IMPROVEMENT OF SOIL MOISTURE PREDICTION THROUGH AMSR-E DATA ASSIMILATION

by

Alok Kumar Sahoo A Dissertation Submitted to the Graduate Faculty of George Mason University in Partial Fulfillment of The Requirements for the Degree of Doctor of Philosophy Computational Sciences and Informatics

Committe chau.

Date: 04/28/08

Dr. Menas Kafatos, Dissertation Director

Dr. Paul R. Houser, Committee Member

Dr. Ruixin Yang, Committee Member

Dr. Paul A. Dirmeyer, Committee Member

Dr. Dimitrios Papaconstantopoulos, Department Chairperson

Dr. Peter Becker, Associate Dean for Graduate Programs, College of Science

Dr. Vikas Chandhoke, Dean, College of Science

Spring Semester 2008 George Mason University Fairfax, VA Improvement of Soil Moisture Prediction through AMSR-E Data Assimilation

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Alok K. Sahoo Master of Science George Mason University, 2006

Director: Menas Kafatos, Professor Department of Computational and Data Sciences

> Spring Semester 2008 George Mason University Fairfax, VA

Copyright 2008 Alok K. Sahoo All Rights Reserved

DEDICATION

This dissertation is dedicated to my loving parents who have sacrificed a lot with smiling faces and my sister and brother-in-law for their encouragement to achieve my goals.

ACKNOWLEDGEMENTS

I would like to thank my Ph. D. supervisor Dr. Menas Kafatos, who has spent his precious time and effort directing this dissertation work. It would have been a much harder journey without his encouragement and patience.

I am also very thankful to my committee members Dr. Paul Houser, Dr. Paul Dirmeyer and Dr. Ruixin Yang for their technical guidance and recommendations during this research and for being always available and open for discussions. Their thoughtful views and support were invaluable throughout this dissertation work.

My sincere appreciation goes to the people at the Hydrological Sciences Branch, NASA Goddard Space Flight Center, where I got an opportunity to spend two summer terms, for giving me suggestions and creating special interests in me towards this research area. The support provided by the LIS team, Dr. Xiwu Zhan, Dr. Rolf Reichle, Dr. Christa Peters Lidard and Dr. Randy Koster is highly appreciated.

Special thanks go to Dr. Eric Wood and his lab members for giving me an opportunity to spend a summer term at Princeton University and use their facilities to do part of this research work.

I would like to thank people at Center for Research on Environment and Water and Center for Ocean land And Atmosphere Studies for providing their support and facilities to perform this research work.

I acknowledge all my friends and faculties at Center for Earth Observing and Space Research; Earth Systems and GeoInformation Sciences and College of Sciences for their support and cooperation and making my research years at GMU a memorable one.

I would like to express my deepest gratitude to my parents Prafulla Kumar Sahoo and Sabitri Sahoo for being very supportive during these toughest years and for supporting each my career decision. Their immense help and moral support have been inspirational though out my life.

Last but not least I would like to thank my sister Prachi, my brother-in-law Bhabani and my other family members for understanding me and encouraging me to reach my career goals and standing beside me in my rough times.

TABLE OF CONTENTS

	Page
List of Tables	ix
List of Figures	X
List of Abbreviations	xiii
Abstract	xv
1. Introduction	1
1.1 Fundamental Concepts	2
1.1.1 Global Water Cycle	2
1.1.2 Global Energy Budget	4
1.1.3 Surface water and Energy Balance	6
1.2 Soil Moisture and Its Real World Applications	8
1.2.1 Definition of Soil Moisture.	8
1.2.2 Application of Soil Moisture	8
1.3 Statement of the Problem and Motivation	12
1.4 Objectives	14
1.5 Statement of Work	15
1.5.1 Evaluation of the Observed Soil Moisture Data	16
1.5.2 Evaluation of the Model Simulation Results	16
1.5.3 Assimilation of Soil moisture Observations into a Land Surface	Model17
1.5.4 Application of the Assimilated Soil Moisture Product	
1.6 Organization of the Dissertation.	18
2. In-Situ Soil Moisture Data Analysis	
2.1 Introduction	20
2.2 Description of the Study Area and In-Situ Data Sets	23
2.2.1 Study Area.	
2.2.2 LREW In-Situ Observation Data	
2.3 Temporal Characteristics	
2.4 Spatial Characteristics	
2.4.1 Geostatistical Analysis	
2.5 Conclusions	
3. Evaluation of Current Satellite Soil Moisture Products: AMSR-E & LS	MEM43
3.1 Introduction	44
3.2 Soil Moisture Retrievals and Radiative Transfer Theory	49
3.3 Description of Study Area and Data Sets	
3.3.1 Study Area	
3.3.2 LREW In-Situ Observation Data	
3.3.3 AMSR-E Soil Moisture Retrieval Data	

	3.3.4 LSMEM Radiative Transfer Model Soil Moisture Data	55
	3.4 Results and Analysis	62
	3.5 Discussions.	
	3.6 Conclusions	78
4.	A Study of Land Surface Processes: Land Surface Model Inter-Comparison	
	Study	80
	4.1 Introduction	81
	4.2 Land Surface Models	83
	4.2.1 LIS Architecture	85
	4.2.2 HySSiB Model	85
	4.2.3 Noah Model	87
	4.2.4 CLM	88
	4.3 Description of Study Area and In-Situ Data Sets	91
	4.3.1 Study Area	91
	4.3.2 LREW In-Situ Observation Data	91
	4.3.3 SCAN Data	91
	4.4 Input Data and Model Setup	92
	4.4.1 Forcing Data	92
	4.4.2 Parameter Data	96
	4.4.3 Model Initialization	97
	4.4.4 Model Simulation	97
	4.5 Results and Discussion.	
	4.5.1 Scale Issues and Pre-processing of Results	98
	4.5.2 Results and Analysis.	100
	4.5.2.1 Soil Moisture	100
	4.5.2.2 Water Cycle Variables	108
	4.5.2.3 Energy Cycle variables	
	4.5.3 Impact of Scaling on Soil Moisture	115
	4.6 Conclusions	118
5.	Assimilation of LSMEM Retrieved Soil Moisture into Noah Land Surface	
	Model	122
	5.1 Introduction	
	5.2 Literature Review.	
	5.3 Ensemble Kalman Filter (EnKF)	
	5.4 Model Description, Experimental Design in LIS & Study Area	
	5.4.1 The EnKF Module in LIS.	
	5.4.2 The Noah Land Surface Model	131
	5.4.3 LSMEM Soil Moisture Observations	
	5.4.4 Study Area	133
	5 4 5 LREW In-Situ Observation Data	134
	5.5 Results and Discussion.	134
	5.5.1 Soil Moisture	
	5.5.2 Water Cycle Variables	144
	5.5.3 Energy Cycle Variables	147

	5.5.4 Sensitivity Study	153
	5.5.4.1 Sensitivity of EnKF Algorithm to Model Initialization	154
	5.5.4.2 Sensitivity of EnKF Algorithm to Model Spin-up	158
	5.5.5 1-D versus 3-D Data Assimilation.	162
	5.5.6 Model Hydrology	164
	5.6 Conclusions	165
6.	Applications of Research Results	170
	6.1 Introduction	170
	6.2 Soil Moisture – Vegetation Growth Relationship	172
	6.2.1 Datasets Used	174
	6.2.1.1 Land Cover Data	174
	6.2.1.2 Soil Moisture Data	174
	6.2.1.3 NDVI Data	175
	6.2.1.4 EVI Data	176
	6.2.1.5 FVC Data	176
	6.2.2 Results and Discussion	176
	6.2.2.1 Correlation Results for Concurrent Data	178
	6.2.2.2 Lag Correlation Results	180
	6.2.3 Conclusions	185
	6.3 Study of Extreme Events	186
	6.3.1 Definition and Classification	186
	6.3.2 Pre-conditions of Flooding	188
	6.3.3 Terminologies Used in this Study	188
	6.3.4 May 6-13, 2003 Flood of Tennessee, Alabama and Georgia – A Case	
	Study	190
	6.3.4.1 Datasets Used	190
	6.3.4.1.1 Soil Moisture Data	190
	6.3.4.1.2 Precipitation Data	191
	6.3.4.1.3 Dartmouth Flood Observatory Information	191
	6.3.4.2 Discussions	192
	6.3.5 Conclusions	194
7.	Summary, Conclusions and Research Extensions	198
	7.1 Summary and Conclusions	198
	7.1.1 Characteristics of the In-Situ Soil Moisture Observations	198
	7.1.2 Remote Sensing Soil Moisture Observations	199
	7.1.3 Soil Moisture Simulations from Land Surface Models	200
	7.1.4 Soil Moisture Data Assimilation Results	201
	7.1.5 Applications of Research Results	202
	7.2 Limitations.	203
	7.2.1 Limited Availability of the In-Situ Observations	203
	7.2.2 Non-Compatibility of Spatial and Temporal Scales	203
	7.2.3 Choice of Land surface Model.	204
	7.2.4 Bias Corrections to Satellite Observations	205
	7.3 Research Extensions.	205

7.3.1 Requirement of more Data Validation Study	205
7.3.2 Fraternal Twin Experiment.	
7.3.3 3-D Data Assimilation Study	
7.3.4 Multi-model Ensemble Data Assimilation Study	
7.3.5 Coupled Data Assimilation Study	
7.3.6 Operational Assimilated Global Soil Moisture Product	
List of References	209

LIST OF TABLES

Table	Page
3.1: Summary of Input datasets used for the LSMEM Model	60
3.2: Summary of Soil Moisture Retrieval Methods	73
4.1: Basic differences among the three Land Surface models	84
4.2: Initial model setup for the GLDAS/LIS model runs	98
4.3: Skill of hourly and daily time series simulations of soil moisture	118
5.1: A brief description of all the model simulations performed for this study	134
5.2: The recovery time for the dry and wet runs with respect to the intermediate	
runs for all the four soil moisture layers for both the openloop and EnKF	
simulations	154
5.3: The recovery time for the cold start runs with respect to the spin-up run for all	the
four soil moisture layers for the EnKF simulation	159
6.1: Vegetation class values and their descriptions	175
6.2: Information about the May 6-13, 2003 Flood of Tennessee, Alabama and	
Georgia	193

LIST OF FIGURES

Figu	Page
1.1:	The Hydrologic Cycle
1.2:	The Earth's annual global mean energy budget
1.3:	Energy and water balance at the land surface
2.1:	Little River Experimental Watershed (LREW), Tifton, Georgia
2.2:	Time series of daily averaged soil moisture from some stations and the mean
	precipitation for the year 2003
2.3:	Scatter plot of temporally averaged mean soil moisture versus standard
	deviation
2.4:	Temporal autocorrelation function of soil moisture versus time lag
2.5a:	Time series of daily averaged spatial mean soil moisture, standard deviation and
	coefficient of variation
2.5b:	Scatter plot of standard deviation of moisture content versus mean moisture
	content
2.5c:	Scatter plot of coefficient of variation versus mean moisture content
2.6:	Scatter plot of cross correlation of soil moisture time series from stations versus
	distance (km) between stations
2.7a:	Time series of daily averaged spatial mean soil moisture, sill and range
2.7b:	Scatter plot of sill versus daily averaged spatial mean moisture content
2.7c:	Scatter plot of range versus daily averaged spatial mean moisture constant40
3.1:	Flowchart of the LSMEM Soil Moisture Retrieval
3.2:	Plots of daytime versus night time soil moisture from (a) in-situ measurements,
	(b) AMSR-E and (c) LSMEM retrievals for the year 2003
3.3:	(a) Time series plot of the in-situ measured, AMSR-E and LSMEM retrieved
	soil moisture at the AMSR-E ascending overpass time along with daily average
	precipitation and irrigation. Scatter plot of the corresponding (b) in-situ and
	AMSR-E soil moisture, (c) in-situ and LSMEM soil moisture data
3.4:	Same as Figure 3.3, but at the AMSR-E descending overpass time72
3.5:	VIC surface temperature validation with the in-situ surface temperature at
	AMSR-E (a) ascending and (b) descending overpass time
4.1:	Scatter Plots of hourly atmospheric forcing from NLDAS and SCAN
	measurements at SCAN site
4.2:	Time series plot of daily precipitation (mm/day) from NLDAS and SCAN
	measurements at SCAN site for the year 2003
4.3:	Daily soil moisture time series plots from Noah (10 cm layer), CLM (2 cm top
	layer), HySSiB (2 cm top layer) in-situ measurements (5 cm layer) and
	Arithmetic Model Ensemble Mean from Noah, CLM and HySSiB100

4.4:	Daily soil Moisture scatter plot, (a) Noah versus In-Situ, (b) CLM versus In-
	situ, (c) HySSiB versus In-situ and (d) Arithmetic Model Ensemble Mean
	versus In-situ
4.5:	(a) Mean Bias, (b) RMSD and (c) Correlation of Noah, CLM, HySSiB and
	Arithmetic Model Ensemble Mean soil moisture
4.6:	Spatial difference image of NLDAS precipitation and corresponding model
	simulated soil moisture results during July 16 – July 19, 2003107
4.7:	Comparison of Water Cycle variables simulated by Noah, CLM and HySSiB111
4.8:	Comparison of Energy Cycle variables simulated by Noah, CLM and HySSiB 115
4.9:	Same as Figure 4.4, but for hourly soil moisture data
5.1:	A schematic diagram of the Ensemble Kalman Filter
5.2:	Daily top layer soil moisture time series plots from Noah (10 cm layer) open
	loop and EnKF simulations, in-situ measurements (5 cm layer) and LSMEM
	retrievals (~ 1 cm)
5.3:	Scatter plots for the Noah (a) 'openloop_spinup' and (b) 'enkf_spinup' results
	versus the in-situ observations of the daily top layer soil moisture data138
5.4:	(a) Time averaged percentage soil moisture improvement in 'enkf_spinup'
	results over the 'openloop_spinup' results. (b) A snap shot of the corresponding
	LSMEM soil moisture observation map is shown for reference
5.5:	Time series of the spatially averaged RMSEs in surface soil moisture for the
	'openloop_spinup' and 'enkf_spinup' simulations140
5.6:	Time averaged improvement metric for the surface soil moisture
5.7:	Time averaged spatially distributed variance of the normalized innovation (no
	unit) over the watershed
5.8:	Daily time series for the moisture for subsurface layers from the
	'openloop_spinup' and 'enkf_spinup' simulations146
5.9:	Daily time series plots of water cycle variables from the 'openloop_spinup' and
	'enkf_spinup' simulations
5.10:	Daily time series plots of energy cycle variables from the 'openloop_spinup'
	and 'enkf_spinup' runs
5.11:	Sensitivity results of EnKF to the model initialization conditions for four layers
	at station RG31
5.12:	Sensitivity results of EnKF to the model spin-up for the four layers at station
	RG31
5.13:	Spatial correlation of the innovation among different grid locations against their
	separation distances (a) the grid locations only in X- and Y-directions and (b) in
	any random direction
5.14:	Spatially averaged daily soil moisture increment for four layers
6.1:	Different land cover (vegetation types) locations considered in this study177
6.2:	Correlations between three vegetation indices and concurrent (a) top layer and
	(b) root zone soil moisture during the vegetation growing season for different
	vegetation types
6.3:	Lag correlations between the three vegetation indices with the root zone soil
	moisture for different vegetation classes

6.4:	Classification of extreme climate events	189
6.5:	The May 6-13, 2003 Flood boundary map over Tennessee, Alabama and	
	Georgia on the USA map	192
6.6:	Daily precipitation maps (mm/day) for May 4-8, 2003 period (left panel) and the corresponding daily soil moisture maps (% vol/vol; right panel)	.196

LIST OF ABBREVIATIONS

ALMA	Assistance for Land Modeling Activities
AMSR-E	Advanced Microwave Scanning Radiometer -EOS
ARM	Atmospheric Radiation Measurement
ARS	Agricultural Research Service
AVHRR	Advanced Very High Resolution Radiometer
BATS	Biosphere–Atmosphere Transfer Scheme
CART	Cloud and Radiation Test Bed
CLASIC	Cloud and Land Surface Interaction Campaign
CLM	Community Land Model
CLSM	Catchment based Land Surface Model
CV	coefficient of variation
DEM	Digital Elevation Model
DFO	Dartmouth Flood Observatory
EDAS	Eta Data Assimilation System
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
EOS	Earth Observing System
ESTER	Electronically Scanned Thinned Array Radiometer
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization
FVC	Fractional Vegetation Cover
GLDAS	Global Land Data Assimilation System
GOES	Geostationary Operational Environmental Satellite
GSFC	Goddard Space Flight Center
GSMDB	Global Soil Moisture Data Bank
GSWP	Global Soil Wetness Project
HySSiB	Hybrid Simplified Simple Biosphere Model
IAP	Institute of Atmospheric Physics
LAI	Leaf Area Index
LIS	Land Information System
LREW	Little River Experimental Watershed
LSM	Land Surface Model
LSMEM	Land Surface Microwave Emission Model
LSS	Land Surface Scheme
MODIS	Moderate Resolution Imaging Spectroradiometer

National Aeronautics and Space Administration
National Center for Environmental Prediction
Normalized Difference Vegetation Index
North American Land Data Assimilation System
Natural Resources Conservation Service
National Snow and Ice Data Center
Optimized Production Inputs for Economic and Environmental
Enhancement
Oregon State University
Push Broom Microwave Radiometer
Probability Distribution Function
Plant Functional Type
Project to Inter-compare Land-surface Parameterization Schemes
Polarization Ratio
Radio Frequency Interference
Rain Gauge 43
Root Mean Square Difference
Radiative Transfer Model
Stem Area Index
Soil Climate Analysis Network
Southern Great Plains 1997
Simple Biosphere Model
Soil Moisture Experiment 2003
Soil Moisture Experiment 2004
Soil Moisture Experiment 2005
Scanning Multichannel Microwave Radiometer
Soil Moisture and Ocean Salinity
Simplified Simple Biosphere Model
Special Sensor Microwave/Imager
Sea Surface Temperature
State Soil Geographic
TRMM Microwave Imager
Tropical Rainfall Measuring Mission
University of Maryland
US Department of Agriculture
US Geological Survey
Variable Infiltration Capacity
Vegetation Water Content

ABSTRACT

IMPROVEMENT OF SOIL MOISTURE PREDICTION THROUGH AMSR-E DATA ASSIMILATION

Alok Kumar Sahoo, Ph.D.

George Mason University, 2008

Dissertation Director: Dr. Menas Kafatos

This dissertation is aimed at evaluating the soil moisture estimation from satellites as well as land surface models and improving it using a data assimilation technique. The entire study was conducted over the Little River Experimental Watershed, Georgia for the year 2003; one of the four selected watersheds to validate the current AMSR-E satellite soil moisture data. Soil moisture data from a comprehensive in-situ observation network at this watershed were first used to study the spatial and temporal soil moisture characteristic of the watershed. There was a high degree of spatial and temporal correlation among different measurement stations which was required to validate other datasets with lower spatial and temporal frequency. Hence, those in-situ observations were treated as ground truth to validate other soil moisture datasets in this dissertation. A satellite based soil moisture product was generated from AMSR-E satellite brightness temperature data using the LSMEM radiative transfer model. This research product was found to be statistically better than the current AMSR-E soil moisture product when both

the datasets were compared against the in-situ observations. Similarly, three land surface models pertaining to different model physics and parameterization were simulated to generate soil moisture over the watershed. There was quite a bit of disagreement among model soil moisture results which was also reflected in other water and energy cycle variables since they were mostly controlled by soil moisture. Noah model soil moisture was found to be better than those of other two models even though it had a constant positive bias. When the LSMEM soil moisture observations were assimilated into the Noah land surface model using the EnKF algorithm, the Noah model predictions got improved significantly. This was confirmed by calculating the improvement metric over the Noah openloop simulations. The EnKF algorithm was found to be sensitive to the model initialization and spin-up conditions. In the end, the assimilated soil moisture results were used to demonstrate two real world applications. It was found that the relationship between the winter/spring soil moisture and vegetation during growing season was different for different vegetations types. This assimilated sol moisture map was also able to show the spatial and temporal extent of the 2003 May flooding event over Tennessee, Alabama and Georgia accurately. The conclusion chapter discusses the limitations we faced during this research work and many research extensions that can be performed to this research work. This assimilated soil moisture shows lot of promise for real world applications. This product can operationally be produced at finer spatial and temporal scales which is required for any kind of real world applications.

Chapter 1. Introduction

Water is a very vital substance that sets the Earth apart from the rest of the planets in our solar system. Water is a necessary ingredient for the development and nourishment of life. Water can naturally exist in all three forms of a substance (gas, liquid and solid). Most of the Earth's water is in the global oceans (approximately 96 %). Other than that, water exists in the atmosphere as water vapor, lakes, rivers, soil, groundwater, glacier, polar ice and permanent snow.

The earth system is changing and water is at the heart of both the causes and the effects of climate change. Soil moisture is the basic link between the hydrologic cycle and energy budget. In order to understand the partitioning of heat and water fluxes from land surface system to the atmosphere, the land surface state variables (especially soil moisture) need to be accurately estimated and studied. Soil moisture study is important to understand its role in agriculture, forest ecology, water resource management, weather and climate change, ecosystem management and drought and flood monitoring.

Information about soil moisture can be obtained through in-situ and remote observations and land surface modeling. Data assimilation is a technique which merges many different observations together with an estimate of the land surface variables provided by a land surface model. Data assimilation provides estimates in space and time when we have no observations and improves the accuracy of the estimate over any estimate from an individual source. Data assimilation is relatively new to the hydrology though it has drawn a lot of attention from the hydrologic scientists in the last decade. This thesis presents an optimal data assimilation technique, called Ensemble Kalman Filter (EnKF) for soil moisture data assimilation. The soil moisture observations for the data assimilation are taken from Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) satellite. The land surface model for this research is chosen from a set of three land surface models (with different model parameterizations and schemes) based on the performance of the models. The feasibility of this approach is tested over the Little River Experimental Watershed (LREW) in Georgia. Observed soil moisture data from this watershed are used for the verification of the assimilation results. A real life application of the data assimilation results has also been demonstrated in this research work.

1.1. Fundamental Concepts

1.1.1. Global Water Cycle

The water cycle (also known as hydrologic cycle) is the continuous movement of water on, above and below the surface of the Earth in a closed path (Figure 1.1). This water cycle recycles the earth's valuable water supply among the ocean, land and atmosphere. The Sun's energy in the form of light and heat controls the hydrologic cycle processes. Water can change states among liquid, vapor and ice inside the water cycle. There are five main processes in the water cycle: (a) condensation, (b) precipitation, (c) infiltration, (d) run-off and (e) evapo-transpiration. Condensation is the process when the water vapor condenses on particles in the atmosphere to form cloud. This happens when the air temperature drops due to the upward rise of the air from the Earth's surface. The wind transports and distributes this cloud across the globe. Precipitation is the process when the water is released from the cloud in liquid or solid form (rain, snow, hail, sleet etc.). Precipitation begins after the condensed water becomes heavier to stay in the atmosphere. Hence the water is transported from the atmosphere to the Earth's surface through precipitation. Infiltration is the process where a portion of the precipitation seeps into the ground below the Earth's surface. The infiltrated water constitutes the ground water or subsurface water. The amount of water that infiltrates depends on soil, rock and vegetation types on the surface, soil saturation, slope and permeability. Run-off happens when the soil is saturated with water and can not hold any more precipitation falling on the surface. It can also happen due to the snow melting in the cold regions. Run-off eventually falls into streams, creeks and rivers which finally flow down usually to the oceans. Evapo-transpiration takes the water from the ground and vegetation back to the atmosphere again to complete the water cycle. When the solar energy hits the Earth's surface, it heats the water and causes the water to evaporate to the atmosphere. Vegetation takes water from the Earth's surface through their roots and routes them to the atmosphere through their leaves which is called transpiration. Some water also gets evaporated before reaching the ground or from run-off while flowing in rivers. Sublimation is also similar to evaporation which takes place in the cold regions when the ice turns directly to water vapor.



Fig. 1.1: The Hydrologic Cycle (taken from Mather, 1974).

1.1.2. Global Energy Budget

The global energy budget explains how the incoming and outgoing energy to and from this earth-atmosphere system is distributed between different processes (Lydolph, 1985). Figure 1.2 shows a schematic diagram of the global mean annual energy budget. The incoming solar radiation is in the ultraviolet, visible and infrared region of the spectrum (also known as incoming shortwave radiation) whereas the Earth's emitted radiation (from the surface and atmosphere) is greater than 3 microns in the infrared region of the spectrum (also known as outgoing longwave radiation). The incoming solar radiation (342 W m⁻²) is either reflected directly back into space (107 W m⁻²), absorbed by the atmosphere (67 W m⁻²) or absorbed by the Earth's surface (168 W m⁻²). Some of this

absorbed heat by the Earth's surface is returned to the atmosphere as sensible heat and latent heat flux. Sensible heat flux is the transfer of the heat energy from the Earth's surface to the atmosphere by convection and conduction. Latent heat flux is related to the movement of water from the Earth's surface to the atmosphere when the water changes its state among solid, liquid and vapor. The sensible and latent heat flux are absorbed by the atmosphere. The emitted outgoing longwave radiation by the Earth's surface (350 W m⁻²) is mostly absorbed and radiated back by the atmosphere (324 W m⁻²) to the Earth's surface again.



Fig. 1.2: The Earth's annual global mean energy budget (taken from Houghton et al. (editor), 1996; Kiehl and Trentberth, 1997). Units are in W m⁻².

1.1.3. Surface Water and Energy Balance

The surface water and energy balance describes the relationship between incoming and outgoing water and energy at the Earth's surface respectively. Figure 1.3 shows a schematic diagram of water and energy balance at the land surface.

The water balance at the surface layer (in the non-snow condition) can mathematically be expressed as (Peixoto and Oort, 1992):

$$P = E + R + C_w \frac{\Delta w}{\Delta t} + miscellaneous \tag{1.1}$$

where, P = precipitation, E = evaporation, R = runoff (both surface and subsurface runoff), $C_w =$ water holding capacity of the surface, $\Delta w =$ change in the degree of saturation of the surface, $\Delta t =$ time interval, $C_w \frac{\Delta w}{\Delta t} =$ change in amount of water in the soil layers, *miscellaneous* = moisture for conversion of plant sugar, consumption by living beings etc.

The precipitation brings water down to the Earth's surface from the atmosphere where as the right hand side variables in Equation (1.1) help the incoming water move on the Earth surface and send back to the atmosphere again. The water balance at the Earth's surface remains fairly constant over long period of time. For a land surface model, precipitation is derived from observations or a general circulation model (GCM) and provided as an input. The land surface model determines the right hand side variables. The energy balance at the surface layer (in the non-snow condition) can mathematically be expressed as (Critchfield, 1974):

$$S_{w}^{\downarrow} + L_{w}^{\downarrow} = S_{w}^{\uparrow} + L_{w}^{\uparrow} + H + \lambda E + C_{p}\Delta T + miscellaneous$$
(1.2)

where, S_w^{\downarrow} = incoming solar radiation, L_w^{\downarrow} = incoming longwave radiation, S_w^{\uparrow} = outgoing shortwave radiation, L_w^{\uparrow} = outgoing longwave radiation, H = sensible heat flux, λ = latent heat of vaporization, E = evaporation rate, λE = latent heat flux, C_p = heat capacity of the surface, ΔT = change in the surface temperature over certain time interval, $C_p \Delta T$ = soil heat flux, *miscellaneous* = energy associated with soil water freezing, plant chemical energy etc.



Fig. 1.3: Energy and water balance at the land surface (taken from Lockwood, 1974).

The left hand side of the energy balance equation is the total incoming energy where as the right hand side is the total outgoing energy at the Earth's surface. As we notice from Equation (1.2), the incoming energy equals the outgoing energy at the Earth's surface on a longer time scale. Similar to the water balance variables, the left hand side parameters are derived from a typical climate model or other observations and provided as input to the land surface model. The right hand side variables are determined by the land surface model.

1.2. Soil Moisture and Its Real World Applications

1.2.1. Definition of Soil Moisture

There is no clear definition exists for soil moisture (Dirmeyer, 2004). Soil moisture can be defined in many ways depending on its application. Soil moisture is defined as water stored in a farm scale for crop production from agriculture point of view. The hydrologists define it as a precondition in the watershed scale to affect the surface runoff, source of contamination and aquifer recharge. For the meteorologists, soil moisture is defined over a large area which can interact with the atmosphere to affect the precipitation over land (Lawford, 1992). We follow the soil moisture is expressed as a dimensionless ratio of mass or volume of water by total mass or volume of soil which contain this water. This ratio is reported as decimal fraction or percentage after multiplying by 100 (Yu et al., 1993).

1.2.2. Application of Soil Moisture

Soil moisture is a very important component of the hydrologic cycle. It has a great impact on climate change over land. It plays the same role over land as sea surface temperature plays over the ocean. It has a long memory (order of months) of storing the atmospheric signature/energy transferred to it through precipitation, in turn transferring them back to the atmosphere through evaporation and affecting the climate (Martinez-Fernandez and Ceballos, 2003). Soil moisture helps determine the redistribution of rainfall into surface runoff and subsurface run off (Delworth and manabe, 1988; Pauwels et al., 2002). It controls the surface feedback of energy and water to the atmosphere. It divides the outgoing energy into sensible heat and latent heat fluxes. Soil moisture impacts the soil erosion, soil aeration, distribution and growth of vegetation, soil microbial activity, the concentration of toxic substances, the movement of nutrients in the soil to the roots and weather prediction at a local to regional scale (Koster et al., 2004). Soil moisture has many real life applications which make it a very important parameter to measure and study. Some of the soil moisture applications are discussed here:

Drought and flood monitoring – Hydrologic drought and flooding are closely related to the amount of soil moisture available in a specific region. The deficiency of root zone soil moisture for a considerable period of time leads to hydrologic drought (scarcity of water in ponds, lakes, rivers) and then later on to agricultural drought (no vegetation). Contrast to that, if precipitation happens for a longer period of time, then the soil gets saturated with water leading to flooding. Sometimes, the instant heavy precipitation does not give enough time for the precipitable water to percolate through the soil and hence generates high run-off and flash flooding. **Fire risk management** – Many trees die in the regions with very low soil moisture. They provide an excellent fuel (dry trees and wood) for wild land fires. So, the regions with low soil moisture are highly prone to wild land fires.

Soil conservation – Soil erosion is a major problem for the crop system since they provide food for the growth of crops. Soil moisture holds the soil tightly and controls the soil erosion, soil transportation by many factors such as river, stream, wind etc.

Agriculture and crop system modeling – Agriculture is highly dependent on soil moisture and irrigation, especially in summer season when the rainfall is scanty. It is very important to know the amount of moisture in the root zone before planting any crop. That can help farmers to determine whether they want to plant a crop type which intakes lot of water, but yields food very fast or a crop type which intakes less water and takes longer time to yield food.

Civil Engineering – Soil water makes soil heavier and softer to hold any big structural property. So, it is important to know the moisture amount and variation of the soil water over years. This can help to modify the engineering structures such as dams, bridges, roads, etc. that are properly suitable for the soil conditions.

Hydrologic modeling – It is important to quantify the change in soil water due to natural and anthropogenic causes. Since soil moisture is a major component of water cycle, it will be very helpful to study water cycle processes, precipitation and discharge pattern analysis by knowing the amount of soil water.

Climate and weather modeling - In a climate system, soil moisture acts as a water reservoir. Soil moisture coming from the precipitation or snow melting during spring

time is stored in soil and is evaporated during the summer time when the soil becomes warmer due to strong solar radiation. So, soil moisture acts as a connection between spring season and summer season processes to drive the climate. Knowing the amount of soil moisture helps to understand the seasonal processes and model the climate system.

Similarly, the weather is also controlled by the soil moisture. Soil moisture determines the amount of water available to be transported to the atmosphere though evapo-transpiration. Soil moisture controls the amount of sensible and latent heat fluxes going to the atmosphere. The local weather such as intensity and genesis of severe storms and heating and cooling of the atmosphere depend on the amount of soil moisture being evaporated to the atmosphere.

Ecosystem and forest modeling – Ecosystem balance is very important for all the living beings. Trees require water from soil to survive. In a closed environment such as forests if the available soil water is not sufficient, then all the trees compete among themselves to survive. Soil moisture also controls the nutrients and chemicals uptake by the trees and sediment transportation by the rivers. Soil moisture is a major cause of deforestation. All the living beings in the ecosystem are also dependent on soil water, especially in the arid environment where people look for aquifers to find some water to survive. Quantifying the soil moisture helps to model the ecosystem accurately.

Watershed and reservoir management – Watershed is a portion of the land where all the water beneath it or draining off it goes to a common place such as streams, lakes or rivers. The watershed is highly controlled by the moisture available in the root zone of the soil. Also, the water recharge from and percolation to the ground below a reservoir is a very important component of the reservoir's water storage. It is important to know the soil properties and the amount of water it can hold for a proper management of watersheds and reservoirs.

1.3. Statement of the Problem and Motivation

Soil moisture is a critical hydrology parameter whose spatial and temporal variability has direct applications in agriculture, forest ecology, civil engineering, water resources management, and crop system modeling as discussed in the previous section. Its incorrect specification leads to the erroneous assessment of the other hydrology and energy cycle parameters. Scientists are trying to understand the physical relationship among the soil moisture and other ecosystem components (e. g. precipitation, runoff, elevation, soil type, vegetation etc.) through field campaigns, land surface models as well as remote sensing to estimate the soil moisture accurately in a global scale.

