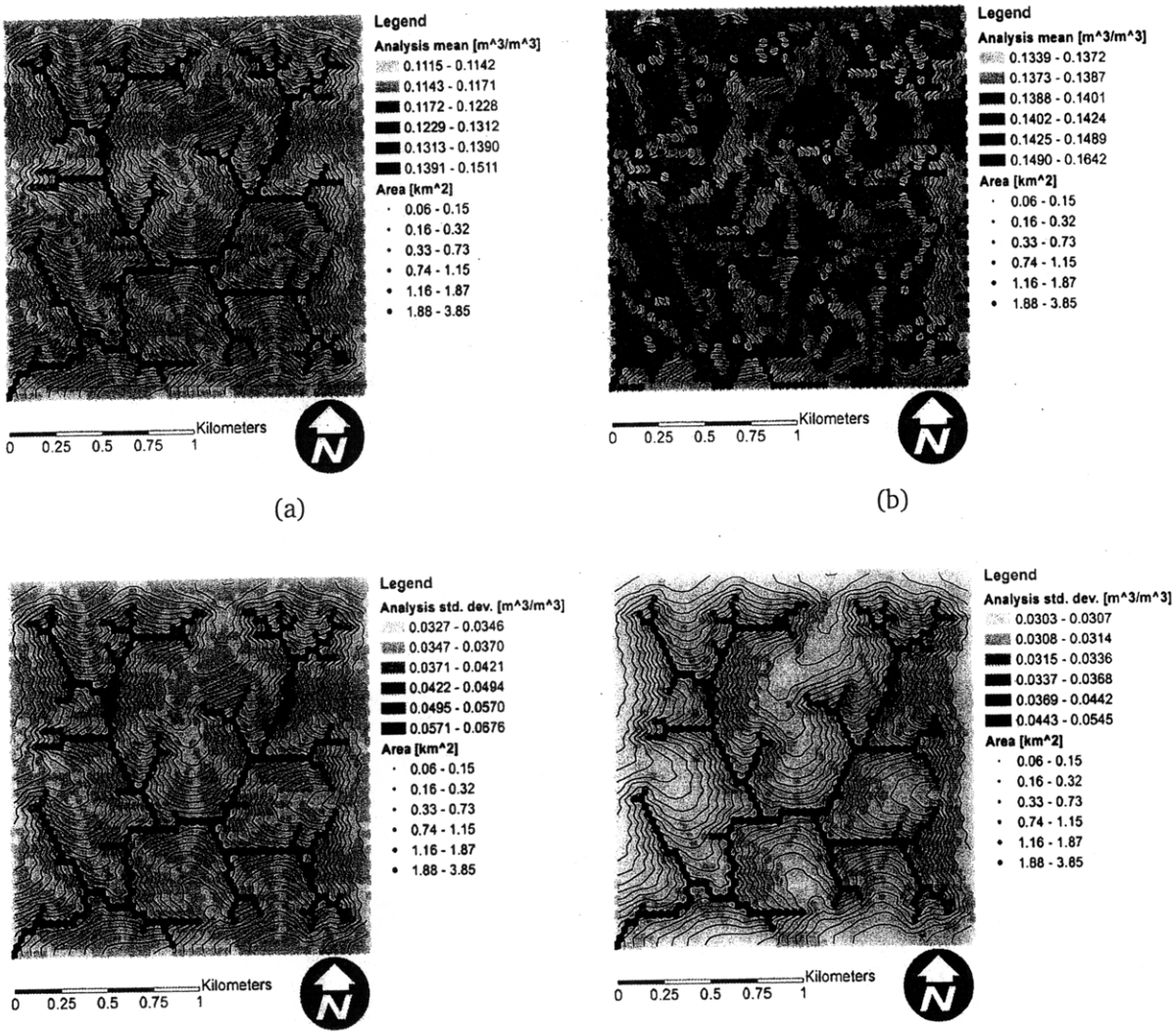


Figure D-18: Analysis output, filter-forecast cycle 9. (a) Ensemble mean profile moisture, (b) ensemble mean near-surface moisture, (c) ensemble standard deviation profile moisture, (d) ensemble standard deviation near-surface moisture

D.1.3 Change across the update

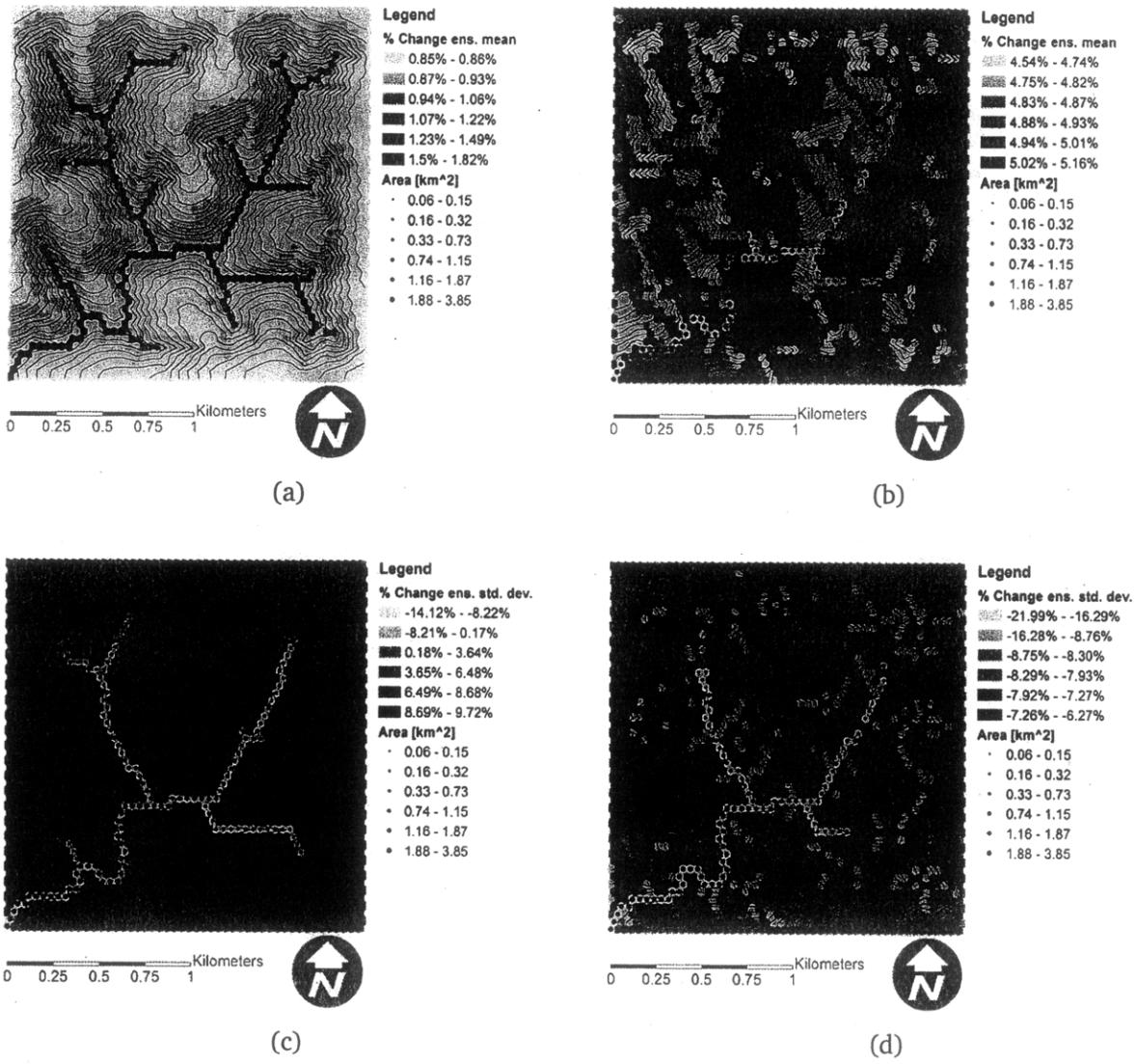


Figure D-19: Filter-forecast cycle 1 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation

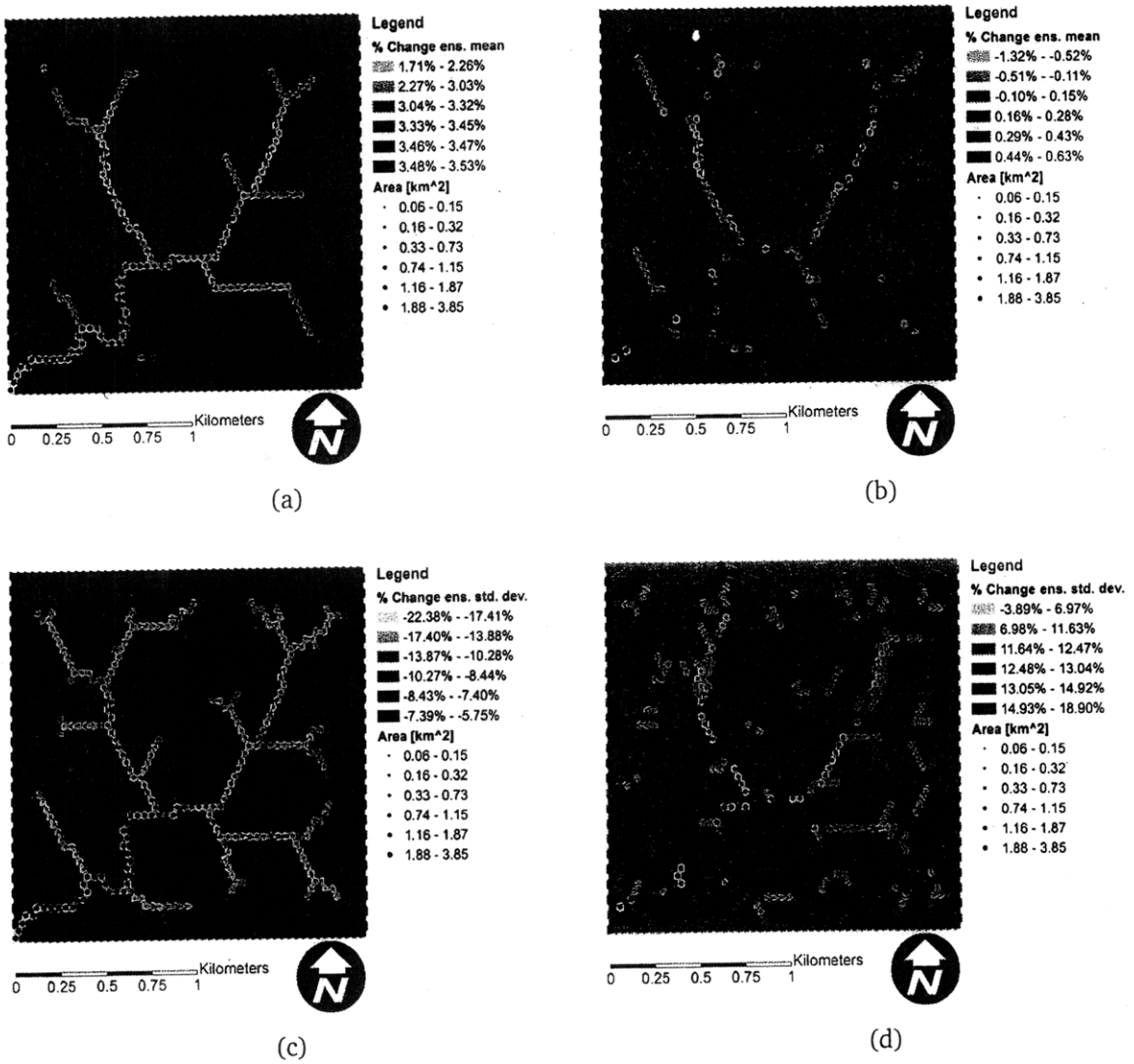


Figure D-20: Filter-forecast cycle 2 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation

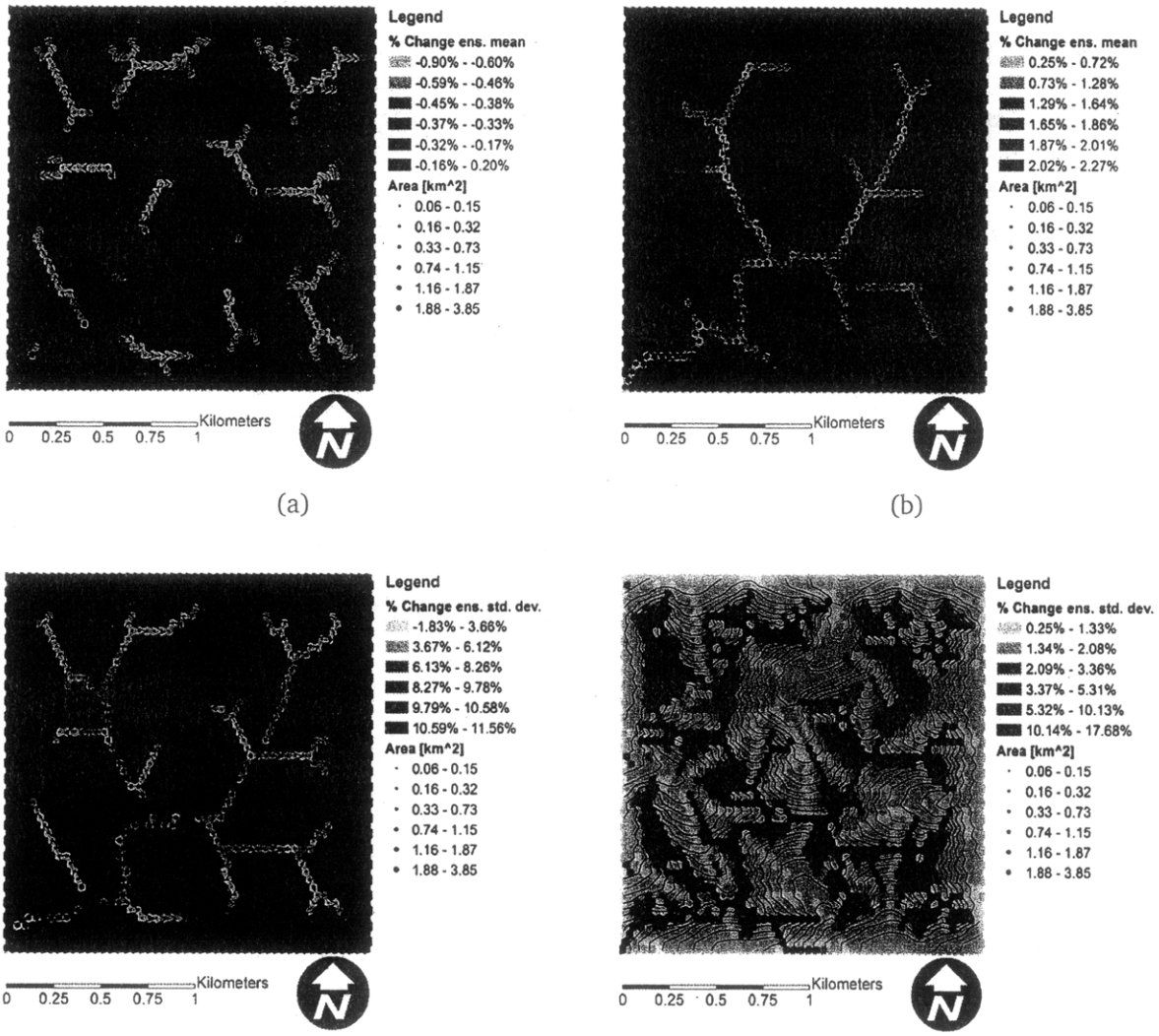


Figure D-21: Filter-forecast cycle 3 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation

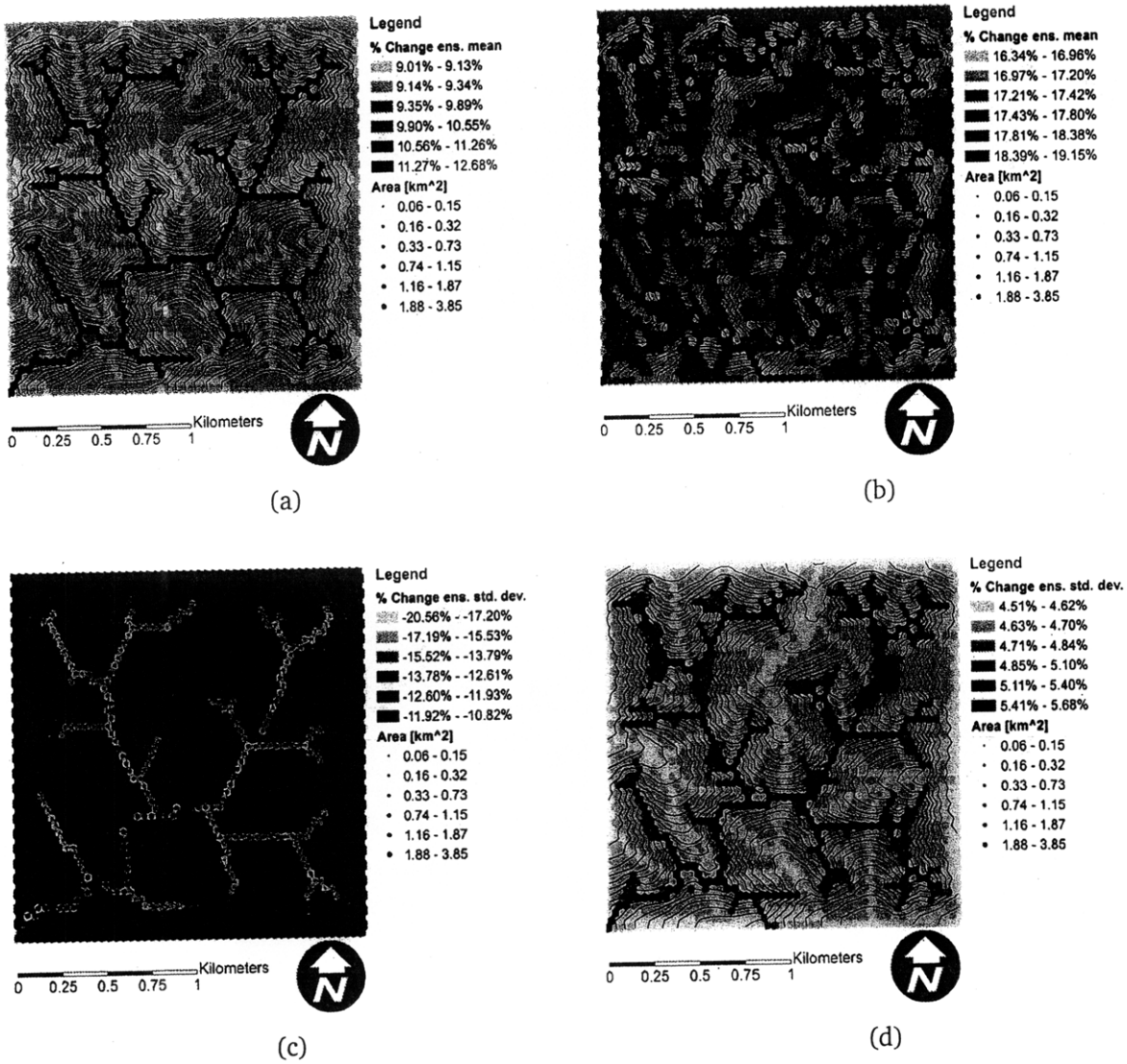
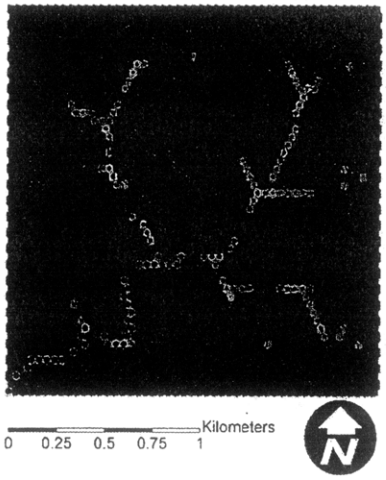
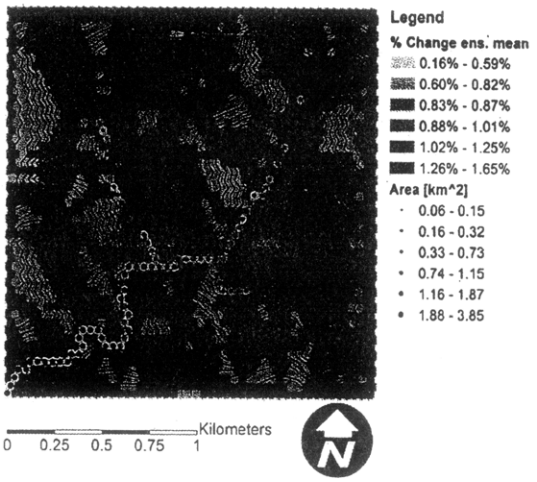


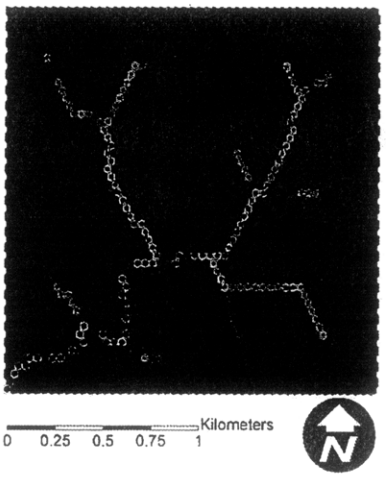
Figure D-22: Filter-forecast cycle 4 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation



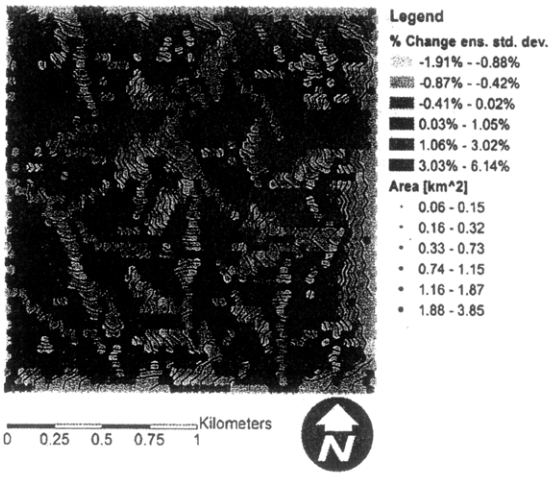
(a)



(b)



(c)



(d)

Figure D-23: Filter-forecast cycle 5 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation

