Assimilation of coupled microwave/thermal infrared soil moisture profiles into a crop modeling system

Vikalp Mishra
ASSIMILATION OF COUPLED
MICROWAVE/THERMAL INFRARED SOIL MOISTURE
PROFILES INTO A CROP MODELING SYSTEM

by

VIKALP MISHRA

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
The Joint Civil Engineering Program of
The University of Alabama in Huntsville
The University of Alabama at Birmingham
to
The School of Graduate Studies
of
The University of Alabama in Huntsville

HUNTSVILLE, ALABAMA

2017
In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or, in his/her absence, by the Chair of the Department or the Dean of the School of Graduate Studies. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this dissertation.

Vikalp Mishra

10/18/2017
(date)
DISsertation APPROval FORM

Submitted by Vikalp Mishra in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering and accepted on behalf of the Faculty of the School of Graduate Studies by the dissertation committee.

We, the undersigned members of the Graduate Faculty of The University of Alabama in Huntsville, certify that we have advised and/or supervised the candidate of the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering.

Dr. James F. Gruse 10/18/17 Committee Chair

Dr. John R. Macikalski 10/30/17 (Date)

Dr. Ashraf Z. Al-Hamdan 10/18/17 (Date)

Dr. Robert B. Griffin 10/18/17 (Date)

Dr. Jason T. Kirby 10/31/2017 (Date)

Dr. Sherif Ishak 11/2/2017 (Date)

Dr. Shankar Mahalingam 11/2/17 (Date)

Dr. David Berkowitz 11/17/17 (Date)

Department Chair

College Dean

Graduate Dean
The main focus of this study is to utilize Microwave (MW) and Thermal Infrared (TIR) based satellite derived soil moisture (SM) estimates within a robust, state-of-the-art crop model for obtaining gridded, high spatio-temporal estimates of crop yield and water stress, particularly during drought conditions. The study is unique in many aspects: primarily, it integrates MW and TIR based techniques to provide SM estimation from the surface (0-5 cm) downward to the rooting depth (2 m); further, for the first time, the Principle of Maximum Entropy (POME)-derived SM profiles developed to drive a gridded crop model. This study allowed us to assess the robustness of agricultural decision support system in remote areas where ground-based SM and weather observations are not typically available for the crop model to perform adequately. Our previous study shows that satellite derived TIR based SM estimates can be made to run the crop model with reasonable success in terms of crop yields as compared to reported yields; yet further research was needed in order to attain higher accuracy at regional spatial scales. We present here: a) detailed validation of the POME model; b) disaggregation of MW SM estimates; c)
developing MW/TIR coupled SM profiles using the POME model; and d) application of the developed profile into a crop model via Ensemble Kalman filter. Coupling of MW and TIR SM estimates, downscaling of surface SM and assimilation into crop model are the key components of this study. NASAs Land Information system (LIS) gridded SM estimates from the Noah Land Surface Model (LSM), as well as ground based SM observations from operational Natural Resource Conservation Services (NRCS) SCAN sites within the study region were used for profile validation. Yield comparisons were made against NASS reported yields at county level.
Abstract Approval: Committee Chair
Dr. James F. Cruise
10/18/17

Department Chair
Dr. Sherif Ishak
11/2/17

Graduate Dean
Dr. David Berkowitz
11/3/17
ACKNOWLEDGMENTS

I would like to thank my advisors, Dr. James F. Cruise and Dr. John R. Mecikalski, for showing faith in me and patiently guiding me through the course of my research. I would like to thank the committee members Dr. Griffin, Dr. Al-Hamdan and Dr. Kirby for their useful insights and feedbacks. I would like to thank the extremely friendly and helpful office staff at Earth System Science Center and Civil Department. This work has been greatly benefited by useful discussions with the colleagues at UAH. I would like to thank NASA for providing financial support through fellowship which greatly helped me carrying out this work. Finally, I would like to express my gratitude to my family members and friends for their unwavering support. Not but not the least, I would like to thank my wife Niru, without her constant support this would not have been possible.
# Table of Contents

List of Figures \hspace{16cm} xiii

List of Tables \hspace{16cm} xvii

Chapter

1 Introduction \hspace{16cm} 1

1.1 Aims and Purposes of the Study \hspace{16cm} 3

2 Modeling Soil Moisture Profiles in Irrigated Fields by Principle of Maximum Entropy \hspace{16cm} 5

2.1 Introduction \hspace{16cm} 5

2.2 Soil Moisture States and Irrigation \hspace{16cm} 8

2.3 Entropy Applied to Soil Moisture States \hspace{16cm} 11

2.4 Soil Moisture profile Development Models \hspace{16cm} 12

2.4.1 Development of Soil Moisture Profiles from POME \hspace{16cm} 12

2.4.2 Physical Soil Moisture Profile Model \hspace{16cm} 15

2.5 Study Area \hspace{16cm} 16

2.6 Methodology \hspace{16cm} 17

2.6.1 Irrigation Phase Simulation \hspace{16cm} 20

2.6.2 Drying Phase Simulation \hspace{16cm} 22

2.7 Results and Discussions \hspace{16cm} 25
2.7.1 Varification of Entropy Profiles ........................................ 25
  2.7.1.1 SCAN site varification ........................................... 25
  2.7.1.2 Entropy Validation with Derived Parameters ................. 29
2.7.2 Application of the POME model to an irrigation cycle ........... 32
2.7.3 Comparision to Physical Model ...................................... 35
2.8 Conclusion .............................................................. 41