Global soil moisture observation is a high priority. AMSR-E (Advanced Microwave Scanning Radiometer), aboard the EOS (Earth Observing System)-AQUA produces global soil moisture data from 6 and 10 GHz frequency radiometer channels. Since the 6 GHz observations are contaminated due to the Radio Frequency Interference (RFI) over select regions (Li et al., 2004), the AMSR-E soil moisture algorithm was modified to only use the 10 GHz channel. But this global soil moisture data produce very uncharacteristic soil moisture values as compared to that of the model and field observations and are still under validation stage (Njoku et al., 2003; Zhan et al., 2004).

Many land surface model groups are participating in model inter-comparison studies (e.g. Project for Inter-comparison of Land-surface Parameterization Schemes (PILPS experiment)) to better understand model physics and parameterizations. Assimilation is another technique which is getting popular within many model groups these days to find the model representation that is most consistent with the observations so that a global product can be generated from that model. In essence, data assimilation merges a range of diverse data fields with a model prediction to provide that model with the best estimate of the current state of the natural environment so that it can then make more accurate predictions. The application of data assimilation in hydrology has been limited to a few one-dimensional, largely theoretical studies primarily due to the lack of sufficient spatially-distributed hydrologic observations. Ensemble data assimilation addresses the probabilistic aspect of prediction and analysis (Zupanski et al., 2006).

Few dedicated soil moisture field experiments (SMEX02, Iowa; SMEX03, Georgia, Alabama, Oklahoma and Brazil; and SMEX04, Arizona) have also been conducted recently to validate the AMSR-E satellite soil moisture data product. But, these field experiments are confined to very small geographic areas of the globe.

Though satellite data and models are plentiful, an accurate global scale soil moisture data product is elusive. Improving the synergistic use of satellite data, model forcing, model physics and data processing techniques should improve our knowledge and take us closer to achieve the goal of producing a global soil moisture product accurate enough to be useful in end-user solutions.

The motivation for this doctoral research work starts with the fact that there is no accurate global state of the art soil moisture data available even though research is going on for more than two decades. Apart from that, most of the researchers have assimilated either ground based or airborne soil moisture observations; or some other satellite observed forcing parameters; or performed synthetic twin experiments to estimate the soil moisture state. This may be due to the non-availability of reliable satellite observed soil moisture data. But with the launch of soil moisture dedicated AMSR-E instrument aboard EOS-AQUA satellite in 2002, we are getting plentiful satellite observed soil moisture data now. There is some soil moisture measuring satellite missions coming up in near future too (e.g. Soil Moisture and the Ocean salinity (SMOS)). So, it would be of great interest to the scientific community to assimilate the current AMSR-E soil moisture observations to land surface models to estimate the merged soil moisture product and validate this newly merged product. The performance of the assimilation algorithm to the model spin-up and initialization conditions has not been fully tested yet. That motivates us to include some assimilation algorithm sensitivity studies in this research work.

1.4. Objectives

The motivation for this research has already raised few interesting science questions in context with the data assimilation technique. This thesis work is going to verify the science question: roles of model spin-up and model initialization conditions on the performance of the EnKF technique. The scientific objectives of this doctoral dissertation are to improve the soil moisture prediction by assimilating the satellite observations into a land surface model and use that assimilated product for real life applications such as soil moisture and vegetation growth relationship and flood monitoring. We seek to:

- Better understand how different retrieval approaches are being used to generate soil moisture products from satellite observations
- Better understand how the land surface model complexity and physics contribute to the land surface simulations for the land surface processes and soil moisture results
- Apply a data assimilation algorithm to merge the best available soil moisture observation data and land surface model to produce an assimilated soil moisture product and check the sensitivity of the assimilation algorithm to the model spin-up and initialization conditions.
- Use the assimilated soil moisture product to study soil moisture and vegetation growth relationship and flood monitoring over USA.

The hypothesis behind this assimilation study is that the data assimilation algorithm will perform better for the model spin-up and proper initialization cases than no spin-up and extreme model initializations for satellite-model merged soil moisture retrieval as they do stand alone. But the above mentioned hypothesis has not been assessed by the hydrologic science community for soil moisture retrieval through data assimilations so far.

1.5. Statement of Work

To address the proposed tasks, this study involves a data assimilation technique and a data inter-comparison study. The systematic research steps are described as follows:

1.5.1. Evaluation of the Observed Soil Moisture Data

Here we will investigate the quality of the existing AMSR-E satellite observed soil moisture along with another AMSR-E soil moisture product derived by a different retrieval approach by comparing with the in-situ observations. The following tasks have been designed to accomplish this objective:

- Process the AMSR-E Level-3 soil moisture product over the Little River Experimental Watershed, Georgia for 2003.
- Use AMSR-E brightness temperature and NLDAS (North American Land Data Assimilation) surface temperature and derive different soil moisture product by another retrieval approach adopted in the Land Surface Microwave Emission radiative transfer Model (LSMEM) for 2003.
- Evaluate the quality of both the observed soil moisture products by comparing them with the in-situ observations and draw comparison statistics.

1.5.2. Evaluation of the Model Simulation Results

Here we will evaluate the land surface model simulated soil moisture results. From the model results differences, we will try to understand how different model incorporates different land surface processes to produce soil moisture. The following tasks will be performed to accomplice this objective:

- Create respective restart files for the Noah, CLM (Community Land Model) and HySSiB (Hybrid Simplified Simple Biosphere model) models by spinning-up the models three times using the five year NLDAS forcing data from 1998 to 2002 (total 15 years of spin-up).
- Run the Noah, CLM and SSiB model simulations over the Little River Experimental Watershed, Georgia using the NLDAS forcing and restart files for 2003.
- Evaluate the quality of the model simulation results by comparing them with the insitu observations and try to understand the differences among the comparison results.

1.5.3. Assimilation of Soil Moisture Observations into a Land Surface Model

We will apply and investigate the assimilation algorithm performance and sensitivity for soil moisture estimate. To accomplice this, we:

- Assimilate the best available soil moisture observations into the best available land surface model (after evaluating the results in the previous steps) using Ensemble Kalman Filtering (EnKF) algorithm over the Little River Experimental Watershed for 2003.
- Check the performance of the EnKF algorithm by comparing the model assimilated results with the model open loop results (generated in the previous step) and the observations.

• Perform the EnKF algorithm sensitivity studies to the model spin-up and initialization conditions.

1.5.4. Application of the Assimilated Soil Moisture Product

Here we will look at a real life application of the assimilated soil moisture data over USA. For this step, we:

- Generate an operational assimilated soil moisture product using the previous task over USA for 2002-2006.
- Study the soil moisture and vegetation growth relationship and flood monitoring over USA during 2002-2006.

1.6. Organization of the Dissertation

Chapter One gives a brief introduction, discusses the problem statement, motivation, objective and scope of this research work. Chapter Two provides a description of the research location and performs the spatio-temporal analysis of the in-situ observed soil moisture over the research location. The objective of Chapter Three is to find a good and reliable soil moisture dataset for the data assimilation study. Chapter Three describes the soil moisture retrieval procedure from satellite observations using a radiative transfer model. It also verifies the radiative transfer model results along with the current operational soil moisture data against the in-situ observations and discusses the performance of both the satellite soil moisture products. Chapter Four focuses on to find a good land surface model out of three chosen models for the data assimilation study.

Chapter Four also describes the models and compares the three model soil moisture results with the in-situ observations and comments on the comparison statistics. Chapter Five covers the data assimilation study. It first explains the EnKF data assimilation (DA) algorithm. Then it verifies the improvement of the model soil moisture estimates by comparing the non-DA and DA results. This chapter also tests the sensitivity of the EnKF algorithm to the model spin-up and initialization conditions. Chapter Six describes couple of real life applications of the assimilated results derived in Chapter Five. It looks at the soil moisture and vegetation growth relationship and flood monitoring over USA from 2002 to 2006. Finally, the results obtained from all the chapters are analyzed in Chapter Seven. It also discusses the limitations encountered in this research work and provides some directions for the research extensions.
Chapter 2. In-Situ Soil Moisture Data Analysis

Summary: The usefulness of intermediately-spaced in situ soil moisture observations for validation of other soil moisture data products is ascertained. As a part of the Advanced Microwave Scanning Radiometer (AMSR-E) – Earth Observing System (EOS) soil moisture calibration and validation project, a network of Steven-Vitel hydra probe soil moisture instruments have been installed at rain gauge sites in Little River Experimental Watershed (LREW) since 2001 to monitor soil water continuously. High resolution soil moisture data from 14 in-situ stations for 2003 have been used in this study to characterize the temporal and spatial variability in the watershed. The time series of the in-situ soil moisture data show little seasonality for 2003. Higher soil moisture autocorrelation at each individual site and the spatial cross correlation between stations found in this watershed provide useful information to validate other coarse temporal and spatial variability increases whereas the range decreases with the increase of soil moisture.

2.1 Introduction

The spatial variability of soil moisture is influenced by large scale atmospheric forcing (mostly precipitation) and small scale land surface variability (topography, soil texture,

vegetation etc.) (Vinnikov et al., 1999). Knowledge of spatial and temporal distribution of soil moisture is essential to study the influence of soil moisture on the land and atmosphere processes at different scales.

This chapter documents the temporal and spatial analysis of a dense network of in-situ soil moisture observations over the Little River Experimental Watershed (LREW), Georgia. Soil moisture measurements at local scale have scientific applications such as (i) understanding the local scale land and atmosphere processes, physical laws, empirical relationships associated with soil moisture; (ii) understanding the impacts of vegetation, surface temperature, topography, and soil texture on soil moisture variations; (iii) validation of satellite measurements and model simulation results. For these kinds of applications, a dense network of soil moisture in-situ measuring stations is required. Collection of such soil moisture datasets has been initiated to study the above mentioned applications (Georgakakos and Baumer, 1996; Vinnikov et al., 1996). The Global Soil Moisture Data Bank (GSMDB; Robock et al., 2000) is one such project which has soil moisture measurements from networks of stations over many regions of the globe. These datasets have been widely used for research, applications and validation studies (Robock et al., 1997; Robock et al., 1998; Entin et al., 2000; Reichle et al., 2004; Prigent et al., 2005). Many short term soil moisture field experiments have also been conducted including Portos 91/93 over Avignon, France, Washita'92 and Washita'94 over the Little Washita Watershed, Oklahoma; SGP'97 and SGP'99 over the U.S. Southern Great Plains; the Soil Moisture Experiments (SMEX02 over Iowa, SMEX03 over Georgia, Alabama and Southern Great Plains, SMEX04 over Arizona, SMEX05 over Iowa); the

Cloud and Land Surface Interaction Campaign (CLASIC) 2007 over the Southern Great Plains.

Vinnikov et al. (1999) considered the soil moisture measuring stations over Illinois, with an average station distance of 93 km (coarse scale), and analyzed the optimal design of surface networks of observation for mesoscale (~ 30 km) to climate scale models. Similar to Vinnikov et al. (1999), others have used the coarse spatial scale network stations for soil moisture study (Robock et al., 1997; Prigent et al., 2005). In contrast to the above coarse scale soil moisture study, De Lannoy et al. (2006) considered the in-situ measurements with average station distance of few hundreds of meters (very fine scale) in the Optimized Production Inputs for Economic and Environmental Enhancement (OPE) experimental field in Beltsville, Maryland operated by US Department of Agriculture (USDA). They studied the spatial and temporal variation of soil moisture studies have been conducted by others too (Western et al., 1998; Famiglietti et al., 1999; Grayson et al., 2002; Western et al., 2004; Ryu and Famiglietti, 2005).

This chapter differs from yet compliments that previous work by considering the in-situ measurements with an intermediate spacing of 3 to 9 km (fine to medium scale). This study has been conducted to assess the spatial and temporal soil moisture variations for a medium-sized watershed and to verify whether the available data from these in-situ observations can be useful to validate microwave satellite soil moisture estimate with a typical grid size of ~ 25 km.

In the next section, the field site and the datasets are described. The temporal and spatial characteristics of the soil moisture observations are analyzed in section 2.3 and section 2.4 respectively. The conclusions are summarized in section 2.5.

2.2 Description of the Study Area and In-Situ Data Sets

2.2.1 Study Area

The Little River Experimental Watershed (LREW) located near Tifton, Georgia (Figure 2.1) is one of four designated watersheds selected to calibrate and validate the Advanced Microwave Scanning Radiometer (AMSR-E) – Earth Observing System (EOS) satellite soil moisture observations. Hence, we have very high temporal and spatial resolution insitu soil moisture observations available to study the variations over this watershed. This watershed encompasses 334 km² area. The main watershed includes seven gauged subwatersheds ranging in size from 3 to 115 km². This watershed is in the headwaters of the Suwannee River Basin that begins in Georgia and empties into the Gulf of Mexico. The Little River is a tributary of the Withlacoochee River; one of the two main tributaries of the Suwannee River. The LREW has very flat topography with broad flood plains that is poorly defined by stream channels (Sheridan, 1997).



Fig 2.1: Little River Experimental Watershed (LREW), Tifton, Georgia (from Cashion et al., 2005). The in-situ soil moisture measuring instruments have been installed at some of the rain-gauge sites.

Vegetation in the watershed is a mixture of row-crop agriculture, pasture and forage production, upland and riparian forest. The major crops are peanuts and cotton, tobacco, corn, soybeans, melons and some vegetable crops are also grown in the watershed. Coverages are approximately 36% forest, 40% crops, 18% pasture, with the remaining area in wetlands and residential areas. Swamp hardwoods with thick vegetation occur along the stream edges (Bosch et al., 2006). Extensive land use information and physical characteristics of this LREW watershed have been described in

Williams (1982), Perry et al. (1999) and Sheridan and Ferreira (1992). The dominant soil type is sandy loam consisting of sandy surface layer and loamy subsoil. Most of the soils are well drained and they have fairly low water holding capacities with porosities ranging from 0.1 to 0.3 (Hubbard et al., 1985).

The area experiences long, hot, humid summers, and short, mild winters. The average annual precipitation is approximately 1200 mm. Precipitation in this region is unevenly distributed and typically occurs in short duration high intensity thunderstorms with relatively small spatial extent during the summer months (Bosch et al., 1999).

2.2.2 LREW In-Situ Observation Data

There is a network of 35 tipping bucket precipitation gauges located within the LREW which record the cumulative rainfall every 5 minutes. There is one Soil Climate Analysis Network (SCAN) site also located within the watershed. The spacing between neighboring precipitation gauges varies from three to eight km (Figure 2.1). As a part of the AMSR-E calibration and validation project, a network of Steven-Vitel hydra probe soil moisture instruments (http://www.stevenswater.com/soil_moisture_sensors/index.aspx) have been installed at some rain gauge sites since 2001 to monitor soil water continuously at 5 cm, 20 cm and 30 cm depths (Cashion et al., 2005). The detailed description of the soil moisture measuring sites can be found in Bosch et al. (2006). Soil water measurements are taken every half hour at these sites to conform to the SCAN data. This watershed was also a part of the Soil Moisture Field Experiment conducted in June and July 2003 (SMEX03).

The field observation data were provided by the United States Department of Agriculture - Agricultural Research Service (USDA-ARS) located at Beltsville, MD (Jackson et al., 2006). The data include instantaneous soil moisture from top 5 cm, instantaneous soil temperature and cumulative precipitation data at every 30 minute interval for the year of 2003 from 14 individual stations located within the LREW (Figure 2.1). The data also included some statistics (e.g. mean and standard deviation) of instantaneous soil moisture, instantaneous soil temperature and cumulative precipitation over all the 14 stations in each 30 minute interval. Limited quality control and quality assurance have been carried out by USDA-ARS. Arithmetic averages and averages based on nearest neighbor weighting are done based on the same set of sensors and several sensors have been eliminated from this averaging by USDA during the quality control because of poor or suspicious performance. The detailed description of these in-situ data can be found in Jackson et al. (2006, 2007). Only the soil water data from the top 5 cm soil layer have been used in this study.

2.3 Temporal Characteristics

Figure 2.2 shows the daily averaged soil moisture for four scattered stations out of the 14 available in the watershed. The daily averaged station mean precipitation data are also shown in the same figure for reference. There is no clear seasonal cycle seen at this resolution. All the stations show similar behavior including the correspondence to the precipitation events. In spring there are higher values of soil moisture because of long duration persistent heavy rainfall events during that season. In summer, there are many

short duration thunderstorms, as described by Bosch et al. (1999) and potential evapotranspiration is very high. Because of this and because the sandy loam soil holds very little water, summer soil moisture often drops to very low levels between precipitation events.

Figure 2.3 shows the variation with the temporal standard deviation of soil moisture of the temporal average for each of the 14 stations. The temporal statistics were calculated from the 30 minutes interval station observations for the year 2003. The temporal statistics describe the overall wetness and the variation in the wetness due to the meteorological conditions and soil properties. The range of mean soil moisture varies between 4 and 26 % vol/vol. Similarly, standard deviation of soil moisture ranges between 2 to 7 % vol/vol. Stations RG43 (Rain Gauge) and RG08 are the lower and higher extremes respectively for temporal mean as well as standard deviation. We find a fairly linear monotonic relationship suggesting that the stations with higher mean soil moisture values also exhibit higher standard deviation for 2003. This indicates that the wetter sites are more sensitive to the local climatic state (e. g. precipitation).



Fig 2.2: Time series of daily averaged soil moisture (% vol/vol) from some stations and the daily mean precipitation (mm/day) for the year 2003.

The autocorrelation was calculated at varying time lags for the daily averaged soil moisture values at all 14 sites. Figure 2.4 shows the autocorrelations versus the time lag in days. A constant line of e⁻¹ is also shown in the figure. The temporal autocorrelation length is determined as the time lag when the autocorrelation function crosses the e⁻¹ line. The temporal autocorrelation length ranges from 4 (for RG43) to 22 (for RG63) days. It can be inferred from this graph that the drier sites (e.g. RG43) have shorter autocorrelation duration than wetter sites (e.g. RG63, RG08). The time series numerically become stationary between 12 (for drier sites) and 66 (for wetter sites) days, which is

nearly three times the autocorrelation length. Such a high range for autocorrelation length among the sites could be attributed to the different stochastic rainfall, evapo-transpiration events and drainage rates at each site. The minimum 4-day autocorrelation lag time found here is important when considering the utility of satellite observations. Satellite revisit times are approximately 3 days; so satellites should capture most of the signature of soil moisture variations indicated by in-situ observations.



Fig. 2.3: Scatter plot of temporally averaged mean soil moisture (% vol/vol) versus standard deviation (% vol/vol) from all the 14 stations for the year 2003.



Fig. 2.4: Temporal autocorrelation function of soil moisture versus time lag for all 14 stations.

2.4 Spatial Characteristics

Recently, many satellite observed and model simulated soil moisture products have become available. In-situ observed soil moisture is the best data to validate such products. So, the spatial characteristics of in-situ observations are very important both to understand the physical hydrologic processes of the region as well as to perform the validation studies.



Fig. 2.5a: Time series of daily averaged spatial mean soil moisture (% vol/vol, left axis), standard deviation (% vol/vol, left axis) and coefficient of variation (right axis) derived from 14 stations.

Figure 2.5a shows the daily averaged spatial mean soil moisture time series of 14 stations for 2003. Note that the soil moisture time series for a few of the individual stations that contributed to this time series are shown in Figure 2.2. Figure 2.5a also shows the spatial standard deviation and the coefficient of variation (CV) corresponding to the spatial soil moisture mean time series. The coefficient of variation (CV) is a measure of dispersion of a probability distribution. It is defined as the ratio of the sample standard deviation to the sample mean (relative variability). When the mean value is very

small, the coefficient of variation is very sensitive to change in the standard deviation values. It can be very clearly noticed that the general behavior of the standard deviation of daily averaged soil moisture is very similar to that of the spatial mean soil moisture. The standard deviation peaks correspond to the peaks of mean soil moisture. The standard deviation varies between 0.5 to 7.5 % vol/vol whereas the mean soil moisture varies between 4 to 22 % vol/vol. Seasonality can be seen in the standard deviation values. The coefficient of variation shows an opposite behavior to the mean soil moisture. It can be inferred from this figure that the coefficient of variation is mostly controlled by rapid variation of daily mean soil moisture curve as compared to the standard deviation since the range of the mean soil moisture content is nearly 3 times greater than the range of the standard deviation. Similar results were found by Famiglietti et al. (1999) during the Southern Great Plains 97 (SGP97) Experiment. This can be attributed to the spatial differences in drainage rates, which implies different soil properties.

Figure 2.5b shows the scatter plot of the daily averaged mean soil moisture to the standard deviation of the soil moisture corresponding to Figure 2.5a. A positive correspondence is evident between the spatial mean and standard deviation values even though there is a good deal of scatter in the relationship. This supports similar results found by Hills and Reynolds (1969); Henninger et al. (1976) and Robinson and Dean (1993). The coefficient of variation (CV, relative variability) versus daily averaged spatial mean moisture content is shown in Figure 2.5c. Relative variability clearly decreases with increasing moisture content which is also noticed in the time series plot of Figure 2.5a. The range of the scatter depends on the relative variability of the spatial

mean and standard deviation since the coefficient of variation is calculated from them. This is consistent with the earlier findings of Bell et al. (1980) and Owe et al. (1982).



Fig. 2.5b: Scatter plot of standard deviation (% vol/vol) of moisture content versus mean moisture content (% vol/vol).

The cross-correlation information for soil moisture datasets between different sites at a certain spatial distance is important to understand the horizontal distribution of water in the watershed and how the processes at different sites are correlated. For a soil moisture comparison study, this cross correlation information is very helpful when the averaged in-situ measurements are compared with the coarse resolution satellite



Fig. 2.5c: Scatter plot of coefficient of variation versus mean moisture content (% vol/vol).

observations and model simulations. Figure 2.6 shows the cross correlations for soil moisture among all possible pairs of stations versus their separation distances. The cross correlation was calculated by taking the soil moisture time series from each pair of stations for 2003 with zero lag time. It is evident that the cross correlation between two sites decreases considerably (exponentially of order ~ 2) with the increase of their separation distance. But the correlation is still reasonably high (correlation coefficient ~ 0.35) between sites with a separation distance of more than 30 km. This result is very encouraging since most of the microwave satellite soil moisture observations (e.g.

AMSR-E) have 25 km by 25 km grid size and these sites in this watershed can be spatially averaged to validate the satellite observations.



Fig. 2.6: Scatter plot of cross correlation of soil moisture time series from stations versus distance (km) between stations.

2.4.1 Geostatistical Analysis

To assess the spatial correlation, a geostatistical method known as omni-directional variogram analysis was conducted for the data from the 14 stations. Geostatistical methods have previously been used by many authors (e.g., Mohanty et al., 2000; Entin et al., 2000; Western et al., 2004) to characterize the spatial pattern of soil moisture. The

variogram explains the semivariance values against the separation distance between two sites. Semivariance is a measure of dissimilarity of the soil moisture data between any two stations separated by a certain spatial distance and can mathematically be expressed as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[\theta(u_{\alpha} + h) - \theta(u_{\alpha}) \right]^2 \quad \dots \dots (2.1) \quad \text{(Isaaks and Srivastava, 1989)}$$

where, $\theta(u_{\alpha}) =$ soil moisture as a function of spatial location u_{α} , also known as the tail variable; $\theta(u_{\alpha} + h) =$ soil moisture at a lag (spatial distance) h, also known as the head variable, N(h) = number of pairs at lag h (± lag tolerance).

The semivariance for the soil moisture data for each pair of stations was calculated and all the semivariance values for all pair of stations were plotted against distance (variogram plot) at each instant in time (daily scale in this case). After trying several variogram models to fit through the scatter data and reviewing studies in the literature (Western et al. 1998; De Lannoy et al., 2006), we decided to use exponential variogram model to our soil moisture data. The exponential variogram can be mathematically shown as:

$$g(h) = c \left[1 - \exp\left(\frac{-3h}{a}\right) \right] \qquad \dots \dots (2.2) \qquad \text{(De Lannoy et al., 2006)}$$

where, g(h) = the fitted variogram, c = sill, a = range, h = lag distance. We neglect the nugget in this variogram model since the nugget represents small scale variability and relates the variance between pairs of stations with very small distances (Western et al. 2004). In our case, the station distances are relatively high and the nugget values are

found to be negligible in variogram plots for these stations. A detailed description of sill, range and nugget is given in Western et al. (2004).



Fig. 2.7a: Time series of daily averaged spatial mean soil moisture (% vol/vol, left axis), sill (left axis, % vol/vol) and range (correlation length) (m, right axis) for the year 2003.

The soil moisture over this watershed is found to be approximately stationary since the variograms reach the plateau (close to the estimated variance) for most of the days. The sill and range values were calculated for each day by least square minimization of the RMSE error between the observed and the fitted variogram. Figure 2.7a shows the

time evolution of the sill and range as well as the daily spatial mean soil moisture for 2003. Figures 2.7b and 2.7c show the corresponding scatter plots of the sill and range against the spatial mean of soil moisture respectively. There is a great deal of scatter in both Figures 2.7b and 2.7c because we are dealing with high resolution daily data over a



Fig. 2.7b: Scatter plot of sill (% vol/vol) versus daily averaged spatial mean moisture content (% vol/vol).

one-year period. The sill ranges between 5 to 47 % vol/vol and the correlation length ranges between 10.5 and 11.8 km in the watershed for 2003. These plots indicate that

both the sill and the range are related to the soil moisture spatial mean. From Figure 2.7a, it is evident that the sill (~ estimated variance) is higher during the spring when the watershed received more precipitation. This could be due to the differences in the soil moisture responses at different sites to the local precipitation events and different drainage rate. Figure 2.7b indicates that the sill is positively correlated with the spatial mean moisture with correlation coefficient of 0.66. In contrast to the sill, the correlation length (range) tends to be longer during the dry conditions as evident from Figures 2.7a and 2.7c. The correlation length shows a strong negative relationship with the spatial mean moisture with a correlation coefficient of -0.78. Beyond the correlation length, the correlation between the moisture of two stations becomes minimal. That could be attributed to differential precipitation and evapo-tanspiration rates at different sites in this watershed. Similar results (positive correlation between sill and spatial mean and negative correlation between range and spatial mean) were found by Western et al. (2004) for the Point Nepean sites during his study.

2.5 Conclusions

The purpose of this study was to statistically characterize the observed soil moisture from 14 in-situ stations over the Little River Watershed that fall within a 25 km by 25 km microwave satellite grid. The soil moisture measurements from the top 5 cm were available every 30 minutes for 2003. No strong seasonality was found in the soil moisture time series from the stations, but heavy continuous rainfall produced somewhat higher soil moisture during the spring. The mean soil moisture form all the stations is generally

low (range between 4 to 22 %vol/vol) for the whole year because of the predominance of well drained sandy loam soil. The stations with higher temporal mean soil moisture also



Fig. 2.7c: Scatter plot of range (m) versus daily averaged spatial mean moisture constant (% vol/vol).

showed higher temporal standard deviation. The autocorrelation time scale was found to range from 4 to 22 days with longer scales where soils drain more slowly. This autocorrelation indicates that less frequent temporal in-situ sampling (less than daily sampling) should be good enough to capture the soil moisture variability at each site and

can be useful to validate the satellite soil moisture products. The general behavior of the spatial mean and standard deviation of soil moisture from the 14 sites was similar, but the range of the spatial mean was higher than that of the spatial standard deviation for 2003. The coefficient of variation (relative variability) was negatively related to the spatial mean and was controlled mostly by the behavior of the spatial mean. The correlation between the soil moisture time series from in-situ stations was reasonably high, even though they are more than 30 km apart. This spatial correlation is very encouraging, given the footprint we expect from satellite soil moisture observations. These data should be effective to validate corresponding satellite soil moisture data to study the evolution of spatial patterns and the sill and correlation length were calculated from daily variograms. The range of sill lied between 5 to 47 % vol/vol, whereas the correlation length were found to be positively and negatively correlated with the mean spatial soil moisture respectively.

Our study was limited to the spatial and temporal characterizations of the in-situ soil moisture. It was not possible to provide any information on the local hydrologic processes and find factors that affected the soil moisture variability since we did not have soil moisture data from deeper soil layers, atmospheric, soil and vegetation data in this watershed. However, we can infer that given the relatively high correlation between watershed mean precipitation and soil moisture at individual stations, much of the interstation variability in decorrelation time scale and mean soil moisture must derive from differences in soil properties. The results from this study are encouraging regarding the validation of satellite observed and model simulated soil moisture products with in-situ observations at this frequency and spatial density.

Chapter 3. Evaluation of Current Satellite Soil Moisture Products: AMSR-E & LSMEM

Summary: An operational global soil moisture data product is currently generated from the observations of the Advanced Microwave Scanning Radiometer (AMSR-E) aboard NASA's Aqua satellite using the retrieval procedure described in Njoku et al. (2003). We have generated another soil moisture dataset from the same AMSR-E observed brightness temperature data using the Land Surface Microwave Emission Model (LSMEM) adopting a different retrieval procedure. This paper focuses on a comparison study of soil moisture estimates from the above two retrieval methods. The soil moisture data from current AMSR-E product and LSMEM are compared with the in-situ measured soil moisture datasets over the Little River Experimental Watershed (LREW), Georgia, USA for the year 2003. The comparison study was carried out separately for the AMSR-E daytime and night time overpasses. The LSMEM method performed better than the current operational AMSR-E retrieval algorithm in this study. Since both the soil moisture retrieval methods use the same radiative transfer algorithm, the differences in the soil moisture results appear to be due to differences in the model parameterizations and retrieval approaches. This study confirms that remote sensing data have the potential to provide useful hydrologic information, but the accuracy of the retrieved geophysical parameters could vary depending on the retrieval approaches. It cannot be concluded from this study whether the soil moisture retrieval by the LSMEM approach will perform better in other geographic, climatic or topographic conditions. Nevertheless, this study sheds light on the effects of different approaches for the retrieval of geophysical parameters, which may be useful for current and future satellite missions.

3.1 Introduction

Soil moisture is a critical element for both global water and energy budgets. Soil moisture controls the redistribution of rainfall into infiltration, surface runoff and evaporation at the earth surface (Delworth and Manabe, 1988; Vinnikov and Yeserkepova, 1991; Wagner et al., 2003). Soil moisture also has a strong effect on surface energy exchange (Prigent et al., 2005). Thus soil moisture trends may have a great impact on climate change over land (Seneviratne et al., 2006; Schär et al., 1999). Likewise, soil moisture is clearly important for the hydrologic applications like flood and drought monitoring, weather forecast, water management and agricultural plant growth.

Quantitative retrieval of accurate global soil moisture is always a challenge because the satellite, model and ground based data; each has its own limitations. Satellite remote sensing data products contain uncertainties due to imperfect instrument calibration and inversion algorithms, geophysical noise, representativeness error, and data transmission breakdowns (Zhan et al. 2004; Eymard et al., 1993). Also, the presence of moderate vegetation obscures the soil moisture signals, impeding accurate satellite measurements. Meanwhile, model-based soil moisture is strongly influenced by the atmospheric forcing components required to drive land surface models (e.g. precipitation,

incoming solar radiation, humidity, air temperature and wind speed); which mostly come from imperfect atmospheric general circulation models (GCMs). Error in land surface parameters like vegetation, soil color, and texture also influence the model soil moisture simulations. So any uncertainties associated with these forcing and land parameters strongly limit the accuracy of global model soil moisture results. It is difficult to characterize the errors associated with these atmospheric forcing data because of the lack of realism between different model schemes (Robock et al., 1998). Most of the land surface schemes produce very different soil moisture values even if forced with the same meteorological forcing data as it can be seen in many model inter-comparison studies, e. g. the Global Soil Wetness Project (GSWP; Dirmeyer et al., 1999); GSWP-2 (Dirmeyer et al., 2006; Guo and Dirmeyer, 2006) and the experiments of the Project to Intercompare Land-surface Parameterization Schemes (PILPS; Henderson-Sellers et al., 1995). There have been many field experiments conducted to measure hydrologic data over land including soil moisture, e.g. the Washita'92 experiment over the Little Washita watershed, Oklahoma (Jackson et al., 1993); SGP'97/99 experiments over Oklahoma (Jackson, 1997); SMEX02 over Iowa (http://hydrolab.arsusda.gov/smex02/); Portos 91/93 over Avignon, France (Wigneron et al., 1995) to name a few. Most of the field experiments use in-situ measurements to understand soil moisture processes at a local scale and to validate satellite and model estimations. These efforts are rife with problems of calibration, data intermittency, and strong limitations on spatial and temporal coverage. The sparseness of in-situ measurements means they do not represent the whole globe (Reichle et al., 2004).