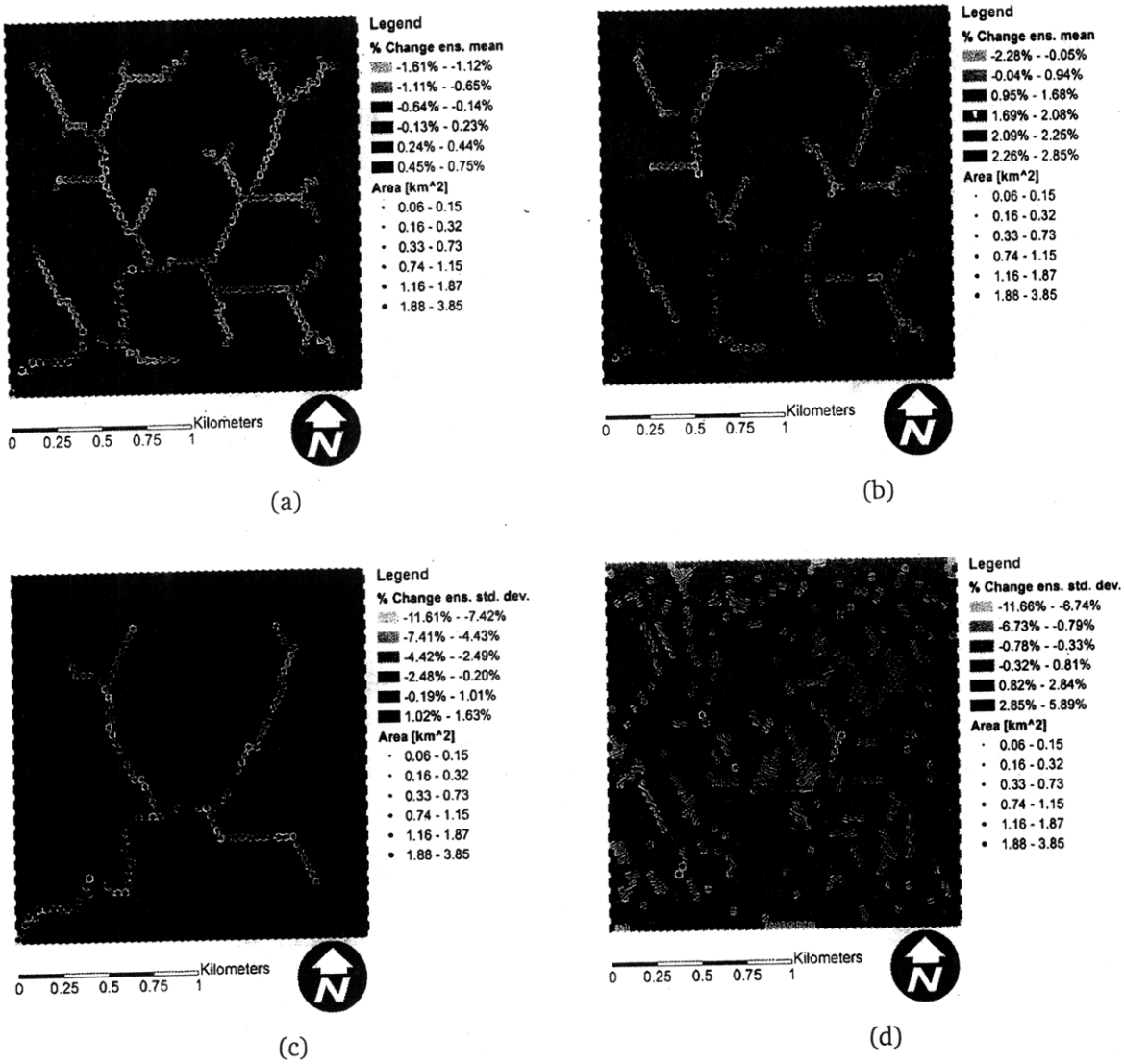


Figure D-24: Filter-forecast cycle 6 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation

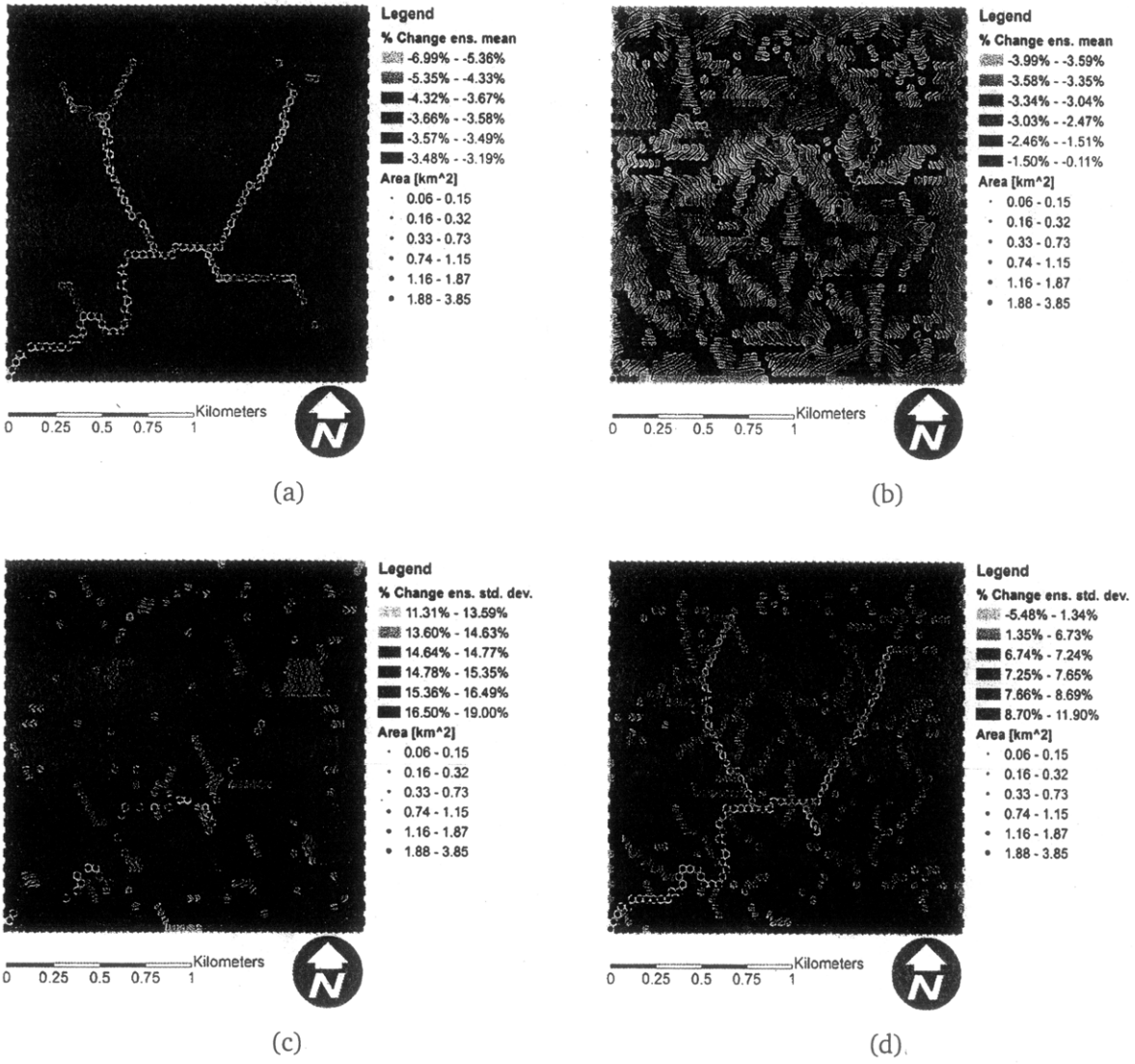


Figure D-25: Filter-forecast cycle 7 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation

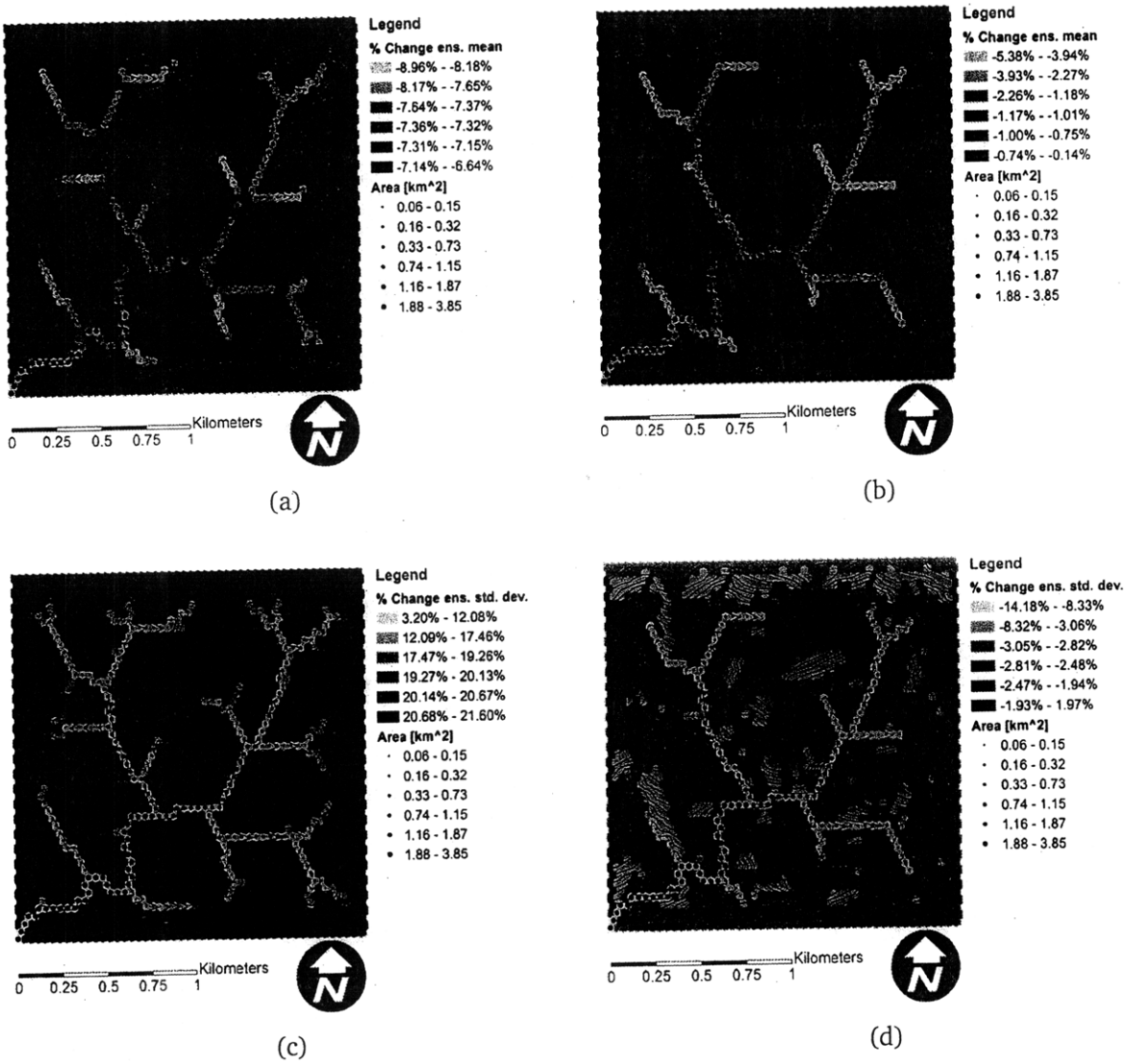


Figure D-26: Filter-forecast cycle 8 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation

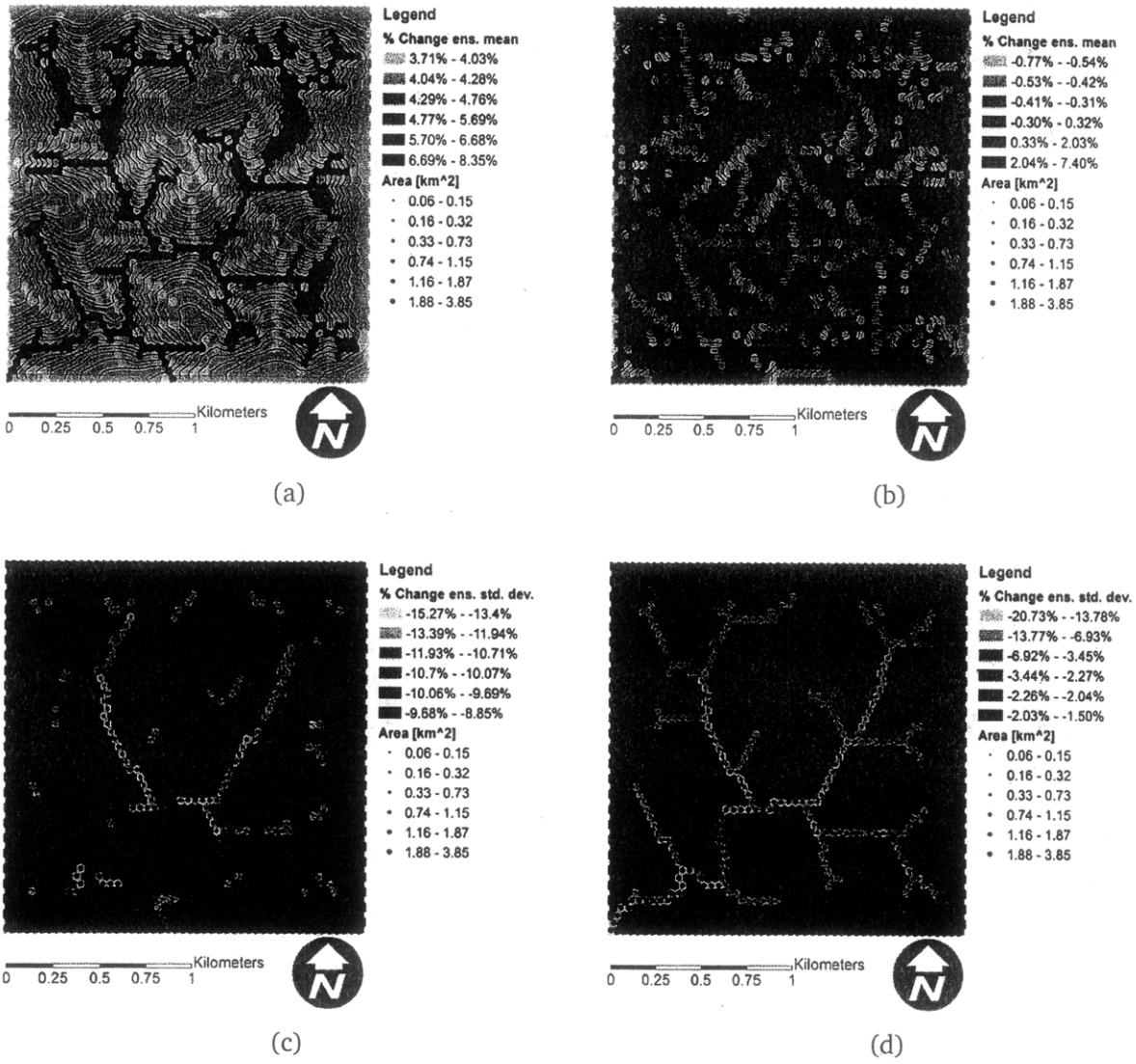
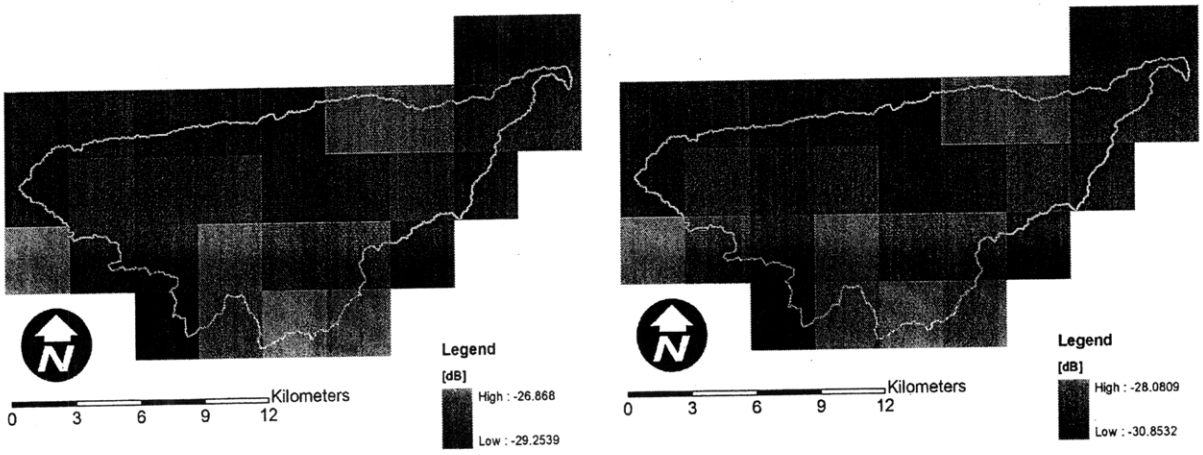


Figure D-27: Filter-forecast cycle 9 (a) percent change in profile moisture ensemble mean, (b) percent change in near-surface moisture ensemble mean, (c) percent change in profile moisture ensemble standard deviation, (d) percent change in near-surface moisture ensemble standard deviation

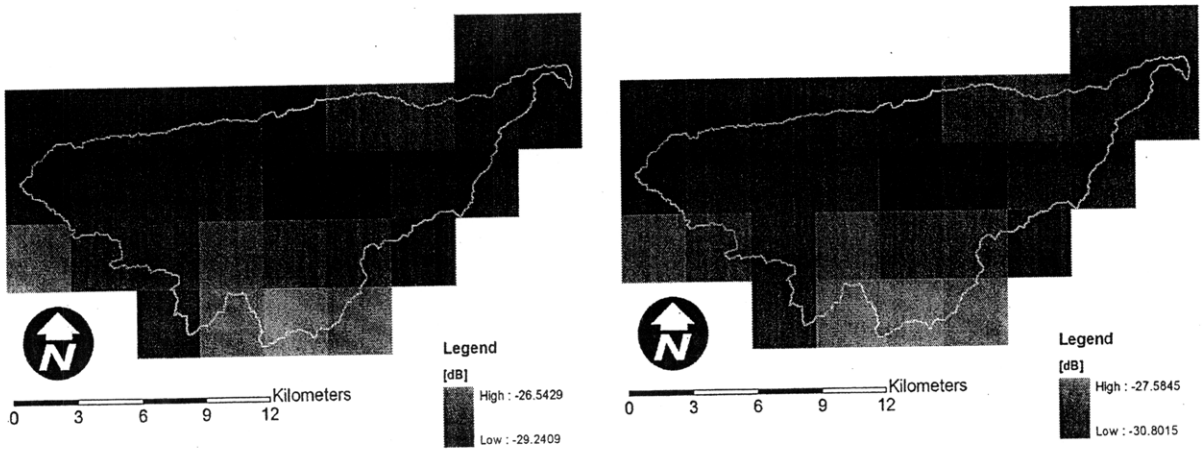
D.2 Expanded results: assimilation of radar backscatter data in WGEW

D.2.1 Synthetic observations



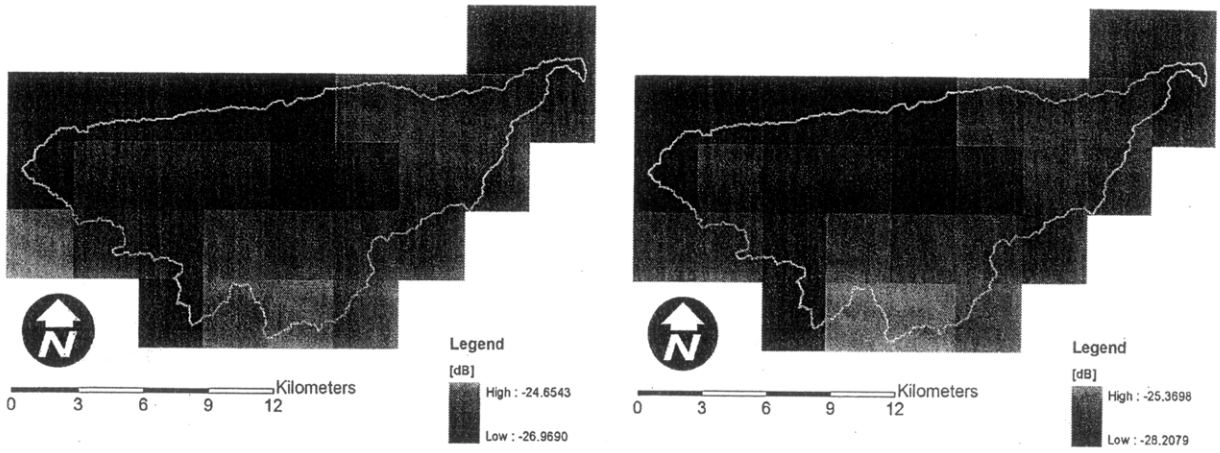
(a) (b)

Figure D-28: Average observations across the four candidate observations in the (a) horizontally- and (b) vertically-copolarized states at the first analysis (72 hr).



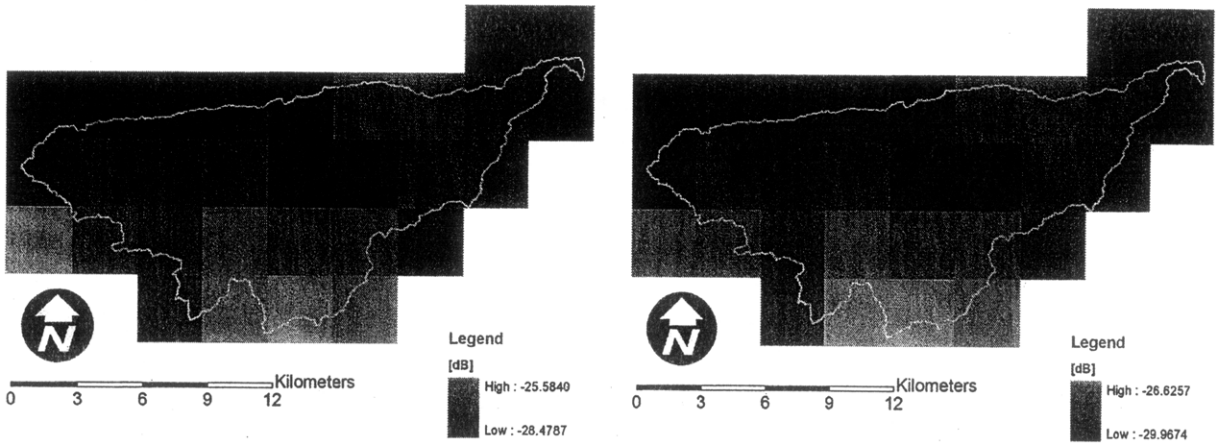
(a) (b)

Figure D-29: Average observations across the four candidate observations in the (a) horizontally- and (b) verticall-copolarized states at the first analysis (144 hr).



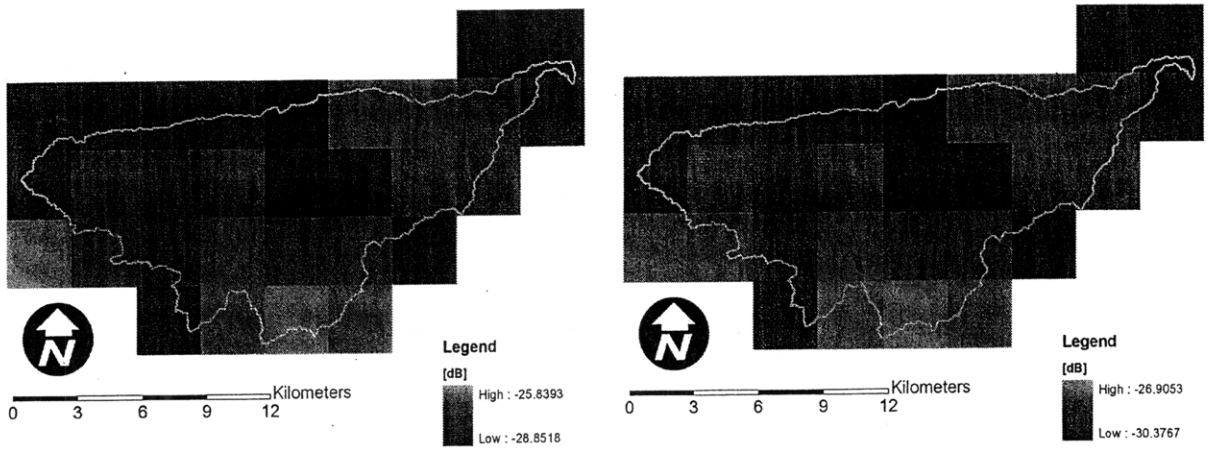
(a) (b)

Figure D-30: Average observations across the four candidate observations in the (a) horizontally- and (b) vertically-copolarized states at the third analysis (216 hr).