The main objective of this paper is the retrieval of soil moisture from a passive microwave instrument in orbit. Passive microwave remote sensing has been widely used to extract soil moisture data (due to different dielectric constant of soil and water) since the launch of the Scanning Multichannel Microwave Radiometer (SMMR; 6.6, 10.7 and 18.0 GHz channels) aboard Seasat and Nimbus-7 in 1978. Vinnikov et al. (1999) concluded that both the polarization difference and emissivity at the horizontal polarization below 18 GHz can be used for soil moisture information over grass or crop with no dense vegetation from his SMMR brightness temperature comparison study with the Global Soil Moisture Data Bank (GSMDB, Robock et al., 2000) data over Illinois, USA. Special Sensor Microwave/Imager (SSM/I; 19.4 GHz channel) and Tropical Rainfall Measuring Mission/Microwave Imager (TRMM/TMI; 10, 19 and 21 GHz channel) have been quite useful in providing proxies for soil moisture measurements in spite of not carrying any dedicated soil moisture mapping sensors. Wen et al., 2005 retrieved soil moisture using dual polarization 19.4 GHz algorithm from SSM/I sensor over corn and soybean fields and found the standard error of estimate of 5.49% over the 3-week field experiment period. They concluded that soil moisture retrieval was feasible using SSM/I data, but the accuracy depended upon the levels of vegetation and atmospheric precipitable water. Gao et al. (2006) used a single polarization radiative transfer model to derive soil moisture from TRMM/TMI over the Southern United States from 1998 to 2002. Bindlish et al. (2003) also used a single TMI X-band frequency radiative transfer model to produce soil moisture data over the Southern United States. Lee and Anagnostou (2004) combined the data from active precipitation sensor radar with the passive microwave data from TRMM/TMI and TRMM/PR to retrieve soil moisture from TRMM sensor. Pursuit of all these approaches has culminated with the launch of a passive microwave sensor at frequencies useful for the retrieval of soil moisture, namely the Advanced Microwave Scanning Radiometer (AMSR) aboard ADEOS II and AMSR– Earth Observing System (AMSR-E) aboard the Aqua satellite in 2002. Soil moisture is an official product from AMSR-E and is being continuously generated since 2002 using a multi-frequency radiative transfer algorithm of Njoku et al. (2003). Paloscia et al. (2006) retrieved AMSR-E soil moisture using an algorithm based on a simplified radiative transfer (tau-omega) model. McCabe et al. (2005a, b) also retrieved soil moisture from AMSR-E using a single frequency channel radiative transfer algorithm which was previously used by Gao et al. (2004a, 2006) for soil moisture retrievals from Electronically Scanned Thinned Array Radiometer (ESTAR) observations and TRMM sensors respectively.

Most of the above research studies have indicated the sensitivity of the soil moisture estimation to dense vegetation at higher frequencies. Recent research (Pellarin et al., 2003; Gao et al., 2004b) suggests that the 1.4 GHz (L-band) channel is optimal for soil moisture retrieval due to its deeper penetration through the earth's soil and lesser sensitivity to surface roughness and vegetation. Also the atmospheric effects are very negligible at this frequency. So, the next generation passive microwave satellite programs are incorporating 1.4 GHz channel for soil moisture mapping, e. g. European SMOS (Soil Moisture and Ocean Salinity) mission (Kerr et al., 2001) which is scheduled for launch in 2008.

This paper discusses the estimation of soil moisture from AMSR-E 10.7 GHz frequency channel using a forward model. The effect of dense vegetation will somewhat be there in the soil moisture results at 10.7 GHz frequency. However, that is the lowest frequency AMSR-E channel information available right now though theoretically not the best channel for soil moisture estimation. So, the effect of vegetation can not be totally removed for any AMSR-E soil moisture estimation method. This paper also compares this forward model soil moisture results with the current operational AMSR-E soil moisture product and ground based soil moisture measurements. The current operational soil moisture product uses an "inverse method" for multi-parameter (soil moisture and vegetation water content) retrievals whereas the radiative transfer model in this study uses an "iterative parameter fitting to a single channel single polarization forward brightness temperature model" for single parameter (soil moisture) estimation. The watershed considered for this study is one of the four selected watersheds to calibrate and validate the current operational AMSR-E soil moisture product and it contains moderate vegetation which is confirmed from the surface type map used by the operational product as well as by calculating the MODIS vegetation water content scaled for a 25 km grid. So, the effect of vegetation at this watershed is not very significant and the soil moisture estimation is possible here. However, the effect of dense vegetation is an issue for global soil moisture estimation. Hence, a polarization ratio method is discussed later in the discussion section to account for the dense vegetation and mask those areas where the operational soil moisture estimation is not possible by this forward model. The layout of this chapter is as follows: Section 3.2 highlights key aspects of radiative transfer theory.

Section 3.3 describes the geographic site and different data products used for this study. Section 3.4 presents the results of our analysis. Section 3.5 discusses the retrieval procedures; provides some possible explanations for the results and gives a brief description on the operational aspect of soil moisture retrieval. Conclusions follow in section 3.6.

3.2 Soil Moisture Retrievals and Radiative Transfer Theory

The current operational AMSR-E soil moisture product is based on an inverse soil moisture method discussed in Njoku and Chan (2006). The other product considered in this study uses a forward model for soil moisture estimation. Both the inverse and forward models are based on simplified radiative transfer theory and assumption for minimal influence of atmospheric contribution. So, it's necessary to revisit the relevant radiative transfer theory in the context of this paper.

The earth's brightness temperature (T_B) observed at the top of the atmosphere (TOA) at a given incidence angle and frequency (as a satellite observes) is a contribution of signals from soil, vegetation, standing water, snow cover and atmosphere. So, the satellite measured brightness temperature $(T_{B,p})$ can be expressed (Njoku et al., 2003) as:

$$T_{B,p} = T_{au} + e^{-\tau_{at}} [T_{b,p} + (1 - \varepsilon_p)T_{ad}]$$
(3.1)

where the subscript p denotes either vertical or horizontal polarization, T_{au} is the upwelling atmospheric emission, τ_{at} is the atmospheric opacity along the viewing path, \mathcal{E}_p is the combined effective surface emissivity of vegetation, bare soil and open water

(by Kirchoff's law, $\mathcal{E}_p = 1 - r_p$, assuming that the transmissivity is negligible and r_p is the surface reflectivity), T_{ad} is the downwelling atmospheric and space-background emission at the top of the vegetation, $T_{b,p}$ is the effective surface brightness temperature of the combination of vegetation, bare soil and open water within a satellite pixel at the top of the vegetation, and can be defined as (Gao et al., 2006):

$$T_{b,p} = (1 - C_v - C_w)T_{bs,p} + C_v T_{bv,p} + C_w T_{bw,p}$$
(3.2)

where $T_{bs,p}$ is the bare soil brightness temperature, $T_{bv,p}$ is vegetation covered soil brightness temperature, $T_{bw,p}$ is the water brightness temperature, C_v is the fraction of vegetation coverage and C_w is the fractional coverage of water within a satellite pixel.

 $T_{bs,p}$, $T_{bv,p}$ and $T_{bw,p}$ can be expressed (Mo et al. 1982; Kerr and Njoku, 1990) as:

$$T_{bs,p} = \mathcal{E}_{s,p} T_s \tag{3.3}$$

$$T_{bv,p} = \varepsilon_{s,p} T_s e^{-\tau_c} + T_c (1 - \omega_p) (1 - e^{-\tau_c}) (1 + (1 - \varepsilon_{s,p}) e^{-\tau_c})$$
(3.4)

$$T_{bw,p} = \mathcal{E}_{w,p} T_w \tag{3.5}$$

where T_s is the effective soil temperature (the effective temperature is the weightedaverage temperature over the microwave penetration depth in the medium), $\mathcal{E}_{s,p}$ is the emissivity of the bare soil, T_c is the canopy (vegetation) temperature, τ_c is the vegetation opacity, ω_p is the vegetation single scattering albedo, $\mathcal{E}_{w,p}$ is the water emissivity and T_w is the water temperature. The Equations (3.1) to (3.5) provide the basis of radiative transfer theory. For the satellite soil moisture estimations, these equations are simplified using a few important assumptions. At C and X-band channels, the atmospheric contribution is relatively small (Drusch et al., 2001). So, the atmospheric component in Equation (3.1) is assumed constant for the atmospheric correction in the soil moisture derivation. Also, many radiative transfer models ignore the difference between the surface temperature (T_s) and canopy temperature (T_c) assuming they are equal. Multiple scattering from the surface is also neglected in most of the soil moisture estimation models.

The vegetation opacity τ_c depends on incidence angle (θ), vegetation water content (w_c) and the vegetation structure parameter (b-parameter (b_p), an empirical variable) that is a function of frequency (γ) and vegetation type. It can mathematically be expressed as (Njoku and Chan, 2006):

$$\tau_c = \frac{b_p(\gamma) f(w_c)}{\cos \theta}$$
(3.6)

where $f(w_c)$ is a function of vegetation water content (w_c) .

The presence of vegetation cover adds a source of error to the soil moisture retrieval. Njoku and Li (1999) have studied the effect of vegetation on soil moisture estimation and have concluded that the satellite soil moisture estimation is not very reliable with vegetation water contents greater than 1.5 kg m⁻². The sensitivity of brightness temperature to soil moisture also decreases with increasing frequency in the presence of vegetation. This has been shown in Prigent et al. (2005) by comparing the satellite data with the in-situ observations. So, many radiative transfer models use

brightness temperature data from lower frequencies (lower than 18 GHz frequency) to estimate soil moisture.

The rough surface reflectivity ($r_{s,p}$) depends on the surface dielectric constant (Schmugge, 1990) and roughness (Njoku et al., 2003). This can mathematically be shown as (Njoku and Chan, 2006; Wang and Choudhury, 1981):

$$r_{s,p} = [(1-Q)r_{o,p} + Qr_{o,q}]\exp(-h)$$
(3.7)

where $r_{o,p}$ is the smooth soil reflectivity, p, q = two orthogonal polarizations, h = surface roughness parameter which is a function of wave number (k) and RMS surface height (s), Q is a constant and function of s and horizontal correlation length (l).

The dielectric constant of water is ~ 80 whereas for soil it is ~ 4 . This is the basis of differentiating water from soil (Ulaby et al., 1986; Wang and Schmugge, 1980; Dobson et al., 1985). This difference is also detectable by passive microwave sensors (Njoku and Kong, 1977).

3.3 Description of Study Area and Data Sets

3.3.1 Study Area

A detailed description of the study area is described in Section 2.2.1. Figure 2.1 shows an image of the study area along with the in-situ measuring stations.

3.3.2 LREW In-Situ Observation Data

A detailed description of the in-situ observation datasets is described in Section 2.2.2.

3.3.3 AMSR-E Soil Moisture Retrieval Data

AMSR-E is a passive microwave radiometer launched aboard NASA's Aqua Satellite (Parkinson, 2003; http://nsidc.org/daac/amsre/index.html). The local crossing time of AMSR-E is 0130 LST (Descending pass) and 1330 LST (Ascending pass). This instrument measures microwave radiation (brightness temperatures) at 6 frequencies ranging from 6.9 to 89.0 GHz (both horizontal and vertical polarized radiation at each frequency for a total of 12 channels; Kawanishi et al., 2003). The AMSR-E C- (6.9 GHz) and X-band (10.7 GHz) channels are strongly related to land surface soil moisture variable (Njoku et al., 2003). However, the C-band brightness temperature measurements have been affected by Radio Frequency Interference (RFI) near populated urban areas (Li et al., 2004).

The current official AMSR-E land products have gone through many significant modifications since its earlier version described in Njoku (1999) and Njoku et al. (2003). The current algorithm uses an inverse model described in Njoku and Chan (2006) and retrieves multiple land surface variables from multi-frequency channel brightness temperature data. This algorithm uses only 10.7 and 18.7 GHz V and H polarization data and does not calculate the land surface temperature because of Radio Frequency Interference (RFI) contamination in the 6.9 GHz channels. It uses polarization ratio (PR) of multiple channels instead of single channel brightness temperature (T_B) data because PR eliminates or reduces surface temperature effect on the algorithm for the vegetation

water content and soil moisture retrieval (Kerr and Njoku, 1990). PR is the difference between the vertical and horizontal brightness temperature values at a given frequency divided by their sum (Njoku et al. 2003).

By considering only the unpolarized light, neglecting scattering albedo and $T_s \approx T_c$, and using Equations (3.6) and (3.7), Equation (3.4) can be simplified to

$$T_{bv,p} = T_s \{ 1 - [1 - Q)r_{o,p} + Qr_{o,q}] \exp(-\alpha g) \}$$
(3.8)

where α = a coefficient and g is a single parameter to account for the vegetation and roughness parameter together which is expressed as:

$$\alpha g = h + \frac{2b_p w_c}{\cos \theta} \tag{3.9}$$

The algorithm first computes the variable g using an approximated version of Equation (3.8) as shown here (Njoku and Chan, 2006):

$$g = \frac{1}{\beta \alpha} \ln \left\{ \frac{A(1 - 2Q)}{\zeta} \right\}$$
(3.10)

where A is a function of soil moisture and expressed as $A = (e_{o,v} - e_{o,h})/(e_{o,v} + e_{o,h})$, $e_{o,v}$ and $e_{o,h}$ are smooth soil V- and H-polarization emissivities, ζ is the PR of brightness temperatures, β is a coefficient.

This 'g' variable is then used as a correction factor in soil moisture computation which is calculated using the deviation of the PR of the 10.7 GHz channel from a baseline value. The baseline values are fixed from the monthly minima values at each

grid

(http://nsidc.org/data/docs/daac/ae land l2b soil moisture.gd.html#derivtechnique).

Daily Level-2B and Level-3 land products are available from the National Snow and Ice Data Center (NSIDC) from June 18, 2002. The level 2 land products are composited daily to make global maps (Level-3 land product), separating ascending and descending passes so that diurnal effects can be evaluated. Soil moisture is not retrievable where significant fractions of snow cover, frozen ground, dense vegetation, precipitation, open water, or mountainous terrain occur within the sensor footprint (Njoku et al., 2003). The products are generated on an earth-fixed grid with ~25- km nominal grid spacing (Kawanishi et al., 2003).

The daily AMSR-E Level-3 land surface data (referred to as the AE_Land3 product) were collected from NSIDC (Njoku, 2004; <u>http://nsidc.org/data/amsre/;</u> 2 files per day pertaining to ascending and descending pass separately) for the period January 1 to December 31, 2003. The AMSR-E data were processed to reproject the data from the 25 km ease-grids to 0.25 degree lat-lon grids and to extract the top layer daily soil moisture from the HDF-EOS files.

3.3.4 LSMEM Radiative Transfer Model Soil Moisture Data

The Land Surface Microwave Emission Model (LSMEM) used in this study is based on the radiative transfer theory described in section 3.2 which uses the equations from Kerr and Njoku (1990). This model uses single frequency single polarization brightness temperature data to derive soil moisture, as opposed to the multi-frequency retrieval of
multiple parameters applied by Njoku et al. (2003). The other two most important parameters, surface temperature and vegetation water content have been provided as input data to this model. The LSMEM model uses an iterative procedure to find the numerical solution by minimizing the differences between the observed ($T_{B_i}^{obs}$) and the computed ($\phi_i(x)$) brightness temperature for soil moisture which is mathematically expressed in Equation (3.11). The LSMEM model has performed consistently well in estimating surface soil moisture using observed brightness temperature data from different sensors like ESTAR (Gao et al., 2004a), TRMM/TMI (Gao et al., 2006) and AMSR-E (McCabe et al. 2005a; 2005b).

For LSMEM (Gao et al., 2004):

$$\chi = T_{B_i}^{obs} - \phi_i(x) \tag{3.11}$$

where i = 10.7 GHz frequency channel (one channel) and $x = \{m_v\}, m_v = \text{soil moisture}.$

The smooth wet soil dielectric constant was calculated in the model after Wang and Schmugge (1980). The smooth wet soil reflectivity was derived from the dielectric constants using the Fresnel expressions. Then the surface roughness was given a constant value of 0.3 in the model (Choudhury et al., 1979) which is typical for a medium rough surface. This is a widely used approach to account for the surface roughness in the calculation of the brightness temperature (Drusch et al., 2004). Again the measurements over an AMSR-E footprint scale will average many terrain types; hence a constant surface roughness can be a good approximation to calculate the brightness temperature (Njoku et al., 2003). Finally, the rough soil emissivity ($\mathcal{E}_{s,p}$) was calculated from the smooth soil reflectivity and soil dielectric constants using the semi-empirical formulation of Wang and Choudhury (1981).

The spatial distribution of the vegetation water content (VWC) at 1 km resolution was calculated using the relationship among the MODIS (Moderate-resolution Imaging Spectrometer) based LAI (Leaf Area Index), foliar and stem biomass and their relative water content as described in Rodell et al. (2005). The vegetation water content ranged between 0 to 1.1 kg m⁻² in January (winter) and 0 to 4.06 kg m⁻² in July (summer) over the whole watershed. Vegetation may introduce some error in the soil moisture estimation. But when we rescaled those 1 km resolution VWC data to AMSR-E 0.25 degree grid, the vegetation water content was less than 1.5 kg m⁻² over the watershed throughout the year. The b-parameter data were not available for individual vegetation cover types. A constant value of 0.7 at X-band was assigned for the b-parameter based on the Figure 4 of Jackson and Schmugge (1991). Then the vegetation opacity/optical depth (τ_c) in LSMEM was derived using the vegetation water content and b-parameter. The vegetation single scattering albedo was given a constant value of 0.07 (according to Ulaby et al. (1983) and Pampaloni and Paloscia (1986).

The soil texture (sand fraction, clay fraction and bulk density) was derived from the State Soil Geographic (STATSGO) database (Miller and White, 1998). The water fractional coverage and the vegetation fractional coverage were taken from the 1-km MODIS land cover data (Hansen et al., 2000) and the Normalized Difference Vegetation index (NDVI) data using the method described by Chang and Wetzel (1991).

Both the soil and vegetation temperature were taken from the surface temperature simulations of the Variable Infiltration Capacity (VIC) land surface scheme (Liang et al., 1994; 1999) since VIC has a single surface layer. The model was run at one hour time step with North American Land Data Assimilation System (NLDAS) forcing input data. NLDAS incorporates in-situ gauge, radar and satellite observations over the NCEP Eta Data Assimilation System (EDAS) baseline analysis to produce the forcing data over the North America (Cosgrove et al., 2003). NLDAS data include air temperature and specific humidity at 2 m height, wind speed at 10 m height, surface pressure, downward shortwave and longwave radiation, convective available potential energy, skin temperature, total and convective precipitation and photosynthetically active radiation. The input parameters for VIC include the vegetation (land cover) and soil (texture, color) data. The University of Maryland's (UMD) 1 km global land cover product (Hansen et al., 2000) was used as land cover input. This dataset has a total of 13 land cover classes excluding water bodies. The land-sea mask was also generated from this vegetation classification map. The surface temperature data from VIC simulations only matching to the AMSR-E overpass times were considered for the LSMEM model input. The VIC model has performed well in many previous model inter-comparison and validation studies (Mitchel et al., 2004).

The input variables required for the LSMEM run have been summarized in Table 3.1 and a detailed description of these variables can be found in Gao et al. (2006). Figure 3.1 shows the flowchart of the LSMEM forward model for soil moisture estimation. The sensor information (for AMSR-E 10.7 GHz frequency channel), state variables and state

and atmospheric contributions were provided as input data to the model as shown in The LSMEM model was simulated starting with an initial guess of Figure 3.1. antecedent soil moisture condition. The LSMEM predicted brightness temperature was compared with the AMSR-E 10.7 GHz observed brightness temperature and the initial guess of soil moisture was increased iteratively until the model predicted and satellite observed brightness temperatures converged to within a certain threshold value. Even though we knew that the sensitivity of 6.9 GHz signal for soil moisture detection is higher than that of the 10.7 GHz frequency, we used AMSR-E 10.7 GHz frequency brightness temperature data for LSMEM forward model because the 6.9 GHz frequency data have been affected by the Radio Frequency Interference (RFI) (Njoku et al., 2005). Also the horizontal polarization signal is more sensitive to the soil moisture than vertical polarization (Njoku and Li, 1999). So, we preferred horizontal polarization signal than that of the vertical polarization at 10.7 GHz frequency for LSMEM estimation. The penetration depth of the 10.7H GHz channel is small (may be less than 1 cm depth). So, the LSMEM soil moisture information is also from less than the top 1 cm soil layer (\sim skin soil moisture).

Table 3.1: Summary of Input datasets used for the LSMEM Model

Input Data	Parameters	Value	Data source	Reference
Sensor Information	Incidence Angle	54.8°		Njoku et al. (2003)
	Modeling Frequency	10.65 GHz		Njoku et al. (2003)
AMSR-E Observation	Brightness Temperature (T _B)	Level 3 Global Geophysical Retrieval	National Snow and Ice Data Center (NSIDC)	Njoku et al. (2003)
Atmospheric Contribution	Optical depth Emitted T _B	0.014 6.0 K	MOLTS/radiative transfer	Drusch et al. (2001)
Surface parameters	Sand and Clay fraction; soil bulk density	Spatially distributed constants	STATSGO	Miller and White (1998)
	Soil surface roughness	0.3		Choudhury et al. (1979)
	Water coverage	Spatially distributed constants	MODIS	Hansen et al. (2000)
Vegetation parameters	Vegetation coverage	Spatially distributed, monthly values	NLDAS greenness fraction	Chang and Wetzel (1991)
	Vegetation water content	Spatially distributed, monthly values	Calculated from MODIS LAI and land cover types	Rodell et al. (2005)
	Vegetation b- parameter	Constants based on classification		Jackson and Schmugge (1991)
	Vegetation single scattering albedo	0.07		Pampaloni and Paloscia (1986); Ulaby et al. (1983)
State variable	Surface temperature	Spatially distributed, hourly values	VIC land surface model output	Liang et al. (1994,1999)



Fig. 3.1: Flowchart of the LSMEM Soil Moisture Retrieval (from Gao et al., 2004a)

3.4 Results and Analysis

This section presents the comparison results carried out among the watershed in-situ observations, current AMSR-E soil moisture and LSMEM soil moisture data. Before showing the comparison results, it's important to discuss issues associated with spatial, temporal and vertical resolution when comparing datasets from different sources. The spatial resolution of the AMSR-E and LSMEM soil moisture results considered here are at 0.25^{0} by 0.25^{0} resolution. However, the in-situ observations are from stations (point locations). We assumed here that the spatial average of observed data from 17 in-situ sites could represent the AMSR-E and LSMEM 0.25^{0} by 0.25^{0} grid reasonably well. The assumption was based on the high spatial correlation found among in-situ observation sites in Chapter 2. Many previous satellite and in-situ soil moisture comparison studies have been carried out with as little as one station available within a satellite pixel (Vinnikov and Yeserkepova, 1991; Entin et al., 2000; Prigent et al., 2005; Reichle et al., 2004). By comparison, we thought our assumption of representing the AMSR-E pixel with the spatially averaged data from 17 in-situ sites is reasonable.

The AMSR-E satellite overpass frequency over any geographic region in the midlatitudes area is in every 2 to 3 days (Njoku et al., 2003). The LSMEM soil moisture from AMSR-E could be derived only on the AMSR-E overpass days. The instantaneous measurements were considered at the hour that matched closely with the AMSR-E overpass time. For precipitation, we used the in-situ cumulative precipitation data over the previous 24 hours from the satellite overpass time instead of an instantaneous precipitation rate. AMSR-E soil moisture data were available separately for the ascending and descending passes. So, we also considered the other two datasets corresponding to the ascending and descending pass of AMSR-E. That gave us an opportunity to compare the data at daytime versus night time to check the consistency of the datasets/approaches. The in-situ observations were available for the year 2003 only, so the comparison studies are only for that year. We thought the comparison at such a high temporal resolution for one year is reasonable enough to draw any conclusive statistics on these results. In many previous studies (Prigent et al., 2005; Reichle et al., 2004; Cashion et al., 2005), researchers have used monthly averaged multi-year soil moisture data because (i) high temporal resolution field observation datasets were not available, (ii) they addressed the comparison in the long term and large scale in a 'climate' sense and (iii) they avoided instantaneous data due to the strong variability and noise associated with short time scales. Since our study focuses solely on high temporal and small spatial scales, we used the daily instantaneous data.

The AMSR-E (top ~ 1 cm soil layer), LSMEM (top ~ 1 cm) and in-situ soil moisture (top 5 cm) data come from different soil depths. We used only the volume percentage soil moisture values in this study. Calvet et al. (1999) and Wigneron et al. (1995) extensively studied the vertical profile of soil moisture and found that the surface soil moisture (from ~ 0.5 cm layer) was well correlated with the 10 cm soil layer moisture even at short time scales. Prigent et al. (2005) compared the soil moisture datasets from different depths after they found that the soil moisture variability from top 5 cm was highly correlated with that from the top 20 cm using the Global Soil Moisture Data Bank. It is a common practice to compare the soil moisture data from somewhat different depths because of the lack of soil moisture data from equivalent depths from different sources. We made a similar assumption to Prigent et al. (2005) when comparing the three datasets in this study.

We decided to perform the comparison separately for the daytime and night time satellite overpasses. Figure 3.2a shows the scatter plot of the daytime against night time in-situ measured soil moisture data. A 1:1 line is shown in the plot for reference. We can clearly see that the observed soil moisture data are well spread within range from 5 to 25 % vol/vol. Most of the points fall on and around the 1:1 reference line. Those are the days of little or no precipitation/irrigation. There are, however, many outliers in this plot. Those outliers occur because of heavy precipitation or irrigation between night time and daytime measurements as confirmed from the 30 minute precipitation/irrigation measurements (not shown in this plot). This gives us confidence in the in-situ measurement instruments and datasets. Figures 3.2b and 3.2c show similar plots as in Figure 3.2a, but for the AMSR-E and LSMEM results respectively. In both the plots, it can be seen that the points also fall mainly on and around 1:1 line. But in the case of AMSR-E (Figure 3.2b), the dynamic range of the dataset is very low, within 10 to 18 % vol/vol values throughout the year. Also, it has few outliers, suggesting that these retrieval values are not sensitive to the precipitation/irrigation events. For LSMEM, we get both high dynamic range (5 to 25 % vol/vol) and many outliers corresponding to the precipitation/irrigation events. These outliers indicate that the LSMEM results are sensitive to the precipitation/irrigation events that happen between the AM and PM overpasses (0130 and 1330 LST).





Fig. 3.2: Plots of daytime versus night time soil moisture from (a) in-situ measurements, (b) AMSR-E and (c) LSMEM retrievals for the year 2003. Daytime and night time data correspond to the AMSR-E ascending (1330 LST) and descending (130 LST) overpass time respectively.

Figure 3.3a shows the daily time series of the daytime AMSR-E, LSMEM and insitu instantaneous soil moisture for 2003. The x-axis is the time axis and the left y-axis represents the volumetric soil moisture as a percentage. There are many missing days of

AMSR-E soil moisture data and hence for LSMEM as well. That is because there are no AMSR-E satellite overpasses on those days. The LSMEM time series closely follows the in-situ observations. In contrast, AMSR-E soil moisture data show limited variability. The level of agreement of AMSR-E soil moisture with the other two datasets is highest in winter months (January to April and December). Figure 3.3a also shows the corresponding in-situ cumulative precipitation/irrigation from the previous 24 hours in millimeters on the right y-axis. We can clearly identify the soil moisture peaks in the insitu and the LSMEM soil moisture datasets corresponding to the precipitation peaks, although the in-situ peaks are higher. Stronger response of in-situ soil moisture to the precipitation/irrigation could be because the measurements of both the parameters have been carried out at the same stations or because of shortcomings in the LSMEM estimation algorithm. But the AMSR-E soil moisture data usually do not respond strongly to the precipitation events. It suggests that the AMSR-E algorithm is less sensitive to atmospheric forcing like precipitation. Figure 3.3b shows the scatter plot of the in-situ soil moisture versus AMSR-E corresponding to the data in Figure 3.3a. ±1 standard deviation of in-situ measurements is also shown in this figure. It can be noticed that the level of agreement between these two datasets is low except on a few days mainly during winter (as noted for Figure 3.3a). The RMS error was found to be 8.1 % vol/vol with a correlation coefficient of 0.61. Figure 3.3c is similar to Figure 3.3b, but for in-situ and LSMEM soil moisture data. Here, a significant level of agreement between the two datasets can be seen, although the LSMEM soil moisture has consistent dry bias. The RMS error is 3.31 % vol/vol with a correlation of 0.78.







Fig. 3.3: (a) Time series plot of the in-situ measured, AMSR-E and LSMEM retrieved soil moisture at the AMSR-E ascending overpass time along with daily average precipitation and irrigation for the year 2003. Scatter plot of the corresponding (b) in-situ and AMSR-E soil moisture, (c) in-situ and LSMEM soil moisture data. Horizontal bars in the scatter plot show ± 1 standard deviation of the in-situ measurements. A 1:1 line is included in the scatter plots for reference.

Figures 3.4a, 3.4b and 3.4c show plots similar to Figures 3.3a, 3.3b and 3.3c respectively, but for night time. The cumulative precipitation peaks are smaller in Figure 3.4a than in Figure 3.3a. Likewise the night time in-situ soil moisture peaks are lower in Figure 3.4a. So, our explanation in Figure 3.2a that the outliers are due to the precipitation/irrigation events between the night time and daytime measurements is validated. The comparison among the night time data is very similar to what we found in Figure 3.3. The RMS error for AMSR-E retrieval is 6.8 % vol/vol with a correlation coefficient of 0.54 whereas the RMS error for the LSMEM estimation is found to be 3.5 % vol/vol with a correlation coefficient of 0.81. Our comparison results for LSMEM results agree with the previous comparison study performed by McCabe et al. (2005b) over the Walnut Creek catchment in Iowa. They found 4.1 % vol/vol RMS error with correlation coefficient of 0.87 for the LSMEM results, but their comparison study was limited to only 8 days and they used an average of day and night satellite overpasses. Our results imply that the LSMEM derived soil moisture performed better than the AMSR-E retrieved soil moisture irrespective of the time of estimation.

3.5 Discussion

In this section, we try to give possible explanations for the differences found in AMSR-E and LSMEM soil moisture results. We also discuss the possibility of operational production of LSMEM soil moisture from the AMSR-E data. Both the approaches use the same radiative transfer equations and physical models based on soil and vegetation emissivity as described in section 3.2, but the current operational AMSR-E algorithm







Fig. 3.4: Same as Figure 3.3, but at the AMSR-E descending overpass time.

uses an inverse approach whereas the LSMEM model uses a forward model approach. A summarized description of the current AMSR-E and LSMEM approaches is provided in Table 3.2. As we stated earlier, the current AMSR-E retrieval procedure retrieves multiple parameters (soil moisture and vegetation water content) using multi-frequency brightness temperature data, whereas the LSMEM method generates a single parameter (soil moisture) using a single frequency and single polarization brightness temperature

data. Since both the approaches use AMSR-E 10.7 or higher frequencies for soil moisture estimation, the effect of vegetation is a concerning factor in both the cases. Practically, it is not possible to completely reduce the effect of vegetation and atmosphere on soil moisture estimation at this high frequency channel. Nevertheless, the goal in this study is to estimate LSMEM soil moisture with whatever best (10.7 GHz) frequency brightness temperature data available from satellites right at the moment and compare that estimation with the current operational soil moisture product and discuss the comparison results.

	Current AMSR-E Retrieval	LSMEM Retrieval
Physical	Radiative transfer equations	Radiative transfer
basis	for the AMSR-E	equations for the
	channels 10.7 and 18.7 GHz.	AMSR-E 10.7H GHz
		channel
Inversion	Numerical solution of the	Numerical solution of
Approach	simultaneous equations based on	one equation based on
	inversion of microwave radiative	the iteration technique
	transfer theory for multiple parameter	for single parameter
	retrieval	retrieval
Input data	AMSR-E 10.7 H/V and 18.7 H/V	AMSR-E 10.7H
	brightness temperature; sensor (viewing	brightness temperature,
	angle, frequency, polarization),	surface temperature,
	atmospheric (opacity, temperature lapse	vegetation water content
	rate precipitable water, cloud liquid	(for details, see Table
	water, air temperature), vegetation	3.1)
	(single scattering albedo, b-parameter,	
	vegetation type) and soil (roughness	
	coefficient, bulk density, soil texture and	
	type) parameters.	
Output(s)	Surface soil moisture and vegetation	Surface soil moisture
	water content	

The limitations and sources of error for the current AMSR-E retrievals have been extensively discussed in Njoku and Chan (2006). We provide a brief summary of those here in this section. The footprints are different for different AMSR-E channels. Hence, the multi-channel sensor footprints are co-registered and processed to similar spatial resolution for the production of AMSR-E land data products since the current operational algorithm uses multiple channels/polarizations. The error in co-registration for multiple channels might add some error to the current AMSR-E operational algorithm. The relative calibration biases between different channels and different polarizations in a single channel can be an important factor in this multi-channel soil moisture retrieval. Many assumptions and simplifications such as ignoring the single scattering albedo; the dependency of b-parameter on vegetation types and the assumption of single polarization for vegetation attenuation and emission can be sources of error for this retrieval approach. The value of b-parameter used in this AMSR-E retrieval is derived at 1.4 GHz at field scale which can be different at satellite scales and frequencies.

The surface temperature is not an input to the current AMSR-E algorithm. Even though the PR reduces the effect of surface temperature on the algorithm considerably, a quantitative assessment is required to know how significant the surface temperature is in the soil moisture estimation. The current algorithm does not depend on any ancillary data whereas the LSMEM estimation depends significantly on the input surface temperature and the vegetation water content. Thus, the error in the input variables can introduce large errors in the LSMEM soil moisture results. Since we have in-situ measured surface temperature data, we compared the input VIC model instant surface temperature to the insitu observations at the time of satellite overpass. Figure 3.5 shows comparison plots for the surface temperature for the daytime and night time satellite overpasses. As expected, the absolute values as well as the variability of surface temperature are lower in the night time (Figure 3.5b) as compared to those of the daytime (Figure 3.5a). There is significant agreement between the VIC model simulated and in-situ measured surface temperature in both daytime and night time, except when temperatures drop below 280 K. For the daytime, we found the RMS error in VIC surface temperature as 3.31 K with correlation coefficient of 0.89. For the night time, these values were 2.62 K and 0.94 respectively.





Fig. 3.5: VIC surface temperature validation with the in-situ surface temperature for the year 2003 at AMSR-E (a) ascending and (b) descending overpass time. Horizontal bars in the scatter plot show ± 1 standard deviation of the in-situ measurements. A 1:1 line is included in the scatter plots for reference.

These statistics agree well with previous studies carried out by Mitchell et al. (2004). They found RMS errors ranging from 3.3 to 4.3 K in the VIC surface temperature when compared to Geostationary Operational Environmental Satellite (GOES) and Atmospheric Radiation Measurement (ARM) Cloud and Radiation Test Bed (CART) surface temperatures over the southern Great Plains of USA. The high degree of accuracy in VIC surface temperature contributes towards the accuracy of LSMEM soil moisture estimation.

LSMEM model can be used to estimate soil moisture operationally from AMSR-E at continental and global scale. Gao et al. (2006) addressed the influence of precipitation, vegetation and snow on the operational production of LSMEM soil moisture from TRMM satellite at continental scale and have used few quality flags to get rid of those problems. Similar kind of quality flags can be used in this case to derive LSMEM soil moisture from AMSR-E data operationally since both the satellites carry the same frequency channel (~ 10.7 GHz Channel). Soil moisture estimation is not possible during the time when precipitation is falling (liquid or solid). A precipitation threshold during the hour of satellite overpass can be used to mask those satellite pixels. For vegetation, Gao et al. (2006) used the 10.7 GHz polarization ratio (vertical to horizontal; $T_{b,v}/T_{b,h}$) criteria to access the vegetation/land cover since this ratio is almost independent of surface temperature and it only depends on land cover conditions. They found that this ratio was low and varied slightly over the forested areas. They also found reasonable consistency between the monthly averaged spatial maps of this polarization ratio to the monthly averaged vegetation water content. They concluded that the regions with monthly mean polarization ratio below 1.02 and standard deviation less than 0.005 were covered by dense vegetation and hence soil moisture estimation was not possible. Similar criteria can be used in case of AMSR-E after an extensive validation. Lastly, the daily

frozen soil and snow classification map from National Snow and Ice Data Center (NSIDC) can be used to mask the snow covered regions during the winter season.

3.6 Conclusions

Remote sensing data have the potential to provide insightful information for hydrological studies. Availability of such a high spatial and temporal scale in-situ soil moisture measurements for an extensive period of time holds the key for such kind of comparison studies. There are definitely issues such as inconsistencies in spatial scale (point measurements versus grid scale observations) and vertical resolution (5 cm soil moisture versus top less than 1 cm (skin surface) soil moisture), which create mismatches between the in-situ and satellite soil moisture datasets. We must keep those things in mind while analyzing the results. Nevertheless, the current operational AMSR-E retrieved soil moisture in this study did not perform as well as the LSMEM estimated soil moisture over the well-instrumented Little River Experimental Watershed, Georgia. The differences between the AMSR-E and LSMEM approaches are mostly due to differences in various simplifications and assumptions made for variables in the radiative transfer equations and the soil and vegetation based physical models and the accuracy of the input surface temperature datasets for the LSMEM forward model approach. The coregistration of sensor footprints for similar spatial resolution and relative calibration biases for each channel might produce a source of error in the AMSR-E operational algorithm. The dynamic surface temperature data are not input for the AMSR-E algorithm. On the other hand, the LSMEM results significantly depend on the accuracy of input soil moisture and vegetation water content data. The relative performance among the methods in this study may be dependent on geography, climate and topographic conditions, but the reasons for those differences should be robust. The superior performance of the LSMEM soil moisture data in previous studies (Gao et al., 2004a, 2006; McCabe et al., 2005a, 2005b) is consistent with results in this study.

This study has provided a qualitative as well as quantitative assessment of the AMSR-E retrieved and LSMEM derived soil moisture data. This study also discusses the radiative transfer theory and the different approaches adopted for soil moisture estimations. We hope this will be helpful for the future soil moisture satellite missions (e. g. SMOS; Kerr et al., 2001) and improve understanding of the causes of differences between the soil moisture estimation approaches.

Chapter 4. A Study of Land Surface Processes: Land Surface Model Inter-Comparison Study

Summary: Land surface schemes have been widely used to understand the processes and scales over the land surface. In this study, three different land surface models (HySSiB, Noah and CLM) were simulated on the NASA GSFC's LIS platform at 1 km resolution over the Little River Experimental Watershed, Georgia and the simulated results were analyzed to address the local scale land-atmosphere processes. All the three models simulated the soil moisture in space and time realistically. Noah model produced higher soil moisture whereas CLM got lower soil moisture with many dry down phases. CLM and HySSiB models were over sensitive to the atmospheric events. Different vertical discretizations of the model layers affected the soil moisture results in all the three models. The arithmetic model ensemble mean soil moisture performed reasonably well even at individual in-situ measurement sites.

We found that different model schemes partitioned the incoming water and energy differently and hence produced different results for the water and energy budget parameters. In CLM, the energy and water budget parameters were very closely connected to the soil moisture (e.g. evaporation, latent and sensible heat) change. HySSiB produced very high surface runoff and very low subsurface runoff. Noah model did not produce much surface and subsurface runoff at the cost of high surface soil moisture. We did not find much variability in Noah latent heat, sensible heat and ground heat fluxes. From soil moisture data assimilation point of view, the mean bias removed Noah soil moisture was found to be better than other datasets.

4.1 Introduction

Land surface processes play an integral and substantial part in both global water and energy budgets. Soil moisture is a critical element in all land surface processes. It controls the surface and sub-surface runoff; it supplies moisture to the atmosphere; and helps determine the Bowen ratio (Dirmeyer, 1995). It also acts as a water reservoir for the land surface hydrologic cycle and controls the water uptake by the vegetation above the ground (Vinnikov and Yeserkepova, 1991). Quantifying soil moisture is important for atmospheric scientists, hydrologists, agriculture scientists as well as the policy makers who try to mitigate natural disasters such as flooding and drought.

Yet quantifying soil moisture is very difficult. There is no clear definition of soil moisture. Different researchers and professions look at it from different prospective and define it in many different ways (Dirmeyer, 2004). Nor is there a global dataset of observed soil moisture available because the field observations are very scanty, satellites cannot see below the soil surface or through vegetation to measure it and the current state-of-the-art land surface models are not capable of representing the complex land surface physics to simulate it accurately (Reichle et al., 2004). Even if a land surface

model (LSM) were sufficiently accurate, it requires complete and accurate meteorological data as input – this is also lacking over most of the globe.

Model inter-comparison studies are one of the ways to overcome the problems associated with any single model, since results from the model inter-comparison studies are not biased by any single one and have been found to be superior to any of the individual model results (Gao and Dirmeyer, 2006). As a result, there have been many model inter-comparison studies conducted in the last decade to simulate different land surface variables and address specific science problems related to the land surface hydrology. The Project for Inter-comparison of Land Surface parameterization Schemes (PILPS) is one such model inter-comparison project which was established in 1992 and has evaluated the parameterization of energy and water fluxes to/from the landatmosphere interface using many land surface schemes (e.g. Henderson-Sellers et al., 1995). The Global Soil Wetness Project (GSWP) I and II are other such model intercomparison projects to assess the quality and performance of different land surface schemes (LSS) and estimate land surface variables on global scales (Dirmeyer et al., 1999 and Dirmeyer et al., 2006). Another such land surface scheme inter-comparison project focused on river hydrology and snow simulation was carried out by Boone et al. (2004) over the Rhone River.

Does the complexity of the LSM parameterizations contribute to model performance differences for land surface simulations? Does the model ensemble mean soil wetness from different LSMs perform reasonably better compared to any single model as concluded by Gao and Dirmeyer (2006)? To answer the above questions, this paper focuses on three different LSMs with different model parameterizations and analyzes the soil moisture simulation results from them over the Little River Experimental Watershed (LREW), Georgia. This paper is the second in the series; in the first work (Sahoo et al., 2008a) we compared the AMSR-E satellite retrieved soil moisture results over LREW using two different retrieval methods with the field observed soil moisture data. The first paper was focused on assessment of the observed remote sensing data. In contrast, the present paper focuses on land surface modeling and tries to look further at the roles of model complexity, forcing datasets and land surface conditions in order to answer the differences in the land surface simulation results. Here, we integrate the LSMs generating hourly output at 1 km spatial resolution. The availability of high quality and fine spatial and temporal (every 30 minutes) field observed datasets in this study is advantageous to perform this comprehensive comparison study. Our ultimate objective is to perform soil moisture data assimilation using observations and improved model simulation results. Hence, more emphasis has been given to soil moisture in this paper and the model performance has been evaluated based on that.

Description of the models and their physical processes are briefly summarized in section 4.2. The study area and field observation datasets are described in section 4.3. Brief description about the model input datasets and the model setup are given in Section 4.4. The model simulation results and discussion are presented in section 4.5. Finally, the conclusions are provided in section 4.6.

4.2 Land Surface Models

There are three LSMs used in this study. They are i) the Hybrid Simplified Version of the Simple Biosphere Model (HySSiB; Mocko et al. 1999; Sud and Mocko 1999); ii) the National Center for Environment Prediction (NCEP) Noah Land Surface Model Version 2.7.1 (Noah 2.7.1, hereafter Noah; Ek et al., 2003) and iii) the National center for Atmospheric Research (NCAR) Community Land Model Version 3 (CLM3.0, hereafter CLM; Dai et al., 2003). A summary of the model differences is listed in Table 4.1. HySSiB and Noah are "second generation" LSMs, primarily concerned with calculation of the surface energy and water balances. CLM is "third generation" in that it also maintains a carbon budget and explicitly represents the controls that photosynthesis exerts on water and energy in the soil-vegetation atmosphere system. As a basic requirement for any model inter-comparison study, all the three models were simulated on the same platform/environment (NASA's Land Information System (LIS)) and forced by identical atmospheric forcing and land surface parameters.

	HySSiB	Noah	CLM
No. of Soil	3	4	10
Layers			
Soil Layer	2, 150, 350 cm	10, 40, 100, 200 cm	1.8, 4.5, 9.1, 16.6,
Boundaries			28.9, 49.3, 82.9, 138.3, 229.6, 342.3 cm
Model Physics	Uses mass conservation law, Vertical discretized Darcy's Law	Uses mass conservation law, diffusive form of Richard's Equation	Water conservation at canopy and soil layer, vertical discretized Darcy's Law
Reference	Mocko and Sud, 2001	Ek et al., 2003	Dai et al., 2003

Table 4.1: Basic differences among the three Land Surface models

4.2.1 LIS Architecture

The NASA/GSFC Land Information System (LIS; Kumar et al. (2006); http://lis.gsfc.nasa.gov/) is built upon Global (GLDAS; Rodell et al., 2004) and North American (NLDAS; Mitchell et al., 2004) Land Data Assimilation Systems (http://ldas.gsfc.nasa.gov). LIS features a high performance and flexible design, provides infrastructure for data integration and assimilation, and operates primarily on an ensemble of land surface models for execution over user-specified regional or global domains. The LIS software is designed within an object-oriented framework, with explicit abstract interfaces defined for customization and extension to different applications. LIS is a flexible and expandable system and it can be customized to incorporate more user defined land surface schemes, atmospheric forcing and land surface parameter datasets. All the land surface models (LSMs) in LIS simulate energy and water variables (e.g. soil moisture (both liquid and frozen), soil temperature, skin temperature, runoff) and fluxes (e.g. evaporation and transpiration) at 1-km (fine) to 25km (coarse) spatial resolutions, and at one-hour or shorter temporal resolutions (Zhan et al., 2004). LIS also follows the Assistance for Land Modeling Activities (ALMA; Polcher et al., 2000) convention, which is a common data and metadata standard used among the land surface community to denote energy and water variables. The version 4.3.2. of LIS is used in this study.

4.2.2 HySSiB Model

HySSiB is a biophysical model designed to simulate land surface processes realistically and calculate radiation absorption, reflection, and provide fluxes of momentum, and sensible and latent heat (Mocko and Sud, 2001). The original version of the HySSiB (known as SSiB; Xue et al., 1991) got its lineage from the Simple Biosphere Model (SiB) (Sellers et al. 1986) with reduced physical parameters and improved computational efficiency. A major difference of HySSiB from the original SSiB of Xue et al. (1991) is its distinct conceptual architecture of snow pack and radiative transfer through snow. It also includes an orography-based surface runoff scheme and interaction with a water table below the third soil layer. HySSiB includes 3 soil layers with lower boundaries at 2 cm, 150 cm and 350 cm below the surface. HySSiB considers the rooting depth at 100 cm below the surface (Oliveira et al., 2006). It has one canopy and one snow layer. It has eight prognostic variables, namely soil wetness in 3 soil layers, water stored on canopy and ground and temperature at the canopy, ground surface and deep soil layers (Xue et al., 1996). The equation for canopy interception is based on conservation of water. It uses a finite difference approximation and a discretization of Darcy's law for vertical flow of water between soil layers. Soil parameters are a function of a small set of soil types. The drainage of water out of the bottom layer includes the water due to gravitational percolation and baseflow suggested by Liston et al. (1994). The temperature of the canopy is based on the energy conservation equation whereas the surface and deep soil temperatures are solved using the force-restore method (Deardoff, 1978). The mass and energy transfers between land surface and atmosphere are represented using a resistance formulation. The water stress term includes the stomatal resistance to the atmosphere, soil

water potential and vapor pressure deficit. The stomatal resistance in this model is based on Jarvis (1976). The soil water potential is taken from the empirical relationship of Clapp and Hornberger (1978). The resistance between the canopy and the reference height is based on similarity theory. The resistance to the water transfer from the surface soil layer to the canopy layer includes an aerodynamic resistance and a soil surface resistance. The aerodynamic resistance is based on its relationship to the Richardson number (Xue et al., 1996).

4.2.3 Noah Model

The Noah LSM gets its lineage from the OSU LSM originally developed in the 1980's at Oregon State University (Mahrt and Pan 1984). It has been upgraded and extended by the National Center for Environmental Prediction (NCEP) and its collaborators (Chen et al., 1996). This model has been validated through model inter-comparison studies; both in coupled (Betts et al., 1997; Ek et al., 2003) and uncoupled (Wood et al., 1998; Schlosser et al., 2000; Robock et al., 2003) studies. It has been implemented in operational weather and climate models due to its moderate complexity and computational efficiency. This model has a vertical soil profile that extends two meters below the surface. This vertical profile is partitioned into 4 soil layers with lower boundaries at 10 cm, 40 cm, 100 cm and 200 cm below the surface. The rooting depth of Noah model is fixed at 100 cm, which includes top three soil layers. It has one snow layer and one canopy layer. The prognostic variables include soil moisture and temperature in soil layers, water stored on the canopy and the snow stored on the ground (Chen and Dudhia, 2001). The physics of vertical water mass movement between the soil layers is governed by the mass conservation law and the diffusive form of the Richard's law whereas the infiltration is governed by a conceptual parameterization for the sub-grid treatment of precipitation and soil moisture (Schaake et al., 2004). At the bottom of the soil layers, drainage is only due to gravitational percolation as the hydraulic diffusivity is zero. The total evaporation includes the direct evaporation from the top soil layer, the evaporation from the canopy intercepted water and transpiration. The surface skin temperature is determined from surface energy balance equation representing combined ground-vegetation surface. The soil layer temperature is solved using the Crank-Nicholson scheme. The ground heat flux is determined using diffusion equations for soil temperature (Chen and Dudhia, 2001). This study includes the community version of the one-dimensional Noah model, version 2.7.1.

Noah uses a vegetation lookup table for static vegetation parameters such as minimum canopy resistance, solar radiation term for canopy resistance, vapor pressure deficit, threshold snow depth, roughness length and leaf area index. Noah uses a soil lookup table for static soil parameters. In this case, we used soil types map described in Zobler (1986).

4.2.4 CLM

CLM is a community model that combines features from the Land Surface Model (LSM) of Bonan (1996), the Biosphere–Atmosphere Transfer Scheme (BATS) of Dickinson et al. (1986) and the 1994 version of the Chinese Academy of Sciences Institute of

Atmospheric Physics LSM (IAP94; Dai and Zeng, 1997). The current CLM includes 10 soil layers with telescoping layer boundaries approximately at 1.8 cm, 4.5 cm, 9.1 cm, 16.6 cm, 28.9 cm, 49.3 cm, 82.9 cm, 138.3 cm, 229.6 cm, and 342.3 cm below the surface. It has one canopy layer and one to five snow layers depending on the snow depth. CLM focuses on biogeophysics of the land surface and includes vegetation dynamics and river routing modules. The water intercepted by canopy is calculated from a mass balance equation. The water flow within the snow layers is by an explicit scheme which permits a portion of liquid water over the holding capacity of snow to percolate into the underlying layer. The water flow from the bottom of the snow layer is available for infiltration into soil and for runoff (Dai et al., 2003). CLM runoff includes the surface and baseflow. CLM uses the conceptual TOPMODEL approach to parameterize the surface run-off and a discretized version of Darcy's law for vertical downward flow of water within soil layers (Oleson et al., 2004). The baseflow includes bottom drainage, saturation excess and the subsurface lateral runoff. The ground albedo includes the soil and canopy albedo (and snow over snow surface). It applies a two-stream radiative transfer approximation for canopy albedo (Sellers and Dorman, 1987). Soil albedo is a function of soil color and moisture. The snow albedo is a function of snow age, grain size, solar zenith angle, pollution and the amount of fresh snow. The vapor flux between the reference height and the canopy is calculated iteratively using Monin-Obukhov similarity theory (Bonan et al., 2002). Total evapotranspiration includes the evaporation from the canopy intercepted water, transpiration through vegetation and direct evaporation from the ground. The canopy temperature is calculated by solving the foliage

energy conservation equation using the Newton-Raphson iteration method. The soil and snow heat transfer is based on the heat diffusion equation. The heat flux at the surface is calculated using the energy balance equation at the surface where as the heat flux at subsurface is described by the Fourier law for heat conductance and the heat flux value is zero at the bottom of the soil column.

CLM addresses sub-grid variability through the use of tiles. Each grid cell is divided into any number of tiles, each tile consisting of a single land cover type. Each vegetation land cover includes up to four plant functional types (PFTs; Bonan et al., 2002). Energy and water balances are calculated separately for each tile at each time step. The tiles interact directly with the mean atmospheric grid condition over the respective tiles, but do not interact with each other. The values over a grid box are areally weighted averages of all the tiles inside the grid (Dai et al., 2003).

Like Noah, CLM uses a vegetation lookup table for time invariant vegetation parameters. These include the ratio of momentum roughness length to canopy top height, ratio of displacement height to canopy height, characteristic of leaf dimension, photosynthetic pathway, maximum rate of carboxylation at 25 C, slope of conductance to photosynthesis relationship and quantum efficiency at 25 C, visible and infra-red reflectance and transmittance from leaf and stem, leaf orientation index and rooting distribution parameter. This study includes the 3.0 version of the CLM.

All three models used for this study conserve energy and water at each time step. CLM is the most complex model of the three because of its finer vertical soil and snow layer resolution, tiling structure, subsurface lateral runoff and direct calculation of carbon processes.

4.3 Description of Study Area and In-Situ Data Sets

4.3.1 Study Area

A detailed description of the study area is described in Section 2.2.1 Figure 2.1 shows an image of the study area along with the in-situ measuring stations.

4.3.2 LREW In-Situ Observation Data

A detailed description of the in-situ observation datasets is described in Section 2.2.2.

4.3.3 SCAN Data

SCAN is a nationwide comprehensive soil moisture system which have been collecting and providing soil moisture and soil temperature as well as precipitation, solar radiation, air temperature, specific humidity, wind speed and direction data for longer periods of time than field experiments area capable of (Schaefer and Paetzold, 2001). The measured data are first automatically validated against the preset limits and then manually checked. Measurements of soil moisture at 5, 10, 20, 50 and 100 cm soil depths are taken wherever possible. This network is distributed mostly over the agriculture areas of USA. The data can be obtained from the USDA Natural Resources Conservation Service website (NRCS; http://www.wcc.nrcs.usda.gov/scan/). As mentioned earlier,
there is only one SCAN site available within the LREW watershed (Station 2027, Little River, Georgia, 31.50° N and 83.55° W). Since this SCAN site provides most of the meteorological forcing data, we used this SCAN site data to validate the North American Land Data Assimilation System (NLDAS) forcing data needed to drive LSMs. These NLDAS forcing data have been used as input for all land surface model simulations carried out in this study.

4.4 Input Data and Model Setup

4.4.1 Forcing Data

We did not have adequate observed forcing data from the field experiments to drive the model simulations. So North American Land Data Assimilation System (NLDAS) forcing data were used for the model simulations here. NLDAS incorporates in-situ gauge, radar and satellite observations over the NCEP Eta Data Assimilation System (EDAS) baseline information to produce the forcing data over the North America (Cosgrove et al., 2003). NLDAS data include air temperature and specific humidity at 2 m height, wind speed at 10 m height, surface pressure, downward shortwave and longwave radiation, convective available potential energy, skin temperature, total and convective precipitation and photosynthetically active radiation. Two important NLDAS atmospheric forcing variables, precipitation (Higgins et al., 2000) and the incoming solar radiation (Pinker et al., 2003), have been generated using observations only. NLDAS data provide hourly measurements from September 30, 1996 to present. Luo et al. (2003)

performed a comprehensive validation study for the NLDAS datasets using the station observations over the Southern Great Plains (SGP). They found a high bias in the NLDAS downward shortwave radiation, but low biases in the downward longwave radiation; hence cancellation provides a lower bias in total incoming radiation. They also found most meteorological fields other than wind speed agreed very well with observations. Most importantly, they found the differences between LSM simulations with the input NLDAS forcing versus station observations were not primarily due to the atmospheric forcing, but the differences among physics between the models.

The findings of Luo et al. (2003) give us confidence that the NLDAS atmospheric forcing datasets will be adequate for our study here. Since we have a single SCAN site available within our study region and the SCAN instrument provides meteorological datasets along with the soil moisture data, we also performed a validation study for the NLDAS forcing data. Figure 4.1 shows the scatter plots of the hourly meteorological datasets (downward shortwave radiation, precipitation, air temperature and wind speed) from NLDAS and SCAN for the year 2003. Downward longwave radiation was not available from the SCAN site. The bias, root mean square difference (RMSD) and the correlation have also been given for the comparison of each forcing variable. In general, the forcing variables are in good agreement with the SCAN observed data. The most notable difference between the two datasets is in the downward shortwave radiation. It has a relatively large RMSD and a notable high bias relative to the SCAN shortwave radiation data (Figure 4.1a). Luo et al. (2003) found similar behavior of the NLDAS incoming shortwave radiation and attributed those differences with the in-situ



Fig. 4.1: Scatter Plots of hourly atmospheric forcing from NLDAS and SCAN measurements at SCAN site for the year 2003, (a) Downward shortwave radiation (W/m^2) , (b) Total Precipitation (mm/hr), (c) Air temperature (K) and (d) Wind speed (m/s).

observations to morning cloud cover conditions. The SCAN hourly precipitation datasets show higher precipitation rates than the NLDAS hourly precipitation rates for heavy precipitation events (Figure 4.1b). This is because NLDAS derives hourly precipitation from daily totals. We also drew daily time series for precipitation comparisons. The time series plot of daily precipitation from NLDAS and SCAN shows a very good match between both the datasets (with RMSD 0.78 mm/hr) except for few large daily precipitation rates (Figure 4.2). The air temperature (Figure 4.1c) datasets match very well with the SCAN instrument observations. Wind speed (Figure 4.1d) datasets for NLDAS and SCAN vary considerably and but the differences are not as systematic as is the case with the incoming shortwave radiations. These findings agree well with those of Luo et al. (2003).



Fig. 4.2: Time series plot of daily precipitation (mm/day) from NLDAS and SCAN measurements at SCAN site for the year 2003.

4.4.2 Parameter Data

The primary land surface parameters are concerned with vegetation (land cover, greenness) and soil (texture, color, etc.). For land cover, we use the University of Maryland's (UMD) 1 km global land cover product (Hansen et al., 2000). This dataset has a total of 13 land cover classes excluding water bodies. The land-sea mask was also generated from this vegetation classification map. For soil, the sand, silt and clay fraction and soil color data were generated at 1 km resolution from the original FAO (Food and Agriculture Organization) 5 minute resolution global soil maps. These above datasets were used for the simulations of all the models. LIS uses GTOPO30 Digital Elevation Model (DEM) elevation data from US Geological Survey (USGS). LIS does the elevation correction by adjusting the forcing data whenever the elevation differs between LIS and the atmospheric model which produced the atmospheric forcing data. All three models were integrated using the same soil and vegetation parameter datasets, but the procedures used to estimate model parameters were model specific. The Noah model required additional parameters such as a quarterly albedo climatology, monthly greenness fraction climatology, maximum snow albedo, and bottom temperature without elevation correction. Similarly, CLM required canopy height and vegetation classification lookup tables, Leaf Area Index (LAI) and Stem Area Index (SAI) datasets. For the calculation of the LAI climatology, the Moderate Resolution Imaging Spectroradiometer (MODIS) 1 km LAI data were preferred over the Advanced Very High Resolution Radiometer (AVHRR) 16 km LAI data because of the better spatial resolution in the earlier one. The MODIS LAI data were collected from Boston University (Yang et al., 2006). The SAI

climatology was calculated from LAI data using the methods described in Sellers et al. (1996) and Los et al. (2000).

4.4.3 Model Initialization

Improper model initialization can produce erroneous model output results. We adopted one of the model initialization methods described in Rodell et al. (2005). NLDAS input forcing data are available from October 1996. We use five years of NLDAS input forcing data (from January 1997 to December 2002) to spin up the model state variables by looping three times through the 5 years of forcing data (a total 15 years of spin-up). Then the climatology state was calculated from the mean of the outputs of five January months of the last five years to produce initialization conditions for 1 January 2003. This approach is one of the better ways of initializing a land surface model to reduce the occurrence of unrealistic extremes in the initialization (Rodell et al. 2005).

4.4.4 Model Simulation

All three models were integrated retrospectively from 1 January to 31 December 2003 at 1 km spatial resolution over a region bounded by 83.38°-84°W longitude and 31.11°-31.88°N latitude (62 x 77 domain). The model time step was 15 minutes with model output saved every hour. The NLDAS retrospective forcing was hourly, so forcing variables were interpolated to 15-min intervals. The solar zenith angle interpolation to avoid

the error introduced by a simple linear interpolation scheme. Table 4.2 shows the details of the model setup.

Land Surface Model (LSMs)	Noah, CLM, HySSiB			
Base Forcing	NLDAS			
Land Cover Type	UMD global 1 km land cover map			
Soil Classification Map	FAO			
Maximum number of tiles per grid	13			
Time step of the run	15 minutes			
Latitude Range	31.11° N to 31.88° N			
Longitude range	84.00° W to 83.38° W			
Output Data Resolution	1 km			
Output interval to write the output files	1 hour			
Output data format	Binary			

Table 4.2: Initial model setup for the GLDAS/LIS model runs

4.5 Results and Discussion

4.5.1 Scale Issues and Pre-processing of Results

The simulation results from the three models have been compared for all simulated variables except the surface soil moisture where we also used the in-situ observations. It is important to discuss the issues associated with spatial, temporal and vertical resolutions when we are comparing soil moisture datasets from different sources. The field measurements were carried out at point scale whereas the models simulated the land surface processes averaged over 1 km spatial grids. So, instead of comparing the point observations with the grid averaged model simulations, we created a composite average of all the observed data from all the stations in the watershed and compared this composite soil moisture with the averaged soil moisture from all the corresponding grid

points for each LSM. A similar kind of approach has also been used when data were from multiple sources (e.g., Vinnikov and Yeserkepova, 1991; Entin et al., 2000; Robock et al., 2003; Schaake et al., 2003; Reichle et al., 2004; Prigent et al., 2005). Also, this region is topographically very flat and the soil moisture spatial variation is not large over this region (Cashion et al., 2005). Thus, spatial averaging of soil moisture over this watershed is reasonable for this study. We had complete soil moisture observations from 8 measurement stations for the year 2003. So, we also used corresponding 8 model grids for this comparison study. For all other model simulated variables, we also used the spatial averaging for same 8 grids to reduce any model uncertainties at local grid scale.

The model simulation outputs were at 1 hour interval and the in-situ soil moisture observations were at 30 minutes interval. But for energy cycle variables, we used daily averaged values and for precipitation we used daily accumulated values instead of instant values for the comparison study. This avoids the strong variability and noise associated at the hourly time scale.

The in-situ soil moisture data are from top 5 cm layer. The top soil layer depth for HySSiB model is 2 cm (surface layer). Similarly Noah model has 10 cm and CLM model has ~ 2 cm surface layer depths. For the comparison study here, we used only the volume percentage soil moisture values. Previous studies have shown very high vertical correlations of soil moisture variability within top 20 cm soil layer (Wigneron et al., 1995; Calvet et al., 1999; Prigent et al., 2005). It is a common practice to compare the soil moisture data from different depths in most of the comparison studies because of the lack of soil moisture data from equivalent depths from different sources. So, we made a

similar assumption for soil moisture variability in top soil layers as Prigent et al. (2005) and compared the datasets from various layer depths mentioned here.



Fig. 4.3: Daily soil moisture time series plots from Noah (10 cm layer), CLM (2 cm top layer),HySSiB (2 cm top layer) in-situ measurements (5 cm layer) and Arithmetic Model EnsembleMean from Noah, CLM and HySSiB over the LR Watershed.

4.5.2 Results and Analysis

4.5.2.1 Soil Moisture

Figure 4.3 shows the daily time series plots of the top layer soil moisture simulation results from the LSMs (10 cm for Noah, 1.8 cm for CLM and 2 cm for HySSiB) along

with the observed LR Watershed in-situ (5 cm) soil moisture data for the year 2003. This figure also includes the multi-model mean soil moisture, which was calculated taking the arithmetic mean of the three models. It is very clear from this plot that the daily soil moisture peaks from all sources match very well corresponding to daily precipitation peaks (shown in Figure 4.2). A few interesting points can be noted from this figure. First, the models and observations show higher soil moisture during the spring season of the year because of the consistent rainfall events during the spring over this watershed. Second, HySSiB and Noah simulate higher soil moisture values throughout the year compared to observations irrespective of different top soil layer thicknesses for the two models. Third, the CLM soil moisture values are very sensitive to the precipitation forcing as compared to those of other two models and it dries down to a minimum top layer soil moisture threshold value very fast after a precipitation event is over (~ couple of days). This high sensitivity of CLM soil moisture data is because the model soil layer structure has a thin top soil layer. However, HySSiB soil moisture estimates do not show the same behavior as those of the CLM model at daily temporal time scale even though it has an equally thin top layer. Fourth, the model ensemble mean soil moisture performs reasonably well with less bias than any individual model.

To understand the behavior of all the soil moisture estimates, we use scatter plots of the model daily soil moisture simulations along with the model ensemble mean against in-situ observations (Figure 4.4). A 1:1 line is also shown in each figure for reference. All the comparison statistics for the estimates are calculated with respect to in-situ observations. Figure 4.4a indicates a high systematic bias (8.78 % vol/vol) for Noah. HySSiB also shows high mean bias (9.92 % vol/vol) compared to the observations (Figure 4.4c), but it is not as systematic as that of Noah model. Because of these high biases, Noah and HySSiB show very high RMSD values; 8.92 and 10.62 % vol/vol for Noah and HySSiB model respectively. Yet Noah and HySSiB show high correlations (0.90 and 0.81 for Noah and HySSiB model respectively) with in-situ observations. In contrast to the other two models, CLM (Figure 4.4b) has a lower mean bias (-4.61 vol/vol) and RMSD (6.19 % vol/vol). The lower threshold value for CLM top layer soil moisture (~ 2 % vol/vol) is frequently reached (Figure 4.4b), which is noticeable during the dry down phases in the soil moisture time series plot (Figure 4.3).

The multi-model estimates show the lowest mean bias (2.70 % vol/vol) and RMSD (3.66 % vol/vol). The negative bias in CLM model estimates almost cancels the positive biases in the Noah and HySSiB model estimates, keeping the multi-model mean bias very low. The correlation values for all datasets range from 0.80 (CLM model) to 0.90 (Noah model). The multi-model mean estimates show comparable correlation value (0.86) as those of the individual model estimates. Since some of the models have systematic high bias even though they have high correlations, we removed the mean biases from all the model simulations and recalculated the RMSD for all the datasets. The RMSD got reduced considerably for the Noah (from 8.92 to 1.60) and HySSiB (from 10.62 to 3.77) soil moisture results in this case (figures not shown in this paper). Moreover, the Noah model estimates scattered around the 1:1 reference line indicating perfect model estimates.



Fig. 4.4: Daily soil Moisture scatter plot averaged over 8 in-situ stations in the LR Watershed,(a) Noah versus In-Situ, (b) CLM versus In-situ, (c) HySSiB versus In-situ and (d) ArithmeticModel Ensemble Mean versus In-situ.

We also looked at the performances of the individual model and multi-model soil moisture estimates at individual measuring stations. Figure 4.5 shows a comparison of the skill scores of the original model simulated soil moisture products along with the multi-

model mean for 8 individual watershed stations and the station-averaged dataset. These skill scores are calculated based on hourly soil moisture values for the year 2003 in contrast to the above graphs where we used daily averaged soil moisture values. We will discuss more about the hourly soil moisture estimates later in a separate section. As can be seen, the multi-model mean gives bias (3 to 8 % vol/vol), RMSD (5 to 10 % vol/vol) and correlation (0.6 to 0.8) skills as good or better than those of the individual model results at each individual observation station as well as for the station average. CLM shows a lower magnitude of bias than the multi-model mean at half the stations and lower RMSD at three stations. Noah shows superiority of time series correlation at all stations. Since any single model is not consistent at all individual sites for all the skill scores, the multi-model mean across all the models can potentially provide the best overall estimates for multiple applications. Moreover, the multi-model mean is very easy to calculate and it can ameliorate systematic error associated with any individual model because of the averaging of the simulation results for different models with fundamentally different statistics. This finding agrees well with Gao and Dirmeyer (2006) who used many different ensemble schemes to combine simulation results from multi-models and concluded that the arithmetic model ensemble mean could be a better representative than any individual model simulation results for soil wetness estimates.