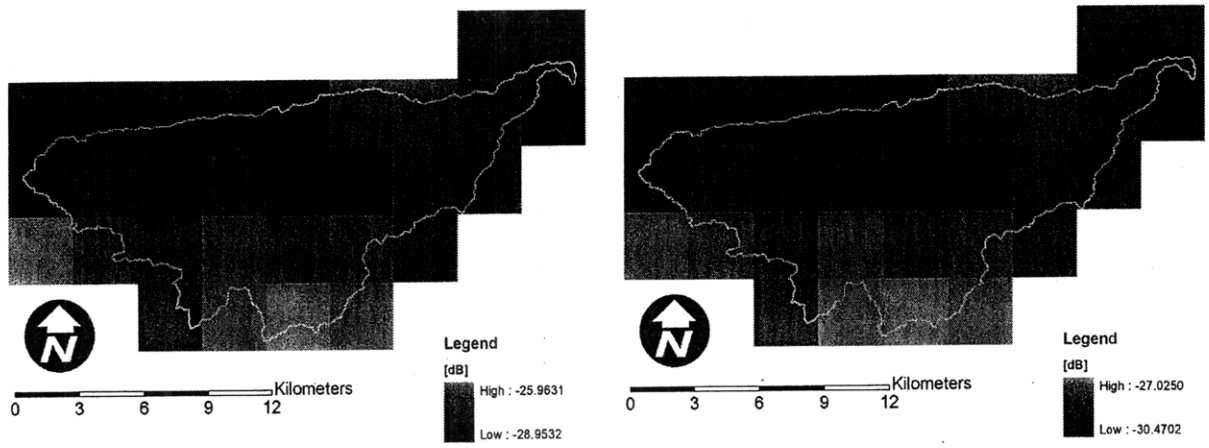
(a) (b)

Figure D-31: Average observations across the four candidate observations in the (a) horizontally- and (b) vertically-copolarized states at the fourth analysis (216 hr).



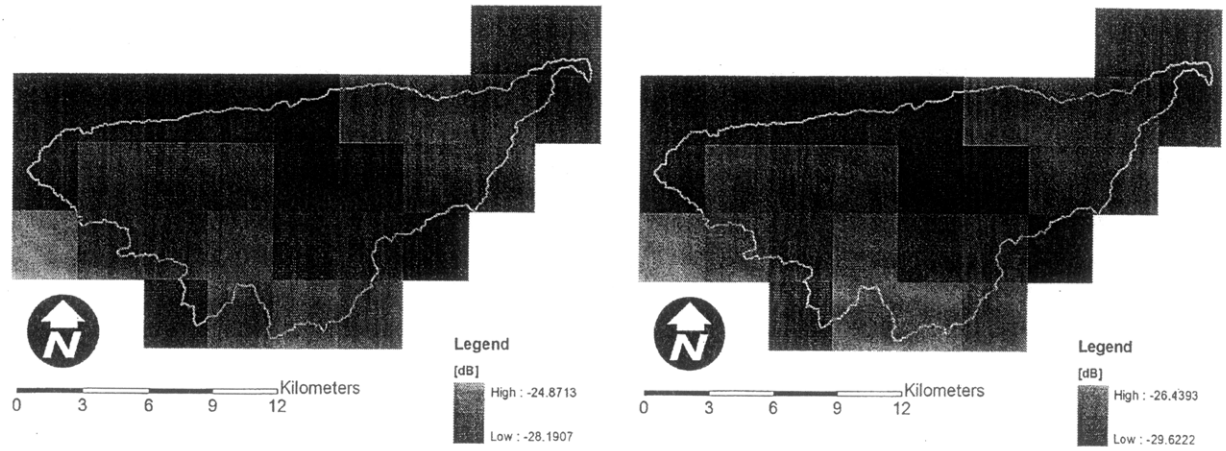
(a) (b)

Figure D-32: Average observations across the four candidate observations in the (a) horizontally- and (b) vertically-copolarized states at the fifth analysis (360 hr).



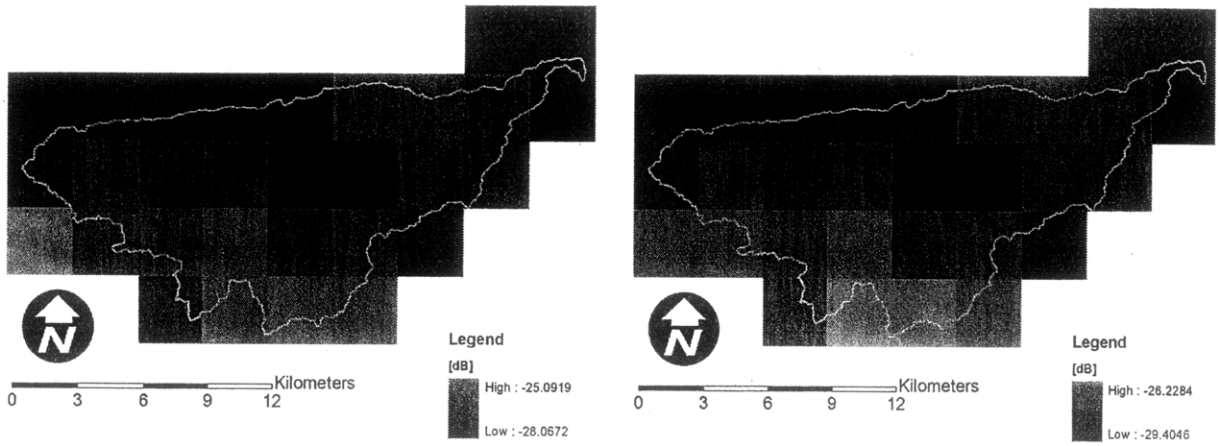
(a) (b)

Figure D-33: Average observations across the four candidate observations in the (a) horizontally- and (b) vertically-copolarized states at the sixth analysis (432 hr).



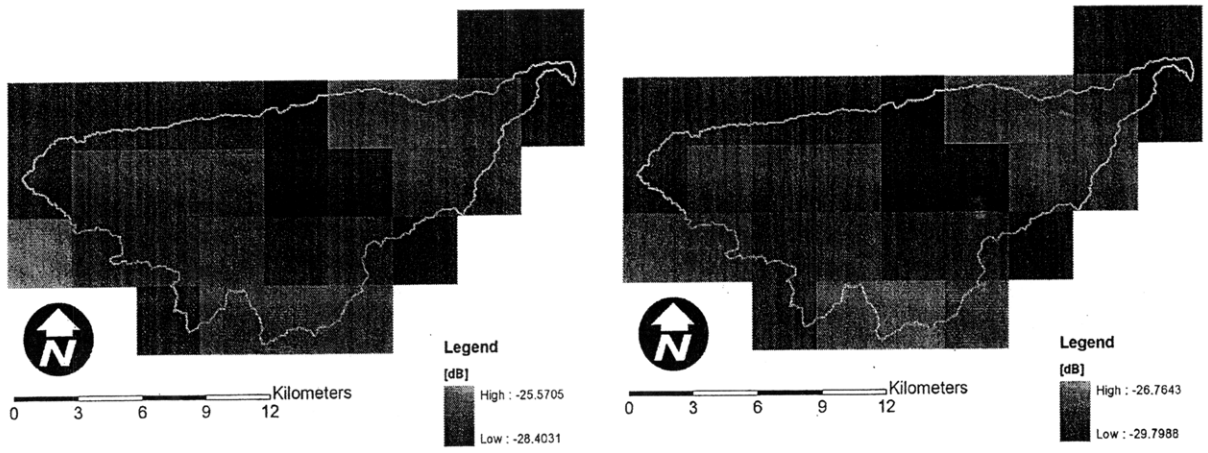
(a) (b)

Figure D-34: Average observations across the four candidate observations in the (a) horizontally- and (b) vertically-copolarized states at the seventh analysis (504 hr).



(a) (b)

Figure D-35: Average observations across the four candidate observations in the (a) horizontally- and (b) vertically-copolarized states at the eight analysis (576 hr).



(a) (b)

Figure D-36: Average observations across the four candidate observations in the (a) horizontally- and (b) vertically-copolarized states at the ninth and final analysis (648 hr).

D.2.2 Analysis results

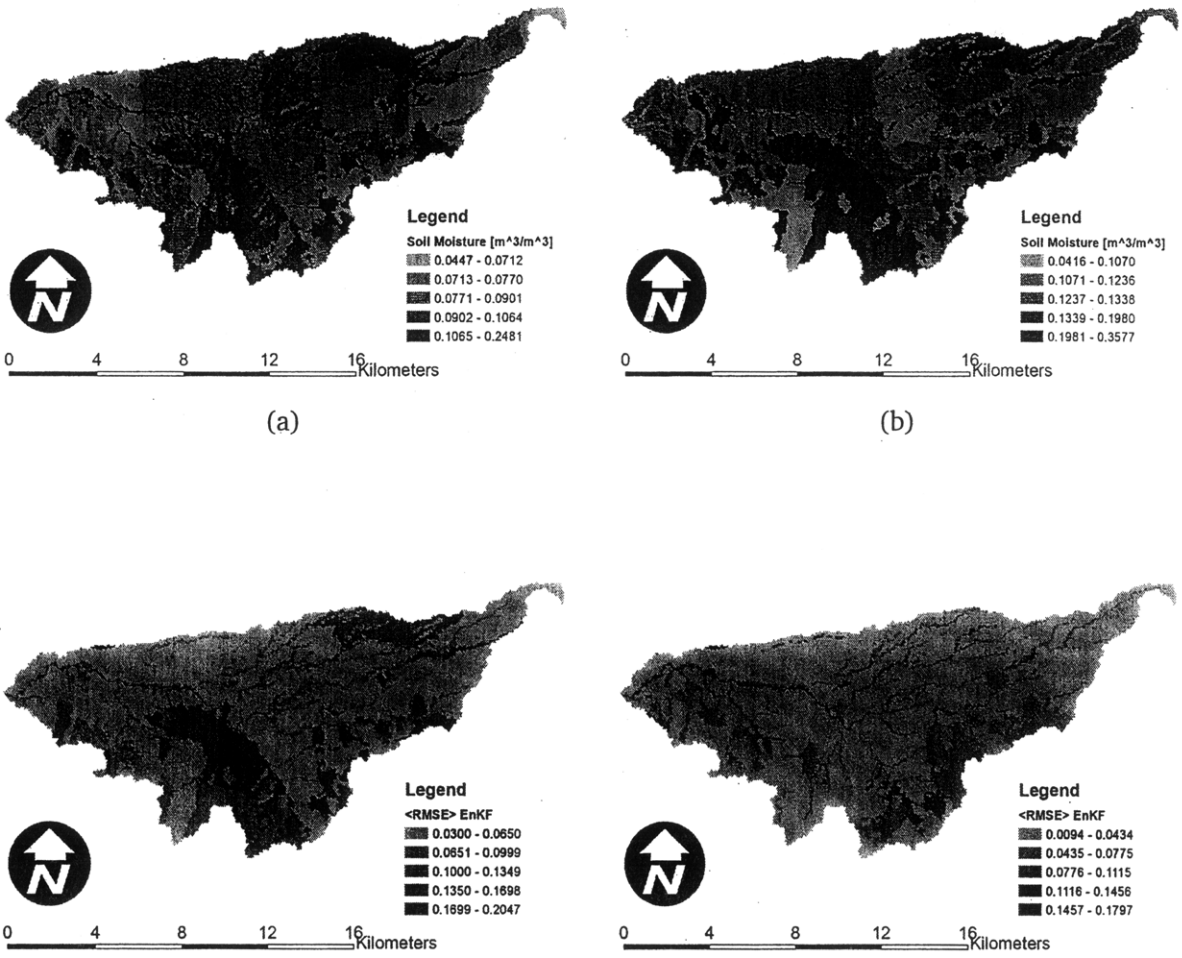


Figure D-37: At the first analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.

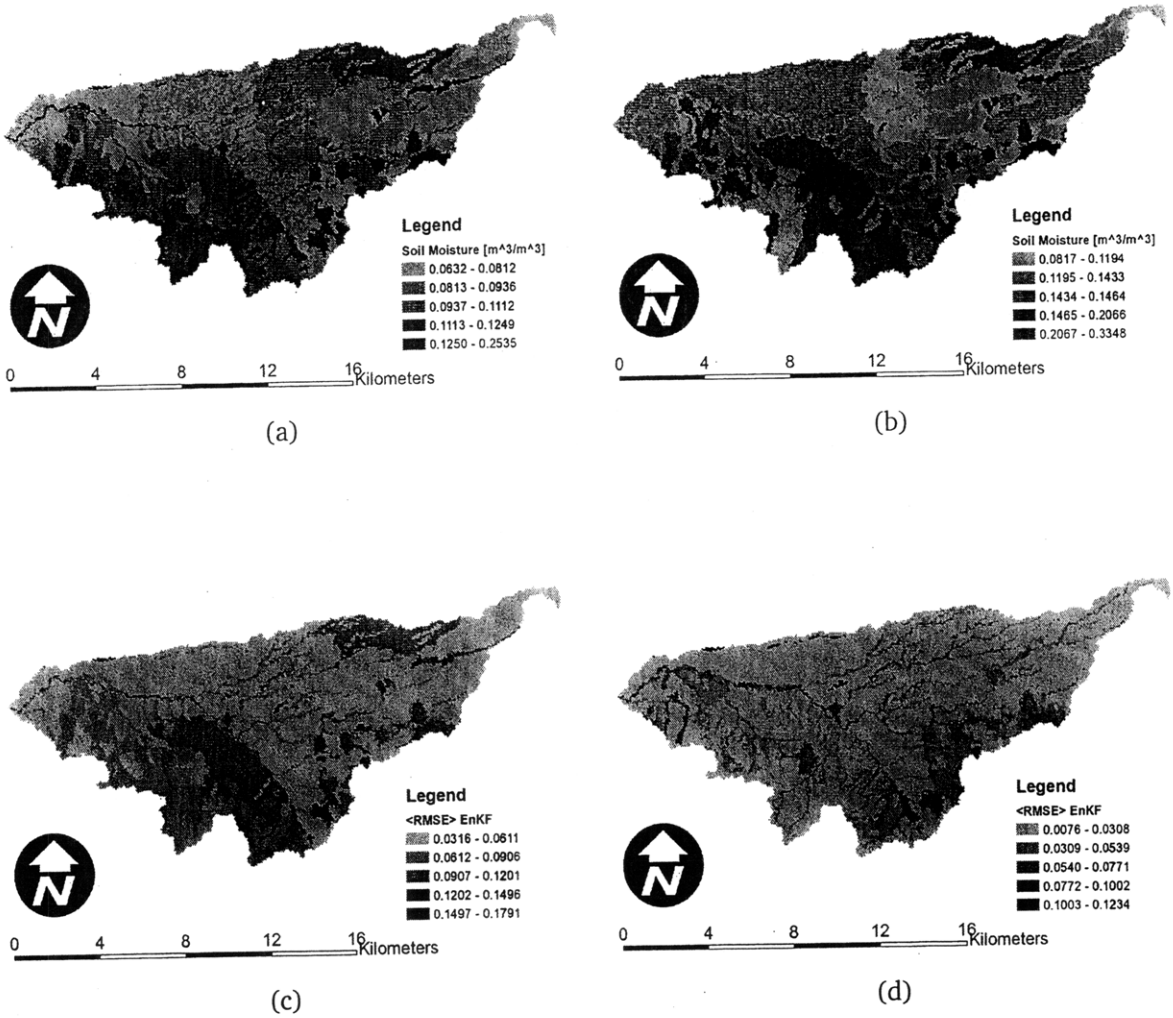
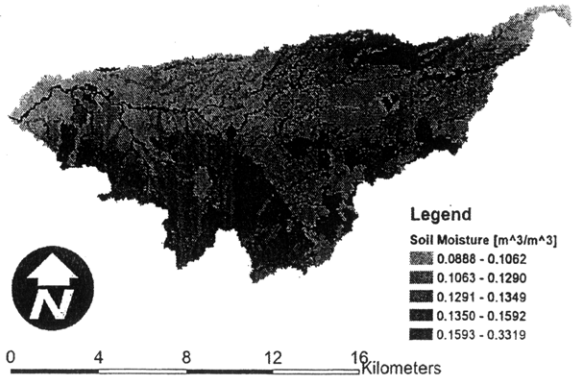
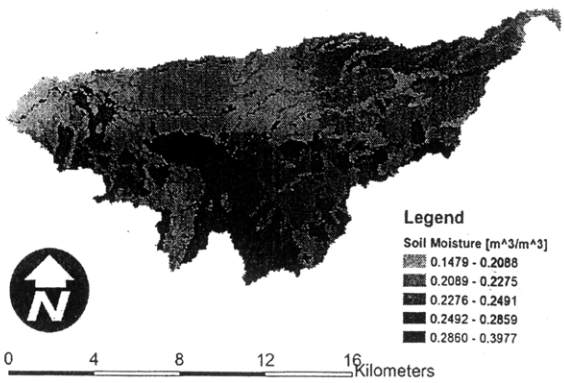


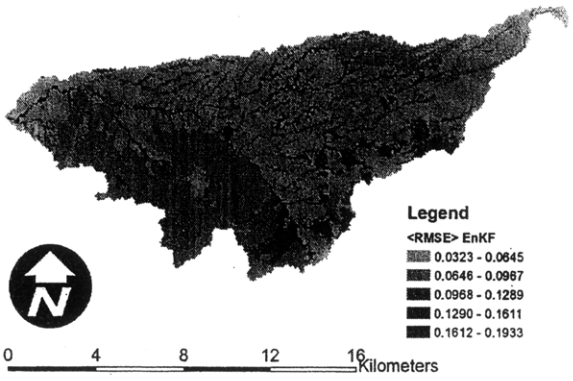
Figure D-38: At the second analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.



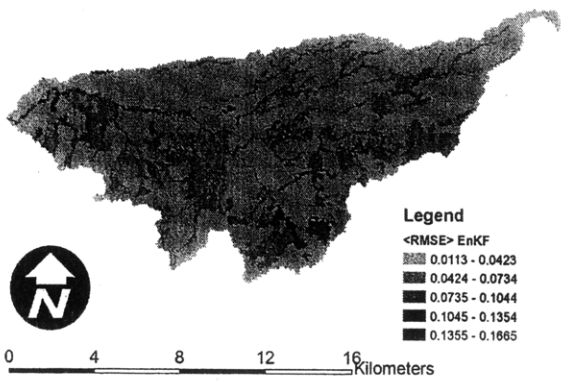
(a)



(b)



(c)



(d)

Figure D-39: At the third analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.

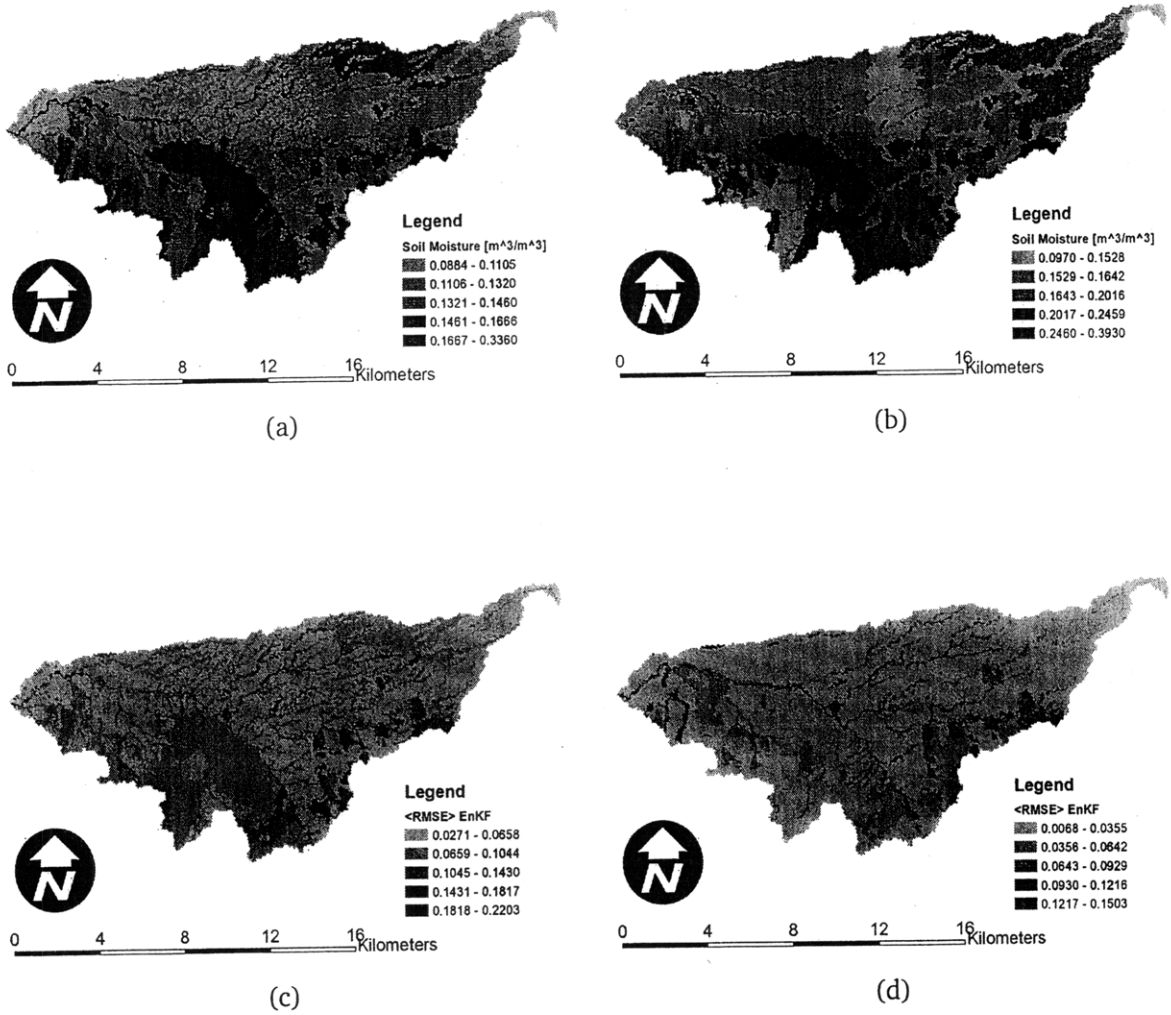
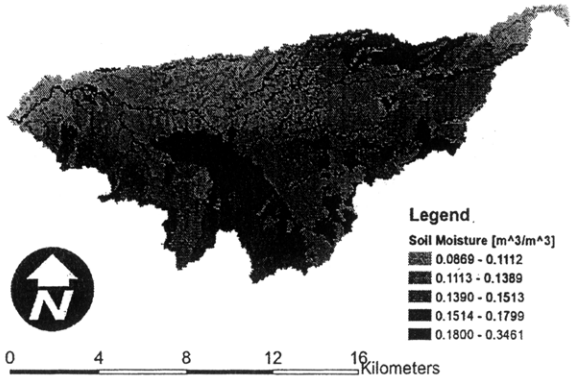
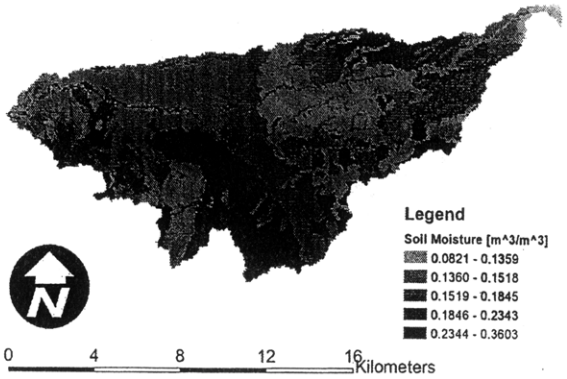


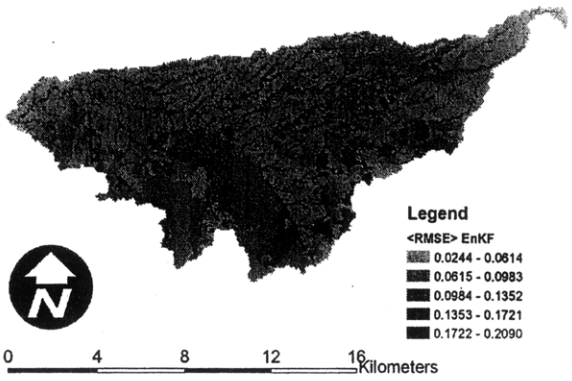
Figure D-40: At the fourth analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.