To analyze the spatial distribution of soil moisture and precipitation–soil moisture coherence patterns, we chose a time period of wetting and drying dynamics (July 16 to July 19) over the Little River Experimental Watershed area (approximately 50 km by 75 km area). Figure 4.6 shows daily NLDAS precipitation forcing (column 1) and



Stations





Fig. 4.5: (a) Mean Bias, (b) RMSD and (c) Correlation of Noah, CLM, HySSiB and Arithmetic Model Ensemble Mean soil moisture for 8 in-situ stations and station-averaged dataset.

corresponding daily soil moisture difference images from Noah (column 2), CLM (column 3) and HySSiB (column 4). All these spatial images are in 1 km spatial resolution. Same color scale has been used for easier visual recognition of the changes in the spatial images. The white region in the precipitation image for July 19 represents the zero precipitation area. Looking at the precipitation panels, there is heavy precipitation in the south-east part on July 17; in the north-east and central parts on July 18 and complete dry down phase everywhere on July 19 over the watershed. The spatial pattern of soil



Fig. 4.6: Spatial difference image of NLDAS precipitation and corresponding model simulated soil moisture results during July 16 – July 19. There are spatial coherences between the precipitation events and model soil moisture results.

moisture for all the three models corresponds very well to the spatial distribution of the precipitation events with few exceptions. HySSiB does not show very distinct soil moisture patterns as we see from Noah and CLM though the changes in soil moisture due to precipitation events is still visible in HySSiB simulations. But the soil moisture difference values are higher (both positive and negative (more than 7 % vol/vol)) in HySSiB as compared to the other two models. This indicates that the HySSiB model top layer holds lot of water after precipitation events before it is removed through surface runoff.

4.5.2.2 Water Cycle Variables

Figure 4.7 shows daily time series plots of other water cycle parameters (surface runoff (Figure 4.7a), subsurface runoff (Figure 4.7b) and evaporation (Figure 4.7c)) from the three models for the year 2003 spatially averaged over the 8 in-situ stations in the LR watershed. Each daily surface runoff (Qs) peak (Figure 4.7a) for all three models corresponds well to each of the precipitation events (Figure 4.2) indicating direct response of model surface runoff to the precipitation events. Subsurface runoff peaks can be seen corresponding to only heavy precipitation days, especially during frequent precipitation in the spring season. This indicates that all the models have high water holding capacity to retain some water in their soil layers and they produce sub-surface runoff only during high precipitation events, losing water through surface runoff (Figure 4.7a). At the same time, we have noticed the high top layer soil moisture estimates for

HySSiB during precipitation events (Figure 4.3). Hence HySSiB produces very low subsurface runoff (Figure 4.7b). Contrast to that, Noah model has almost no surface runoff (Figure 4.7a) but it has very high top layer soil moisture (Figure 4.3). Since Noah top layer is thick (10 cm), it accommodates greater infiltration and produces less surface runoff than HySSiB. In contrast, CLM shows very different characteristics for hydrologic variables estimations. The dry down of CLM soil moisture is compensated by intermediate surface runoff and very high subsurface runoff. This high subsurface runoff during precipitation events could be because CLM subsurface runoff has more pathways than the other models

The evaporation patterns are also very different among the three models (Figure 4.7c). All models exhibit the expected seasonal pattern of evaporation, with higher evaporation during the summer and lower evaporation during the winter season. On short time scales, CLM evaporation is very sensitive to the precipitation patterns. Along with the surface and subsurface runoff, evaporation also removes some CLM top layer soil water immediately during the precipitation events. HySSiB shows a similar kind of evaporation pattern as those for CLM. Noah shows comparatively low variability in evaporation estimates than other two model models. Hence the evaporation is less from the top layers of CLM and HySSiB and they show many dry-down events during the lower precipitation season (August-October). Moreover, the transpiration through vegetation is also shut down due to low root zone soil moisture for the CLM and HySSiB models. The Noah model holds more water in the top layer due to the thicker top layer and the evaporation still goes on for the Noah model when the other two models show

dry down phases, hence the Noah model shows an opposite behavior to that of the other two models (e. g. Sep 8 to Sep 14, Sep 15 to Sep 21, Sep 26 to Oct 6 in Figure 4.7c).







Fig. 4.7: Comparison of Water Cycle variables simulated by Noah, CLM and HySSiB, (a) Surface runoff (mm/day), (b) Subsurface runoff (mm/day) and (c) Evaporation (mm/day).

4. 5.2.3 Energy Cycle Variables

Figure 4.8 shows the daily time series plots of the energy budget parameters for the same in-situ stations and time period as in Figure 4.7. The incoming energy at the earth's surface is indicated with positive sign and the outgoing energy is negative. All models show very similar net solar radiation patterns for the whole time period (Figure 4.8a). Since downward shortwave radiation is the same for all models, the net solar radiation indirectly represents the albedo values used by the models. The net solar radiation values from all the three models are distinctly different during the spring season (April-June). CLM produces the lowest net surface solar radiation during spring season. This implies

that CLM exhibits higher surface albedo than the other two models. During the other seasons, the values are similar to one another. In the case of the net surface longwave radiation, Noah and HySSiB match very closely (Figure 4.8b). CLM estimates match the other two models except during the phases of top layer soil moisture dry down. During these days, CLM shows increasing soil and surface temperature, which contributes significantly to the lower net longwave radiation.

Figure 4.8c and Figure 4.8d represent the daily latent and sensible heat flux. These fluxes reciprocate each other for all three models. Variability of latent and sensible heat flux depends on the amount of energy and water available at the earth's surface. Latent and sensible heat flux variability is very high for CLM as compared to the other two models for this time period. For reasons stated previously, CLM produces relatively high latent heat flux and low sensible heat flux for any precipitation event. On the other hand, Noah has the lowest variability. HySSiB exhibits relatively high variability in heat fluxes. Figure 4.8e shows the ground heat flux variability from all the three models. Ground heat flux is a residual in the energy budget and contributes to changes in subsurface soil temperature. Noah exhibits the least variability in the ground heat flux, while CLM shows high variability. Since all these land surface models close the energy budget at each time step, the variability of the ground heat flux depends on the variability of other energy budget parameters. In winter months, the ground loses heat to the atmosphere in all three models. This indicates that the total incoming energy flux is less than the total outgoing energy flux in winter season; spring and summer show the opposite situation.











Fig. 4.8: Comparison of Energy Cycle variables simulated by Noah, CLM and HySSiB, (a) Net Shortwave Radiation ($W/m^2/day$), (b) Net Longwave Radiation ($W/m^2/day$) and (c) Latent Heat Flux ($W/m^2/day$), (d) Sensible Heat Flux ($W/m^2/day$) and (e) Ground Heat Flux ($W/m^2/day$).

4.5.3 Impact of Scaling on Soil Moisture

Boone et al. (2004) discussed the impact of spatial scaling on model simulated water and energy cycle parameters. In this section, we discuss temporal scaling of the model soil moisture simulations. As shown in section 5.2.1, there exist systematic biases in some LSMs on daily time scales. Our objective in this section is to look at the original hourly soil moisture model outputs in contrast to the daily soil moisture simulations and provide our critical comments from data assimilation point of view. Figure 4.9 shows the scatter plot of the hourly soil moisture data from 3 individual models and their arithmetic ensemble mean against the in-situ observations for the year 2003. The corresponding daily soil moisture scatter plots were shown in Figure 4.4. Compared to the daily soil moisture data, the Noah model hourly soil moisture data show very similar behavior with systematic high bias and high correlation (Figure 4.9a). The lower threshold bound can very clearly be seen for CLM hourly soil moisture data in Figure 4.9b. CLM model hourly soil moisture data also exhibit similar characteristics as those of the daily soil moisture data with much scatter that does not appear to reflect a simple systematic bias. In the case of the HySSiB model, the hourly simulated soil moisture data show a very clear lower threshold boundary around 7 % vol/vol which is not apparent in the daily soil moisture scatter plots due to the scaling of hourly values to daily values (Figure 4.9c). This characteristic of HySSiB at hourly scale resembles that of the CLM model. This confirms that the sensitivity of the top layer soil moisture to the precipitation events is due to the soil layer parameterization since we see this complete dry down phase in CLM and HySSiB model (both have 2 cm top soil layer), but not in case of Noah (10 cm soil layer) model. HySSiB seems to exhibit the diurnal cycle of surface soil moisture incorrectly, but does well at the daily time scale. CLM has problems with both diurnal and synoptic dry-down phases. The hourly multi-model mean (Figure 4.9d) shows similar behavior as for the mean daily data (Figure 4.4d). For all the soil moisture products, the mean bias and RMSD are higher (except CLM mean bias) and the correlation is lower at hourly scale than those at daily scale. This is expected since the uncertainty for soil moisture at hourly scale is supposed to be higher than that at the daily scale. Table 4.3

shows the statistics from these comparisons, along with the results after a simple bias correction for each model. Bias correction reduces RMSD by 70% or more when dry-down phases are well modeled.



Fig. 4.9: Same as Figure 4.4, but for hourly soil moisture data.

	Hourly				Daily			
							Bias	
			Bias	RMSD			correct	RMSD
			corrected	reductio			ed	reductio
	Bias	RMSD	RMSD	n	Bias	RMSD	RMSD	n
Noah	8.82	9.02	1.88	79%	8.78	8.92	1.60	82%
CLM	-3.32	6.30	5.36	15%	-4.61	6.19	4.13	33%
HySSiB	10.44	11.86	5.63	53%	9.92	10.62	3.17	70%
Multi-								
model	5.31	6.47	3.69	43%	2.70	3.66	2.47	33%
Mean								

Table 4.3: Skill of hourly and daily time series simulations of soil moisture

4.6 Conclusions

In this paper, we performed comparison studies of water and energy cycle variables simulated by three different land surface models. The offline simulations were conducted using HySSiB, Noah and CLM land surface models driven by NLDAS atmospheric forcing data over the Little River Watershed, Georgia region from October 1996 to September 2003. The model simulation results for the year 2003 were compared to the in-situ observations. Important differences among the three land surface models were the complexity of the models, the top soil layer thickness and the layer parameterizations in each model. When NLDAS precipitation forcing was compared with the corresponding SCAN measured meteorological parameters, we found reasonable agreement among the datasets from the two sources. Model simulations at 1 km spatial resolution and hourly temporal resolution were good enough to look at the model responses to individual precipitation events and compare simulation results from the three models.

All the three land surface models simulated soil moisture realistically and exhibited close correspondence in soil moisture results to each precipitation event in space and time. The model layer parameterization played a major role in soil moisture simulation results at hourly scale. CLM was found to be overly sensitive to climatic conditions. CLM and HySSiB both overestimated the magnitude of surface soil moisture variations on dry-down time scales. The scatter plots of the individual model and the arithmetic model ensemble mean soil moisture against the in-situ observed soil moisture were conducted irrespective of the location of the data sites. The model mean performed well at the hourly and daily scale. The multi-model mean soil moisture also provided comparable skill scores with those of any individual model even at individual in-situ measuring sites, while avoiding some of the systematic problems exhibited by individual models. Our results agree well with the results of Gao and Dirmeyer (2006).

The models showed discrepancies in partitioning the precipitation water into soil moisture, surface runoff, infiltration and evaporation terms. CLM compensated low top layer soil moisture by high surface and subsurface runoff. On the other hand, Noah maintained high top layer soil moisture by near zero surface and lower subsurface runoff. HySSiB produced high top layer soil moisture and surface runoff, but very low subsurface runoff. All models exhibited very similar results in net solar radiation and longwave radiations. CLM simulated higher albedo as well as higher surface temperatures. Model physics played a critical role in partitioning the outgoing energy into latent heat and sensible heat fluxes pertaining to different climatic conditions. Models did not agree well in partitioning of latent and sensible heat flux.

We had some difficulties in performing this kind of comparison study. First, there were no in-situ observations available for most of the water and energy cycle variables to compare with the model simulation results. Second, we found a high bias in the downward shortwave radiation from NLDAS. We did not check how these high forcing biases affected each model simulation results in this study, but we believe that the agreement or disagreement among the model simulated results is mostly due to the different treatments of land surface processes by different schemes From the point of view of soil moisture data assimilation, high time-resolution simulations with good quality soil moisture estimates and comparable observation measurements are required. From the hourly scatter plots, it can be seen that Noah soil moisture with the mean bias removed can serve as the best model for a data assimilation study. Noah appears to be the least likely model to "fight against" the assimilation of observations. However, when we look at the Noah model parameterization, it has a 10 cm thick top soil layer where as most of the available remote sensing soil moisture products are from only the top 2 cm of soil. This inconsistency may be an important factor to consider while performing the data assimilation. In the case of CLM and HySSiB, the model soil discretizations agree better with the character of remote sensing observations, but these model results are not as well behaved compared to the Noah model. It is hard to say at this point which plays a more important role in data assimilation: better model simulation behavior or the choice of model soil layer discretization. This question can be answered in future studies by performing some data assimilation tests with these model simulations.

Modeling studies of local scale land surface processes help us to understand the land surface with the goal of realizing its local scale applications in water resource management, climate prediction and disaster mitigation. The results produced here also motivate us to look further into land surface model complexity and physics, external factors like atmospheric forcing and land surface parameters, and to estimate the relative contribution of each factor on these local scale processes in our future study.

Chapter 5. Assimilation of LSMEM Retrieved Soil Moisture into Noah Land Surface Model

Summary: A data assimilation module has recently been incorporated in the Land Information System (LIS) modeling framework at NASA GSFC. In this study, we use the LIS data assimilation module to assimilate the current AMSR-E satellite observed soil moisture data into the Noah land surface model to estimate the surface soil moisture over the Little River Experimental Watershed (LREW). It is found that the Ensemble Kalman Filter (EnKF) assimilation algorithm improved the model soil moisture prediction results constraining them by the observations. The assimilated soil moisture affects the model estimate of other water and energy budget variables. Few sensitivity studies have also been performed on the assimilation algorithm to verify the sensitiveness of the algorithm to the model spin-up and initialization conditions.

5.1 Introduction

Soil moisture and Sea Surface Temperature (SST) are two major initialization parameters for the seasonal to inter-annual climate prediction models. Soil moisture is used for other applications such as water resource management, agriculture production and hazard mitigation study. It also influences the hydrologic cycles from local to global scales. Earth observing satellites have revolutionized our understanding and prediction of the Earth system over the last 30 years though these data have not been widely used in Hydrology. One of the major applications of the satellite remote sensing in the hydrologic modeling environment is to constrain the time varying model state variable such as soil moisture for accurate model prediction. This paper here focuses on the above application of remote sensing data in hydrological sciences. Soil moisture data assimilation has been studied and proved to attenuate model errors and produce better results. Apart from that, the assimilation technique produces results at model temporal (1 hour) and spatial resolutions (1 km) which are finer than any typical satellite measurements (~ 3 days and 25 km). Assimilation data can also be retrieved over space and time where and when satellite observations are not available.

This chapter here discusses the assimilation study of Land Surface Microwave Emission Model (LSMEM) soil moisture derived from AMSR-E (Advanced Microwave Scanning Radiometer-Earth Observing System) satellite observations into the Noah land surface model existing in the Land Information System (LIS) framework using the Ensemble Kalman Filter (EnKF) technique.

5.2 Literature Review

Data assimilation in Hydrologic Science is very new though Charney et al. (1969) first suggested combining different datasets in an explicit dynamic model leading to many different data assimilation techniques. A number of past works have assimilated observed

soil moisture into model in an effort to get useful insight of data assimilation in hydrologic sciences and understand how data assimilation can be useful to improve the quality of hydrology data.

Evenson (1994) was one of the first few researchers who worked on data assimilation technique and he presented how to forecast error statistics in a non-linear sequential data assimilation method using Monte Carlo technique.

Burgers et al. (1998) described the analysis scheme in the Ensemble Kalman Filter method.

Houser et al. (1998) assimilated PBMR derived soil moisture into TOPLATS model using 4 different data assimilation techniques (Direct insertion method, Statistical correlation, Newtonian nudging and optimum interpolation) and illustrated that Newtonian Nudging method performed the best.

Walker and Houser (2001) performed synthetic twin experiment using Catchment model and Kalman filter and illustrated that assimilating near surface soil moisture, errors in the soil moisture forecast due to poor initialization can be removed.

Reichle et al. (2002a) assimilated L-band (1.4 GHz) microwave radio-brightness observations into land surface model and investigated the effect of ensemble size and non-Gaussian forecast error on the performance of the Ensemble Kalman filter technique.

Margulis et al. (2002) assimilated L-band brightness temperature using Ensemble Kalman Filter technique and showed that the estimated soil moisture was in reasonable accuracy with the ground measurements over Southern Great Plains 1997 (SGP97 field experiment). Reichle et al. (2002b) examined the performance of extended (EKF) and Ensemble (EnKF) Kalman Filters for soil moisture assimilation and found out that the EKF and EnKF with four ensemble members are equally accurate, but EnKF has high flexibility and performance for soil moisture initialization.

Walker and Houser (2004) addressed satellite mission accuracy, repeat time and spatial resolution through a synthetic twin experiment and found that the soil moisture observation better than 5% vol/vol, 1 to 5 day repeat time and spatial resolution less than the model resolution are required to have positive impact on data assimilation.

Reichle and Koster (2005) assimilated Scanning Multichannel Microwave Radiometer (SMMR) soil moisture retrieval into NASA Catchment Model for 1979-87. They validated the assimilated soil moisture results against ground based observations in Eurasia and USA from Global Soil Moisture Data Bank and concluded that assimilated soil moisture is better than satellite and model data alone.

Moradkhani et al. (2005) used dual EnKF approach to estimate both the states and parameters simultaneously and demonstrated it through streamflow forecasting using rainfall-runoff model.

Ni-Meister et al. (2005) studied error characterization associated with the catchment-based land surface model (CLSM) and Scanning Multichannel Microwave Radiometer (SMMR) soil moisture estimation over Eurasia since unbiased model predictions and observations are key assumptions in any data assimilation technique. They emphasized that accurate error assessment in model and observation estimation is important to improve the prediction of skills in data assimilation method.

Wilker et al. (2005) used error propagation theory on precipitation perturbations to calculate model soil moisture errors. They assimilated 2-m temperature and humidity and L-band brightness temperature with this error distribution to improve soil moisture forecast over Southern Great Plains Field Experiment Site (SGP97).

Pan and Wood (2006) assimilated observed soil moisture, Bowen Ratio and streamflow into Variable Infiltration Capacity (VIC) model over Oklahoma Mesonet using two-step Ensemble Kalman Filter approach. The first step was the standard Kalman filter and the second step was another Kalman filter which distributed the imbalance from the first step and was constrained by the water balance.

Crow et al. (2006) indicated through synthetic identical twin experiments that the inappropriate assumption of the source and magnitude of model errors could degrade the performance of the Ensemble Kalman Filter (EnKF). They found that dual parameter assimilation was more robust with incorrect model error assumptions.

Zupanski et al. (2006) addressed the impact of ensemble initiations to data assimilation algorithms. They found that the initial correlations of random perturbations in ensemble data assimilation improved the RMS error rate of convergence against the uncorrelated random perturbations. It implied computational savings though the sensitivity depended on the correlation length scale.

Baek et al. (2006) illustrated the effect of parameterization for representing forecast model bias in the Ensemble Kalman Filter Technique.

De Lannoy et al. (2006) assessed the model uncertainties through ensemble generations for soil moisture forecast using Community Land Model (CLM2.0). They

found that during the extreme drought and precipitation periods, the ensemble density function (pdf) deviates far from normality and the model behaved very nonlinearly. In that case, the pdfs could not be assumed Gaussian, hence the optimal solution of forecast is difficult to achieve.

5.3 Ensemble Kalman Filter (EnKF)

Ensemble Kalman Filter (EnKF) is an optimal sequential data assimilation method. EnKF has recently been used by many researchers (Walker and Houser, 2001; Reichle and Koster, 2005) for the hydrologic data assimilation study. It has better performance over the other data assimilation method such as Extended Kalman Filter (EKF) and it avoids the expensive integration of the state error covariance matrix by propagating an ensemble of state vectors from which the error covariance is calculated (Reichle et al., 2002b). However the computational efficiency depends on the number of ensembles used in the EnKF. A schematic diagram of the EnKF is shown in Figure 5.1.

Any hydrologic data assimilation system typically requires a state of the art land surface model, observation data of the variable of interest and a data assimilation algorithm. A nonlinear hydrologic model can mathematically be expressed as:

$$x_{k+1} = f_k x_k + w_k (5.1)$$

where x_k is the prognostic state variable vector at time 'k', f(.) is the nonlinear forward operator which mostly includes the forcing data and w_k is the model error which includes all the uncertainties in the model or the forcing data.


Fig. 5.1: A schematic diagram of the Ensemble Kalman Filter (EnKF; after Reichle et al. (2002b)).

The observations at any time 'k+1' can be put together into an observation vector y_{k+1} . The relationship of this observation vector to the model state vector can linearly be expressed as:

$$y_{k+1} = H_{k+1}(x_{k+1}) + v_{k+1}$$
(5.2)

where H_{k+1} is the linear operator and v_{k+1} is the uncertainties associated with the observations due to measuring instruments or representativeness of the observations in the model space.

There are few assumptions regarding the errors in Equations (5.1) and (5.2) required for the optimal data assimilation algorithms: w_k and v_{k+1} are (i) gaussian in nature with mean zero and covariance vectors Q_k and R_{k+1} respectively; (ii) random (white noise), i.e. they are uncorrelated in time and (iii) mutually independent.

EnKF assimilation method consists of two steps, (a) the forecast or propagation step and (b) the update step. During the forecast step, the EnKF propagates an ensemble of state vectors in parallel (each state vectors represent a realization of the Equation (5.1)) from the initial state estimate at time 'k' (x_k^{i+}) to give a model forecast at time 'k+1' (x_{k+1}^{i-}) which can be represented as:

During the update stage, the EnKF first calculates the diagnosed state error covariance (P_{k+1}^-) and Kalman gain (K_{k+1}) from the spread of the model ensemble forecasts and the observation error covariance using the equations described in Equations (5.4)-(5.7).

$$P_{k+1}^{-} = \frac{1}{N-1} Z_{k+1} Z_{k+1}^{T}$$
(5.4)

$$K_{k+1} = P_{k+1}^{-} H_{k+1}^{T} [H_{k+1} P_{k+1}^{-} H_{k+1}^{T} + R_{k+1}]^{-1}$$
(5.5)

$$Z_{k+1} = [x_{k+1}^{1-} - x_{k+1}^{-}, \dots, x_{k+1}^{N-} - x_{k+1}^{-}]$$
(5.6)

where,

$$x_{k+1}^{-} = \frac{1}{N} \sum_{i=1}^{N} x_{k+1}^{i-}$$

$$Z_{k+1}^{T} = \text{the transpose of } Z_{k+1}$$
(5.7)

EnKF then updates each model ensemble member prediction separately using the model forecast states and the observations through Equation (5.8).

$$x_{k+1}^{i+} = x_{k+1}^{i-} + K_{k+1} \left[y_{k+1} - H_{k+1} x_{k+1}^{i-} + v_{k+1}^{i} \right] \qquad i = 1, 2, \dots, N$$
(5.8)

The accurate estimation of the state error covariance depends on the size of the ensemble members. The spread of the ensemble decreases and hence the uncertainty associated with the ensembles also decreases after the update stage. The EnKF state estimate at any time step is the mean of the ensemble members.

5.4 Model Description, Experimental Design in LIS & Study Area

5.4.1 The EnKF Module in LIS

Land Information System (LIS) is a land surface modeling framework developed at the NASA Goddard Space Flight Center (GSFC). LIS provides a lot of flexibility to the land surface modelers because of its object oriented programming and the plug and play architecture and expandable system. It includes a core which is the primary software component and many extensible software components integrated to the core to handle the land surface models (LSMs), input forcing and parameter datasets and output datasets. This LIS system right now supports multiple LSMs, input datasets. LIS can also be

customized and expanded by the users with more user defined LSMs and datasets (Kumar et al., 2006).

Recently, a data assimilation (DA) module has been added to the LIS system. This DA module includes three basic components for the observations, land surface models and the DA algorithms. For the EnKF DA algorithm, a perturbation component has also been included to account for the model states and the error characteristics. The observation component reads and processes observation datasets of the variable to be assimilated (e.g. soil moisture, snow water equivalent). The land surface component takes care of the model predictions of the model prognostic variables before data assimilation. Then the DA algorithm component holds various DA algorithms (EnKF, EKF (Extended Kalman Filter), DI (Direct Insertion)) used for the data assimilation process. For a comprehensive discussion of this LIS DA module, please refer to Kumar et al. (2008).

A typical cycle of an EnKF data assimilation step in LIS includes (i) running the user defined land surface model and creating a land surface state variable dataset (from any time step 'k' to the next time step 'k+1'; forecast step), (ii) reading in the user provided observation data for the variable of the interest and creating an observation state (at time step 'k+1') and (iii) calculating the Kalman gain and updating the land surface model state at that time step (at time step 'k+1'; update step).

5.4.2 The Noah Land Surface Model

The Noah LSM gets its lineage from the OSU LSM originally developed in the 1980's at Oregon State University (Mahrt and Pan 1984). It has been upgraded and extended by

the National Center for Environmental Prediction (NCEP) and its collaborators (Chen et al., 1996). Noah model includes four soil layers, a vegetation layer and a snow layer. Noah model prognostic variables include soil moisture and temperature in soil layers, water stored on the canopy and the snow stored on the ground (Chen and Dudhia, 2001). A detailed description of the Noah model is provided in Sahoo et al. (2008b).

We performed a multi-model soil moisture comparison study over the Little River Experimental Watershed for 2003 using the HySSiB (Hydrologic improvement of Simplified Simple Biosphere model), Noah and CLM (Community Land Model) land surface models. Despite different model physics and parameterizations among the three models, the Noah model soil moisture retrievals outperformed those of the other two models (Sahoo et al., 2008b). That provided us some confidence to choose the Noah model over other models for this data assimilation study.

5.4.3 LSMEM Soil Moisture Observations

The Land Surface Microwave Emission Model (LSMEM) is an iterative forward model which is based on the radiative transfer theory described in Kerr and Njoku (1990). We used this model to retrieve soil moisture at 10.7H GHz frequency using the AMSR-E observed brightness temperature data. This model has been previously used to retrieve soil moisture from other sensors such as Electronically Scanned Thinned Array Radiometers (ESTER; Gao et al., 2004); Tropical Rainfall Measuring Mission (TRMM/TMI; Gao et al., 2006) and AMSR-E (McCabe et al., 2005a; 2005b). A detailed description of the LSMEM model has been provided in Sahoo et al. (2008a).

AMSR-E overpass is twice a day at 130 LST (ascending pass) and 1330 LST (descending pass). Hence LSMSM soil moisture could also be retrieved twice a day. We evaluated the LSMEM soil moisture retrieval along with the current operational AMSR-E soil moisture product against the in-situ observations over the Little River Experimental Watershed for 2003. LSMEM provided better soil moisture results as compared to that of the AMSR-E (Sahoo et al., 2008a). Hence, we chose LSMEM soil moisture retrieval for this data assimilation study.

We did some preprocessing of the LSMEM soil moisture data before using for the data assimilation. The Noah land surface model was simulated at fine spatial scale (1 km resolution) over the watershed, but the LSMEM soil moisture was at 25 km resolution. So, we regridded the observation data corresponding to the model grid resolution. We made sure that the observation data were in the correct unit (% vol/vol) required by the LIS DA module. Also, we provided a metadata file with the observations which was needed for the DA module. The metadata file included some statistical information about the observation like minimum and maximum values, observation error etc.

5.4.4 Study Area

A detailed description of the study area is described in Section 2.2.1 Figure 2.1 shows a map of the study area along with the in-situ measuring stations. A description of the physio-climatic conditions and land use patterns has been provided in Sahoo et al. (2008a).

5.4.5 LREW In-Situ Observation Data

A detailed description of the in-situ observation datasets is described in Section 2.2.2.The description of the in-situ data and their processing has also been provided in Sahoo et al. (2008a).

Table 5.1: A brief description of all the model simulations performed for this study

Experiment Name	Description of the Experiment				
openloop_spinup	Model open loop simulation after 15 year model spin-up				
enkf_spinup	Model EnKF (data assimilation) simulation after 15 year				
	model spin-up				
enkf_int	Model EnKF (data assimilation) simulation with cold start				
	(30 % vol/vol model initial guess)				
enkf_dry	Model EnKF (data assimilation) simulation with cold start				
	(3 % vol/vol model initial guess)				
enkf_wet	Model EnKF (data assimilation) simulation with cold start				
	(49 % vol/vol model initial guess)				
openloop_int	Model open loop simulation with cold start (30 % vol/vol				
	model initial guess)				
openloop_dry	Model open loop simulation with cold start (3 % vol/vo				
	model initial guess)				
openloop_wet	Model open loop simulation with cold start (49 % vol/vol				
	model initial guess)				

5.5 Results and Discussion

The input meteorological forcing and the model prognostic variables were perturbed using perturbation routines to create ensembles required for the EnKF DA algorithm. The ensemble size was set to 12 for computational efficiency. Kumar et al. (2008) have shown that the RMSE errors for surface and root zone soil moisture do not decrease much with the increase of ensemble size above 10. The model prediction at any time step was calculated as the mean of these ensembles. The model was simulated at one hour time step, but the data assimilation was performed twice a day when the observations were available (at 130 and 1330 LST). Table 5.1 gives a brief description of the model experiments performed for this study.

5.5.1 Soil Moisture

Figure 5.2 shows the daily top layer soil moisture for 2003 from 'openloop_spinup' run, 'enkf_spinup' run, in-situ observations and the LSMEM retrievals averaged over the watershed. It can clearly be noticed that there is a large persistent bias between the 'openloop_spinup' run and the LSMEM observations used for data assimilation. The EnKF algorithm does a good job by updating the model prediction to the observations and bringing the model predictions closer to the observations in the 'enkf_spinup' run. We assume the in-situ observations as ground truth in this study. It is clearly evident from this figure that the 'enkf_spinup' results are within a close range of the in-situ observations. Figure 5.3 shows the scatter plot for the Noah 'openloop_spinup' and 'enkf_spinup' results against the in-situ observations of the daily top layer soil moisture data for 2003. A 1:1 line is shown in each scatter plot for reference. There is a consistent high positive mean bias for the 'openloop_spinup' run (8.78 % vol/vol). The Root Mean Square Deviation (RMSD) is also very high (8.92 % vol/vol) because of high

bias in the 'openloop_spinup' case. At the same time, the correlation coefficient is also high (0.90) for the 'openloop_spinup' case irrespective of high mean



Fig. 5.2: Daily top layer soil moisture time series plots from Noah (10 cm layer) open loop and EnKF simulations, in-situ measurements (5 cm layer) and LSMEM retrieval (~ 1 cm).

bias and RMSD indicating a very consistent variability of the results with the in-situ observations. In contrast to the 'openloop_spinup' run, the 'enkf_spinup' run shows very low negative mean bias (-0.02 % vol/vol) as well as very low RMSD (2.47 % vol/vol). The correlation coefficient for the 'enkf_spinup' run (0.80) is lower than that of the open loop run. This is evident from the large scatter of the soil moisture in the 'enkf_spinup'

case. It is worth mentioning here that we did not train the satellite observations with the in-situ observations which we used for the data assimilation study. Otherwise, the data assimilation results might have produced better statistics when compared with the in-situ observations in Figures 5.2 and 5.3. Figure 5.4a shows the spatial map of the temporally averaged absolute improvement values for the 'enkf_spinup' run over the 'openloop_spinup' run. The absolute improvement was calculated using the Equation 5.9. There is an absolute improvement in the 'enkf_openloop' run over the entire watershed. The absolute improvement values range from 51 to 78 %vol/vol. The absolute improvement image shows abrupt changes in the values across the vertical lines at -83.75 and -83.5 longitude. Similar kind of abrupt changes across the longitude -83.75 and -83.5 has also been noticed in later plots. Those vertical lines correspond to the grid boundaries for the LSMEM soil moisture observations at 25 km resolution. Figure 5.4b shows a snap shot of the LSMEM soil moisture observations to confirm the grid boundaries.

$$absolute improvement = \frac{enkf_spinup-openloop_spinup}{LSMEM-openloop_spinup} \times 100$$
(5.9)

To access the performance of the EnKF algorithm, the Root Mean Square Errors (RMSEs) are calculated for both the 'openloop_spinup' and 'enkf_spinup' runs with respect to the in-situ observations (ground truth). Figure 5.5 shows the time series of the spatially averaged RMSE results over the whole watershed. The open loop run shows very high consistent RMS error throughout the experiment period with range from 12 to



Fig. 5.3: Scatter plots for the Noah (a) 'openloop_spinup' and (b) 'enkf_spinup' results versus the insitu observations of the daily top layer soil moisture data for 2003. A 1:1 line is shown in each scatter plot for reference.



Fig. 5.4: (a) Time averaged percentage soil moisture improvement in 'enkf_spinup' results over the 'openloop_spinup' results for the LREW. The image shows abrupt changes in the values across the vertical lines at -83.75 and -83.5 longitude. Those vertical lines correspond to the grid boundaries for the LSMEM soil moisture observations at 25 km resolution. (b) A snap shot of the corresponding LSMEM soil moisture observation map is shown for reference.

20 % vol/vol. The 'enkf_spinup' simulation immediately reduces the RMSE for the surface soil moisture and show systematic improvement over the open loop run throughout the experiment period. However, the range of the RMSE for the 'enkf_spinup' run is higher (from 2 to 17 % vol/vol). Both the simulations show very similar characteristic RMSE time series with higher values during the Spring and Summer season. It indicates that all the high RMSE values are associated with high precipitation events/wet ground conditions.



Fig. 5.5: Time series of the spatially averaged RMSEs (Root Mean Square Errors) in surface soil moisture for the 'openloop spinup' and 'enkf spinup' simulations over the whole watershed.

A corresponding improvement metric was calculated from the RMSE differences of the 'openloop_spinup' and 'enkf_spinup' runs temporally averaged over the year 2003 to evaluate the improvements in spatially distributed soil moisture values. This kind of improvement metric has been used by Kumar et al. (2008) for their data assimilation study. Figure 5.6 shows the improvement metric map for the surface soil moisture. The positive values show the improvement for the 'enkf_spinup' simulation over the 'openloop_spinup' run. The improvement is evident over the whole watershed. The improvement is higher on the west side as compared to the east side of the watershed. Especially, the south-east side shows very less improvement.

The innovation is defined as the difference between the model forecast and the corresponding satellite observation at any instant of time. Normalized innovation (innovation normalized with their expected covariance) was calculated in LIS as a measure of the performance of EnKF algorithm. The innovation statistics were spatially and temporally averaged over the watershed for the whole year of 2003. The averaged normalized innovation mean was found to be -5.23. The non-zero normalized innovation mean suggests that there was a large bias between the observations and the model prediction. This bias should be addressed carefully to have better model assimilation results. The variance and the lag one autocorrelation were found to be 1.91 and 0.510 respectively. Kumar et al. (2008) have calculated the same statistics for their synthetic



Fig. 5.6: Time averaged improvement metric (RMSE('openloop_spinup')-RMSE('enkf_spinup')) for the surface soil moisture. The positive values show the improvement for the 'enkf_spinup' simulation over the 'openloop_spinup' run. The vertical discontinuities are also very clear at -83.75 and -83.5 longitudes. These are again the 25 km grid boundaries of the LSMEM observation data. Please refer to figure 5.4b to see a snap shot of LSMEM observation data.

data assimilation study in LIS and have provided detailed descriptions for the innovation statistics parameters. In this case, mean has larger negative value where as the variance and the lag one autocorrelation have higher positive values as compared to those of Kumar et al. (2008). The higher values in our case are expected as we use the real satellite observations compared to the synthetic observations of Kumar et al. (2008). Figure 5.7 shows the time averaged spatially distributed variance of the normalized innovation. The variance is higher on the eastern side (especially on the south-east side) of the watershed and is far from unity. This map corresponds very well to the improvement metric map shown in Figure 5.6. As mentioned in this section, this might be because of the uncertainties associated with the real observations. One of those errors could be the representativeness error because of different spatial and vertical resolutions for the observation (25 km spatial resolution and top \sim 1 cm vertical sensing depth) and the model (1 km spatial resolution and 10 cm top layer).



Fig. 5.7: Time averaged spatially distributed variance of the normalized innovation (no unit) over the watershed.

5.5.2 Water Cycle Variables

Figure 5.8 shows the time series for the moisture in the layer 2 (Figure 5.8a), layer 3 (Figure 5.8b) and layer 4 (Figure 5.8c) for the 'openloop_spinup' and 'enkf_spinup' simulations. All these moisture variables are prognostic variables in Noah model. Moisture values for all the three layers are very similar for both the simulations during the early period of the simulation. It takes around 8 days, 20 days and 40 days for the 2nd, 3rd and 4th layer respectively to respond to the change in soil moisture results due to the EnKF simulations. This is expected as the deeper layers respond to the surface layer very late. The magnitude of response is different for each layer. The maximum difference between the 'openloop_spinup' and 'enkf_spinup' soil moisture values are 15, 10 and 4 % vol/vol for the 2nd, 3rd and 4th layer respectively. But the skills of the soil moisture time series are similar for both the 'openloop_spinup' and 'enkf_spinup' and 'enkf_spinup' cases.

Figure 5.9 shows daily time series plots of surface runoff (Figure 5.9a), subsurface runoff (Figure 5.9b) and evaporation (Figure 5.9c) averaged over the watershed for 2003. The surface runoff is strikingly identical for both the simulations. The surface runoff peaks correspond to individual precipitation peaks (not shown here in the figure) throughout the year. It suggests that the surface runoff in Noah model does not get affected much by the change in surface soil moisture during the EnKF simulations. It is because the precipitation water is getting removed through surface runoff before the rest of the precipitation water enters to the soil. Hence, the surface runoff is not directly getting affected due to the change in soil moisture by data assimilation. However, the





Fig. 5.8: Daily time series for the moisture in the layer 2 (Figure 5.8a), layer 3 (Figure 5.8b) and layer 4 (Figure 5.8c) from the 'openloop_spinup' and 'enkf_spinup' simulations averaged over the watershed for 2003.

subsurface runoff shows different magnitude for both the 'openloop_spinup' and 'enkf_spinup' simulations though the pattern is similar. For both the cases, the subsurface runoff peaks correspond to the continuous high precipitation events during the spring season (not shown in the figure).

The difference in the evaporation values for the 'openloop_spinup' and 'enkf_spinup' run is quite evident in Figure 5.9c. As expected, the seasonal cycle is present in the both the evaporation time series curves with higher evaporation during summer months and lower evaporation during the winter months. The high frequency variability in the evaporation time series directly corresponds to the soil moisture change for both the simulations. The top layer soil moisture values are adjusted towards the lower satellite observation values during the 'enkf_spinup' simulations. Hence, the evaporation is lower in case of the 'enkf_spinup' simulation because of less availability of water in the top layer for evaporation to take place.

5.5.3 Energy Cycle Variables

Figure 5.10 shows the daily time series plots of net shortwave (Figure 5.10a), net longwave (Figure 5.10b), latent heat (Figure 5.10c), sensible heat (Figure 5.10d) and ground heat (Figure 5.10e) fluxes averaged over the watershed for 2003. The incoming energy is considered positive where as the outgoing energy is considered negative in this study. The net radiation is calculated as the difference between incoming and outgoing energy. A clear seasonal cycle is present in the net shortwave radiation time series for both the 'openloop_spinup' and 'enkf_spinup' simulations with higher peaks during the summer months due to higher incoming solar radiation. Both the net shortwave time series show very similar values except for the summer months (June -September). The lower soil moisture and the high incoming solar radiation contribute to larger net shortwave radiation for the EnKF case in summer.





Fig. 5.9: Daily time series plots of surface runoff (Figure 5.9a), subsurface runoff (Figure 5.9b) and evaporation (Figure 5.9c) from the 'openloop_spinup' and 'enkf_spinup' simulations averaged over the watershed for 2003.

The net longwave radiation values for the 'openloop_spinup' and 'enkf_spinup' cases match very well except for the summer months (June – September) as we notice in the case of net shortwave radiation. The surface temperature increases with high incoming solar radiation and relatively dry surface layer during the summer months for the 'enkf_spinup' case. Hence, the outgoing longwave radiation increases leading to less net longwave radiation in summer. Contrast to that, the 'openloop_spinup' run produces









Fig. 5.10: Daily time series plots of net shortwave (Figure 5.10a), net longwave (Figure 5.10b), latent heat (Figure 5.10c), sensible heat (Figure 5.10d) and ground heat (Figure 5.10e) fluxes from the 'openloop_spinup' and 'enkf_spinup' simulations averaged over the watershed for 2003.

higher surface soil moisture in the top layer which keeps the surface cooler. Hence the net longwave radiation is higher in this case.

The latent heat flux is nothing but just the manifestation of the evaporation. So, the latent heat flux time series is exactly similar to that of the evaporation time series. The seasonal cycle is present in the latent heat flux time series. The availability of soil moisture in the surface layer directly controls the latent heat flux. The lower prediction of the top layer soil moisture by the 'enkf_spinup' run produces lower latent heat flux than that of the 'openloop_spinup' run throughout the year. On the other hand, the sensible heat flux shows the opposite relationship for both the model simulations. The sensible heat flux for the 'enkf_spinup' run is higher than that of the 'openloop_spinup' run throughout the year. The sensible heat compliments the latent heat flux, but the sensible heat flux show peaks in different seasons for different runs. The sensible heat flux peak for the 'openloop_spinup' run is in spring season (April – May) whereas the corresponding peak for the 'enkf_spinup' run is in summer season (June – July). The ground heat flux is treated as a residual of the energy budget in the model. The ground heat flux shows the peak in spring season in both the cases. The ground heat flux values are very similar in both the model simulations throughout the year. This indicates that the change in top layer soil moisture in both the model simulations readjusts the outgoing energy and water cycle parameters from the surface and does not directly affect the ground heat flux.

5.5.4 Sensitivity Study

This section explains two sensitivity studies performed on the EnKF algorithm with different model initialization conditions. A description of all the designed model simulations is provided in Table 5.1. All the simulations were run for 2003 with one hour time step. The run domain was over the Little Rive Experimental Watershed with 1 km spatial resolution. For all the sensitivity studies, we chose a single location (Station RG31) instead of the spatially averaged model simulation results.

Table 5.2: The recovery time for the dry and wet runs with respect to the intermediate runs for all the four soil moisture layers for both the openloop and EnKF simulations.

Open Loop Simulations							
Experiment time	Recovery time with respect to the openloop_int simulation results						
	Layer 1	Layer 2	Layer 3	Layer 4			
openloop_dry	64 days	80 days	365+ days	365+ days			
openloop_wet	6 days	46 days	174 days	365+ days			
EnKF Simulations							
Experiment	Recovery time with respect to enkf_int simulation results						
name							
	Layer 1	Layer 2	Layer 3	Layer 4			
enkf_dry	4 days	61 days	243 days	365+ days			
enkf_wet	6 days	158 days	204 days	365+ days			

5.5.4.1 Sensitivity of EnKF Algorithm to Model Initialization

Model initialization plays a major role in controlling the evolution of the model parameters through time. Model initialization has been a great concern in the modeling community. Here we tested the evolution of model soil moisture parameters in the open loop and EnKF cases given different model soil moisture initialization condition. We chose three different initial soil moisture conditions for this study: (a) dry start (3 % vol/vol), (b) intermediate start (30 % vol/vol) and (c) wet start (49 % vol/vol). Therefore, we performed six model simulations, three each for the open loop and EnKF cases. The soil moisture values for the intermediate start were taken as reference in both the simulations and the results from dry and wet starts were compared to those reference values. Table 5.2 shows the recovery time for the dry and wet runs with respect to the

intermediate runs for all the four soil moisture layers for both the openloop and EnKF simulations.

Figure 5.11a shows the time series of the soil moisture from all six simulations for the model top layer. We assimilated the top layer soil moisture observations directly to the first layer of the model. As expected, the model top layer soil moisture should directly be controlled by the observation values. The results in Figure 5.11a support that. After a few model time steps, the top layer soil moisture predicted exactly same values for the EnKF case irrespective of different initialization conditions. Hence for the EnKF case, the initialization condition did not have much impact on the evolution of top layer soil moisture. Contrast to the EnKF case, the model initialization had an impact on the soil moisture evolution for the open loop case. The soil moisture for the wet start reaches the saturation immediately and the moisture was removed from the top layer through drainage. So, the top layer soil moisture decreased quickly and followed the soil moisture time series of the intermediate initialization case. The soil moisture values for the dry initialization case increased slowly with subsequent precipitation events. It took almost 64 days for the dry case to recover and reached close to the soil moisture values for the intermediate initialization.

Figure 5.11b shows the soil moisture time series for the 2nd layer for all the six simulations. The impact of observations on the sub-surface model layers depends on the strength of the vertical coupling of soil moisture among different layers and the impact decreases with depth (Kumar et al., 2008). For the Noah model 2nd layer, the initialization condition has an impact on the soil moisture evolution for the EnKF as well as the open

loop cases. In the EnKF case, it took almost 61 days and 158 days for the dry and wet initialization conditions respectively to catch the soil moisture values for the intermediate initialization condition. In the open loop case, the behavior of the 2nd layer soil moisture time series was exactly same as that of the 1st layer though it took 80 days for the dry case here to catch the other two soil moisture time series. Similarly, it took 46 days for the wet initialization case.

Figure 5.11c shows the soil moisture time series for the 3rd layer. The impact of observation in the EnKF case decreased further for the 3rd layer. We can notice that the soil moisture for the dry initialization condition in both the EnKF and open loop cases never got updated till mid-February until the persistent spring precipitation events occurred. The same behavior can be noticed for the 4th layer in Figure 5.11d. This suggests the weak vertical coupling of the soil moisture between layers in the Noah model which has also been mentioned by Kumar et al. (2008). For the 3rd layer, soil moisture evolution showed similar behavior as that for the 2nd layer though the recovery time was longer due to weak vertical coupling. The soil moisture for the dry and wet initialization conditions caught the soil moisture time series for the intermediate case after 200 days (in September) in the EnKF case. In the open loop case, the soil moisture for the dry initialization never reached the other two soil moisture values during the whole one year of the model simulation. The soil moisture for the wet initialization took almost 174 days to catch the soil moisture time series for the intermediate initialization.

Figure 5.11d shows the soil moisture evolution for the 4th layer. The impact of model initialization condition on the EnKF and open loop simulations is highly evident in



Fig. 5.11: Sensitivity results of EnKF to the model initialization conditions for the (a) Layer 1, (b) Layer 2, (c) Layer 3 and (d) Layer 4 at station RG31 for 2003.

this figure. The soil moisture time series for both the wet and dry initialization conditions and for both the model simulation cases (EnKF and open loop) never recovered and hence never caught the corresponding soil moisture values for the intermediate initialization condition in the whole one year study period.

This sensitivity study clearly shows the discrepancies in the soil moisture results in both the EnKF and open lop cases due to different model initialization conditions. The impact of initialization condition on the EnKF and open loop soil moisture simulations is higher in the deeper layers as compared to those in the surface layer.

5.5.4.2 Sensitivity of EnKF Algorithm to Model Spin-up

Model spin-up is a very important factor to avoid any bias of the model to an abrupt initial condition. All our previous EnKF and open loop simulations have been carried out with 15 year model spin up except the above sensitivity study where the simulations have been done with cold start mode and different model soil moisture initialization conditions. Here in this case, we compared the soil moisture results generated after the model spin-up and from the cold start conditions only in the EnKF simulation case. We designed four EnKF experiments for this sensitivity study. One experiment was the original EnKF simulation which used 15 year model spin-up. The other three model EnKF experiments used three different cold start conditions. The experiments have been described in Table 5.1. In this sensitivity study, the EnKF results from the spin-up were used as reference data and the results from other experiments were analyzed with respect

to the spin-up results. Table 5.3 shows the recovery time for the cold start runs with respect to the spin-up run for all the four soil moisture layers for the EnKF simulation.

Figure 5.12a shows the top layer soil moisture time series for all the four model simulations. It is clearly evident from this figure that all the model simulations produced same soil moisture values through out the experiment period irrespective of different cold start conditions. The soil moisture time series from all the cold start runs took as little as 1 to 10 days recovery time to catch the EnKF spin-up results. This is expected since the soil moisture observations were directly assimilated to the top layer model soil moisture. Hence, the top layer model soil moisture was mostly controlled by the observation values.

Table 5.3: The recovery time for the cold start runs with respect to the spin-up run for all the four soil moisture layers for the EnKF simulation.

EnKF Simulations							
Experiment name	Recovery time with respect to enkf_spinup simulation results						
	Layer 1	Layer 2	Layer 3	Layer 4			
enkf_int	1 day	2 days	10 days	25 days			
enkf_dry	10 days	115 days	244 days	365+ days			
enkf_wet	4 days	172 days	205 days	365+ days			



Fig. 5.12: Sensitivity results of EnKF to the model spin-up for the (a) Layer 1, (b) Layer 2, (c) Layer 3 and (d) Layer 4 at station RG31 for 2003.

Figure 5.12b shows the time series for the 2nd layer soil moisture for al the four experiments. One of the cold start model runs had 30 % vol/vol initial soil moisture condition (intermediate wetness) which was reasonably close to the initial soil moisture condition found after the 15 year model spin-up. Hence the soil moisture time series for the 'enkf_int' experiment recovered very quickly even for the 2nd layer and followed the corresponding soil moisture time series for the 'enkf_spinup' experiment rest of the experiment time period. But, the soil moisture time series for the 'enkf_dry' and the 'enkf_wet' experiments took 115 and 172 days respectively before reaching the soil moisture time series for the 'enkf_spinup' experiment. This was because the soil moisture initialization conditions for both these experiments were far off from that of the 'enkf_spinup' experiment and hence took more time to recover.

Figure 5.12c and 5.12d show the soil moisture time series from all the four experiments for the model 3rd and 4th layer respectively. The soil moisture results for both these deeper layers show very similar behavior as we find for the 2nd layer. But the recovery time for the cold start experiments is longer as we go to deeper layers. For the deepest layer, the soil moisture time series for the 'enkf_dry' and 'enkf_wet' experiments never recovered during the whole experiment time period. This was because the observed soil moisture from top few centimeters soil layer did not have much impact on the deeper soil layers during the EnKF simulations. This also indicates that the vertical coupling between the soil layers for this Noah model is not very strong.

This sensitivity study clearly shows that the EnKF results are strongly affected by the model spin-up conditions. So, model spin-up is necessary especially for the accurate deeper layer soil moisture estimates in the EnKF case.

5.5.5 1-D versus 3-D Data Assimilation

1-D EnKF algorithm only updates the vertical soil layer moisture profiles (1-D column) at individual points in space by vertically aggregating the objective function over different soil layers during the data assimilation, whereas 3-D EnKF updates the moisture in vertical layers as well as in the horizontal space during the data assimilation. So, the 3-D filter would account for the horizontal as well as vertical distribution of moisture in soil layers. 3-D data assimilation has some advantages over the 1-D vertical data assimilation. For example, (a) if the observations are available over a partial region of the study area, then through 3-D data assimilation the model soil moisture prediction can be updated over the region where the observations are not available if there exists a systematic relationship of the innovation (satellite observation – model forecast) values among different regions over the study area; (b) if there is any systematic constant error present in the observations (such as instrument error), then the 3-D filter could account for that error during the data assimilation. Figure 5.13 shows the temporal correlation of the innovation (satellite observation – model forecast) among different grid locations for 2003 against their separation distance over the LREW. It is easily noticed that a consistent relationship exists among the grid locations in X- and Y-directions (Figure



Fig. 5.13: Spatial correlation of the innovation (satellite observation – model forecast) among different grid locations for 2003 against their separation distance over the LREW (a) the grid locations only in X- and Y-directions and (b) the grid locations in any random direction (Grid locations in X- and Y-directions are not included in this plot).
5.13a) though this relationship is not very clear in the random directions (Figure 5.13b). So, the relationship in the innovation values in the X- and Y-directions can be used to update the model predictions in X- and Y-directions whenever the satellite observations are not available. But at the same time, 3-D assimilation is computationally very expensive as compared to the 1-D data assimilation. In our data assimilation study here, we used a 1-D data assimilation filter. So, we directly assimilated the satellite soil moisture observations into the Noah model top layer, but updated three sub-surface layers along with the surface layer whenever the observation was available (through vertical 1-D filter). We did not perform any update on the horizontal or lateral direction (3-D filter). There are few reasons for that: (a) our AMSR-E satellite observations were available over the whole watershed. So, we did not require a 3-D filter to update the soil moisture field in horizontal direction; (b) the Noah model does not include the lateral moisture flow physics. So, even if we try to update the soil moisture field in horizontal direction due to the data assimilation, the Noah model physics will not propagate the update information between the grid cells. But it will definitely be interesting to apply a 3-D EnKF algorithm to verify the usefulness of this filter over the 1-D EnKF algorithm.

5.5.6 Model Hydrology

The temporal characteristic of the model ensemble members for the top layer soil moisture remained similar to the open loop simulation after perturbing the input forcing variables. The assimilated soil moisture showed the space-time variability realistically for the top layer. But the same characteristic was not found for the deeper layers. The dry-out

and wetting phenomena were not simulated properly for the deeper layers during the data assimilation. This is because the deeper layers are mostly controlled by the model physics and the model physics dominated over the data assimilation in controlling the behavior of the deeper layer soil moisture. The vertical propagation of the assimilated innovation was not sufficient to update the deeper soil layers in the model quickly. This can be noticed in the daily soil moisture increment plots for all the four soil layers in Figure 5.14. It was not because of the filter parameters, but because of the strength of the vertical coupling between soil layers in any specific model. The deeper layers were mostly benefited by the data assimilation in the top layer since the deeper layers were only controlled by simple model dynamics. But it has been shown by Kumar et al. (2008) that the vertical coupling is weak in the Noah model. Hence the deeper layers took longer time to respond to the changes in the top layer soil moisture due to the data assimilation. In the horizontal space, the model does not simulate the lateral flow. Hence, there should be some correlated errors between different vertical model columns. The horizontal propagation of the update due to the 3-D data assimilation filter might get constrained by the unavailability of the lateral flow physics in the model. But we did not use the 3-D filter in our study. The update in the top layer soil moisture also affected other water and energy cycle variables in the model which were mostly controlled by the model physics and related to the top layer soil moisture through the water and energy cycle relationship.

5.6 Conclusions

This chapter describes the performance of an optimal sequential data assimilation algorithm for the soil moisture estimate. The EnKF algorithm was very successful to



Fig. 5.14: Spatially averaged daily soil moisture increment over the LREW (a) Layer 1, (b) Layer 2, (c) Layer 3 and (d) Layer 4 for 2003.

constrain the model forecast by the observations throughout the study period. It improved the model soil moisture forecast which was evident from RMSD values. The absolute improvement and the improvement metric illustrated that the EnKF algorithm improved the model forecast throughout the watershed though the magnitude of improvement was different in different parts of the watershed. We did not train our LSMEM retrieved soil moisture with the in-situ observations before performing the data assimilation. Otherwise, the improvement in the data assimilation results might have been higher than what we got right now. The normalized innovation mean and standard deviation were also analyzed for this data assimilation experiment. High normalized innovation mean suggests that there was a consistent bias between the model prediction and the observation values used for the data assimilation. So, it is necessary to do the bias correction to remove the bias between the model and the observations before performing such kind of data assimilation study. We used a true 1-D data assimilation filter in this research work. So, all the data assimilation updates were in vertical direction. During the data assimilation, each grid was treated as independent column and there was no interaction between each grid cell with the surrounding grid cells in horizontal space. We found very high spatial correlation for the innovation values among grid cells in X- and Y-directions. This indicates that the 3-D filter might have been more appropriate than the true 1-D filter in our research work.

The change in the model top layer soil moisture values due to the data assimilation affected the other model water and energy cycle variable outputs which were directly or indirectly controlled by the surface layer soil moisture. The model results showed large differences for variables such as evaporation, latent and sensible heat fluxes between the open loop and the EnKF simulations. So, a coupled land-atmosphere data assimilation system is required to understand how the atmospheric variables (e. g. precipitation) respond to such changes in the land to atmosphere feedback terms due to data assimilation and in turn how they affect the land surface variables (e. g. soil moisture).

The sensitivity studies were performed to verify the robustness of the data assimilation algorithm. The results suggest that the top layer soil moisture was mostly not affected by different model initialization conditions and spin-up since the top layer soil moisture was directly constrained by the observations. But the moisture content in the subsurface layers took longer recovery time as we moved deeper from the surface since the effect of the observations through the data assimilation was reduced with distance from the surface. The magnitude of the recovery time depended how dry or wet the model initialization condition was from the mean soil moisture condition of the region and how strong the vertical coupling was among the soil layers in a land surface model. The coupling strength is model dependent, hence the recovery time too. Nevertheless, a near perfect soil moisture initialization condition is required for the model data assimilation to get accurate prediction with less recovery time. This is often difficult to start the model simulations with a perfect soil moisture initialization condition. One guess for that would be to use the mean soil moisture content of the study region as a model initial guess. The other approach would be to use model spin-up to get a better initialization condition. The model spin-up reduces any bias associated with the model due to poor model initialization condition and provides a reasonable model initial guess for better model prediction. The results in this study showed that the data assimilation results with different model initialization condition converged to the assimilation results produced after a model spin-up with different recovery time depending on how far those initial guesses were from the model spin-up produced initial condition.

This study has focused on the performance and sensitivity of the data assimilation algorithm. The results could be model dependent; and the model representation and the choice of model error parameters can affect the results. Nevertheless, the results have raised many issues such as bias correction, coupled data assimilation system, better model initialization conditions which should be carefully considered while performing the data assimilation study to improve the assimilation results.

Chapter 6. Applications of Research Results

6.1 Introduction

Soil moisture is a very important component of the global water and energy cycle and land-atmosphere coupling system. It provides the moisture as well as thermal inertia through its heat storage which are required to drive the global climate system (Famiglietti et al., 1998). The relationship of soil moisture with other global climate system components is very complex and hardly understood so far. Also the quantification of soil moisture variability is a challenge because the variability could range from minutes to months at temporal scale and centimeters to thousands of kilometers at spatial scale. The small scale variability is mostly controlled by the land parameters such as vegetation, topography, soil type and texture whereas the longer scale variability is impacted by the variability of climatic conditions and the amount of water coming in and going out of soil by precipitation and evapotranspiration respectively (Entin et al., 2000). The soil moisture quantification at different temporal and spatial scales has applications in different areas (e.g. vegetation phenology, crop monitoring, extreme event detection: drought and flood monitoring, forest fire detection, watershed management, ecological conservation etc.) and is of interest to different groups of people (e.g. local farmers, agriculture scientists, hydrologists, atmospheric scientists, policy makers). There have been many local scale field experiments carried out to understand the complex physics

related to soil moisture, measure the local scale variability of soil moisture and validate the other soil moisture products from the satellites and land surface models. But global soil moisture estimation through field experiments is not feasible from cost and labor point of view. Similarly the soil moisture data from satellites and models have their own advantages as well as disadvantages (e.g. poor temporal and spatial resolution and temporal frequency in satellite products; poor and simple physics in land surface models and large disparity in results across different models) though we can estimate global soil moisture from these two sources.

Various soil moisture data products have been used in the past to study its effects/applications on different climatic, biological and ecological components. For real world applications, we require continuous operational soil moisture product. This chapter here discusses two soil moisture applications using an in-house model-satellite assimilated soil moisture product which has already been tested with the current satellite soil moisture product and in-situ field observations data over the Little River Experimental Watershed in Georgia and has the potential of producing the operational global soil moisture product. A detailed description of the generation of this assimilated product and the validation study and the quantitative assessment of this product have been provided in Chapter 5. The required input satellite retrieved soil moisture (LSMEM) data and the land surface model (Noah) have been extensively discussed in Chapter 3 and Chapter 4 respectively. This chapter here is only going to show two applications of this assimilated soil moisture product generated in this research study: (i) the soil moisture – vegetation growth relationship and (ii) study of extreme events. For this application

chapter, the study area was extended to the whole USA. So, the assimilated soil moisture was generated over the whole USA. But one thing we need to keep in mind is that the assimilated product has only been validated over the Little River Experimental Watershed, Georgia. We will perform more validation study over different regions in future.

6.2 Soil moisture – Vegetation Growth Relationship

Various global vegetation indices such as Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Fractional Vegetation Cover (FVC) are calculated from satellite observations to monitor the Earth's terrestrial photosynthetic activities.

NDVI is the most commonly used vegetation index. It is defined as (Rouse et al., 1974; Matsushita et al., 2007):

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$$
(6.1)

where, ρ_{red} and ρ_{nir} represent reflectances at the Red (0.6-0.7µm), and Near-Infrared (NIR) wavelengths (0.7-1.1µm), respectively.

By design, NDVI ranges between -1 and 1, but the typical range is between about -0.1 (not very green area) to 0.6 (for a very green area) (Kidwell, 1990). Since the NDVI is a ratio, it cancels out a large proportion of the noise caused by changing sun angles,

topography, clouds or shadows, and atmospheric conditions (Huete and Justice, 1999). However, the NDVI equation is still susceptible to large sources of errors and uncertainties over variable atmospheric and canopy background conditions (Liu and Huete, 1995; Gao, 1996).

While EVI is calculated similar to NDVI, it is based on a feedback based approach which takes care of the distortions in the reflected light caused by the particles in the atmosphere as well as the ground cover below the vegetation. It has very improved sensitivity to high biomass regions and it does not become saturated as easily as NDVI over the highly vegetated areas with large amount of chlorophyll.

EVI is defined as (Liu and Huete, 1995):

$$EVI = G \times \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + (C_1 \times \rho_{red} - C_2 \times \rho_{blue}) + L}$$
(6.2)

where, L is a soil adjustment factor, and C_1 and C_2 are coefficients of the aerosol resistance term which use the blue band to correct aerosol scattering in the red band,

 P_{blue} represents reflectance at the blue (0.45-0.52µm) wavelength and G is a gain factor. Typically, G, C₁, C₂ and L values are taken as 2.5, 6.0, 7.5 and 1 respectively (Huete et al., 2002).

FVC represents the fraction of a pixel covered with vegetation. FVC can be derived from NDVI. There is generally no photosynthesizing vegetation below the lower threshold (value close to zero) of NDVI. So, FVC is assigned 0.0 below the lower

threshold of NDVI. Similarly, the FVC is assigned 1.0 above the upper threshold (value close to 1.0) of NDVI where the pixel is assumed to be totally covered with photosynthesizing vegetation. Between the lower and upper threshold of NDVI, FVC increases approximately as the square of NDVI (Gillies and Carlson, 1995). FVC can be shown as:

$$FVC = 0.0 \quad if N \le T_{1}$$

$$= [(N - T_{1})/(T_{2} - T_{1})] \times [(N - T_{1})/(T_{2} - T_{1})] \quad if T_{1} < N < T_{2}$$

$$= 1.0 \quad if N \ge T_{2}$$
(6.3)

where, N = NDVI value, $T_1 = lower threshold and <math>T_2 = upper threshold$.

FVC is less sensitive to the background noise and it captures the changes in vegetation during the growing season very well (Adegoke and Carlton, 2002).

6.2.1 Datasets used

6.2.1.1 Land Cover Data

The 1 km global land cover data were collected from University of Maryland (Hansen et al., 2000). It is a static dataset. This dataset has a total of 13 land cover classes excluding water bodies (Table 6.1). This dataset was basically used in this study to find out different vegetation cover areas over the USA.

6.2.1.2 Soil Moisture Data

The assimilated soil moisture data from two different soil depths were used in this study. The surface soil moisture was taken from the surface 10 cm (model top layer) of the soil layer. The root zone soil moisture was taken from the top 1 m (model top three soil layers) of the soil layer. Both the datasets ere at 25 km spatial resolution and 12 hour temporal resolution over the USA from 2002 to 2005.

 Table
 6.1:
 Vegetation
 class
 values
 and
 their
 descriptions
 (Source:

Vegetation Class	Description Water					
0						
1	Evergreen Needleleaf Forest					
2	Evergreen Broadleaf Forest					
3	Deciduous Needleleaf Forest					
4	Deciduous Broadleaf Forest Mixed Forest Woodland					
5						
6						
7	Wooded Grassland					
8	Closed Shrubland					
9	Open Shrubland					
10	Grassland					
11	Cropland					
12	Bare Ground					
13	Urban and Built-up					

http://www.geog.umd.edu/landcover/1km-map/meta-data.html)

6.2.1.3 NDVI Data

NDVI data were collected from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite, on board NASA's AQUA (since 2002) satellites, because of the better performance of MODIS NDVI in recent past. The data are global sixteen day composite

files at 5 km spatial resolution. We preferred the 16-day composite data instead of daily data to smooth out the short term variations available in the daily data due to the atmospheric conditions including cloud and aerosol, variations in the soil background noises, surface bidirectional reflectance and sensor registration errors (Rahman et al., 2005). The data were collected for the growing season (April – September) for 2003-05 since we had the assimilated soil moisture data for the complete growing season during those years only.

6.2.1.4 EVI Data

The EVI data were collected from the same source as that of the NDVI data. It is an operational product generated from MODIS along with NDVI data.

6.2.1.5 FVC Data

FVC data were derived from the NDVI data files using the FVC and NDVI relationship mentioned above. The minimum and the maximum values for each NDVI 16-day composite file were considered as the lower and upper threshold values for NDVI which were required for the calculation of FVC values.

6.2.2 Results and Discussion

The land cover map was used to visually find out the regions over USA with distinct vegetation cover types. Since the soil moisture data were at 25 km resolution, we choose study areas with spatial size more than 25 km \times 25 km area for any single land cover

type. We used 80 % as threshold for any particular vegetation type to represent a study area. We find total 8 different study areas with distinct vegetation covers. The study areas are shown in Figure 6.1.



Fig. 6.1: Different land cover (vegetation types) locations considered in this study. The X-axis and Y-axis represent longitudes and latitudes respectively.

To evaluate the sensitivity of the vegetation indices to the soil moisture, we integrated the 12-hourly assimilated soil moisture data over the same 16-day periods as those for the vegetation indices. The seasonal cycle is the dominant signature in both soil moisture and vegetation indices data. So, we deseasonalized all the datasets by taking the difference between the original data and their respective multiyear annual mean data to remove the effect of the seasonal cycle for the sensitivity study.