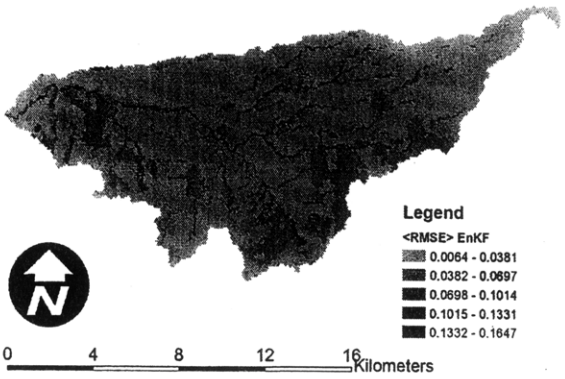
(a)



(b)



(c)



(d)

Figure D-41: At the fifth analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.

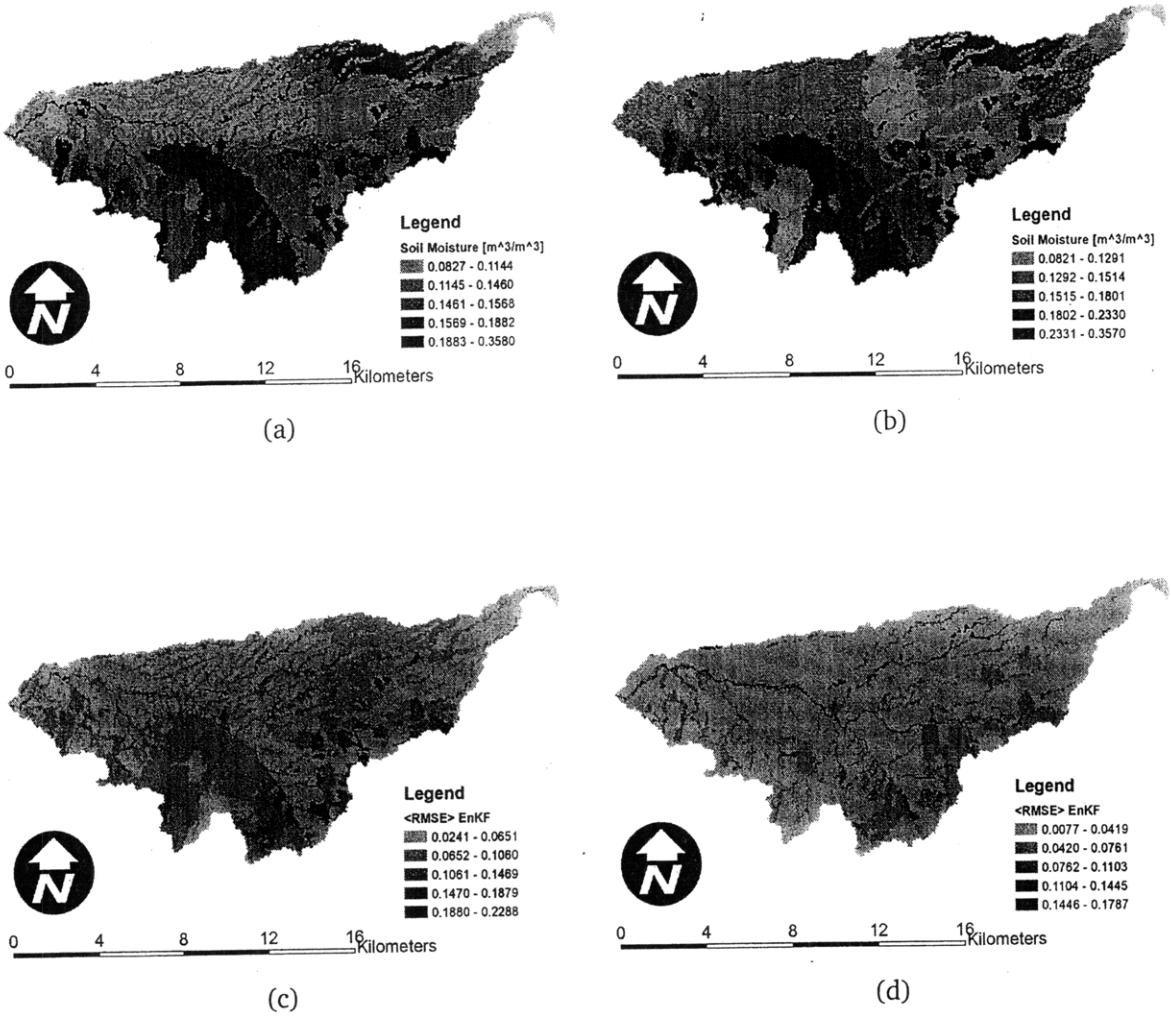
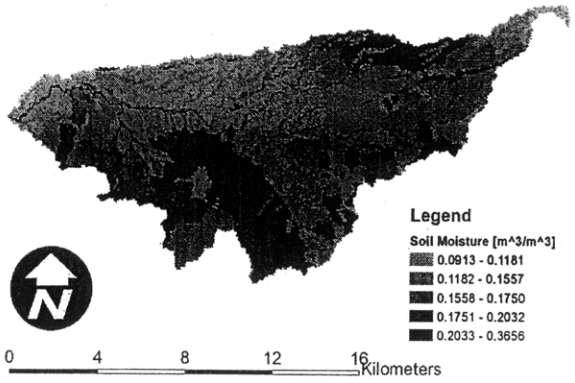
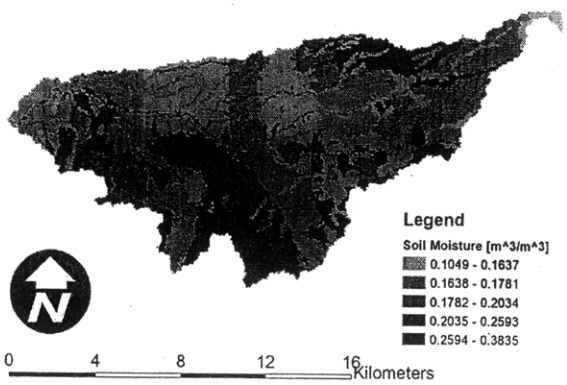


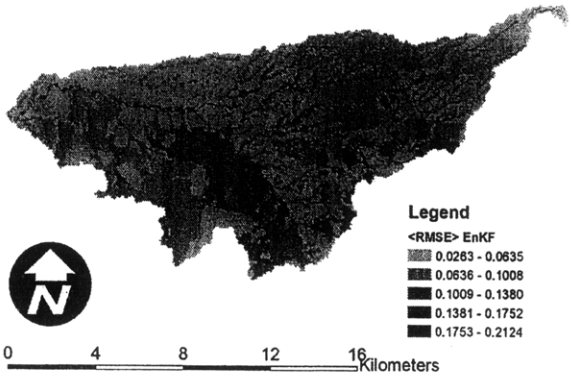
Figure D-42: At the sixth analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.



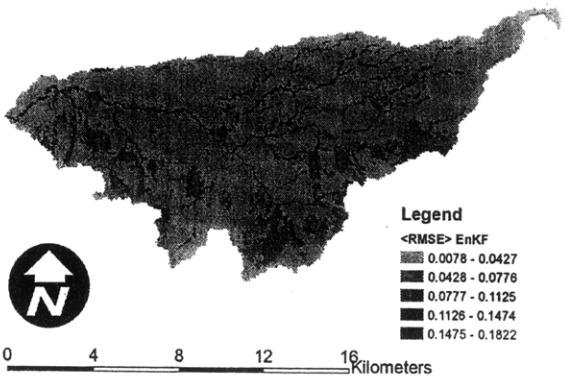
(a)



(b)

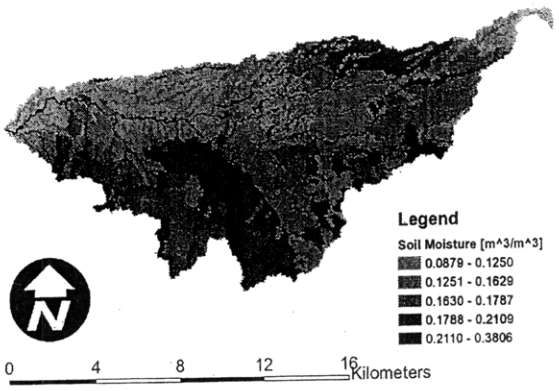


(c)

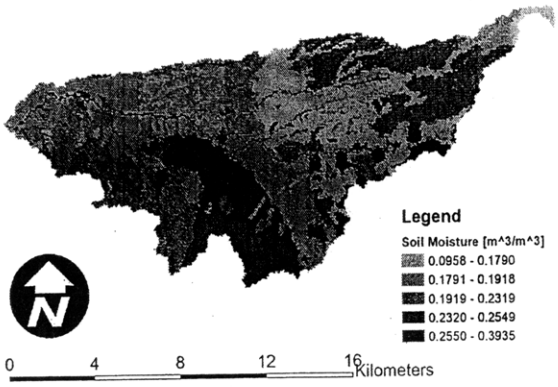


(d)

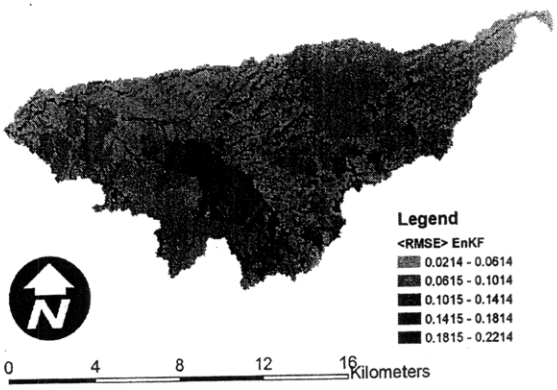
Figure D-43: At the seventh analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.



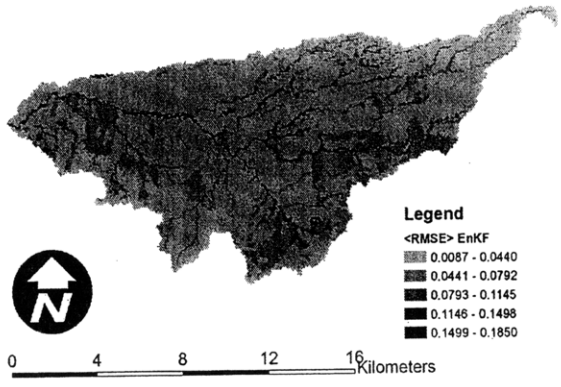
(a)



(b)



(c)



(d)

Figure D-44: At the eighth analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.

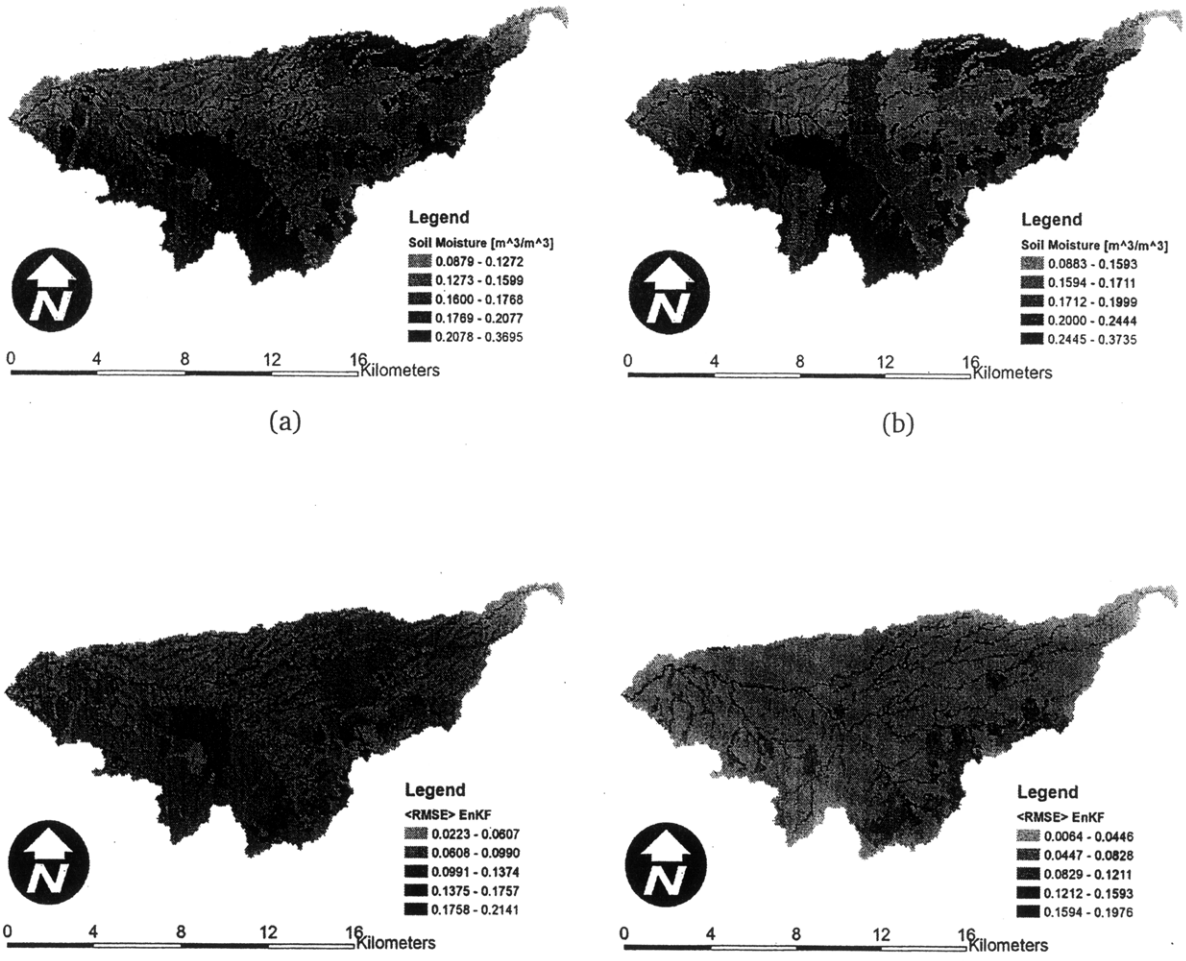
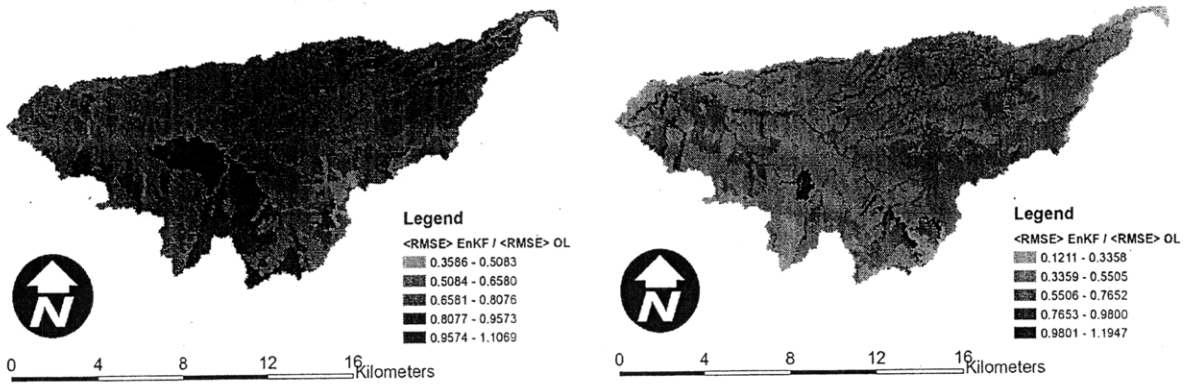


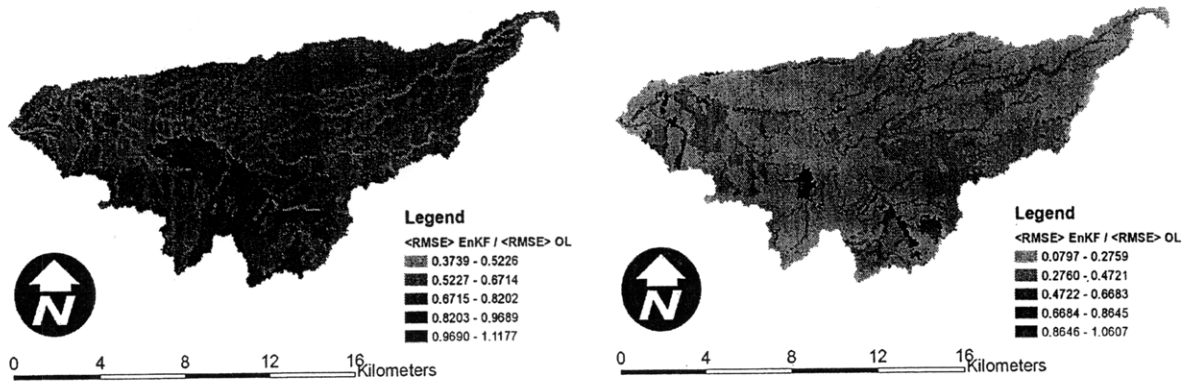
Figure D-45: At the ninth analysis, the average across the four sets of synthetic observations of the: (a) analysis ensemble mean of profile-integrated soil moisture, (b) analysis ensemble mean of near-surface soil moisture, (c) RMSE in the profile-integrated estimate of soil moisture, and (d) RMSE in the near-surface estimate of soil moisture.

D.2.3 EnKF average RMSE relative to OL average RMSE



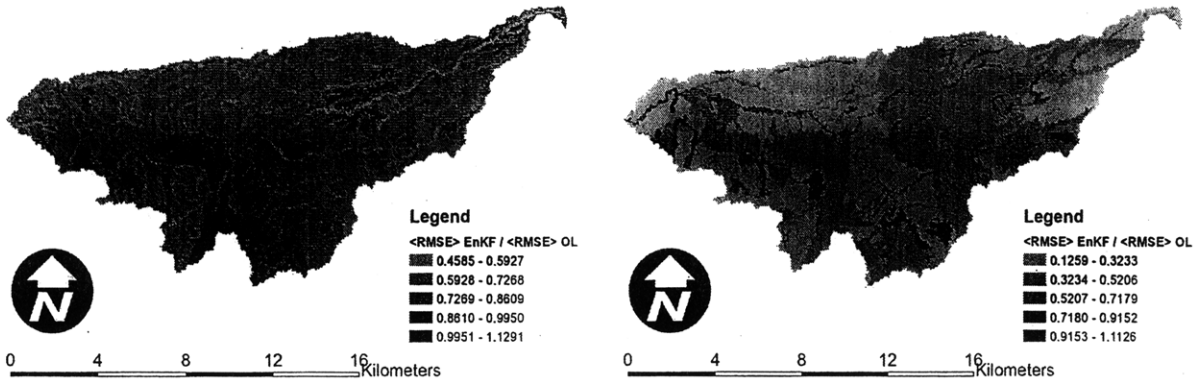
(a) (b)

Figure D-46: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the first analysis.



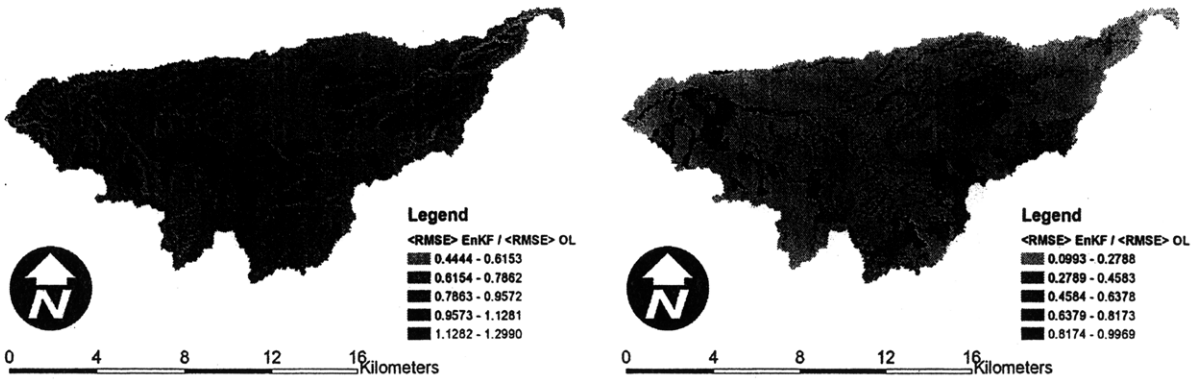
(a) (b)

Figure D-47: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the second analysis.



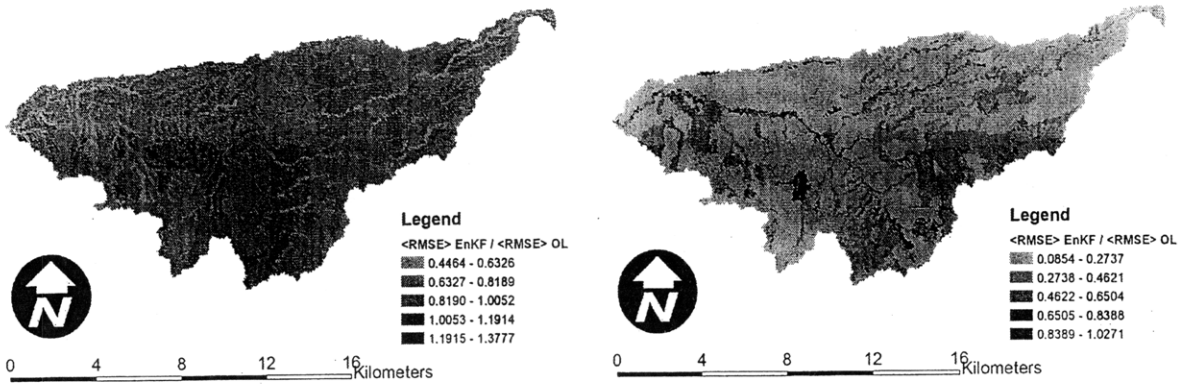
(a) (b)

Figure D-48: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the third analysis.



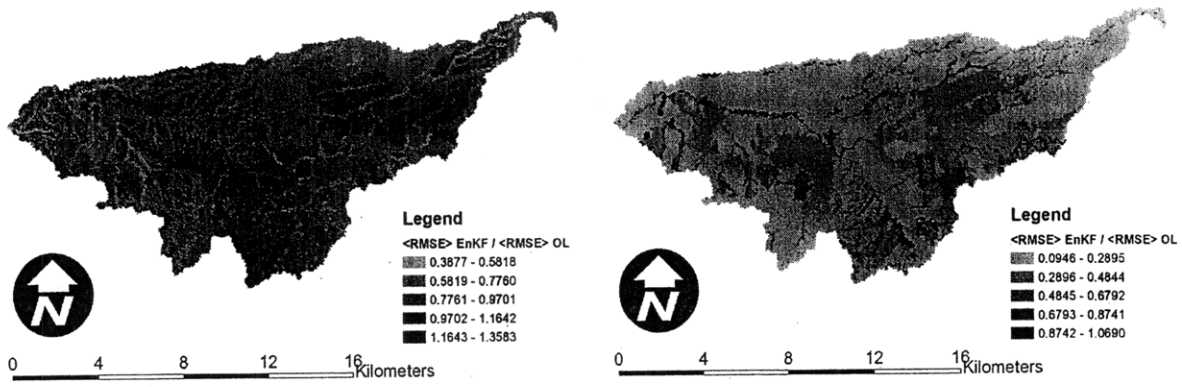
(a) (b)

Figure D-49: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the fourth analysis.



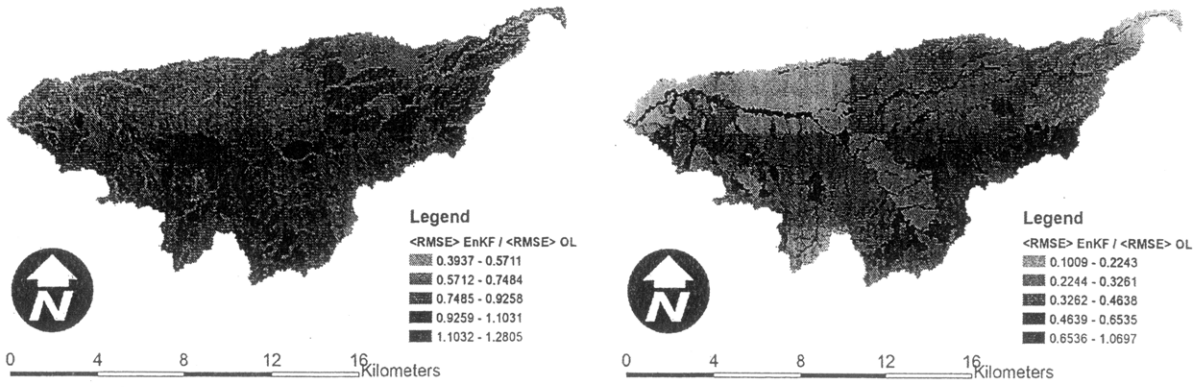
(a) (b)

Figure D-50: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the fifth analysis.



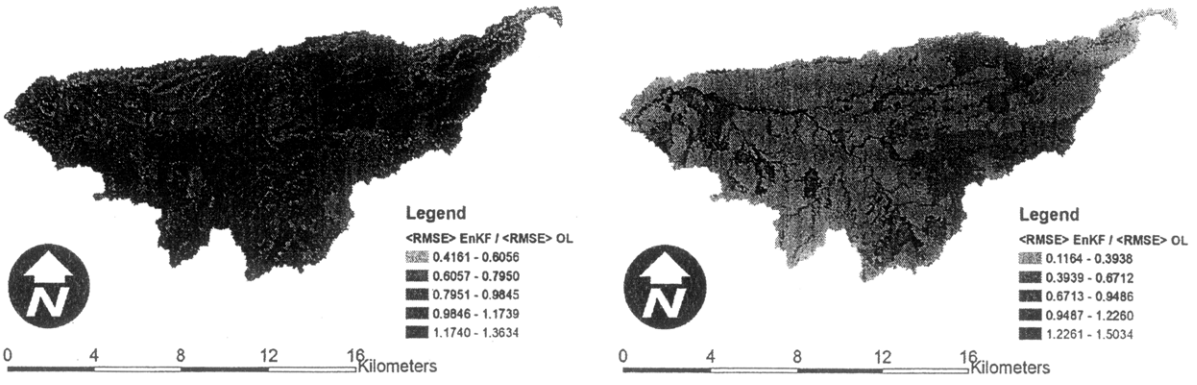
(a) (b)

Figure D-51: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the sixth analysis.



(a) (b)

Figure D-52: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the seventh analysis.



(a) (b)

Figure D-53: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the eighth analysis.

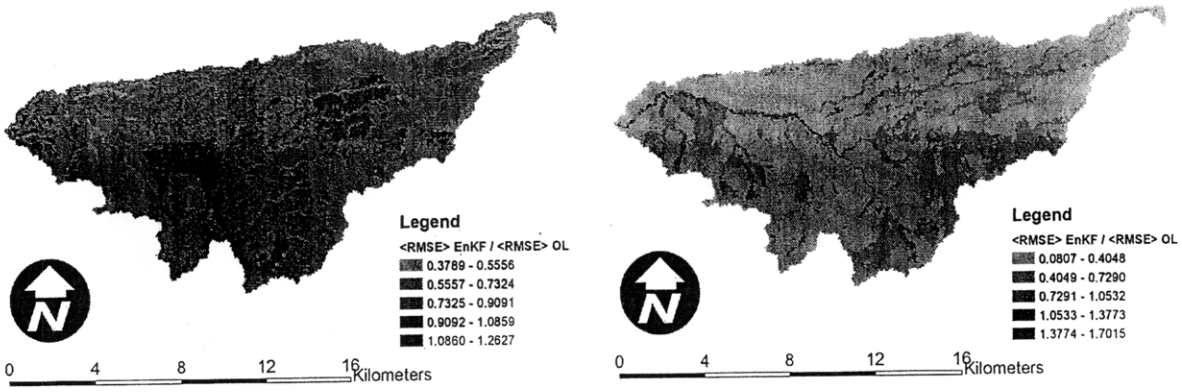
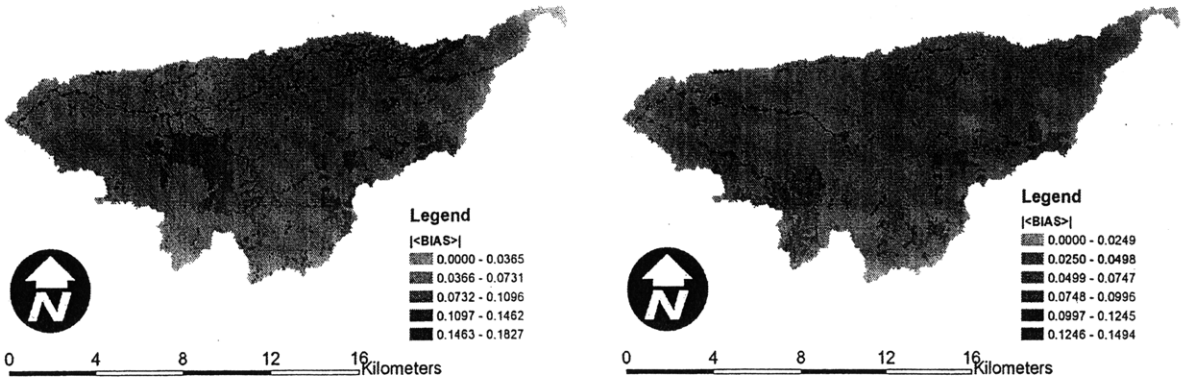


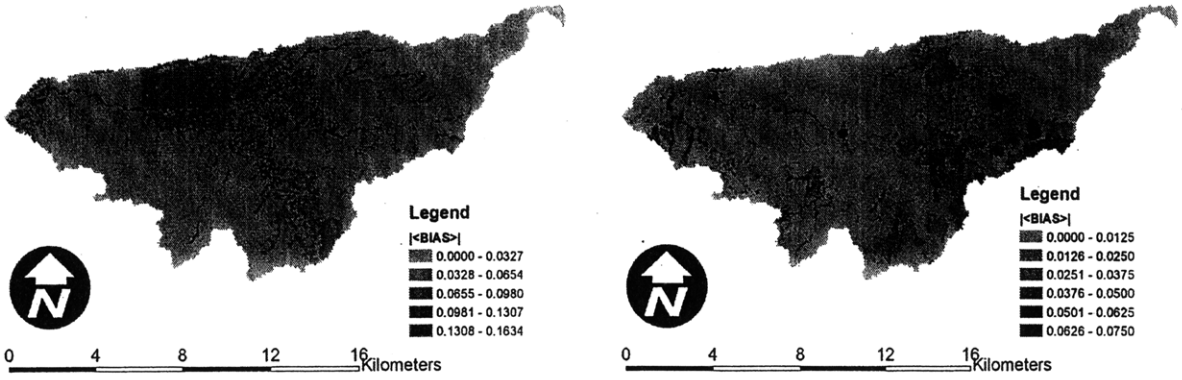
Figure D-54: The local ratio of average RMSE from the EnKF estimate to the OL estimate for the: (a) profile-integrated soil moisture and (b) near-surface estimates at the ninth analysis.

D.3 Bias in EnKF estimates



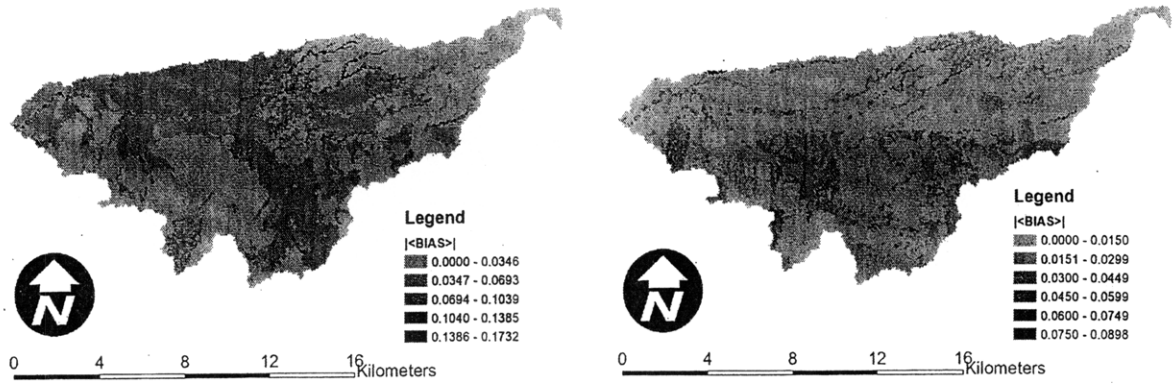
(a) (b)

Figure D-55: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the first analysis.



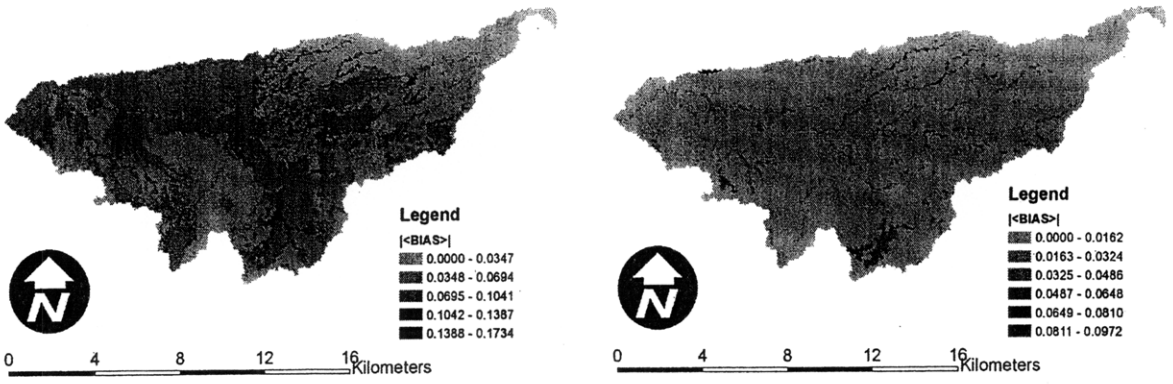
(a) (b)

Figure D-56: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the second analysis.



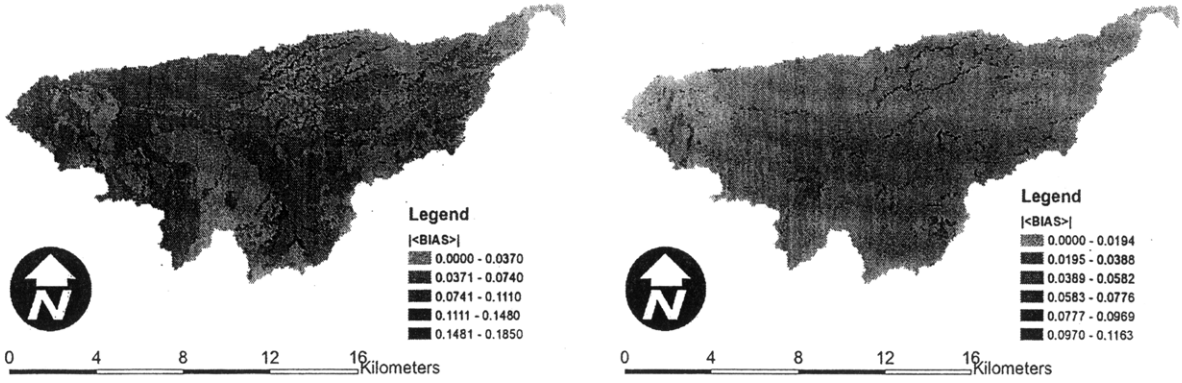
(a) (b)

Figure D-57: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the third analysis.



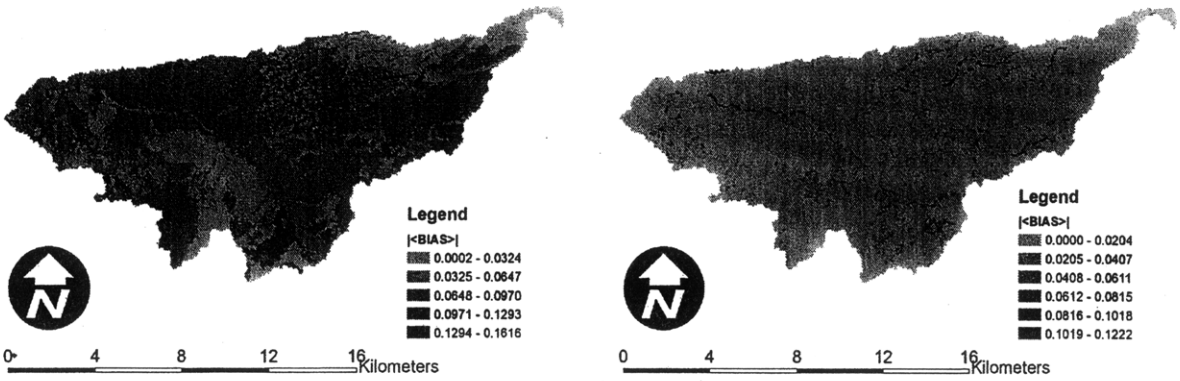
(a) (b)

Figure D-58: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the fourth analysis.



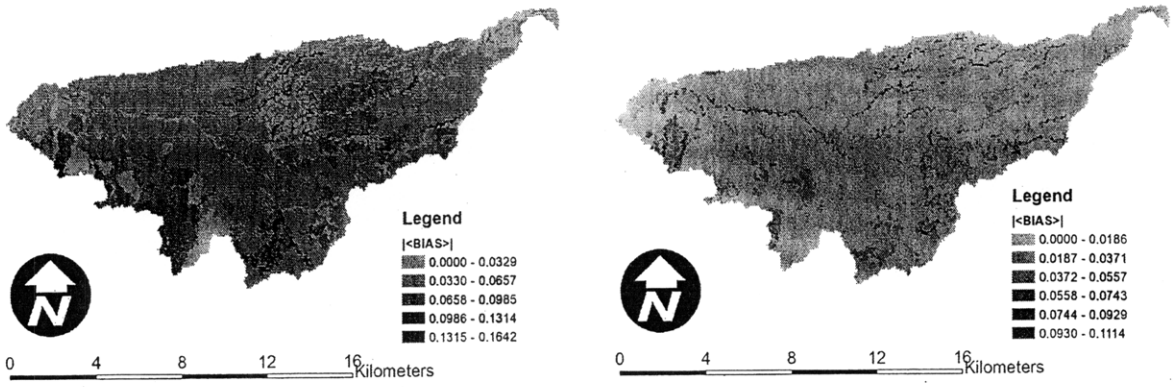
(a) (b)

Figure D-59: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the fifth analysis.



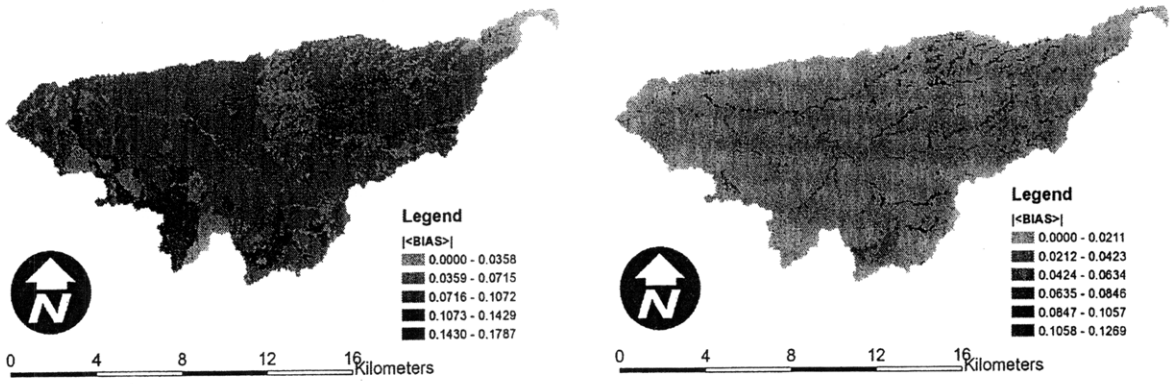
(a) (b)

Figure D-60: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the sixth analysis.



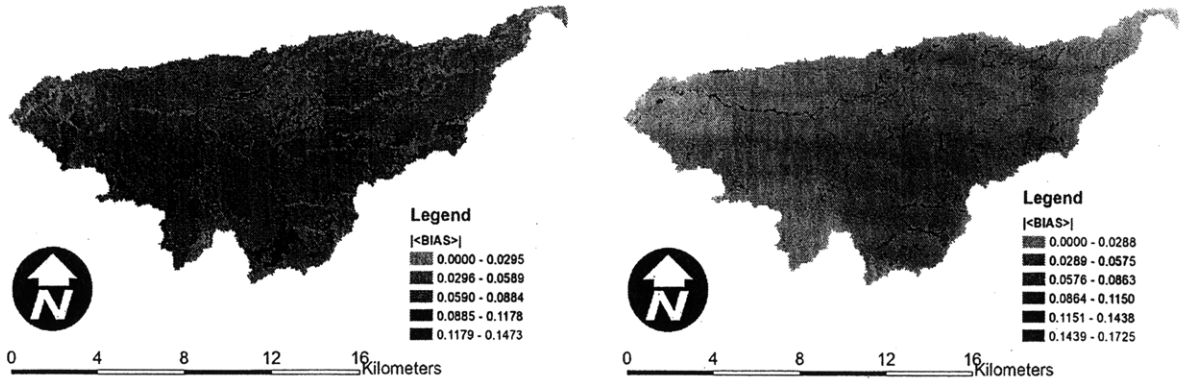
(a) (b)

Figure D-61: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the seventh analysis.



(a) (b)

Figure D-62: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the eighth analysis.



(a) (b)

Figure D-63: The local average bias in the EnKF estimate for the: (a) profile-integrated, and (b) near-surface soil moisture estimate at the ninth analysis.

D.3.1 Pixel-scale diagnostic results

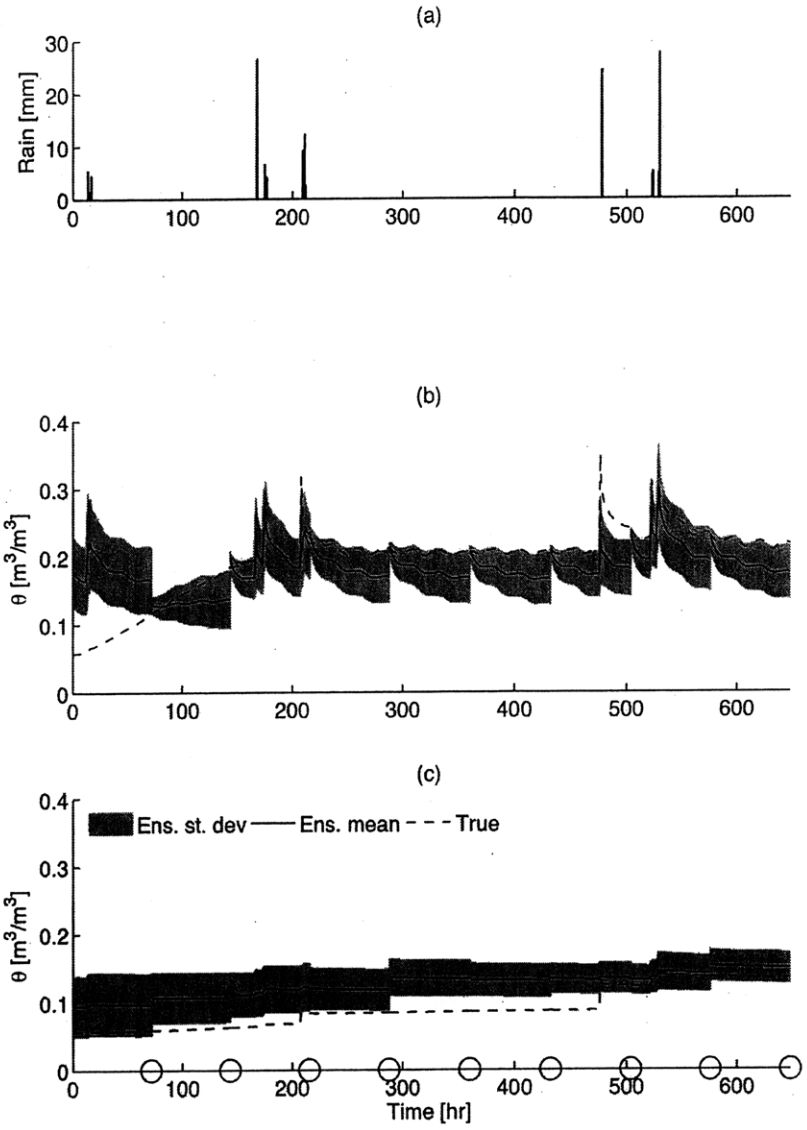


Figure D-64: Example of the pixel-scale behavior near the Lucky Hills experimental site during first of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