6.2.2.1 Correlation Results for Concurrent Data

Figure 6.2a shows the correlation coefficient values for all the three vegetation indices against the surface soil moisture (top 10 cm). The x-axis represents different vegetation types. The concurrent correlation between top layer soil moisture with all the three vegetation indices is very weak with correlation values ranging between 0.04-0.31, 0.11-0.33 and 0.13-0.29 for NDVI, EVI and FVC respectively. The correlation is generally higher for the forests, woodlands and wooded grassland whereas it is lower for the shrubs, grasses and croplands. It is kind of surprising to see higher correlations for the bigger trees such as for forests since the bigger trees should be less susceptible to the soil moisture changes at a shorter time scale. Similar kind of relationship was found by Adegoke and Carlton (2002) and they provided few possible explanations (susceptibility of different vegetation types to the local air temperature stress and differences in hydrological and ecological characteristics for different vegetation types) for such kind of relationship. Figure 6.2b shows the corresponding correlation coefficient values as in Figure 6.2a, but for root zone soil moisture (top 1 meter). Here we find kind of opposite



Fig. 6.2: Correlations between three vegetation indices and concurrent (a) top layer and (b) root zone soil moisture during the vegetation growing season for different vegetation types. Please see Table 6.1 for the description of different vegetation classes.

relationship for different vegetation covers with the root zone soil moisture as compared to Figure 6.2a. The shrubs, grasses and woodlands are highly correlated with the deep layer soil moisture with correlation values ranging between 0.215-0.63, 0.21-0.65 and 0.24-0.66 for the NDVI, EVI and FVC respectively, whereas the correlation values are lower than 0.3 for all the three vegetation indices for forests, woodlands and wooded grasslands. This kind of relationship is expected for concurrent soil moisture and vegetation comparisons at shorter time scales. As compared to the top layer soil moisture, the correlation is generally higher for all the vegetation types with the root zone soil moisture. This could be because the soil moisture variability in the top layer is at very fine temporal scale (hourly to daily) pertaining to the atmospheric conditions and the 16-day composite does not represent the fine scale variability with the 16-day composite top layer soil moisture data.

6.2.2.2 Lag Correlation Results

The objective of looking at the lag correlation is to find out if moisture coming from the atmosphere during the winter and spring season is stored in the soil layers and being used by the vegetation during the vegetation growing season. For this lag correlation, we used only the root zone soil moisture (top 1 meter), since only the deep layers can hold the moisture memory for a longer period of time. The lag correlation was studied for the lag time of zero to 12 weeks. It is important to mention that the lag time here is the difference between the start of the sixteen-day composite soil moisture and the start of the

vegetation growing season. Figure 6.3 shows the lag correlation plots for all the three vegetation indices with the root zone soil moisture data. Each figure represents a different vegetation cover type. The x-axis in all the figures represents lag time for the vegetation growing season to the soil moisture. The results are quite interesting in this figure. First, the forests and woodlands show positive as well as negative correlations with different lag times whereas the smaller trees such as the shrubland and grassland show positive relationship for all the lag times. Second, the FVC shows higher correlation than that of NDVI for most of the crop types even though the FVC is derived from NDVI. This indicates that the FVC is more sensitive to the soil moisture change than the NDVI. Third, the EVI shows higher correlation than the other two vegetation indices for the bigger trees such as forests and woodlands and lower correlation for the smaller trees such as shrubs and grasses.

It can be noticed that the evergreen forest (Figure 6.3a) shows high negative correlation for 0 to 6 weeks lag time, but positive correlation with 8 to 10 weeks of lag time. In general, the evergreen forest shows lower correlation compared to the other vegetation types. This indicates that the evergreen forests are less susceptible to the soil moisture variability. The deciduous and mixed forests (Figure 6.3b) show positive correlation for all the lag times with highest correlation for the lag time of 4 weeks. As the land cover type changes from woodland to wooded grassland (Figure 6.3c – e), the correlation peak for all the vegetation indices with the soil moisture shifts from 6 to 2 weeks lag time. The shrublands (Figure 6.3f) and the grasslands (Figure 6.3g) show high positive correlations with the soil moisture for all the lag times with highest correlations (Figure 6.3g) show high positive correlations with the soil moisture for all the lag times with highest correlations (Figure 6.3g) show high positive correlations with the soil moisture for all the lag times for all the lag times for all the grasslands (Figure 6.3g) show high positive correlations with the soil moisture for all the lag times with highest correlation for the lag times for all the soil moisture for all the lag times for all the soil moisture shifts form 6 to 2 weeks lag time. The shrublands (Figure 6.3f) and the grasslands (Figure 6.3g) show high positive correlations with the soil moisture for all the lag times with highest correlation







Fig. 6.3: Lag correlations between the three vegetation indices with the root zone soil moisture for different vegetation classes. For description of the vegetation classes, please see Table 6.1.

values for 2 to 4 weeks and 4 to 6 weeks lag time respectively. The cropland (Figure 6.3h) shows highest correlations for 4 to 6 weeks lag time. The croplands generally show

lower correlation as compared to those of other smaller vegetation types such as shrublands and grasses. Unlike the shrublands and grasses, the croplands show negative correlation with more than 6 weeks lag time. This could be due to the different crop types practiced by farmers in different seasons and years since some crop types with deep roots are susceptible to root zone soil moisture change and other crop types with shallow roots are not. So, it is important to know what kind of crop types are there in the crop lands. Depending on the type of crops, the correlation results might differ than what we got in this study.

6.2.3 Conclusions

This soil moisture-vegetation growth relationship study provides many interesting results. It is difficult to find many locations (equivalent of a satellite pixel size) for all vegetation classes within a single climatic regime. So, this study performed here could be affected by different climatic conditions since this study was conducted over many different regions of USA and the vegetation growing season could be different in different climatic regimes. Nevertheless, the growing season considered in this study is common for most of the climatic regimes and the results provided here are robust. There was an opposite relationship exists for all the vegetation types with the top layer and root zone soil moisture. The bigger trees such as forests, woodlands and wooded grasslands showed higher correlation whereas the smaller trees such as shrubs and grasses showed lower correlation with the top layer soil moisture and vice versa for the root zone soil moisture. While checked for lag correlation with the root zone soil moisture, all the vegetation

indices showed positive correlation peaks with the soil moisture for zero or small lag time for shrublands and grasslands. These correlation peaks were gradually shifted towards higher lag time for the forests and woodlands.

This study was conducted with 25 km spatial resolution soil moisture data. So, these results suggest that the microwave satellite soil moisture data at coarse spatial (25 km) resolution can be helpful to perform this kind of study and this study can be extended for the whole globe since the satellite data are available for the whole globe. The correlation results in this study does not necessarily suggest that only the soil moisture variability drives the vegetation greenness during the growing season since there are many other atmospheric, hydrologic, ecologic, geologic factors that could affect the vegetation growth and also could affect the soil moisture – vegetation relationship. So, this study will be extended in future to add more parameters to understand this complex ecological relationship between vegetation and other eco-climatic parameters.

6.3 Study of Extreme Events

6.3.1 Definition and Classification

Extreme events such as floods, droughts, storms, hurricanes, cyclones occur naturally in the physical system. There is no clear definition of an extreme event. For example, Extreme climate events are defined by Easterling et al. (2000) as those climate events that cause extraordinary economic and social (loss of life or livelihood) damage. Any natural hazards that have increased in intensity and frequency are defined as extreme events by Wisner et al. (2004). The definition of an extreme event can be expressed in terms of the physical parameter itself (i. e. the relevant class of the event, the threshold of the extremeness, how rare the event is etc.) and its impact on the society.

The meteorological extremes can broadly be classified into (i) Climatological extreme event: An event that falls in the tails of that event's climatologically expected distribution where the cut-off to decide the tail portion of the distribution is somewhat arbitrary; (ii) Forecast extreme event: A climatological extreme event that falls at or below a given forecast probability level (conditional sense) and (iii) User specific extreme event: The weather event that can lead as an extreme event for any specific group of users even though it is not an extreme event in a traditional sense (http://www.emc.ncep.noaa.gov/gmb/ens/target/ens/albapr/albapr.html). On the other hand, the extreme events can be classified based on spatial and temporal extent and the energy concentration. Such a classification is shown in Figure 6.4.

The physical systems which create the non-extreme events also create the extreme events. Extreme events occur only near the edges when the extreme event frequency distribution is studied in a phase space. Many non-linear processes contribute to determine the position of the edge in the frequency distribution. Hence, it is always a challenge to detect extreme events. In the following sections, the research results from this Ph. D. work have been used to detect one such extreme event (flooding) as discussed above.

6.3.2 Pre-conditions for Flooding

Several factors independently or jointly contribute to the flooding event. Few such conditions are:

- Existing condition of the soil prior to the rainfall: The soil saturation before the rainfall event determines how much more rain water the soil layers can hold. Highly saturated condition before the rainfall or storm event is a favorable condition for flooding.
- Intensity of rainfall: Intensity of rainfall is very important for initiation of flood events. A heavy storm is more likely to create a flood event.
- Spatial extent and duration of rainfall: The heavy rainfall events at a very local scale within a short duration of time are major factors for flooding.

6.3.3 Terminologies Used in this Study

• Recurrence Interval: The recurrence interval (sometimes called the return period) is based on the probability that the given event will be equaled or exceeded in any given year. For example: if there is a 1 in 100 chance (1 % probability) that a flood event with a certain magnitude will happen in any given year, then that flood event is said to have 100 year recurrence interval i.e. this flood can be

expected to occur on average with a frequency of once in every 100 years at that given location.

		Spectrum of characteristics						
Criteria		Pervasive events	Intensive events					
Spatial	Area affected Energy release per unit area	Extensive Low	Localized High					
Temporal	Frequency Onset Duration	Frequent (?) Slow Long	Rare (?) Rapid Short					
Event Type		Drought Tornad Windstorm Hailstorm Avalanche Flood						

Fig. 6.4: Classification of extreme climate events (from Kates, 1979; http://www.icsu-scope.org/downloadpubs/scope27/chapter15.html#tab15.1).

• Severity Class: There are three severity classes defined by the Dartmouth Flood Observatory based on the amount of damage and recurrence interval of a flood event. Class 1: Large flood events: significant damage to structures or agriculture; fatalities; and/or 1-2 decades-long reported interval since the last similar event.

Class 2: Very large events: greater than 20 year but less than 100 year recurrence interval, and/or a local recurrence interval of at 10-20 years.

Class 3: Extreme events: with an estimated recurrence interval greater than 100 years.

• Flood Magnitude: The Dartmouth Flood Observatory (DFO) calculates the flood magnitude based on the following equation -

$$Flood Magnitude = \frac{\ln(duration) \times severity \ class \times \sqrt{affected \ region}}{100}$$
(6.4)

where, duration = duration of flooding in days, severity class = defined above in this section and affected region = area affected by flooding in squared kilometers.

6.3.4 May 6-13, 2003 Flood of Tennessee, Alabama and Georgia – A Case Study

6.3.4.1 Datasets Used

6.3.4.1.1 Soil Moisture Data

The assimilated soil moisture data from the root zone (top 1 meter; model top three soil layers) of the soil layer were used in this study. This dataset was at 25 km spatial

resolution and 12 hour temporal resolution over the USA from 2002 to 2005. But we considered the soil moisture data only before, during and after the flood event.

6.3.4.1.2 Precipitation Data

The North American Land Data Assimilation System (NLDAS) precipitation data were used in this study. The precipitation data were available at 1 hour temporal interval over the whole USA from 1997 onwards. In this case also, we considered the precipitation data only before, during and after the flood event. This same precipitation dataset was used as land surface model input to derive the assimilated soil moisture product used in this application too.

6.3.4.1.3 Dartmouth Flood Observatory Information

Dartmouth Flood Observatory at the Dartmouth College, New Hampshire, USA maintains the global archive of large flood events for more than two decades (http://www.dartmouth.edu/~floods/Archives/index.html). The information includes flooding area, date and duration, damage information, magnitude, severity class and some satellite as well as flood inundation maps.



Fig. 6.5: The May 6-13, 2003 Flood boundary map over Tennessee, Alabama and Georgia shown on the USA map (Source: <u>http://www.dartmouth.edu/~floods/images/2003111sum.htm</u>).

6.3.4.2 Discussion

This flooding event happened due to heavy rain from May 5 to May 7. It was one of the worst floods in 30 years at Chattanooga, Tennessee; in 15 years at Huntsville, Alabama and in 13 years at Columbus, Georgia. It was one of the biggest spring rainfall events in last decade there. The Chattahoochee, Cahaba and Tennessee rivers were the major rivers which caused the flooding. Figure 6.5 shows the flood affected regions on the USA map. 3 people died and around 2000 people got displaced and the damage was worth of 17

million US Dollar. Recurrence interval for this flood was 42 years. Table 6.2 shows the details of this flood with a magnitude of 21.7.

Table 6.2: Information about the May 6-13, 2003 Flood of Tennessee, Alabama and Georgia (Source: http://www.dartmouth.edu/~floods/images/2003111sum.htm)

Case	Location	Duration	Dead	Displaced	Damage (million USD)	Flood Type	Recurrence Interval (years)	Severity Class	Magnitude	Affected Region (sq km)
1	Tennessee, Alabama, Georgia	May 6-13, 2003	3	2000	17	Rain	42	2	21.7	272,100

The research results derived from this Ph. D. research work were used to detect this flooding event. Figure 6.6 shows the spatial maps of daily rainfall data and corresponding root zone soil moisture data from May 4 to May 8. The root zone soil moisture values below 33 % vol/vol were masked in the figure since we were dealing with flood event here. It can be noticed that there was no rainfall over Tennessee, Alabama and Georgia on May 4, 2003 (Figure 6.6a). Hence there was no root zone soil moisture over 33 % vol/vol over that region (Figure 6.6f). Then heavy rainfall up to 180 mm/day happened on May 5 (Figure 6.6b). That made the soil layers very saturated on May 5 with root zone soil moisture values more than 33 % vol/vol (Figure 6.6g) and an

ideal condition for a flooding event. The saturated soil could not hold the continuous rainfall that happened for next few days (till May 7); hence led to a heavy flooding event over Tennessee, Alabama and Georgia from May 6 which can be noticed in the soil moisture maps shown for May 6, May 7 and 8 (Figure 6.6h-j). Even if the rainfall stopped on May 7, the flooding was there till May 13 (maps not shown here). These rainfall maps match very well with the rainfall maps from TRMM satellite provided by the Dartmouth Edu/~floods/images/2003111sum.htm). The rainfall and soil moisture spatial and temporal patterns are compatible with the flood information provided at the above website.

6.3.5 Conclusions

The satellite-model assimilated soil moisture maps created from this Ph.D. research work along with the precipitation maps were used in this chapter to study a flooding event. The soil moisture maps also showed the spatial and temporal extents of the flooding event very well. The results here provided some encouragement to use such kind of soil moisture and precipitation maps to study other flooding and extreme events. These datasets can also be used to forecast extreme events. But to do that, many other local land and atmospheric information are required to set the threshold quantitatively to call an event as extreme.





Fig. 6.6: Daily precipitation maps (mm/day) for May 4-8, 2003 period (left panel) and the corresponding daily soil moisture maps (% vol/vol; right panel). The soil moisture values below 33 % vol/vol are masked in the soil moisture maps.

For example, it is important to know the soil texture, types, slope, topography, land cover types, climatic condition at very fine scale which determines how much water the soil at any geographic location can hold before creating a flooding event. Once you have that kind of local information and you know the condition of soil at that specific time and forecast information about any weather event such as storms, heavy precipitation etc., then you can predict the extreme events (flooding) and warn the local people about the extreme events ahead of time.

Chapter 7. Summary, Conclusions and Research Extensions

The knowledge about the spatial and temporal distribution of the global surface and root zone soil moisture is a key to many real world applications. There have been several efforts made to estimate global soil moisture in last three decades from remote sensing, land surface models and in-situ field experiments. In this Ph.D. research, a data assimilation technique was used to merge the remote sensing soil moisture observations into a land surface model predicted soil moisture to create a synthetic soil moisture product. The research work conducted in this Ph.D. study is summarized here.

7.1 Summary and Conclusions

7.1.1 Characteristics of the In-Situ Soil Moisture Observations

The very first question always comes to our mind is 'what we know about the distribution of soil moisture at the research location. It is very helpful if we have any prior knowledge about the characteristics of soil moisture distribution at the research location and if we know whether the in-situ soil moisture data can be comparable with soil moisture products from other sources before using the in-situ data as ground truth. Chapter 2 analyzed the spatial and temporal distribution of soil moisture in the Little River Experimental Watershed. Many interesting results were found from the in-situ observations. We found that the heavy persistent precipitation happened in spring of 2003 which produced higher soil moisture during those months. The summer precipitation events were more discrete which was also mentioned in the previous literatures. The discrete precipitations created frequent soil moisture wetting periods whereas the porous sandy loamy soil and higher evapo-transpiration caused frequent drying events during the summer season. The watershed produced very low surface soil moisture (less than 22 % vol/vol) due to well drained sandy loam soil. The autocorrelation study indicated that the soil moisture memory could range from 4 to 22 days from the driest to wettest sites respectively. But this temporal range is good enough to validate any remote sensing soil moisture product with less frequent observations. Similarly, the spatial correlation between two sites was found to be reasonably high even the sites in the watershed are 30 km apart. That is a very encouraging result to validate satellite observations of a typical resolution of 25×25 km.

These in-situ soil moisture observations were used as ground truth in the entire research work here. That helped us to validate and understand whether the other remote sensing and model soil moisture products were reasonable or not over the research location.

7.1.2 Remote Sensing Soil Moisture Observations

Quality control of the observation data is required before using the data for data assimilation study. Chapter 3 analyzed and validated the current operational AMSR-E satellite soil moisture product. It was found that the current operational product had very
low skill and it did not match very well with the in-situ observations. Hence, a new soil moisture product was generated from the observed AMSR-E brightness temperature using a single frequency radiative transfer model (LSMEM). This new soil moisture product had somewhat better skills and had better correlation with the in-situ observations than that of the current satellite product. That encouraged us to use this new soil moisture product in the data assimilation study in later chapters. In the end of Chapter 3, we also discussed for a possible global operational soil moisture product using this new forward model approach.

7.1.3 Soil Moisture Simulations from Land Surface Models

Similar to the quality control of observations, it is important to choose a land surface model which produces reasonable soil moisture before using that land surface model for data assimilation study. Chapter 4 discussed model results and model validation study. For that purpose, we used three land surface models (HySSiB, Noah and CLM) with different model physics and parameterization scheme incorporated in the same computing platform (LIS at NASA GSFC). All the three land surface models produced soil moisture data within a reasonable range of the in-situ observation data and showed higher spatial and temporal coherence with the model prescribed and observed precipitation data. The HySSiB and Noah models showed higher absolute soil moisture values whereas the CLM showed higher sensitivity to the atmospheric conditions. The Noah soil moisture showed a consistent positive bias though this model results were found to have higher correlation with the field observations than those of other two models. Hence this model was chosen

for the data assimilation study. We also looked at other water and energy cycle parameters simulated by all the three models and the results showed significant disparities among models. This is expected since the variability of all these parameters were mostly tied to the soil moisture changes for the individual model. We also verified the temporal scaling issue towards the end of the chapter and compared daily with hourly model soil moisture data for validation. The hourly soil moisture data showed larger scatters, higher bias, root mean squared error and lower correlation with the in-situ observations as compared to those of the daily soil moisture data. This also highlighted the sensitivity of the different model parameterization schemes to the meteorological forcing (precipitation) and their role in the model soil moisture results. We created a multi-model mean soil moisture data just by taking the arithmetic mean of the soil moisture data from all the three models. This arithmetic mean outperformed all the three individual soil moisture datasets in the validation study.

7.1.4 Soil Moisture Data Assimilation Results

Soil moisture data assimilation is the main objective of this Ph.D. research work. Keeping this objective in mind, the validation research was conducted in Chapter 3 and 4 for the remote sensing and model soil moisture data respectively. Chapter 5 dealt with the soil moisture data assimilation study. Ensemble Kalman Filtering (EnKF) data assimilation algorithm was used for this study. The normalized innovation was found to have high negative bias and variance more than one. This suggested a large difference in the remote sensing and model predicted soil moisture data. The soil moisture results from the data

assimilation and only model run (open loop) were compared with the in-situ observations. There was an absolute improvement of 51 to 78 %vol/vol in the data assimilations run results over that of the open loop run throughout the watershed. The assimilated soil moisture data had comparable correlation, lower bias and root mean squared error as compared to those of the only model run results. Since the LSMEM retrieved soil moisture observation controlled the data assimilated results, the root mean squared error was continuously got reduced for the assimilated results compared with the LSMEM observations with each model time step. We found changes in other water and energy cycle parameter values due to change in soil moisture values after data assimilation. In the end, couple of algorithm sensitivity studies was performed for the model initialization condition and model spin-up. It was found that the assimilation algorithm is sensitive in both the cases.

7.1.5 Application of Research Results

The assimilated soil moisture was used to perform two real world applications. First study was the relationship of soil moisture with the vegetation during the growing season. We used three different vegetation indices and we found different correlation values for different vegetation indices with the assimilated soil moisture for different vegetation types. We also studied the lag correlation of the vegetation indices with the root zone soil moisture and found high positive correlation with the lag time. This suggests that the root zone soil layer holds the soil moisture memory from winter and spring season till the vegetation growing season.

The other application was to study soil moisture change during an extreme event, particularly flooding. We found high coherence for the precipitation and assimilated soil moisture maps with the temporal and spatial extent of a flooding event over Tennessee, Alabama and Georgia during May 6-13, 2003. In this study, we used soil moisture for flood monitoring. But the soil moisture along with good precipitation forecast and condition of the soil prior to heavy precipitation can be used to detect extreme events.

7.2 Limitations

There were many limitations we encountered during this research work. They are discussed here below.

7.2.1 Limited Availability of the In-Situ Observations

The Little River Experimental Watershed is one of the four selected watersheds for AMSR-E satellite soil moisture validation study. So, comprehensive in-situ soil moisture measurements have been conducted at this watershed since 2001. But the data are yet to release to the public. We got just one year of data (2003) from USDA-ARS, Beltsville through personal request. But we think that one year data are not enough to study the soil moisture characteristics of the watershed we performed in Chapter 2. At the same time, it is better to have the in-situ observations for longer period of time to validate the satellite and model soil moisture data.

7.2.2 Non-Compatibility of Spatial and Temporal Scales

There were differences in spatial and temporal scales for different datasets. For example, the in-situ observations were point data from 5 cm soil depth whereas the satellite products were at 25 km spatial resolution and from less than 1 cm soil layer. Similarly different land surface model data were at 1 km spatial resolution but from variable soil depths. So, we used many assumptions for all the validation studies conducted in Chapter 3 and 4. We think the disparity in spatial and temporal scales among different datasets might have impacted the validation results, but it is difficult to say how much the impact is. So, this is a limitation in our validation study.

7.2.3 Choice of Land Surface Model

We considered three land surface models in the model validation and comparison study in Chapter 4 and chose Noah model based on its performance in the comparison study to perform the data assimilation work. But it is hard to tell which model is the best for the data assimilation study. Though the Noah model produced better results, the model parameterization (top soil layer is 10 cm) was not compatible with the remote sensing observations (from top \sim 1 cm) as opposed to other two models (top soil layer is 2 cm). It is difficult to say whether a model with better results and incompatible soil layer structure or a model with comparable results and compatible soil layer structure can contribute more to the data assimilation results. This requires more model experiments with different models before answering the above question. This is a limitation in this research work.

7.2.4 Bias Correction to Satellite Observations

We found a large bias between the temporal moments of the LSMEM derived soil moisture and the Noah model soil moisture results. Such biases between the observation and model results are commonly addressed through a bias correction before performing data assimilation study. One such bias correction is done by scaling the satellite observations to the model's climatology so that the cumulative distribution function (cdf) of the satellite soil moisture and the model soil moisture match (Reichle and Koster, 2004). There is no such bias correction module available in LIS yet. So, we did not perform any bias correction before doing data assimilation in our research. The data assimilation results might be different with bias correction than those without bias correction. This is a limitation in our research.

7.3 Research Extensions

7.3.1 Requirement of more Data Validation Study

The remote sensing and model soil moisture validation study was conducted over a single watershed in this research work. We require more rigorous validation work over different watersheds with different settings of land cover types, soil types, topography and climatic conditions. Since all these components contribute separately or together to the amount and distribution of soil moisture, we will have better understanding on how remote sensing and model results are affected by different ecological conditions. The idea is to

have a robust soil moisture dataset which is good over the whole globe and different ecological settings and seasons.

7.3.2 Fraternal Twin experiment

A common way to test the performance of an assimilation scheme is to perform identical twin experiments. In Identical twin experiment, the same model is used for truth and open loop runs and creating synthetic observations for data assimilation purpose (Kumar et al., 2008). Unfortunately, such experiments can not determine whether the assimilation can remove the systematic error in the model, as all the model realizations (from same model) possess the same error (Killworth et al., 2001). However in fraternal twin experiment, two different models are used under nearly identical conditions for truth run and synthetic observations in one side and open loop run in the other side (Kumar et al., 2008). This provides a unique opportunity to remove the systematic error associated with any single model. We will perform such fraternal twin experiments using two different land surface models as a research extension of this data assimilation work.

7.3.3 3-D Data Assimilation Study

We used a true 1-D data assimilation filter in this research work where the update due to the data assimilation was only in vertical direction (through different layers). But during the data assimilation study, each grid was treated as a single independent column and there was no interaction between different grid cells. The lateral updates of soil layer prognostic variables are also important along with the vertical updates. So, we will use a 3-D data assimilation filter in future to update both horizontal as well as vertical spaces during the data assimilation. 3-D assimilation filter has clear advantages over the 1-D assimilation filter as discussed in Section 5.5 thought it is computationally very expensive.

7.3.4 Multi-model Ensemble Data Assimilation Study

This is another approach where the systematic error associated with any single model can be removed. In this case, more than one model is perturbed simultaneously to create ensembles from multi-models and then all ensemble members are updated individually by the observations at the update stage irrespective of the model. Since, the LIS at NASA GSFC includes multiple models in the same platform; this gives us a great opportunity to perform multi-model ensemble study. We will do this interesting study in future and verify if we can improve the single model data assimilation results.

7.3.5 Coupled Data Assimilation Study

Coupled data assimilation is another direction where we want to work in future. All the data assimilation studies performed in this research work used the off-line model simulations. In that case, we missed the land-atmosphere interaction component. It is important to understand how the change in soil moisture values due to data assimilation study impacts other water and energy cycle parameters; how the atmospheric components react to the change in those outgoing water and energy variables to the atmosphere and in turn how the atmospheric components contribute to the soil moisture results in a coupled

system. In Chapter 5, we discussed briefly about the impact of change in soil moisture values due to data assimilation on other water and energy cycle variables. But we want to run the data assimilation in a coupled mode to understand the whole land-atmosphere feedback system. The soil moisture results will be quite different in this case and may be more realistic than that of the off-line data assimilation system.

7.3.6 Operational Global Assimilated Soil Moisture Product

We will move one step forward from research domain to operational domain. The research results need to be in operational mode to serve the society in a real world. So, we will work on how to expand the small study area considered in this research work to the whole globe. Similarly, the operational results need to be delivered in a timely manner for proper benefit to the society. Before performing this task, we need to do all or some of the research extensions mentioned here in this section (especially Section 7.3.1) to make sure that the data assimilation results are reliable and can serve the society better. Then we will generate a real time operational global soil moisture data assimilation product. We will also work on regional to local operational product depending on the requirements and objectives (e.g. tracking a hurricane, mapping a drought condition in a region, etc.).

LIST OF REFERENCES

List of References

Adegoke, J. O. & Carleton, A. M. (2002). Relations between Soil Moisture and Satellite Vegetation Indices in the U.S. Corn Belt, *Journal of Hydrometeorology*, *3*, 395–405.

Baek, S., Hunt, B. R., Kalnay, E., Ott, E. & Szunyogh, I. (2006). Local ensemble Kalman filtering in the presence of model bias, *Tellus*, *58A*, 293-306.

Bell, K. R., Blanchard, B. J., Schmugge, T. J. & Witczak, M. W. (1980). Analysis of surface moisture variations within large field sites, *Water Resources Research*, *16*, 796–810.

Betts, A. K., Chen, F., Mitchell, K. E. & Janjic, Z. I. (1997). Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta Model using FIFE data, *Monthly Weather Review*, *125*, 2896–2916.

Bindlish, R., Jackson, T. J., Wood, E. F., Gao, H., Starks, P., Bosch, D. & Lakshmi, V. (2003). Soil moisture estimates from TRMM Microwave Imager observations over the Southern United States, *Remote Sensing of Environment*, *85 (4)*, 507-515.

Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., Dickinson, R.E. & Yang, Z.-L. (2002). The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model, *Journal of Climate*, *15*, 3123-3149.

Bonan, G.B. (1996). A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user's guide, NCAR Technical Note NCAR/TN-417+STR, National Center for Atmospheric Research, Boulder, CO.

Boone, A., Habets, F., Noilhan, J., Clark, D., Dirmeyer, P., Fox, S., Gusev, Y., Haddeland, I., Koster, R., Lohmann, D., Mahanama, S., Mitchell, K., Nasonova, O., Niu, G.-Y., Pitman, A., Polcher, J., Shmakin, A. B., Tanaka, K., van den Hurk, B., Verant, S., Verseghy, D., Viterbo P. & Yang Z.-L. (2004). The Rhone-Aggregation Land Surface Scheme Intercomparison Project: An Overview, *Journal of Climate*, *17*, 187-208.

Bosch, D. D., Lakshmi, V., Jackson, T. J., Choi, M. & Jacobs, J. M. (2006). Large scale measurements of soil moisture for validation of remotely sensed data: Georgia soil moisture experiment of 2003, *Journal of Hydrology*, *323*, 120–137.

Bosch, D. D., Sheridan, J. M. & Davis, F. M. (1999). Rainfall characteristics and spatial correlation for the Georgia Coastal Plain, *Transactions of American Society of Agricultural Engineers*, 42(6), 1637-1644.

Burgers, G., van Leeuwen, P. J. & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter, *Monthly Weather Review*, *126*, 1719–1724.

Calvet, J. C. and CoAuthors (1999). MUREX: A land-surface field experiment to study the annual cycle of the energy and water budgets, *Annales Geophysicae*, *17*, 838-854.

Cashion, J., Lakshmi, V., Bosch, D. & Jackson, T. J. (2005). Microwave remote sensing of soil moisture: Evaluation of the TRMM microwave imager (TMI) satellite for the Little River Watershed Tifton, Georgia, *Journal of Hydrology*, *307 (1-4)*, 242–253.

Chang, J. T., & Wetzel, P. J. (1991). Effects of spatial variations of soil moisture and vegetation on the evolution of a prestorm environment: A numerical case study, *Monthly Weather Review*, *119*, 1368–1390.

Charney, J. G., Halem, M. & Jastrow, R. (1969). Use of incomplete historical data to infer the present state of the atmosphere, *Journal of the Atmospheric Sciences, 26*, 1160-1163.

Chen, F. & Dudhia J. (2001). Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 Modeling system. Part II: Preliminary model validation, *Monthly Weather Review*, *129*, 587-604.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q., Ek M. & Betts, A. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations, *Journal Geophysical Research*, vol. 101(D3), pp. 7251-7266.

Choudhury, B. J., Schmugge, T. J., Chang, A. & Newton, R. W. (1979). Effect of surface roughness on the microwave emission from soils, *Journal of Geophysical Research*, *84*, 5699–5706.

Clapp, R. B., & Hornberger, G. M. (1978). Empirical equations for some soil hydraulic properties", *Water Resources Research*, *14*, 601-604.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, O., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D. & Meng J. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *Journal of Geophysical Research*, *108(D22)*, 8842, doi:10.1029/2002JD003118.

Critchfield, H. J. (1974). General Climatology, *Third Edition, PrenticeHall Inc., Englewood Cliffs, NJ*, 446 pp.

Crow, W. T. & Van Loon, E. (2006). Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed soil moisture: *Journal of Hydrometeorology*, 7, 421–432.

Dai, Y. & Zeng, Q. (1997). A land surface model (IAP94) for climate studies. Part I: formulation and validation in off-line experiments, *Advances in Atmospheric Sciences*, *14*, 433-460.

Dai, Y., Zeng X., Dickinson R. E, Baker I., Bonan G. B., Bosilovich M. G., Denning A. S., Dirmeyer P. A., Houser P. R., Keith G. N., Oleson W., Schlosser C.A. & Yang, Z-L (2003). The Common Land Model, *Bulletin of American Meteorological Society*, *84*, 1013-1023.

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N. & Verhoest N. E. C. (2006). Assessment of model uncertainty for soil moisture through ensemble verification, *Journal of Geophysical Research*, *111*, D10101, doi:10.1029/2005JD006367.

De Lannoy, G.J.M., Verhoest, N.E.C., Houser, P.R., Gish, T.J. & Meirvenne, M.V. (2006). Spatial and temporal characteristics of soil moisture in an intensively monitored agricultural field (OPE³), *Journal of Hydrology*, *331*, 719-730.

Deardorff, J. W. (1978). Efficient prediction of a ground surface temperature and moisture with inclusion of a layer of vegetation, *Journal of Geophysical Research*, 83, 1889-1903.

Delworth, T. L. & Manabe, S. (1988). The influence of potential evaporation on the variabilities of simulated soil wetness and climate, *Journal of Climate*, *1*, 523-547.

Dickinson, R. E., Henderson-Sellers, A., Kennedy, P. J. & Wilson, M. F. (1986). Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model, *NCAR Tech. Note, TN-275+STR, Boulder, CO*, 69 pp.

Dirmeyer, P. A. (2004). Soil moisture - muddy prospects for a clear definition, *GEWEX* News, 14 (3), 11-12.

Dirmeyer, P. A. (1995). Problems in initializing soil wetness, *Bulletin of American Meteorological Society*, *76*, 2234-2240.

Dirmeyer, P. A., Dolman, A. J. & Sato, N. (1999). The Global Soil Wetness Project: A pilot project for global land surface modeling and validation, *Bulletin of American Meteorological Society*, 80, 851-878.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki T. & Hanasaki, N. (2006). The Second Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the land surface, *Bulletin of American Meteorological Society*, *87*, 1381-1397.