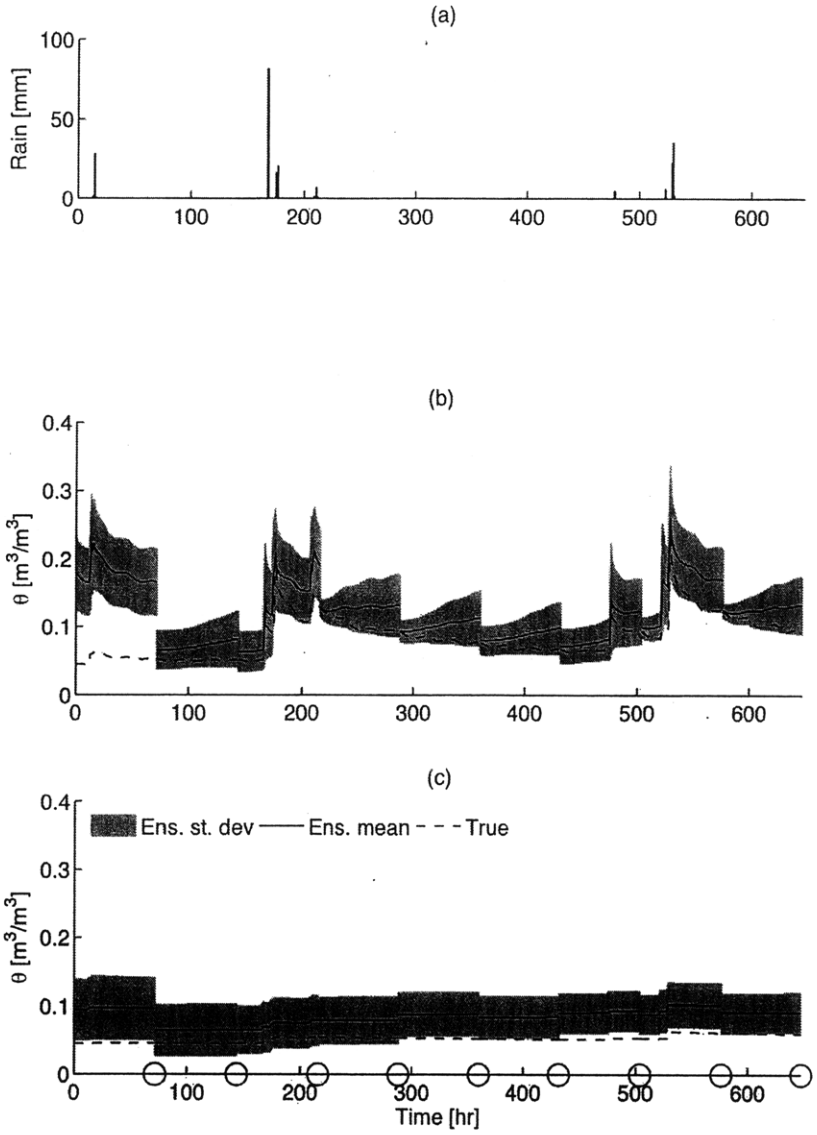


Figure D-65: Example of the pixel-scale behavior near the Lucky Hills experimental site during the first of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

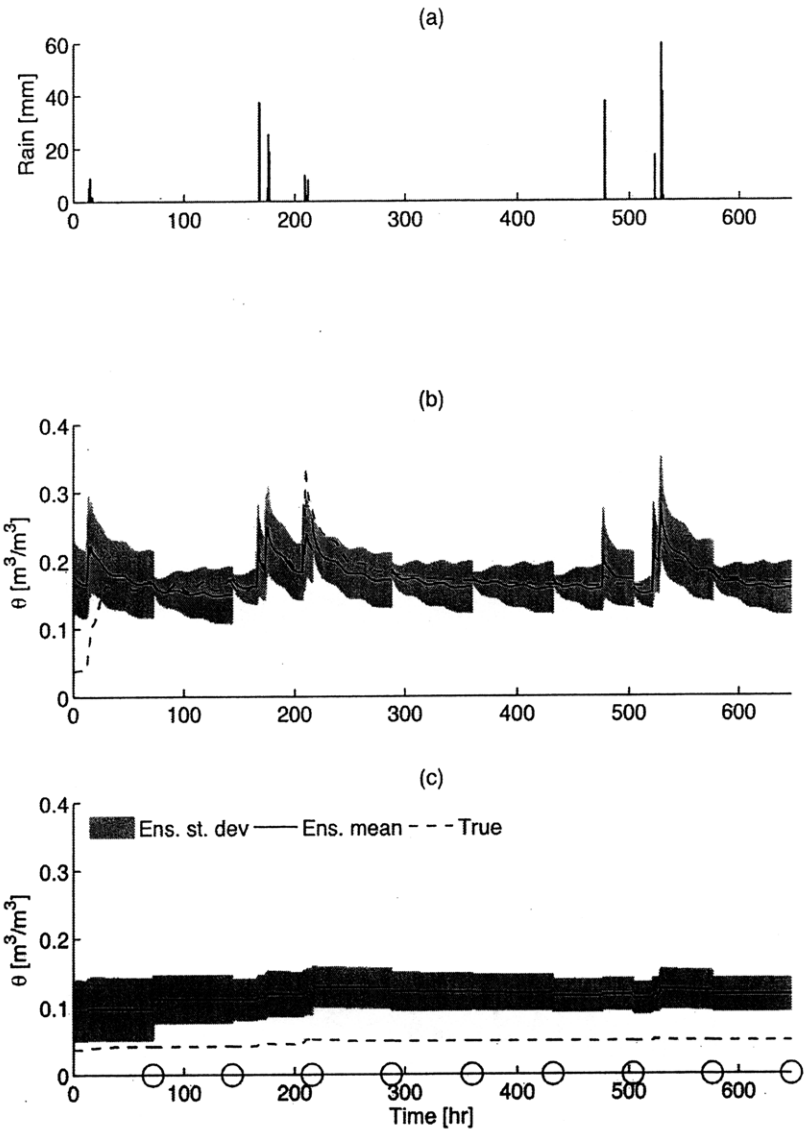


Figure D-66: Example of the pixel-scale behavior near the Lucky Hills experimental site during the second of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

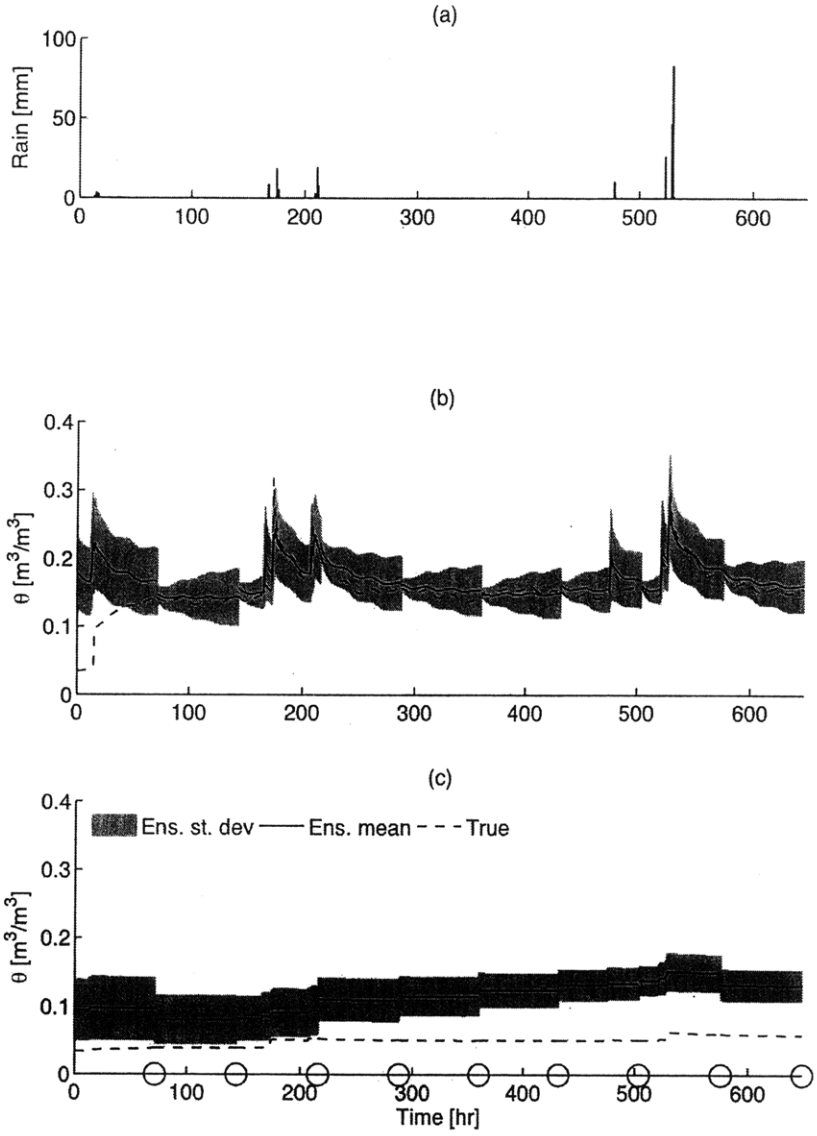


Figure D-67: Example of the pixel-scale behavior near the Lucky Hills experimental site during the fourth of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

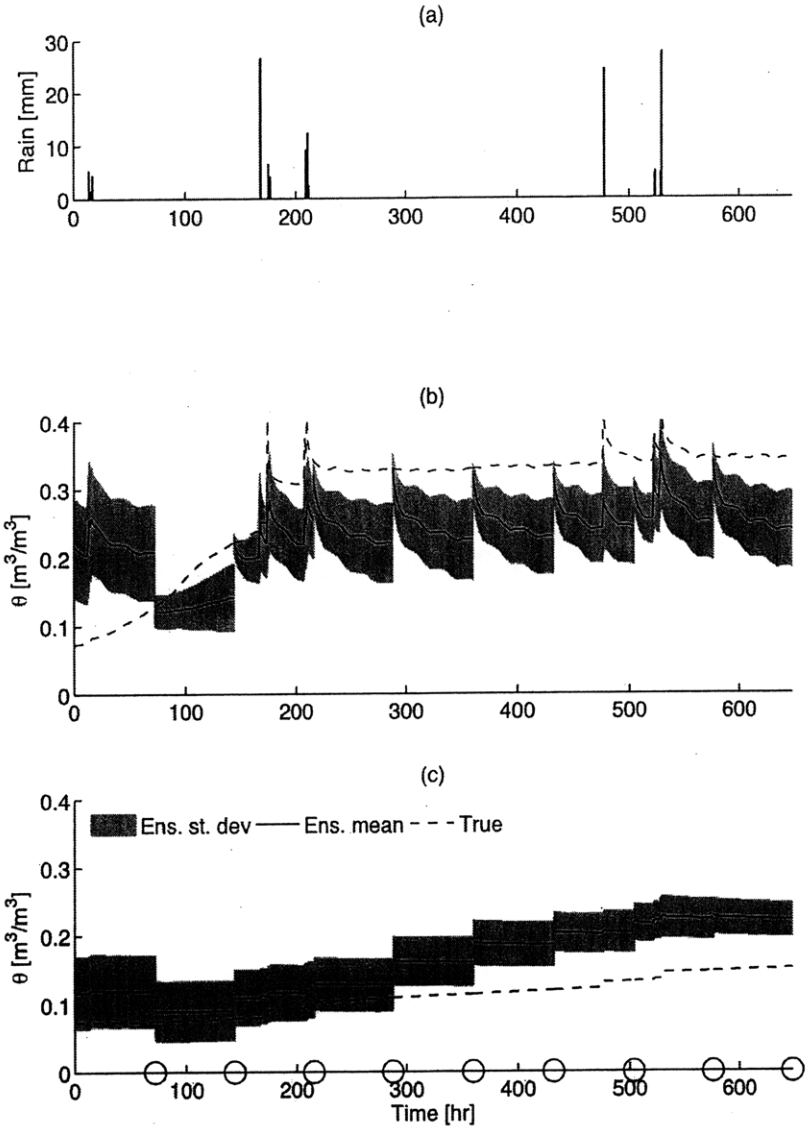


Figure D-68: Example of the pixel-scale behavior near the Kendall experimental site during the first of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

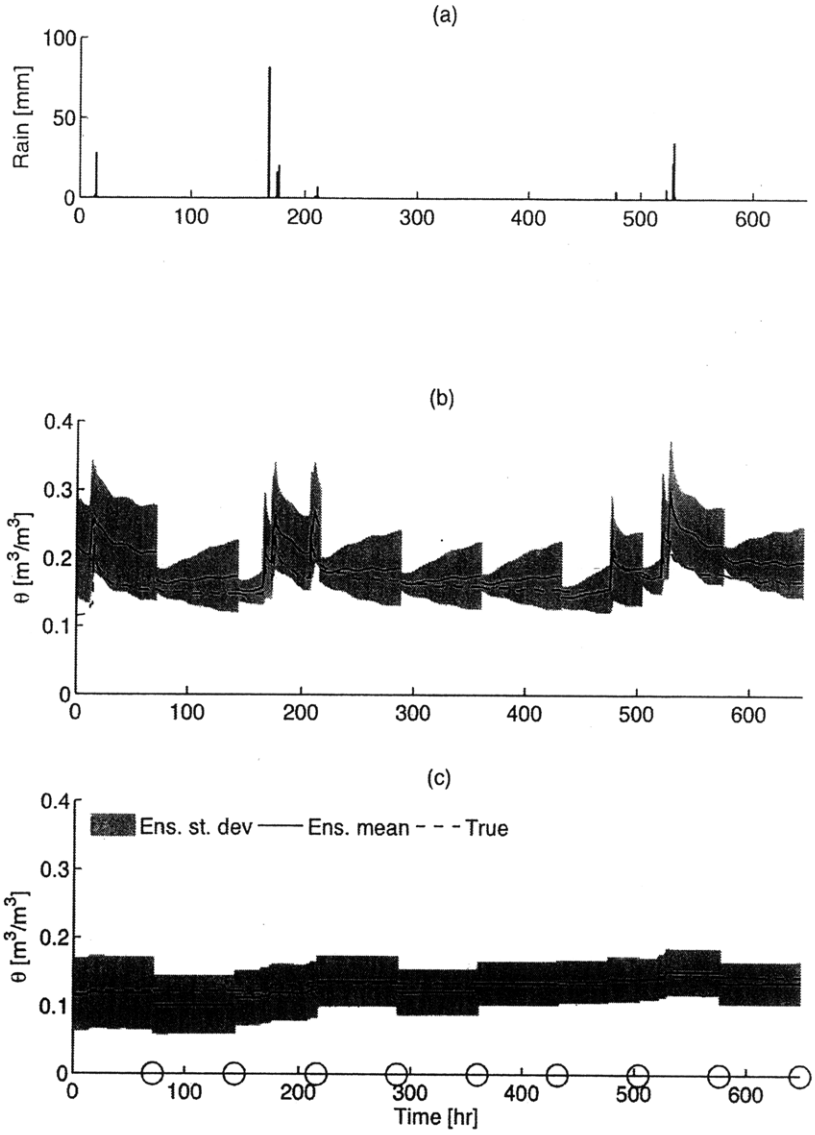


Figure D-69: Example of the pixel-scale behavior near the Kendall experimental site during the second of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

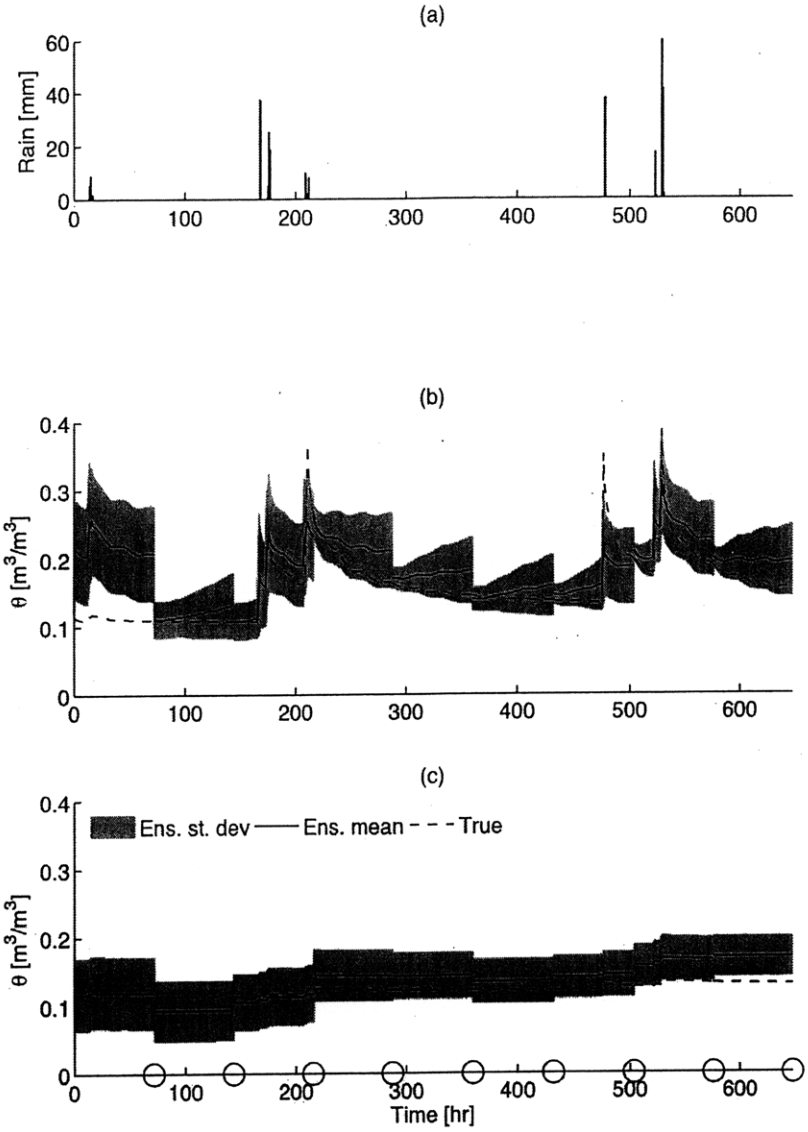


Figure D-70: Example of the pixel-scale behavior near the Kendall experimental site during the third of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

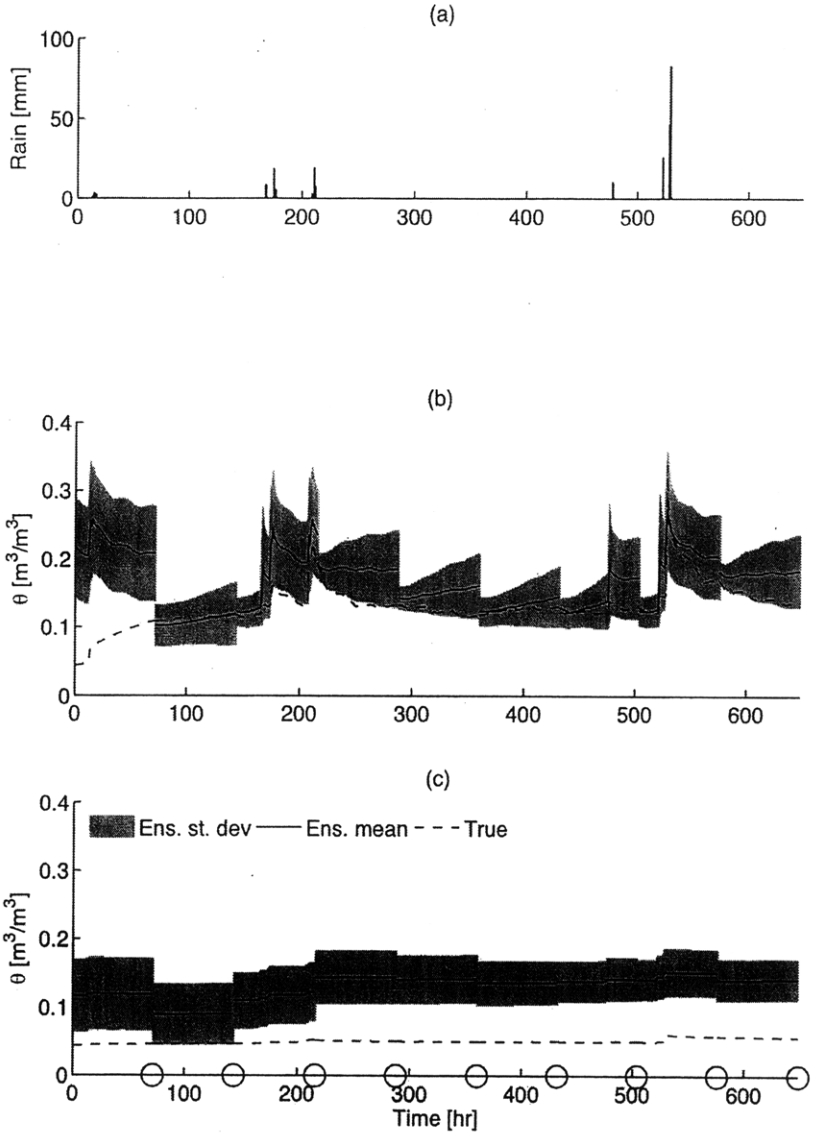


Figure D-71: Example of the pixel-scale behavior near the Kendall experimental site during the fourth of four EnKF experiments, showing the (a) time series of spatially-averaged rainfall during the simulation, (b) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) near-surface soil moisture, and (c) the ensemble mean (solid black line), area within one standard deviation of either side of the mean (grey area), and true (dashed black line) profile-integrated soil moisture.

REFERENCES

ABOITIZ, M., J. W. LABADIE and D. F. HEERMANN. Stochastic soil moisture estimation and forecasting for irrigated fields. *Water Resour. Res.*, 22:pages 180–190 [1986].

AHUJA, L. R., D. K. CASSEL, R. R. BRUCE and B. B. BARNES. Evaluation of spatial distribution of hydraulic conductivity using effective porosity data. *Soil Sci.*, 148:pages 404–411 [1989].

ALTESE, E., O. BOLOGNANI, M. MANCINI and P. A. TROCH. Retrieving soil moisture over bare soil from ERS 1 synthetic aperture radar data: Sensitivity analysis based on a theoretical surface scattering model and field data. *Water Resour. Res.*, 32(3):pages 653–661 [1996].

ANDERSON, J. L. An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, 129:pages 2884–2903 [2001].

ARORA, V. K. and G. J. BOER. A parameterization of leaf phenology for the terrestrial ecosystem component of climate models. *Global Change Biol.*, 11(1):pages 39–59 [2005].

ARRHENIUS, S. On the influence of carbonic acid in the aur upon the temperature of the ground. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science (fifth series)*, 41:pages 237–275 [1896].

ARULAMPALAM, M. S., S. MASKELL, N. GORDON and T. CLAPP. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Processing*, 50:pages 174–188 [2002].

BANIN, A. and A. AMIEL. A correlative study of the chemical and physical properties of a group of natural soils of Israel. *Geoderma*, 3:pages 185–198 [1969].

BISHOP, C. H., B. ETHERTON and S. J. MAJUMDAR. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, 129:pages 420–436 [2001].

BONAN, G. B. A biophysical surface energy budget analysis of soil temperature in the boreal forests of interior Alaska. *Water Resour. Res.*, 27:pages 767–781 [1991].

—. Land-atmosphere interactions for climate system models—Coupling biophysical, biogeochemical, and ecosystem dynamical processes. *Remote Sens. Environ.*, 51(1):pages 57–73 [1995].

—. A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user's guide. NCAR Tech. Note NCAR/TN-417, Natl. Cent. for Atmos. Res., Boulder, Colo [1996].

BONAN, G. B., S. LEVIS, L. KERGOAT and K. W. OLESON. Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochem. Cycles*, 16(2) [2002a]. doi:10.1029/2000GB001360.

BONAN, G. B., K. W. OLESON, M. VERTENSTEIN, S. LEVIS, X. B. ZENG, Y. J. DAI, R. E. DICKINSON and Z. L. YANG. The land surface climatology of the community land model coupled to the NCAR community climate model. *J. Clim.*, 15:pages 3123–3149 [2002b].

BORGA, M., S. D. ESPOSTI and D. NORBIATO. Influence of errors in radar rainfall estimates on hydrological modeling prediction uncertainty. *Water Resour. Res.*, 42(W08409) [2006]. doi:10.1029/2005WR004559.

BOYLE, D. P., H. V. GUPTA and S. SOROOSHIAN. Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods. *Water Resour. Res.*, 36:pages 3663–3674 [2000].

BRAS, R. L. and J. R. CORDOVA. Intraseasonal water allocation in deficit irrigation. *Water Resour. Res.*, 17(4):pages 866–874 [1981].

BROCK, F. V., K. C. CRAWFORD, R. L. ELLIOTT, G. W. CUPERUS, S. J. STADLER, H. L. JOCHNSON and M. D. EILTS. The Oklahoma Mesonet: A Technical Overview. *J. Atmos. Oceanic Tech.*, 12(1):pages 5–19 [1995].

BROOKS, R. H. and A. T. COREY. Hydraulic properties of porous media. Hydrol. Pap. 3, Colo. State Univ. [1964].

BURGERS, G., P. J. VAN LEEUWEN and G. EVENSEN. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, 126:pages 1719–1724 [1998].

BURNASH, R. J. C., R. L. FERAL and R. A. MCGUIRE. A generalized streamflow simulation system: Conceptual modeling for digital computers. Tech. rep., Natl. Weather Serv. and the State of Calif., Dept. Water Resour., Sacramento, CA [1973].

CAHILL, A. T. and M. B. PARLANGE. On water vapor transport in field soils. *Water Resour. Res.*, 34(4):pages 731–739 [1998].

CALVET, J.-C., J. NOILHAN and P. BESSEMOULIN. Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates: A feasibility study based on field measurements. *J. Appl. Meteor.*, 37:pages 371–386 [1998].

CASADEI, M., W. E. DIETRICH and N. L. MILLER. Testing a model for predicting the timing and location of shallow landslide initiation in soil-mantled landscapes. *Earth Surf. Processes Landforms*, 28:pages 925–950 [2003].

CELIA, M. A., E. T. BOULOUTAS and R. L. ZARBA. A general mass-conservative numerical solution for the unsaturated flow equation. *Water Resour. Res.*, 26:pages 1483–1496 [1990].

CHAUHAN, N. S., S. MILLER and P. ARDANUY. Spaceborne soil moisture estimation at high resolution: A microwave-optical/IR synergistic approach. *Intl. J. Remote Sens.*, 24(22):pages 4599–4622 [2003].

CHEN, F., K. MITCHELL, J. SCHAAKE, Y. XUE, H. PAN, V. KOREN, Y. DUAN, M. EK and A. BETTS. Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, 101(D3):pages 7251–7268 [1996a].

CHEN, F., 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. *J. Geophys. Res.*, 101:pages 7251–7268 [1996b].

CHEN, F., 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. *J. Atmos. Sci.*, 58:pages 3204–3223 [2001].

CHEN, J. L., C. R. WILSON, D. P. CHAMBERS, R. S. NEREM and B. D. TAPLEY. Seasonal global water mass budget and mean sea level variations. *Geophys. Res. Lett.*, 25(19):pages 3555–3558 [1998].

CHEN, K. S., S. K. YEN and W. P. HUANG. A simple model for retrieving bare soil moisture from radar-scattering coefficients. *Remote Sens. Environ.*, 54:pages 121–126 [1995].

CHOUDHURY, B. J., T. J. SCHMUGGE, A. CHANG and R. W. NEWTON. Effect of surface roughness on the microwave emission from soils. *J. Geophys. Res.*, 84:pages 5699–5706 [1979].

CLIMATE CHANGE SCIENCE PROGRAM AND SUBCOMMITTEE ON GLOBAL CHANGE RESEARCH. *Our Changing Planet: The FY 2003 U.S. Gloval Change Research Program and Climate Change Research Initiative.* Global Change Research Information Office, Palisades, NY [2003].

CROW, W. T. A novel method for quantifying value in spaceborne soil moisture retrievals. *J. Hydrometeor.*, 8(1):pages 56–66 [2007].

CROW, W. T. and J. D. BOLTEN. Estimating precipitation errors using spaceborne surface soil moisture retrievals. *Geophys. Res. Lett.*, 34(L08403) [2007]. doi:10.1029/2007GL029450.

CROW, W. T., 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. ZHAN. An observing system simulation experiment for Hydros radiometer-only soil moisture products. *IEEE Trans. Geosci. Remote Sens.*, 43(6):pages 1289–1303 [2005].

CROW, W. T., M. DRUSCH and E. F. WOOD. An observation system simulation experiment for the impact of land surface heterogeneity on AMSR-E soil moisture retrieval. *IEEE Trans. Geosci. Remote Sens.*, 38(8):pages 1622–1631 [2001].

CROW, W. T. and E. F. WOOD. The value of coarse-scale soil moisture observations for regional surface energy balance modeling. *J. Hydrometeor.*, 3:pages 467–482 [2002].

—. 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. *Adv. Water Resour.*, 26(2):pages 137–149 [2033]. doi:10.1016/S0309-1708(02)00088-X.

CROW, W. T. and X. ZHAN. Continental-scale evaluation of remotely sensed soil moisture products. *IEEE Trans. Geosci. Remote Sens.*, 4(3):pages 451–455 [2007].

CURTIS, D. C. and P. S. EAGLESON. Constrained stochastic climate simulation. Tech. Rep. 274, Massachusetts Institute of Technology, Dept. Civil Eng., Ralph M. Parsons Lab, Cambridga, MA, USA [1982].

DAI, F. C., C. F. LEE and Y. Y. NGAI. Landslide risk assessment and management: an overview. *Engineering Geol.*, 64:pages 65–87 [2002].

VAN DAM, J. C. and R. A. FEDDES. Numerical simulation of infiltration, evaporation and shallow groundwater levels with the Richards equation. *J. Hydrol.*, 233:pages 72–85 [2000].

DAS, N. N. and B. P. MOHANTY. Root zone soil moisture assessment using remote sensing and vadose zone modeling. *Vaodse Zone J*, 5:pages 296–307 [2006].

DAVENPORT, I. J., J. FERNANDEZ-GALVEZ and R. J. GURNEY. A sensitivity analysis of soil moisture retrieval from the tau-omega microwave emission model. *IEEE Trans. Geosci. Remote Sens.*, 43(6):pages 1304–1316 [2005].

DEBELE, B., R. SRINIVASAN and J. Y. PARLANGE. Accuracy evaluation of weather data generation and disaggregation methods at finer timescales. *Adv. Water Resour.*, 30:pages 1286–1300 [2007]. doi:10.1016/j.advwatres.2006.11.009.

DEBYE, P. *Polar molecules*. Dover [1929].

DEIDDA, R. Rainfall downscaling in a space-time multifractal framework. *Water Resour. Res.*, 36(7):pages 1779–1794 [2000].

DICKINSON, R. E., A. HENDERSON-SELLERS, P. J. KENNEDY and M. F. WILSON. Biosphere-atmosphere transfer scheme (BATS) for the NCAR Community Climate Model. Tech. Note NCAR/TN-275+STR, Natl. Cent. Atmos. Res., Boulder, Colo [1986].

DIRKSEN, C. and S. DASBERG. Improved calibration of time domain reflectometry soil water content measurements. *Soil Sci. Soc. Am. J.*, 57:pages 660–667 [1993].

DIRMEYER, P., X. GAO, M. ZHAO, Z. GUO, T. OKI and N. HANASAKI. Supplement to GSWP-2: Details of the Forcing Data. *Bull. Amer. Meteor. Soc.*, 87:pages S10–S16 [2006].

DOBSON, M. C. and F. T. ULABY. Preliminary evaluation of the SIR-B response to soil moisture, surface roughness, and crop canopy. *IEEE Trans. Geosci. Remote Sens.*, GRS-24(4):pages 510–516 [1986]. doi:10.1109/TGRS.1986.289666.

DOBSON, M. C., F. T. ULABY, M. T. HALLIKAINEN and M. A. EL-RAYES. Microwave dielectric behavior of wet soil – part ii: dielectric mixing models. *IEEE Trans. Geosci. Remote Sens.*, GE-23(1) [1985].

DOUVILLE, H. Validation and sensitivity of the gobal hydrologic budget in stand-alone simulations with the ISBA land-surface scheme. *Climate Dynamics*, 14:pages 151–171 [1998].

DOWNER, C. W., F. L. OGDEN, W. D. MARTIN and R. S. HARMON. Theory, development, and applicability of the surface water hydrologic model CASC2D. *Hydrol. Proc.*, 16:pages 255–275 [2002].

DUAN, Q., S. SOROOSHIAN and V. K. GUPTA. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, 28:pages 1015–1031 [1992].

DUBOIS, P. C., J. VAN ZYL and T. ENGMAN. Measuring soil moisture with imaging radars. *IEEE Trans. Geosci. Remote Sens.*, 33(4):pages 915–926 [1995].

DUNNE, S. and D. ENTEKHABI. An ensemble-based reanalysis approach to land data assimilation. *Water Resour. Res.*, 41(2) [2005]. doi:10.1029/2004WR003449.

—. Land surface state and flux estimation using the ensemble Kalman smoother during the Southern Great Plains 1997 field experiment. *Water Resour. Res.*, 42(W01407) [2006]. doi:10.1029/2005WR004334.

EAGLESON, P. S. *Dynamic Hydrology*. McGraw-Hill [1970].

ELTAHIR, E. A. B. and R. L. BRAS. A description of rainfall interception over large areas. *J. Clim.*, 6:pages 1002–1008 [1993].

ENGMAN, E. T. Applications of microwave remote sensing of soil moisture for water resources and agriculture. *Remote Sens. Environ.*, 35:pages 213–226 [1991].

ENTEKHABI, D., 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 Trans. Geosci. Remote Sens.*, 32:pages 438–448 [1994].

ENTEKHABI, D., 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. PULZ, S. RUNNING, J. 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 Trans. Geosci. Remote Sens.*, 42(10):pages 2184–2195 [2004].

EOM, H. J. and W.-M. BOENER. Scattering from a layered medium connected with rough interfaces: Matrix Doubling method. *IEEE Trans. Geosci. Remote Sens.*, GE-24(6):pages 937–939 [1986].

EVANS, D. L., T. G. FARR and J. J. VAN ZYL. Estimates of surface roughness derived from synthetic aperture radar (SAR) data. *IEEE Trans. Geosci. Remote Sens.*, 30 [1992].

EVENSEN, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99(C5):pages 10,143–10,162 [1994].

—. The ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynam.*, 53:pages 343–367 [2003].

—. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynam.*, 54:pages 539–560 [2004].

FAMIGLIETTI, J. F. and E. F. WOOD. Application of multiscale water and energy balance models on a tallgrass prairie. *Water Resour. Res.*, 30:pages 3079–3093 [1994a].

—. Multiscale modeling of spatially variable water and energy balance processes. *Water Resour. Res.*, 30:pages 3061–3078 [1994b].

FAROUKI, O. T. Thermal Properties of Soils. CRREL Monograph 81-1, US Army Corps of Engineers, Cold Regions Research and Engineering Laboratory, Hanover, NH [1981].

FARR, T. G., P. A. ROSEN, E. CARO, R. CRIPPEN, R. DUREN, S. HENSLEY, M. KOBRICK, M. PALLER, E. RODRIGUEZ, L. ROTH, D. SEAL, S. SHAFFER, J. SHIMADA, J. UMLAND, M. WERNER, M. OSKIN, D. BURBANK and D. ALSDORF. The shuttle radar topography mission. *Rev. Geophys.*, 45 [2007]. doi:10.1029/2005RG000183.

FERRARIS, L., S. GABELLANI, N. REBORA and A. PROVENZALE. A comparison of stochastic models for spatial rainfall downscaling. *Water Resour. Res.*, 39 [2003]. doi:10.1029/2003WR002504.

FREILICH, M. H. and R. S. DUNBAR. A preliminary C-band model function for the ERS-1 AMI instrument. In *Proc. First ERS-1 Symp.*, pages 79–84. Cannes, France [1993].

FRIEDMAN, S. P. A saturation degree-dependent composite spheres model for describing the effective dielectric constant of unsaturated porous media. *Water Resour. Res.*, 34(11):pages 2949–2961 [1998].

FUNG, A. K. *Microwave Scattering and Emission Models and Their Applications*. Artech House, Boston [1994].

FUNG, A. K., Z. LI and K. S. CHEN. Backscattering from a randomly rough dielectric surface. *IEEE Trans. Geosci. Remote Sens.*, 30(2):pages 356–369 [1992].

GALANTOWICZ, J. F., D. ENTEKHABI and E. G. NJOKU. Tests of sequential data assimilation retrieving profile soil moisture and temperature from observed L-Band radiobrightness. *IEEE Trans. Geosci. Remote Sens.*, 37(4):pages 1860–1870 [1999].

GELB, A. *Applied Optimal Estimation*. MIT Press, Cambridge, MA [1974].

GEORGIOU, P. E., D. M. PAPAMICHAIL and S. G. VOUGIOUKAS. Optimal irrigation reservoir operation and simultaneous multi-crop cultivation area selection uding simulated annealing. *Irrig. and Drain.*, 55:pages 129–144 [2006].

GOODRICH, D. C., T. O. KEEFER, C. L. UNKRICH, M. H. NICHOLS, H. B. OSBORN, J. J. STONE and J. R. SMITH. Long-term precipitation database, Walnut Gulch Experimental Watershed, Arizonta, United States. *Water Resour. Res.*, 44(W05S04) [2008].

GOODRICH, D. C., D. A. WOOLHISER and T. O. KEEFER. Kinematic wave routing using finite elements on a triangular irregular network. *Water Resour. Res.*, 27:pages 995–1003 [1991].

GRASSOTTI, C., R. N. HOFFMAN, E. R. VIVONI and D. ENTEKHABI. Multiple-timescale intercomparison of two radar products and rain gauge observations over the Arkansas-Red River basin. *Wea. Forecast.*, 18:pages 1207–1229 [2003].

DE GRIEND, A. A. V. and J.-P. WIGNERON. The b-factor as a function of frequency and canopy type at H-polarization. *IEEE Trans. Geosci. Remote Sens.*, 42(1):pages 1–10 [2004].

GUGLIELMETTI, M., M. SCHWANK, C. MATZLER, C. OBERDORSTER, J. VANDERBORGHT and H. FLUHLER. FOSMEX: forest soil moisture experiments with microwave radiometry. *IEEE Trans. Geosci. Remote Sens.*, 46(3):pages 727–735 [2008]. doi:10.1109/TGRS.2007.914797.

GUPTA, V. K. and S. SOROOSHIAN. Calibration of conceptual hydrologic models: Past, present and future. In *Trends in Hydrology: Research Trends*, pages 329–346. Counc. of Sci. Res. Integration, Trivandrum, India [1994].

GWO, J. P., L. E. TORAN, M. D. MORRIS and G. V. WILSON. Subsurface stormflow modeling with sensitivity analysis using a Latin-Hypercube sampling technique. *Ground Water*, 34(5):pages 811–818 [1996].

HAVERKAMP, R., J.-Y. PARLANGE, R. CUENCA, P. J. ROSS and T. S. STEENHUIS. Scaling of the Richards Equation and Its Application to Watershed Modeling. In ESPOSITO, G., ed., *Scale Dependence and Scale Invariance in Hydrology*, pages 190–223. Cambridge University Press: Cambridge [1998].

HAWK, K. *Climatology of station storm rainfall in the continental United States: Parameters of the Bartlett-Lewis and Poisson rectangular pulses models*. Master's thesis, Massachusetts Institute of Technology [1992].

HELTON, J. C. and F. J. DAVIS. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Eng. Sys. Safety*, 81:pages 23–69 [2003].

HILLEL, D. *Environmental Soil Physics*. Academic Press, Amsterdam [1998].

HINO, M. On-line prediction of hydrologic system. In *Proceedings of the 15th Congress of the International Association for Hydraulic Research*, vol. 4, pages 121–129. International Association for Hydraulic Research, State Hydraulics Printing House, Ankara, Turkey [1973].

HOEBEN, R. and P. A. TROCH. Assimilation of active microwave observation data for soil moisture profile estimation. *Water Resour. Res.*, 36:pages 2805–2819 [2000].

HOLMES, T., M. DRUSCH, J.-P. WIGNERON and R. DE JEU. A global simulation of microwave emission: error structures based on output from ECMWF's operational integrated forecast system. *IEEE Trans. Geosci. Remote Sens.*, 46(3):pages 846–856 [2008]. doi:10.1109/TGRS.2007.914798.

HOUSER, P. R., W. J. SHUTTLEWORTH, J. S. FAMIGLIETTI, H. V. GUPTA, K. H. SYED and D. C. GOODRICH. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resour. Res.*, 34:pages 3405–3420 [1998].

HUFFMAN, G. J., R. F. ADLER, B. RUDOLPH, U. SCHNEIDER and P. KEEHN. Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. *J. Clim.*, 8:pages 1284–1295 [1995].

HUNG, N. Q., M. S. BABEL, S. WEESAKUL and N. K. TRIPATHI. An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydro. Earth Syst. Sci. Discuss.*, 5:pages 183–218 [2008].

IMAN, R. and W. J. CONOVER. A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat.: Simul. Comput.*, B11(3):pages 311–334 [1982].

ISLAM, S., D. ENTEKHABI, R. L. BRAS and I. RODRIGUEZ-ITURBE. Parameter estimation and sensitivity analysis for the Modified Bartlett-Lewis rectangular pulses model of rainfall. *J. Geophys. Res.*, 95(D3):pages 2093–2100 [1990].

IVANOV, V., R. L. BRAS and D. C. CURTIS. A weather generator for hydrological, ecological, and agricultural applications. *Water Resour. Res.*, 43 [2007]. doi:10.1029/2006WR005364.

IVANOV, V. Y. *Effects of dynamic vegetation and topography on hydrological processes in semi-arid areas*. Ph.D. thesis, Massachusetts Institute of Technology [2006].

IVANOV, V. Y., R. L. BRAS and E. R. VIVONI. Vegetation-hydrology dynamics in complex terrain of semi-arid areas: 1. A mechanistic approach to modeling dynamic feedbacks. *Water Resour. Res.*, 44 [2008a]. doi:10.1029/2006WR005588.

—. Vegetation-hydrology dynamics in complex terrain of semi-arid areas: 2. Energy-water controls of vegetation spatio-temporal dynamics and topographic niches of favorability. *Water Resour. Res.*, 44 [2008b]. doi:10.1029/2006WR005595.

IVANOV, V. Y., E. R. VIVONI, R. BRAS and D. ENTEKHABI. Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach. *J. Hydrology*, 298(1-4):pages 80–111 [2004a].

IVANOV, V. Y., E. R. VIVONI, R. L. BRAS and D. ENTEKHABI. Catchment hydrologic response with a fully distributed triangulated irregular network model. *Water Resour. Res.*, 40(W11102) [2004b]. doi:10.1029/2004WR003218.

JACKSON, T. J. and T. J. SCHMUGGE. Vegetation effects on the microwave emission from soils. *Remote Sens. Environ.*, 36:pages 203–212 [1991].

JACKSON, T. J., T. J. SCHMUGGE and J. R. WANG. Passive microwave sensing of soil moisture under vegetation canopies. *Water Resour. Res.*, 18:pages 1137–1142 [1982].

KEANE, R., R. BURGAN and J. VAN WAGTENDONK. Mapping wildland fuels for fire management across multiple scales: Integrating remote sensing, GIS, and biophysical modeling. *INTERNATIONAL JOURNAL OF WILDLAND FIRE*, 10(3-4):pages 301–319 [2001]. ISSN 1049-8001.

KERR, Y., F. SECHERRE, J. LASTENET and J.-P. WIGNERON. SMOS: Analysis of perturbing effects over land surfaces. In *Proc. IGARSS*, vol. 2, pages 908–910 [2003].

KERR, Y. H., P. WALDTEUFEL, J.-P. WIGNERON, J.-M. MARTINUZZI, J. FONT and M. BERGER. Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.*, 39(8):pages 1729–1735 [2001].

KERR, Y. H. and J.-P. WIGNERON. Vegetation models and observations – a review. In CHOUDHURY, B., Y. KERR, E. NJOKU and P. PAMPALONI, eds., *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*. VSP Publishing, Utrecht [1995].

KIM, G. and A. P. BARROS. Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data. *Remote Sens. Environ.*, 83:pages 400–413 [2002a].

—. Space-time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data. *Remote Sens. Environ.*, 81:pages 393–403 [2002b].

KIM, S., Y. TACHIKAWA and K. TAKARA. Applying a recursive update algorithm to a distributed hydrologic model. *J. Hydrol. Eng.*, 12(3):pages 336–344 [2007].

KIRDIASHEV, K. P., A. A. CHUKHLANSEV and A. M. SHUTKO. Microwave radiation of the Earth's surface in the presence of vegetation. *Radio Eng. Electron. USSR*, 24:pages 256–264 [1979].

KITANIDIS, P. K. and R. L. BRAS. Error identification in conceptual hydrologic models. In CHIU, C. L., ed., *Applications of the Kalman Filter to Hydrology, Hydraulics and Water Resources*, pages 325–353. Department of Civil Engineering, University of Pittsburgh, Pittsburgh, PA [1979].

—. Adaptive filtering through detection of isolated transient errors in rainfall-runoff models. *Water Resour. Res.*, 16:pages 740–748 [1980a].

—. Real-time forecasting with a conceptual hydrologic model, 2, Applications and results. *Water Resour. Res.*, 16:pages 1034–1044 [1980b].

KLAZURA, G. E. and D. A. IMY. A description of the initial set of analysis products available from the NEXRAD WSR-88D system. *Bull. Amer. Meteor. Soc.*, 74(7):pages 1293–1311 [1993].

KONG, J. A. *Electromagnetic Wave Theory*. Wiley-Interscience, New York, 2nd ed. [1990].

KORZUN, V. I. World Water Balance and Water Resources of the Earth. In *Studies and Reports in Hydrology*, vol. 25. UNESCO, Paris [1978].

KOSTER, R. D., P. A. DIRMEYER, Z. GUO, G. BONAN, E. CHAN, 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(5687):pages 1138–1140 [2004].

KOSTER, R. D. and M. J. SUAREZ. Energy and Water Balance Calculations in the MOSAIC LSM. NASA Technical Memorandum 104606 9, Natl. Aero. Space Admin. [1996].

KRAMER, P. J. *Water Relations of Plants*. Academic Press, Orlando [1983].

LAAKE, P. E. V. and G. A. SANCHEZ-AZOFEIFA. Mapping PAR using MODIS atmosphere products. *Remote Sens. Environ.*, 94(4):pages 554–563 [2005]. doi:10.1016/j.rse.2004.11.011.

LEENHOUTS, B. Assessment of biomass burning in conterminous United States. *Conservation Ecology*, 2(1) [1998].

LEIJ, F. J., W. J. ALVES, M. T. VAN GENUCHTEN and J. R. WILLIAMS. Unsaturated Soil Hydraulic Database, UNSODA 1.0 User's Manual. Tech. Rep. Report EPA/600/R-96/095, US Environmental Protection Agency, Ada, Oklahoma [1996].

LERMUSIAUX, P. F. J. and A. R. ROBINSON. Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.*, 127:pages 1385–1407 [1999].

LEVIS, S., G. B. BONAN, M. VERTENSTEIN and K. W. OLESON. The Community Land Model's Dynamic Global Vegetation Model (CLM-DGVM): Technical description and user's guide. NCAR Tech. Note NCAR/TN-459+IA, Natl. Cent. for Atmos. Res., Boulder, Colo [2004].

LIANG, X., D. P. LETTENMAIER, E. F. WOOD and S. J. BURGES. A simple hydrologically based model of land surface water and energy fluxes for GCMs. *J. Geophys. Res.*, 99(D7):pages 14,415–14,428 [1994].