Dobson, M. C., Ulaby, F. T., Hallikainen, M. T. & El-Rayes, M. A. (1985). Microwave dielectric behavior of wet soil - part II: dielectric mixing models, *IEEE Transactions on Geosciences and Remote Sensing*, *GE-23*, 35-46.

Drusch, M., Wood, E. F. & Jackson, T. J. (2001). Vegetative and atmospheric corrections for soil moisture retrieval from passive microwave remote sensing data: Results from the Southern Great Plains Hydrology Experiment 1997, *Journal of Hydrometeorology, 2*, 181–192.

Drusch, M., Wood, E. F., Gao, H. & Thiele, A. (2004). Soil moisture retrieval during the Southern Great Plains Hydrology Experiment 1999: A comparison between experimental remote sensing data and operational products, *Water Resources Research*, 40, W0250410, doi:10.1029/2003WR002441.

Easterling, D.R., Evans, J.L., Groisman, P. Y., Karl, T.R., Kunkel, K.E. & Ambenje, P. (2000). Observed Variability and Trends in Extreme Climate Events: A Brief Review, *Bulletin of the American Meteorological Society*, *81(3)*, 417-425

Ek, M., Mitchell, K., Yin, L., Rogers, P., Grunmann, P., Koren, V., Gayno, G. & Tarpley, J. (2003). Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model, *Journal of Geophysical Research*, *108*, doi:10.1029/2002JD003296.

Entin, J. K., Robock, A., Vinnikov, K. Y., Hollinger, S. E., Liu, S. & Namkhai A. (2000). Temporal and spatial scales of observed soil moisture variations in the extratropics, *Journal of Geophysical Research*, *105*, 11,865–11,877.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, 99(C5), 10,143–10,162.

Eymard, L., Bernard, R. & Lojou, J. Y. (1993). Validation of microwave radiometer geophysical parameters using meteorological model analysis, *International Journal of Remote Sensing*, 14 (10), 1945-1963.

Famiglietti, J. S., Rudnicki, J. W. & Rodell, M. (1998). Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas, *Journal of Hydrology, 210,* 259-281.

Famiglietti, J., Devereaux, J., Laymon, C., Tsegaye, T., Houser, P., Jackson, T., Graham, S., Rodell, M. & Van Oevelen, P. (1999). Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern great plains 1997 (SGP97) Hydrology Experiment, *Water Resources Research*, *35* (*6*), 1839–1851.

Gao, B. C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sensing of Environment*, 58, 257-266.

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

Gao, H., Wood, E. F., Jackson, T. J., Drusch, M. & Bindlish, R. (2006). Using TRMM/TMI to retrieve surface soil moisture over the southern United States from 1998 to 2002, *Journal of Hydrometeorology*, *7 (1)*, 23–38.

Gao, X. & Dirmeyer, P. A. (2006). Multi-model analysis, validation and transferability study for global soil wetness products, *Journal of Hydrometeorology*, *7*, 1218-1236.

Gao, X., Dirmeyer, P. A., Guo, Z., Zhao, M., Decharme, B., Niu, G. & Mahanama, S. (2004b). An approach for remote sensing validation of land surface schemes on a global scale, *COLA Tech. Report*, *173*, 42 pp. [available at: ftp://grads.iges.org/pub/ctr/ctr 173.pdf].

Georgakakos, K. & Baumer, O. (1996). Measurement and utilization of on-site soil moisture data, *Journal of Hydrology*, 184, 131–152.

Gillies, R. R. & Carlson, T. N. (1995). Thermal remote sensing of the surface soil water content with partial vegetation cover for incorporation into climate models, *Journal of Applied Meteorology*, *34*, 745–756.

Grayson, R.B., Bloschl, G., Western, A.W. & McMahon, T.A. (2002). Advances in the use of observed spatial patterns of catchment hydrological response, *Advances in Water Resources*, *25*, 1313–1334.

Guo, Z. & P. A., Dirmeyer (2006). Evaluation of GSWP-2 soil moisture simulations, Part I: Inter-model comparison, *Journal of Geophysical Research*, *111*, D22S02, doi:10.1029/2006JD007233.

Hansen, M. C., Defries, R. S., Townshend, J. R. G. & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach", *International Journal of Remote Sensing*, *21*, 1331–1364.

Henderson-Sellers, A., Pitman, A. J., Love, P. K., Irannejad, P. & Chen, T. H. (1995). The project for Inter-comparison of land surface parameterization schemes (PILPS): Phases 2 and 3, *Bulletin of American Meteorological Society*, *76 (4)*, 489-503.

Henninger, D. L., Peterson, G. W. & Engman, E. T. (1976). Surface soil moisture within a watershed: Variations, factors influencing, and relationships to surface runoff, *Soil Science Society of American Journal*, 40, 773–776.

Higgins, R. W., Shi, W. & Yarosh, E. (2000). Improved United States precipitation quality control system and analysis, *Atlas* 7, National. Center For Environmental Prediction/Climate Prediction Center, Camp Springs, MD, 40 pp.

Hills, T. C. & Reynolds, S. G. (1969). Illustrations of soil moisture variability in selected areas and plots of different sizes, *Journal of Hydrology*, *8*, 27–47.

Houghton, J.T. (editor), Meiro Filho, L. G. (editor), Callander, B. A. (editor), Harris, N. (editor), Kattenburg, A. (editor) & Maskell, K. (editor) (1996). Climate Change 1995: The Science of Climate Change: Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change, *Cambridge University Press, Cambridge*.

Houser, P. R., Shuttleworth, W. J., Famglietti, J. S., Gupta, H. V., Syed, K. H. & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resources Research*, *34*, 3405–3420.

Hubbard, R.K., Berdanier, C.R., Perkins, H.F. & Leonard, R.A. (1985). Characteristics of selected upland soils of the Georgia coastal plain, *USDA Agricultural Research Service*, *72*, ARS-37.

Huete, A. R. & Justice, C. (1999). MODIS vegetation index (MOD13) algorithm theoretical basis document, *Version. 3*.

Huete, A. R., Didan, K., Miura, T., Rodriguez, E. P., Gao X. & Ferreira, L.G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sensing of Environment*, **83**, 195–213.

Isaaks, E.H. and Srivastava, R.M. (1989). An Introduction to Applied Geostatistics, *Oxford University Press, New York*, 561 pp.

Jackson, T. J. & Schmugge, T. J. (1991). Vegetation effects on the microwave emission from soils, *Remote Sensing of Environment*, *36*, 203–219.

Jackson, T. J. (1997). Southern Great Plains 1997 (SGP97) hydrology experiment plan, 178 pp. [Available at: <u>http://hydrolab.arsusda.gov/sgp97/explan/]</u>.

Jackson, T. J., Cosh, M. H., Bindlish, R., Goodrich, D., Starks, P., Bosch, D. D. & Seyfried, M. S. (2007). Calibration and Validation of the AMSR-E Soil Moisture Product using In Situ Soil Moisture Networks, *Water Resources Research* (submitted).

Jackson, T. J., Le Vine, D. M., Schiebe, F. R. & Schmugge, T. J. (1993). Large area mapping of soil moisture using passive microwave radiometry in the Washita'92 experiment, *International Geoscience and Remote Sensing Symposium'93, IEEE Publ 93CH3394-6*,1009-1011.

Jackson, T.J., Cosh, M. H., Zhan, X., Bosch, D.D., Seyfried, M.S., Starks, P.J., Keefer, T.O. & Lakshmi, V. (2006). Validation of AMSR-E soil moisture products using watershed networks, *Proceedings of the International Geoscience and Remote Sensing Symposium*, July 31-August 4, 2006, Denver, Colorado, 432-435.

Jarvis, P. G. (1976). The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field, *Philosophical Transactions of the Royal Society B*, 273, 593-610.

Kates, R. W. (1979). The Australian experience: Summary and prospect. In Heathcote, R. L., and Thom, B. G. (Eds.) *Natural Hazards in Australia*, Australian Academy of Science, Canberra, 511-520.

Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H. & Spencer, R. W. (2003). The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's contribution to the EOS for Global Energy and Water Cycle Studies, *IEEE Transactions on Geosciences and Remote Sensing*, *41* (2), 184-194.

Kerr Y. H. & Njoku, E. G. (1990). A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space, *IEEE Transactions on Geosciences and Remote Sensing*, 28, 384–393.

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J.-M., Font, J. & Berger M. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission, *IEEE Transactions on Geosciences and Remote Sensing*, *39*, 1729–1735.

Kidwell, K. B. (Ed.) (1990). Global Vegetation Index User's Guide, U.S. Department of Commerce/National Oceanic and Atmospheric Administration/National Environmental Satellite Data and Information Service/National Climatic Data Center/Satellite Data Services Division, Washington D.C., 45 pp.

Kiehl, J. T. & Trentberth, K. E. (1997). Earth's Annual Global Mean Energy Budget, Bulletin of the American Meteorological Society, 78 (2), 197-208.

Killworth, P. D., Dieterich, C., Provodt, C. Le., Oschlies, A. & Willebrand, J. (2001). Assimilation of altimetric data and mean sea surface height into an eddy-permitting model of the North Atlantic, *Progress In Oceanography*, *48 (2-3)*, 313-335.

Koster, R. D., Suarez, M., Liu, P., Jambor, U., Berg, A., Kistler, M., Reichle, R., Rodell, M. & Famiglietti, J. (2004). Realistic initialization of land surface states: impacts on subseasonal forecast skill, *Journal of Hydrometeorology*, *5*(*6*), 1049–1063.

Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Geiger, J., Houser, P. R., Olden, S., Lighty, L., Eastman, J. L., Dirmeyer, P., Doty, B., Adams, J., Wood, E. F. & Sheffield, J. (2006). LIS — An interoperable framework for high resolution land surface modeling, *Environmental Modeling and Software*, *21*, 1402-1415.

Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., Eylander, J. B. & Houser, P. R. (2008). A land surface data assimilation framework using the land Information System: Description and Application, *Advances in Water Resources*, (in press).

Lawford, R. G. (1992). An overview of soil moisture and its role in the climate system, *Soil Moisture Modeling and Monitoring for Regional Planning, Proceedings of the NHRC workshop held in March 9-10, 1992.*

Lee, K. H. & Anagnostou, E. N. (2004). A combined passive/active microwave remote sensing approach for surface variable retrieval using Tropical Rainfall Measuring Mission observations, *Remote Sensing of Environment*, *92 (1)*, 112–125.

Li, L., Njoku, N., Im, E., Chang, P. & St.Germain, K. (2004). A preliminary Survey of Radio-Frequency Interference over the U.S. in Aqua AMSR-E Data, *IEEE Transactions on Geosciences and Remote Sensing*, *42*, 380-390.

Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. (1994). A simple hydrologically based model of land-surface water and energy fluxes for general-circulation models, *Journal of Geophysical Research*, *99*, 14 415–14 428.

Liang, X., Wood, E. F. & Lettenmaier, D. P. (1999). Modeling ground heat flux in land surface parameterization schemes, *Journal of Geophysical Research*, *104*, 9581-9600.

Liston, G. E., Sud, Y. C. & Wood, E. F., 1994. Evaluating GCM land surface hydrology parameterizations by computing river discharge using a runoff routing model: application to the Mississippi basin, *Journal of Applied Meteorology*, *33*, 394-405.

Liu, H. Q. & Huete, A. R. (1995). A feedback based modification of the NDV I to minimize canopy background and atmospheric noise, *IEEE Transactions on Geosciences and Remote Sensing*, *33*, 457-465.

Lockwood, J. G. (1974). World Climatology: An environmental approach, St. Martin's Press, New York, 330 pp.

Los, S. O., Pollack, N. H., Parris, M. T., Collatz G. J., Tucker, C. J., Sellers, P. J., Malmström, C. M., DeFries, R. S., Bounoua, L. & Dazlich D. A. (2000). A Global 9yr Biophysical Land Surface Dataset from NOAA AVHRR Data, *Journal of Hydrometeorology*, *1*, 183–199.

Luo, L., Robock, A., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Lohmann, D., Cosgrove, B. A., Wen, F., Sheffield, J., Duan, Q., Higgins, R. W., Pinker, R. T. & Tarpley J. D. (2003). Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the southern Great Plains, *Journal of Geophysical Research*, *108(D22)*, 8843, doi:10.1029/2002JD003246.

Lydolph, P. E. (1985). The Climate of the Earth, *Rowman & Allanheld Publishers, New Jersey*, 386 pp.

Mahrt, L. & Pan, H.-L. (1984). A two-layer model of soil hydrology, *Boundary-Layer Meteorology*, 29, 1-20.

Margulis, S. A., McLaughlin, D., Entekhabi, D. & Dunne, S. (2002). Land data assimilation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment, *Water Resources Research*, *38*, 1299, doi:10.1029/2001WR001114.

Martinez-Fernandez J. & Ceballos A. (2003). Temporal stability of soil moisture in a large-field experiment in Spain, *Soil Science Society of America Journal*, 67, 1647-1656.

Mather, J. R. (1974). Climatology: Fundamentals and Applications, *McGraw-Hill Book Company, New York*, 412 pp.

Matsushita, B., Yang, W., Chen, J., Onda, Y. & Qiu, G. (2007). Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-Density Cypress Forest, *Sensors*, *7*, 2636–2651.

McCabe, M. F., Gao, H. & Wood, E. F. (2005a). An evaluation of AMSR-E derived soil moisture retrievals using ground based airborne and ancillary data during SMEX 02, *Journal of Hydrometeorology*, *6*(*6*), 864-877.

McCabe, M. F., Wood, E. F. & Gao, H. (2005b). Initial soil moisture retrievals from AMSR-E: Multi-scale comparison using in situ data and rainfall patterns over Iowa, *Geophysical Research Letter, 32*, L06403, doi:10.1029/2004GL021222.

Miller, D. A. & White, R. A. (1998). A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling, *Earth Interactions*, *2*, 1-26.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H. & Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *Journal of Geophysical Research*, 109, D07S90, doi:10.1029/2003JD003823.

Mo, T., Choudhury, B. J., Schmugge, T. J., Wang, J. R. & Jackson, T. J. (1982). A model for microwave emission from vegetation covered fields, *Journal of Geophysical Research*, *87*, 11 229–11 237.

Mocko, D. M., Walker, G. K. & Sud, Y. C. (1999). New snow-physics to complement SSiB. Part II: Effects on soil moisture initialization and simulated surface fluxes, precipitation, and hydrology of GEOS II GCM, *Journal of the Meteorological Society of Japan*, 77 (1B), 349-366.

Mocko, D.M. & Sud, Y.C. (2001). Refinements to SSiB with an emphasis on snow physics: Evaluation and validation using GSWP and Valdai data, *Earth Interactions*, *5*, 1-31.

Mohanty, B.P., Famiglietti, J.S. & Skaggs, T.H. (2000). Evolution of soil moisture spatial structure in a mixed vegetation pixel during the Southern Great Plains 1997 (SGP97) hydrology experiment, *Water Resources Research*, *36* (*12*), 3675–3686.

Moradkhani, H., Sorooshian, S., Gupta, H. V. & Houser, P. R. (2005). Dual stateparameter estimation of hydrological models using ensemble Kalman filter, *Advances in Water Resources*, 28, 135-147.

Ni-Meister, W., Walker, J. P. & Houser P. R. (2005). Soil moisture initialization for climate prediction: Characterization of model and observation errors, *Journal Geophysical Research*, *110*, D13111, doi:10.1029/2004JD005745.

Njoku, E. (1999). AMSR Land Surface Parameters: Algorithm Theoretical Basis Document (Version 3.0), *NASA Jet Propulsion Laboratory, Pasadena, CA* [Available at: <u>http://eospso.gsfc.nasa.gov/eos_homepage/for_scientists/atbd/docs/AMSR/atbd-amsr-land.pdf</u>]

Njoku, E. (2004). AMSR-E/Aqua daily L3 surface soil moisture, interpretive parms, & QC EASE-Grids, March to June 2004. *Boulder, CO, USA: National Snow and Ice Data Center. Digital media.*

Njoku, E. G. & Kong, J. A. (1977). Theory for passive microwave remote sensing of near-surface soil moisture, *Journal of Geophysical Research*, 82, 3108-3118.

Njoku, E. G. & Li, L. (1999). Retrieval of land surface parameters using passive microwave measurements at 6–18 GHz, *IEEE Transactions on Geosciences and Remote Sensing*, *37*, 79–93.

Njoku, E. G., Ashcroft, P. & Li, L. (2005). Statistics and global survey of radiofrequency interference in AMSR-E land observations, *IEEE Transactions on Geosciences and Remote Sensing*, *43*, 938–947.

Njoku, E.G. & Chan, S. K. (2006). Vegetation and Surface Roughness Effects on AMSR-E Land Observations. *Remote Sensing of Environment, 100 (2)*, 190-199.

Njoku, E.G., Jackson, T. J., Lakshmi, V., Chan, T. K. & Nghiem, S. V. (2003). Soil moisture retrieval from AMSR_E, *IEEE Transactions on Geosciences and Remote Sensing*, 41 (2), 215-229.

Oleson, K. W., Dai, Y., Bonan, G., Bosilovich, M., Dickinson, R., Dirmeyer, P. A., Hoffman, F., Houser, P., Levis, S., Niu, G-Y., Thornton, P., Vertenstein, M., Yang, Z-L. & Zeng, X. (2004). Technical Description of the Community Land Model (CLM), *NCAR Tech. Note, TN-461+STR, Boulder, CO*, 174 pp.

Oliveira, R., Oyama, M. & Nobre, C. (2006). Incorporating hydraulic redistribution (HR) into the Simplified Simple Biosphere model (SSiB), *Proceedings of 8 ICXHMO*, *Foz do Iguacu, Brazil, April 24-28*, 935-938.

Owe, M., Jones, E. B. & Schmugge, T. J. (1982). Soil moisture variation patterns observed in Hand County, South Dakota, *Water Resources Bulletin*, *18*, 949–954.

Paloscia, S., Macelloni, G. & Santi, E. (2006). Soil moisture estimates from AMSR-E brightness temperatures by using a dual-frequency algorithm, *IEEE Transactions on Geosciences and Remote Sensing*, 44 (11), 3135–3144.

Pampaloni, P., & Paloscia, S. (1986). Microwave emission and plant water content: A comparison between field measurements and theory, *IEEE Transactions on Geosciences and Remote Sensing*, *24*, 900–905.

Pan, M. & Wood E. F. (2006). Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter, *Journal of Hydrometeorology*, *7*, 534–547.

Parkinson, C. L. (2003). Aqua: An Earth-Observing Satellite Mission to examine water and other climate variables, *IEEE Transactions on Geosciences and Remote Sensing*, 41 (2), 173-183.

Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., De Troch, F. P. & Troch, P. A. (2002). Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values, *Hydrological Processes*, *16*, 995–1013.

Peixoto, J. P. & Oort, A. H. (1992). Physics of climate, American Institute of Physics, *Springer-Verlag NY Inc.*, 520 pp.

Pellarin, T., Wigneron, J., Calvet, J. & Waldteufel, P. (2003). Global soil moisture retrieval from a synthetic L-band brightness temperature data set, *Journal of Geophysical Research*, *108(D12)*, 4364, doi:10.1029/2002JD003086.

Perry, C. D., Vellidis, G., Lowrance, R. & Thomas, D. L. (1999). Watershed-scale water quality impacts of riparian forest management, *Journal of Water Resources Planning & Management*, *125*, 117-126.

Pinker, R. T. and CoAuthors (2003). Surface radiation budgets in support of the GEWEX Continental Scale International Project (GCIP) and the GEWEX Americas Prediction Project (GAPP), including the North American Land Data Assimilation System (NLDAS) Project, *Journal of Geophysical Research*, *108(D22)*, 8844, doi:10.1029/2002JD003301.

Polcher, J. (2001). GLASS implementation underway, GEWEX News, 10, 9.

Prigent, C., Aires, F., Rossow, W. B. & Robock A. (2005). Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: Relationship of satellite observations to in situ soil moisture measurements, *Journal of Geophysical Research*, *110*, D07110, doi:10.1029/2004JD005087.

Rahman, A. F., Sims, D. A., Cordova V. D. & El-Masri, B. Z. (2005). Potential of MODIS EVI and surface temperature for directly estimating per-pixel ecosystem carbon fluxes, *Geophysical Research Letters*, **32**, L19404, doi:10.1029/2005GL024127.

Reichle R. H., McLaughlin D. & Entekhabi D. (2002a). Hydrologic data assimilation with the ensemble Kalman filter, *Monthly Weather Review*, *130*, 103–114.

Reichle, R. H. & Koster R. D. (2004). Bias reduction in short records of satellite soil moisture, *Geophysical Research Letter*, *31*, L19501, doi:10.1029/2004GL020938.

Reichle, R. H. & Koster R. D. (2005). Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model, *Geophysical Research Letter*, *32*, L02404, doi:10.1029/2004GL021700.

Reichle, R. H., Koster, R. D., Dong, J. & Berg, A. A. (2004). Global soil moisture from satellite observations, land surface models, and ground data: Implication for data assimilation, *Journal of Hydrometeorology*, *5*, 430–442.

Reichle, R. H., Walker, J. P., Houser, P. R. & Koster R. D. (2002b). Extended vs. ensemble Kalman filtering for land data assimilation, *Journal of Hydrometeorology*, *3*, 728–740.

Robinson, M., & Dean, T. J. (1993). Measurement of near surface soil water content using a capacitance probe, *Hydrological Processes*, 7, 77–86.

Robock, A., Luo, L., Wood, E. F., Wen, F., Mitchell, K. E., Houser, P. R., Schaake, J. C., Lohmann, D., Cosgrove, B., Sheffield, J., Duan, Q., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Basara, J. B. & Crawford K. C. (2003). Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season, *Journal of Geophysical Research*, *108(D22)*, 8846, doi:10.1029/2002JD003245.

Robock, A., Schlosser, C. A., Vinnikov, K. Y., Speranskaya, N. A., Entin, J. & Qiu, S. (1998). Evaluation of AMIP soil moisture simulations, *Global Planet Change*, *19*, 181–208.

Robock, A., Vinnikov, K. Y., Govindarajalu S., Entin, J. K., Hollinger, S. E., Speranskaya, N. A., Liu, S. & Namkhai, A. (2000). The Global Soil Moisture Data Bank", *Bulletin of American Meteorological Society*, *81*, 1281-1299.

Robock, Alan, Konstantin Ya. Vinnikov, & C. Adam Schlosser (1997). Evaluation of land-surface parameterization schemes using observations, *Journal of Climate*, *10*, 377-379.

Rodell, M., Chao, B. F., Au, A. Y., Kimball, J. S. & Mc-Donald, K. C. (2005). Global biomass variation and its geodynamic effects: 1982-1998, *Earth Interactions*, 9 (2), 1-19.

Rodell, M., Houser, P. R., Berg, A. A. & Famiglietti J. S. (2005). Evaluation of 10 Methods for Initializing a Land Surface Model, *Journal of Hydrometeorology*, *6*, 146–155.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D. & Toll, D. (2004). The global land data assimilation system, *Bulletin of American Meteorological Society*, 85 (3), 381-394.

Rouse, J.W., Haas, R. H., Schell, J. A. & Deering, D. W. (1974). Monitoring vegetation system in the great plains with ERTS. *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, Greenbelt, USA, NASA SP-351*, 3010-3017.

Ryu, D. & Famiglietti, J. S. (2005). Characterization of footprint-scale surface soil moisture variability using Gaussian and beta functions during the Southern Great Plains 1997 (SGP97) hydrology experiment, *Water Resources Research*, *41*, W12433.1–W12433.13.

Sahoo, A. K., Dirmeyer, P. A., Houser, P. R. & Kafatos, M. (2008b). A study of land surface processes using land surface models over the Little River Experimental Watershed, Georgia, *Journal of Geophysical Research*, (submitted).

Sahoo, A. K., Houser, P. R., Ferguson, C., Wood, E. F., Dirmeyer, P. A. & Kafatos, M. (2008a). AMSR-E soil moisture: Evaluation of retrieval methods using the in-situ data over the Little River Experimental Watershed, Georgia, *Remote Sensing of Environment* (submitted).

Schaake, J. C., Duan, Q., Koren, V., Mitchell, K. E., Houser, P. R., Wood, E. F., Robock,
A., Lettenmaier, D. P., Lohmann, D., Cosgrove, B. A., Sheffield, J., Luo, L., Higgins, R.
W., Pinker, R. T. & Tarpley J. D. (2004). An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS), *Journal of Geophysical Research*, vol. 109, D01S90, doi:10.1029/2002JD003309.

Schaefer, G. L. & Paetzold, R. F. (2001). SNOTEL (SNOwpack TELemetry) and SCAN (Soil Climate Analysis Network), *Proceedings of International Workshop on Automated Weather Stations for Applied in Agriculture and Water Resources Management*, World Meteorological Organization AGM-3 WMO/TD 1074.

Schar, C., Luthi, D., Beyerle, U. & Heise, E. (1999). The soil-precipitation feedback: A process study with a regional climate model, *Journal of Climate*, *12*, 722–741.

Schlosser, C. A., Slater, A. G., Robock, A., Pitman, A. J., Vinnikov, K. Y., Henderson-Sellers, A., Speranskaya, N. A., Mitchell, K. & the PILPS 2(d) contributors (2000).

Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS Phase 2(d), *Monthly Weather Review*, *128*, 301–321.

Schmugge, T. (1990). Measurements of surface soil moisture and temperature, In R.J. Hobbs, and H.A. Mooney (Eds.) *Remote Sensing of Biosphere Functioning*. New York: Springer-Verlag.

Sellers, P. J. & Dorman J. L. (1987). Testing the Simple Biosphere model (SiB) using point micrometeorological and biophysical data. *Journal of Applied Meteorology, 26*, 622-651.

Sellers, P. J., Mintz, Y., Sud, Y. C. & Delcher, A. (1986). A Simple Biosphere Model (SiB) for use within general circulation models, *Journal of Atmospheric Sciences*, 43, 505-531.

Sellers, P. J., Tucker, C. J., Collatz, G. J., Los, S. O., Justice, C. O., Dazlich D. A. & Randall D. A. (1996). A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMS. Part II: The Generation of Global Fields of Terrestrial Biophysical Parameters from Satellite Data, *Journal of Climate*, *9*, 706-737.

Seneviratne, S. I., Luthi, D., Litschi, M. & Schar, C. (2006). Land-atmosphere coupling and climate change in Europe, *Nature*, 443 (7108), 205–209.

Sheridan, J. M. & Ferreira V. A. (1992). Physical characteristic and geomorphic data for Little River Watersheds, Georgia, USDA-ARS, Southeast Watershed Research Laboratory Report 099201, 19 pp.

Sheridan, J.M. (1997). Rainfall-streamflow relations for coastal plain watersheds, *Applied Engineering in Agriculture*, *13 (3)*, 333–344.

Sud, Y. C. & Mocko, D. M. (1999). New snow-physics to complement SSiB. Part I: Design and evaluation with ISLSCP Initiative I datasets, *Journal of Meteorological Society of Japan*, 77, 335-348.

Ulaby, F. T., Moore, R. K., & Fung, A. K. (1986). Microwave Remote Sensing, Active and Passive, vol. III: From Theory to Applications, *Massachusetts: Artech House*, 1120 pp.

Ulaby, F. T., Razani, M. & Dobson, M. C. (1983). Effects of vegetation cover on the microwave radiometric sensitivity to soil moisture, *IEEE Transactions on Geosciences and Remote Sensing*, 21, 51–61.

Vinnikov, K. Y. & Yeserkepova, I. B. (1991). Soil moisture: Empirical data and model results, *Journal of Climate*, *4*, 66-79.

Vinnikov, K. Y., Robock, A., Qiu, S. & Entin, J. K. (1999). Optimal design of surface networks for observation of soil moisture, *Journal of Geophysical Research*, *104*, 19,743-19,750

Vinnikov, K. Y., Robock, A., Qiu, S., Entin, J. K., Owe, M., Choudhury, B. J., Hollinger S. E. & Njoku, E. G. (1999). Satellite remote sensing of soil moisture in Illinois, United States, *Journal of Geophysical Research*, *104*, 4145-4168.

Vinnikov, K. Y., Robock, A., Speranskaya, N. A. & Schlosser, C. A. (1996). Scales of temporal and spatial variability of midlatitude soil moisture, *Journal of Geophysical Research*, *101*, 7163-7174.

Wagner, W., Scipal, K., Pathe, C., Gerten, D., Lucht, W. & Rudolf, B. (2003). Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data, *Journal of Geophysical Research*, *108(D19)*, 4611, doi:10.1029/2003JD003663.

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

Walker, J. P. & Houser P. R. (2004). Requirements of a global near-surface soil moisture satellite mission: accuracy, repeat time, and spatial resolution, *Advances in Water Resources*, *27*, 785-801.

Wang, J. R. & Choudhury B. J. (1981). Remote sensing of soil moisture content over bare field at 1.4 GHz frequency, *Journal of Geophysical Research*, 86, 5277-5282.

Wang, J. R. & Schmugge, T. J. (1980). An empirical model for the complex dielectric permittivity of soils as a function of water content, *IEEE Transactions on Geosciences and Remote Sensing, GE-18*, 288-295.

Wen, J., Jackson, T. J., Bindlish, R., Hsu, A. Y. & Su, Z. B. (2005). Retrieval of soil moisture and vegetation water content using SSM/I data over a corn and soybean region, *Journal of Hydrometeorology*, *6* (6), 854–863.

Western, A.W., Bloschl, G. & Grayson, R.B. (1998). Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment, *Journal of Hydrology*, *205 (1–2)*, 20–37.

Western, A.W., Zhou, S.-L., Grayson, R.B., McMahon, T., Bloschl, G. & Wilson, D. (2004). Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, *Journal of Hydrology*, *286 (1-4)*, 113–134.

Wigneron, J.-P., Chanzy, A., Calvet, J.-C. & Bruguier N. (1995). A simple algorithm to retrieve soil moisture and vegetation biomass using passive microwave measurements over crop fields, *Remote Sensing of Environment*, 51, 331–341.

Wilker, H., Drusch, M., Seuffert, G. & Simmer, C. (2006). Effects of the near-surface soil moisture profile on the assimilation of L-band Microwave brightness temperature, *Journal of Hydrometeorology*, *7*, 433–442.

Williams, R.G. (1982). Little River watersheds land use characteristics, USDA-ARS, SEWRL Lab. Note 098201.

Wisner, B., Blaikie, P., Cannon, T. & Davis, I. (2004). At Risk: Natural hazards, people's vulnerability and disasters (2nd ed.), *New York: Routledge*.

Wood, E. and CoAuthors (1998). The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) Phase 2(c) Red-Arkansas River basin experiment: 1. Experimental description and summary intercomparisons, *Global and Planetary Change*, *19*, 115–136.

Xue, Y., Sellers, P. J., Kinter III, J. L. & Shukla, J. (1991). A Simplified Biosphere Model for Global Climate Studies, *Journal of Climate*, *4*, 345-364.

Xue, Y., Zeng, F. J. & Schlosser, C. A. (1996). Sensitivity to soil properties – a case study using HAPEX-Mobility data, *Global and Planetary Change*, *13*, 183-194.

Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., Privette, J. L., Huemmrich, K. F., Fensholt, R., Sandholt, I., Weiss, M., Ahl, D. E., Gower, S. T., Nemani, R. R., Knyazikhin, Y. & Myneni R. B. (2006). MODIS Leaf Area Index Products: From Validation to Algorithm Improvement, *IEEE Transactions on Geosciences and Remote Sensing*, 44(7), 7, 1885-1898.

Yu, C., Loureiro, C., Cheng, J.-J., Jones, L. G., Wang, Y. Y., Chia, Y. P. & Faillace, E. (1993). Data Collection Handbook to Support Modeling Impacts of Radioactive Material in Soil, *Environmental Assessment and Information Sciences Division, Argonne National Laboratory, Argonne, Illinois, USA* [http://web.ead.anl.gov/resrad/documents/data collection.pdf], 158 pp.

Zhan, X., Houser, P. R., Walker, J. P. & Rodell, M. (2004). Validation of AMSR-E Soil Moisture Product Using Data Assimilation Techniques, *COAA2004, June 28-30, 2004, Beijing*.

Zobler, L. (1986). A world soil file for global climate modeling, *NASA Technical Memorandum* 87802, 32 pp.

Zupanski, M., Fletcher, S. J., Navon, I. M., Uzunoglu, B., Heikes, R. P., Randall, D. A., Ringler, T. D. & Daescu, D. (2006). Initiation of ensemble data analysis, *Tellus*, *58A*, 159-170.

CURRICULUM VITAE

Alok Kumar Sahoo received his Bachelor of Science (Geology Major) from the Khallikote Autonomous College, Berhampur, Orissa, India in 1997. He obtained his Master of Science (Earth Science) from University of Roorkee (now IIT Roorkee), Roorkee, India in 1999. He graduated with Master of Technology degree (Civil Engineering) from IIT Kanpur, Kanpur, India in 2001. He was employed as a senior project assistant in IIT Kanpur for one year before joining George Mason University as a research assistant and Ph.D. student in 2002. He was selected for the 2004 and 2005 Graduate Student Summer Program in Earth System Science (GSSP) offered jointly by NASA Goddard Space Flight Center's Earth-Sun Exploration Division and Goddard Earth Sciences and Technology (GEST) Center of the University of Maryland, Baltimore County. During the summer program, he had an opportunity to work in the Hydrological Sciences Branch, NASA GSFC. He did a research visit to Princeton University in summer 2006 for his Ph.D. research work. He was a member of the Cloud and Land Surface Interaction Field Campaign (CLASIC 2007) which was held in Oklahoma in summer 2007. His research interests include soil moisture estimation using remote sensing, land surface models and data assimilation techniques.