LIVINGS, D. M., S. L. DANCE and N. K. NICHOLS. Unbiased ensemble square root filters. *Physica D: Nonlinear Phenom.*, 237(8):pages 1021–1028 [2008].

LONG, D. G., M. W. SPENCER and E. G. NJOKU. Spatial resolution and processing tradeoffs for HYDROS: Application of reconstruction and resolution enhancement techniques. *IEEE Trans. Geosci. Remote Sens.*, 43:pages 3–12 [2005].

VAN LOON, E. E. and P. A. TROCH. Tikhonov regularization as a tool for assimilating soil moisture data in distributed hydrological models. *Hydrol. Proc.*, 16(2):pages 531–536 [2002].

LOVEJOY, S. and B. B. MANDLEBROT. Fractal properties of rain, and a fractal model. *Tellus*, 37A:pages 209–232 [1985].

MANDLEBROT, B. B. Intermittent turbulence in self-similar cascades: Divergence of high moments and dimensions of the carrier. *J. Fluid Mech.*, 62:pages 331–358 [1974].

MARGULIS, S. A., 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 Resour. Res.*, 38 [2002]. doi:10.1029/2001WR0011114.

MÄTZLER, C. and A. STANDLEY. Technical note: Relief effects for passive microwave remote sensing. *Intl. J. Remote Sens.*, 21(12):pages 2403–2412 [2000].

MCKAY, M. D., R. J. BECKMAN and W. J. CONOVER. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technomet.*, 21(2):pages 239–245 [1979].

MEEHL, G., T. STOCKER, W. COLLINS, P. FRIEDLINGSTEIN, A. GAYE, J. GREGORY, A. KITOH, R. KNUTTI, J. MURPHY, A. NODA, S. RAPER, I. WATTERSON, A. WEAVER and Z.-C. ZHAO. Global Climate Projections. In SOLOMON, S., D. QIN, M. MANNING, Z. CHEN, M. MARQUIS, K. AVERYT, M. TIGNOR and H. MILLER, eds., *Climate Change 2007: The Physical Science Basis. . Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA [2007].

MENABDE, M., D. HARRIS, A. SEED, G. AUSTIN and D. STOW. Multiscaling properties of rainfall and bounded random cascades. *Water Resour. Res.*, 33:pages 2823–2830 [1997].

MENABDE, M., A. SEED, D. HARRIS and G. AUSTIN. Multiaffine random field model of rainfall. *Water Resour. Res.*, 35:pages 509–514 [1999].

MERLIN, O., A. G. CHEHBOUNI, Y. H. KERR, E. G. NJOKU and D. ENTEKHABI. A combined modeling and multispectral/multiresolution remote sensing approach for disaggregation of surface soil moisture: Application to SMOS configuration. *IEEE Trans. Geosci. Remote Sens.*, 43(9):pages 2036–2050 [2005].

MIALON, A., L. CORET, Y. H. KERR, F. SECHERRE and J.-P. WIGNERON. Flagging the topographic impact on the SMOS signal. *IEEE Trans. Geosci. Remote Sens.*, 46(3):pages 689–694 [2008].

MINASNY, B. and D. J. FIELD. Estimating soil hydraulic properties and their uncertainty: the use of stochastic simulation in the inverse modeling of the evaporation method. *Geoderma*, 126:pages 277–290 [2005].

MO, T., B. J. CHOUDHURY, T. J. SCHMUGGE, J. R. WANG and T. J. JACKSON. A model for microwave emission from vegetation-covered fields. *J. Geophys. Res.*, 97(C13):pages 11,229–11,237 [1982].

MO, T., T. J. SCHMUGGE and J. R. WANG. Calculations of the microwave brightness temperature of rough soil surfaces: bare field. *IEEE Trans. Geosci. Remote Sens.*, 25:pages 47–54 [1987].

MONTGOMERY, D. R. and W. E. DIETRICH. A physically based model for the topographic control on shallow landsliding. *Water Resour. Res.*, 30:pages 1153–1171 [1994].

MORADKHANI, H., K.-L. HSU, H. GUPTA and S. SOROOSHIAN. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the parameter filter. *Water Resour. Res.*, 41 [2005]. doi:10.1029/2004WR003604.

MORAN, M. S., W. E. EMMERICH, D. C. GOODRICH, P. HEILMAN, C. D. HOLIFIELD-COLLINS, T. O. KEEFER, M. A. NEARING, M. H. NICHOLS, K. G. RENARD, R. L. SCOTT, J. R. SMITH, J. J. STONE, C. K. UNKRICH and J. WONG. Preface to the special section on Fifty Years of Research and Data Collection: U. S. Department of Agriculture Walnut Gulch Experimental Watershed. *Water Resour. Res.*, 44(W05S01) [2008]. doi:10.1029/2007WR006083.

NARAYAN, U. and V. LAKSHMI. Characterizing sub-pixel variability of low resolution radiometer derived soil moisture using high resolution radar data. *Water Resour. Res.*, 44(W06425) [2008]. doi:10.1029/2006WR005817.

NARAYAN, U., V. LAKSHMI and T. J. JACKSON. High-resolution change estimation of soil moisture using L-band radiometer and radar observations mode during the SMEX02 experiments. *IEEE Trans. Geosci. Remote Sens.*, 44(6):pages 1545–1554 [2006].

NJOKU, E. and D. ENTEKHABI. Passive microwave remote sensing of soil moisture. *J. Hydrol.*, 184:pages 101–129 [1996].

NJOKU, E. G. and J.-A. KONG. Theory for passive microwave remote sensing of near-surface soil moisture. *J. Geophys. Res.*, 82(20):pages 3108–3118 [1977].

NJOKU, E. G., W. J. WILSON, S. H. YUEH, S. J. DINARDO, F. K. LI, T. J. JACKSON, V. LAKSHMI and J. BOLTEN. Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during SGP99. *IEEE Trans. Geosci. Remote Sens.*, 40(12):pages 2659–2673 [2002].

NJOKU, E. G., W. J. WILSON, S. H. YUEH and Y. RAHMAT-SAMII. A large-antenna microwave radiometer-scatterometer concept for ocean salinity and soil moisture sensing. *IEEE Trans. Geosci. Remote Sens.*, 38:pages 2645–2655 [2000].

OH, Y., K. SARABANDI and F. T. ULABY. An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE Trans. Geosci. Remote Sens.*, 30(2):pages 370–381 [1992].

OKI, T. The Hydrologic Cycles and Global Circulation. In ANDERSON, M. G. and J. MCDONNELL, eds., *Encyclopedia of Hydrological Sciences*, vol. 1. Wiley, New York [2005].

OKI, T. and S. KANAE. Global hydrologic cycles and world water resources. *Science*, 313:pages 1068–1072 [2006].

OKIMURA, T. and M. NAKAGAWA. A model for predicting surface mountain slope failture with a digital landform model. *J. Erosion Control Engineering Soc. Japan*, 41:pages 48–56 [1988].

OLESON, K. W. *et al.*. Technical description of the community land model (CLM). NCAR Tech. Note NCAR/TN-461+STR, Natl. Cent. for Atmos. Res., Boulder, Colo [2004].

O'NEILL, P., E. G. NJOKU, T. CHAN, W. CROW, A. HSU and J. C. SHI. Comparison of soil moisture retrieval algorithms using simulated HYDROS brightness temperatures. In *Proceedings of IGARSS 2004*, pages 20–24. IEEE, Anchorage, AK [2004].

OVER, T. M. and V. K. GUPTA. A space-time theory of mesoscale rainfall using random cascades. *J. Geophys. Res.*, 101(D21):pages 26,319–26,311 [1996].

PALOSCIA, S., G. MACELLONI and E. SANTI. Soil moisture estimates from AMSR-E brightness temperatures by using a dual-frequency algorithm. *IEEE Trans. Geosci. Remote Sens.*, 44(11):pages 3135–3144 [2006]. doi:10.1109/TGRS.2006.881714.

PAMPALONI, P. and S. PALOSCIA. Microwave emission and plant water content: a comparison between field measurements and theory. *IEEE Trans. Geosci. Remote Sens.*, 24:pages 900–905 [1986].

PAUWELS, V. R. N., N. E. C. VERHOEST, G. J. M. D. LANNOY, V. GUISSARD, C. LUCAU and P. DEFOURNY. Optimization of a coupled hydrology-crop growth model through the assimilation of observed soil moisture and leaf area index values using an ensemble Kalman filter. *Water Resour. Res.*, 43(W04421) [2007]. doi:10.1029/2006WR004942.

PELLARIN, T., Y. KERR and J.-P. WIGNERON. Global simulation of brightness temperatures at 6.6 and 10.7 GHz over land based on SMMR data set analysis. *IEEE Trans. Geosci. Remote Sens.*, 44(9):pages 2492–2505 [2006]. doi:10.1109/TGRS.2006.874139.

PELLARIN, T., J.-P. WIGNERON, J.-C. CALVET and P. WALDTEUFEL. Global soil moisture retrieval from a synthetic L-band brightness-temperature data set. *J. Geophys. Res.*, 108(D12) [2003]. doi:10.1029/2002JD003086.

PELLENQ, J., J. KALMA, G. BOULET, G.-M. SAULNIER, S. WOOLDRIDGE, Y. KERR and A. CHEHBOUNI. A disaggregation scheme for soil moisture based on topography and soil depth. *J. Hydrol.*, 276:pages 112–127 [2003].

PERSSON, A. and F. GRAZZINI. *Meteorological Bulletin M3.2: User Guide to ECMWF forecast products.* European Center for Medium-Range Weather Forecasting (ECMWF), version 4.0 ed. [2005].

PETERS-LIDARD, C. D., E. BLACKBURN, X. LIANG and E. F. WOOD. The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures. *J. Atmos. Sci.*, 55:pages 1209–1224 [1998].

PETERS-LIDARD, C. D., M. S. ZION and E. F. WOOD. A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes. *J. Geophys. Res.*, 102:pages 4303–4324 [1997].

PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING and B. P. FLANNERY. *Numerical Recipes in C.* Cambridge University Press: Cambridge, 2nd ed. [1992].

RAWLS, W. J., L. R. AHUJA, D. L. BRAKENSIEK and A. SHIRMOHAMMADI. Infiltration and soil water movement. In MAIDMENT, D. R., ed., *Handbook of Hydrology.* McGraw-Hill, New York, NY [1993].

RAWLS, W. J. and D. L. BRAKENSIEK. Prediction of soil water properties for hydrologic modeling. In JONES, E. B. and T. WARD, eds., *Watershed Management in the Eighties*, pages 293–299. Proc. Irrig. Drain. div., ASCE, Denver, CO [1985].

REES, W. G. *Physical Principles of Remote Sensing.* Cambridge University Press, 2nd ed. [2001].

REICHLE, R. H., W. T. CROW and C. L. KEPPENNE. An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resour. Res.*, 44(W03423) [2008]. doi:10.1029/2007WR006357.

REICHLE, R. H., 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:pages 2353–2364 [2001].

REICHLE, R. H., D. B. MCLAUGHLIN and D. ENTEKHABI. Hydrologic data assimilation with the Ensemble Kalman Filter. *Mon. Wea. Rev.*, 130:pages 103–114 [2002].

RESTREPO-POSADA, P. J. and R. L. BRAS. Automatic parameter estimation of a large conceptual rainfall-runoff model: A maximum likelihood approach. Technical Report 267, Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Massachusetts Institute of Technology, Cambridge, MA [1982].

RHENALS, A. E. and R. L. BRAS. The irrigation scheduling problem and evapotranspiration uncertainty. *Water Resour. Res.*, 17(5):pages 1328–1338 [1981].

RICHARDSON, L. F. *Weather prediction by numerical process*. Cambridge University Press [1922].

RODRIGUEZ-ITURBE, I., D. R. COX and V. ISHAM. Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. A.*, 410:pages 269–288 [1987].

—. A point process for rainfall: further developments. *Proc. R. Soc. Lond. A.*, 417:pages 283–298 [1988].

RODRÍGUEZ-ITURBE, I., G. K. VOGEL, R. RIGON, D. ENTEKHABI, F. CASTELLI and A. RINALDO. On the spatial organization of soil moisture fields. *Geophys. Res. Lett.*, 22:pages 2757–2760 [1995].

ROLLINS, M., R. KEANE and R. PARSONS. Mapping fuels and fire regimes using remote sensing, ecosystem simulation, and gradient modeling. *ECOLOGICAL APPLICATIONS*, 14(1):pages 75–95 [2004]. ISSN 1051-0761.

ROSENKRANZ, P. W. Retrieval of temperature and moisture profiles from AMSU-A and AMSU-B measurements. *IEEE Trans. Geosci. Remote Sens.*, 39(11):pages 2429–2435 [2001].

DE ROSNAY, P., J.-C. CALVET, Y. KERR, J.-P. WIGNERON, F. LEMAÎTRE, M. J. ESCORIHUELA, J. M. SABATER, K. SALEH, J. BARRIÉ, G. BOUHOURS, L. CORET, G. CHERELA, G. DEDIEU, R. DURBE, N. E. D. FRITZ, F. FROISSARD, J. HOEDJES, A. KRUSZEWSKI, F. LAVENU, D. SUQUIA and P. WALDTEUFEL. SMOSREX: A long term field campaign experiment for soil moisture and land surface process remote sensing. *Remote Sens. Environ.*, 102:pages 377–389 [2006].

RUNNING, S. W. Is Global Warming Causing More, Larger Wildfires? *Science*, 313(5789):pages 927–928 [2006].

RUTTER, A. J., K. A. KERSHAW, P. C. ROBINS and A. J. MORTON. A predictive model of rainfall interception in forests: 1. Derivation of the model from observation in a plantation of Corsican pine. *Agric. Meteorol.*, 9:pages 367–384 [1971].

SAKOV, P. and P. R. OKE. Implications of the form of the ensemble transforation in the ensemble square root filters. *Mon. Wea. Rev.*, 136(3):pages 1042–1053 [2008].

SANDELLS, M. J., I. J. DAVENPORT and R. J. GURNEY. Passive L-band microwave soil moisture retrieval error arising from topography in otherwise uniform scenes. *Adv. Water Resour.* [2008]. doi:10.1016/j.advwatres.2008.01.012.

SCHAAP, M. and F. LEIJ. Database related accuracy and uncertainty of pedotransfer functions. *Soil Sci.*, 163:pages 765–779 [1998].

SCHERTZER, D. and S. LOVEJOY. Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.*, 92(D8):pages 9693–9714 [1987].

SCHMUGGE, T. J. and T. J. JACKSON. Soil moisture variability. In STEWART, J. B. *et al.*, eds., *Scaling up in Hydrology Using Remote Sensing*, pages 183–192. John Wiley, New York [1996].

SCHUSTER, R. L. and R. W. FLEMING. Economic losses and fatalities due to landslides. *Bull. Assoc. Engineering Geol.*, 23:pages 11–28 [1986].

SCHWANK, M., C. MÄTZLER, M. GUGLIELMETTI and H. FLÜHLER. L-band radiometer measurements of soil water under growing clover grass. *IEEE Trans. Geosci. Remote Sens.*, 43(10):pages 2225–2237 [2005].

SCHWEPPE, F. C. *Uncertain Dynamic Systems*. Prentice Hall, Englewood Cliffs, NJ [1973].

SELLERS, P. J., S. O. LOS, C. J. TUCKER, C. O. JUSTICE, D. A. DAZLICH, G. J. COLLATZ and D. A. RANDALL. A revised land surface parameterization (SiB2) for atmospheric GCMs. 2. The generation of global fields of terrestrial biophysical parameters from satellite data. *J. Clim.*, 9(4):pages 706–737 [1996].

SELLERS, P. J., Y. MINTZ, Y. C. SUD and A. DALCHER. A simple biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.*, 43:pages 505–531 [1986].

SEO, D. J., V. KOREN and N. CAJINA. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *J. Hydrometeor.*, 4:pages 627–641 [2003].

SHARIF, H. O., F. L. OGDEN, W. F. KRAJEWSKI and M. XUE. Numerical simulations of radar rainfall error propagation. *Water Resour. Res.*, 38(8) [2002]. doi:10.1029/2001WR000525.

—. Statistical analysis of radar rainfall error propagation. *J. Hydrometeor.*, 5:pages 199–212 [2004].

SHERMAN, L. K. Stream-flow from rainfall by the unit-graph method. *Eng. News-Rec.*, 108:pages 501–505 [1932].

SHI, J., J. WANG, A. Y. HSU, P. E. O'NEILL and E. T. ENGMAN. Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data. *IEEE Trans. Geosci. Remote Sens.*, 35(5):pages 1254–1266 [1997].

SHIKLOMANOV, I. A., ed. *Assessment of Water Resources and Water Availability in the World*. World Meteorological Institute/Stockholm Environment Institute, Geneva, Switzerland [1997].

SHOOP, S. A. Terrain Characterization for Trafficability. CRREL Report 93-6, U.S. Army Corps of Engineers, Cold Regions Research and Engineering Laboratory [1993].

SHUTTLEWORTH, W. J. Evaporation. Rep. 56, Inst. of Hydrol., Wallingford, UK [1979].

SINGH, V. P. *Kinematic Wave Modeling in Water Resources: Surface Water Hydrology*. John Wiley, Hoboken, NJ [1996].

SMIL, V. *Energy at the Crossroads: Global Perspectives and Uncertainties*. MIT Press [2003].

SOHRABI, T. M., A. SHIRMOHAMMADI and H. MONTAS. Uncertainty in nonpoint source pollution models and associated risks. *Env. Forensics*, 3:pages 179–189 [2002].

SOIL CONSERVATION SERVICE. *Hydrology*. Supplement A to Section 4. U.S. Department of Agriculture, Washington, D.C. [1968].

STRANG, G. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley [1986].

TAIZ, L. and E. ZEIGER. *Plant Physiology*. Sinauer Associates [2002].

THOMA, D. P., M. S. MORAN, R. BRYANT, M. RAHMAN, C. D. HOLIFIELD-COLLINS, S. SKIRVIN, E. E. SANO and K. SLOCUM. Comparison for four models to determine surface soil moisture from C-band radar imagery in a sparsely vegetated semiarid landscape. *Water Resour. Res.*, 42(W01418) [2006]. doi:10.1029/2004WR003905.

TOPP, G. C., J. L. DAVIS and A. P. ANNAN. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resour. Res.*, 16:pages 574–582 [1980].

TSANG, L. and R. W. NEWTON. Microwave emission from soils with rough surfaces. *J. Geophys. Res.*, 87:pages 9017–9024 [1982].

TUCKER, G. E. and R. L. BRAS. Hillslope processes, drainage density and landscape morphology. *Water Resour. Res.*, 34(10):pages 2751–2764 [1998].

TUCKER, G. E., S. T. LANCASTER, N. M. GASPARINI and R. L. BRAS. The Channel-Hillslope Integrated Landscape Development (CHILD) model. In HARMON, R. and W. D. III, eds., *Landscape Erosion and Evolution Modelling*. Kluwer Academic/Plenum Publishers [2001a].

TUCKER, G. E., S. T. LANCASTER, N. M. GASPARINI, R. L. BRAS and S. RYBARCZYK. An object-oriented framework for distributed hydrologic and geomorphic modeling using triangulated irregular networks. *Computers Geosci.*, 21(8):pages 959–973 [2001b].

ULABY, F., M. RAZANI and M. C. DOBSON. Effects of vegetation cover on the microwave radiometric sensitivity to soil moisture. *IEEE Trans. Geosci. Remote Sens.*, 21:pages 51–61 [1983].

ULABY, F. T., R. K. MOORE and A. K. FUNG. *Microwave remote sensing, vol. I-III*. Artech House, Norwood, MA [1986].

VALL-ILOSSERA, M., A. CAMPS, I. CORBELLA, F. TORRES, N. DUFFO, A. MONERRIS, R. SABIA, D. SELVA, C. ANTOLÍN, E. LÓPEZ-BAEZA, J. F. FERRER and K. SALEH. SMOS REFLEX 2003: L-band emissivity characterization of vineyards. *IEEE Trans. Geosci. Remote Sens.*, 43(5):pages 973–982 [2005].

VIVONI, E. R., V. Y. IVANOV, R. L. BRAS and D. ENTEKHABI. Generation of triangulated irregular networks based on hydrological similarity. *J. Hydrol. Eng.*, 27:pages 959–973 [2004].

—. On the effects of triangulated terrain resolution on distributed hydrologic model response. *Hydrol. Proc.*, 19:pages 2101–2122 [2005].

VOGEL, H.-J. and K. ROTH. Moving through scales of flow and transport. *J. Hydrol.*, 272:pages 95–106 [2003].

VOROSMARTY, C. J., P. GREEN, J. SALISBURY and R. B. LAMMERS. Global water resources: Vulnerability from climate change and population growth. *Science*, 289:pages 284–288 [2000].

VRUGT, J. A., C. H. H. DIKS, H. V. GUPTA, W. BOUTEN and J. M. VERSTRATEN. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resour. Res.*, 41 [2005]. doi:10.1029/2004WR003059.

VRUGT, J. A., G. SCHOUPS, J. W. HOPMANS, C. YOUNG, W. W. WALLENDER, T. HARTER and W. BOUTEN. Inverse modeling of large-scale spatially distributed vadose zone properties using global optimization. *Water Resour. Res.*, 40 [2004]. doi:10.1029/2003WR002706.

WANG, J. and R. L. BRAS. Ground heat flux estimated from surface soil temperature. *J. Hydrol.*, 216(3-4):pages 214–226 [1999].

WANG, J. R., E. T. ENGMAN, J. C. SHIUE, M. RUSEK and C. STEINMEIER. The SIR-B observations of microwave dependence on soil moisture, surface roughness, and vegetation covers. *IEEE Trans. Geosci. Remote Sens.*, GRS-24(4):pages 510–516 [1986]. doi:10.1109/TGRS.1986.289665.

WANG, X., C. H. BISHOP and S. J. JULIER. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, 132:pages 1590–1605 [2004].

WEAST, R. C., ed. *CRC Handbook of Chemistry and Physics*. CRC Press, Boca Raton, FL, 63rd. ed. [1983].

WEGMULLER, U., C. MATZLER and E. NJOKU. Canopy opacity models. In CHOUDHURY, B. J., Y. H. KERR, E. G. NJOKU and P. PAMPALONI, eds., *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*, pages 375–388. VSP Publishing, Utrecht [1995].

WESTERN, A. W., R. B. GRAYSON, G. BLÖSCHL, G. R. WILLGOOSE and T. A. MCMAHON. Observed spatial organization of soil moisture and its relation to terrain indices. *Water Resour. Res.*, 35(3):pages 797–810 [1999].

WHITAKER, J. S. and T. M. HAMILL. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, 130:pages 1913–1924 [2002].

WIGMOSTA, M. S., L. W. VAIL and D. P. LETTENMAIER. A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.*, 30:pages 1665–1679 [1994].

WIGNERON, J.-P., P. WALDTEUFEL, A. CHANZY, J.-C. CALVET and Y. KERR. Two-dimensional microwave interferometer retrieval capabilities over land surfaces (SMOS mission). *Remote Sens. Environ.*, 73(3):pages 270–282 [2000].

XU, C., H. S. HE, Y. HU, Y. CHANG, X. LI and R. BU. Latin hypercube sampling and geostatistical modeling of spatial uncertainty in a spatially explicit forest landscape model simulation. *Ecological Modelling*, 185:pages 255–269 [2005].

YAPO, P. O., H. V. GUPTA and S. SOROOSHIAN. Multi-objective global optimization for hydrologic models. *J. Hydrol.*, 204:pages 83–97 [1998].

ZHOU, Y. *Multi-sensor large scale land surface data assimilation via ensemble approaches*. Ph.D. thesis, Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, Cambridge [2005].