3 Development of Soil Moisture Profiles Through Coupled Microwave-
  Thermal Infrared Observations in the Southeastern United States 46
  3.1 Introduction .......................................................... 47
  3.2 Study Area and Data Sources ....................................... 51
    3.2.1 Study Area ..................................................... 51
    3.2.2 Data Sources ................................................... 52
      3.2.2.1 Microwave Surface SM ................................. 52
      3.2.2.2 Thermal Infrared - ALEXI ............................ 53
      3.2.2.3 In-situ Observations .................................. 54
      3.2.2.4 Noah Soil Moisture .................................... 55
  3.3 Methodology .......................................................... 56
    3.3.1 ALEXI Retrievals .............................................. 56
      3.3.1.1 Surface Evaporation ................................... 56
      3.3.1.2 Mean Rootzone Moisture Retrieval .................. 57
    3.3.2 Surface Disaggregation ....................................... 58
      3.3.2.1 Modified SEE Computation ............................ 60
3.3.3 Profile Development ........................................... 61
3.3.4 Temporal Compositing ................................. 63
3.3.5 Evaluation Metrics ........................................... 63
3.4 Results and Discussions ................................. 65
  3.4.1 Comparison with Noah LSM ...................... 65
  3.4.2 Comparison with in-situ Observations .......... 69
  3.4.3 Intercomparison of Noah, POME with In-situ Observations . 73
  3.4.4 Error Characterization ................................. 77
    3.4.4.1 Effect of Disaggregation of AMSR-E MW Data . 79
    3.4.4.2 Effect of Mean SM Inputs ...................... 82
3.5 conclusions ..................................................... 83

4 An Initial Assessment of SMAP Soil Moisture Disaggregation Scheme
Using TIR Surface Evaporation Data over the Continental United States 87
  4.1 Introduction .................................................. 88
  4.2 Background ................................................... 91
  4.3 Data Description ............................................ 94
    4.3.1 SMAP Soil Moisture Data ......................... 94
    4.3.2 ALEXI Surface Evaporation ....................... 95
    4.3.3 NRCS SCAN Observations ......................... 95
  4.4 Methodology ................................................ 96
    4.4.1 Surface SM Disaggregation ..................... 96
5 Assimilation of Coupled Microwave/Thermal Infrared Soil Moisture Profiles into a Crop Model

5.1 Introduction ........................................... 126
5.2 Background ........................................... 129
5.3 Study Area and Data Description ......................... 132
  5.3.1 Study Area ......................................... 132
  5.3.2 Data Sources ....................................... 133
    5.3.2.1 Satellite Derived Soil Moisture Profiles .......... 133
    5.3.2.2 Climate Dataset ................................. 136
    5.3.2.3 Observed Precipitation Data ..................... 136
5.3.2.4 *In-situ* Soil moisture data 137

5.4 Methodology 138

5.4.1 Modification of remotely sensed SM Profiles 138

5.4.1.1 Crop Model Simulation 139

5.4.2 Data Assimilation 140

5.4.2.1 Modeling Strategy 142

5.5 Results 143

5.5.1 Profile Calibration Results 143

5.5.2 Data Assimilation Results 144

5.5.3 Results with Degraded Climate Inputs 148

5.6 Discussion 153

5.7 Conclusion 158

6 Conclusion 161

6.1 Future Directions 164

APPENDIX A: Physically based Mathematical Model 168
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Possible soil moisture profiles [after Al-Hamdan and Cruise (2010)]</td>
</tr>
<tr>
<td>2.2</td>
<td>Study area and SCAN site (2078) location.</td>
</tr>
<tr>
<td>2.3</td>
<td>Flow Chart of wetting front development using POME model</td>
</tr>
<tr>
<td>2.4</td>
<td>Results of derived inflection points at multiple applied water amounts at SCAN site the using moisture redistribution technique</td>
</tr>
<tr>
<td>2.5</td>
<td>Typical wet case POME based profiles against SCAN site observed data</td>
</tr>
<tr>
<td>2.6</td>
<td>Typical dry case POME based profiles against SCAN site observed data</td>
</tr>
<tr>
<td>2.7</td>
<td>Dynamic case POME based profiles against SCAN site observed data</td>
</tr>
<tr>
<td>2.8</td>
<td>Representative dynamic case POME based soil moisture profiles with derived inputs parameters against SCAN site observed data</td>
</tr>
<tr>
<td>2.9</td>
<td>Typical wet case soil moisture profile results during irrigation phase simulation a) Ad- Silt loam soil; b) Dc - Silty clay soil</td>
</tr>
<tr>
<td>2.10</td>
<td>Field map with soil moisture conditions on day 3; 6; 9 and 12 at surface; 20cm; 45cm and 60cm depths</td>
</tr>
<tr>
<td>2.11</td>
<td>Total mass error of the POME model against the model</td>
</tr>
<tr>
<td>2.12</td>
<td>Typical soil moisture profile comparisons between physical model and entropy generated profiles for days 1, 5 and 11 of an irrigation cycle. Vertical dashed lines represents the soil specific effective wilting point (left) and field capacity (right) positions.</td>
</tr>
<tr>
<td>2.13</td>
<td>Box Plot of errors of the POME model against the mathematical model at layer depths</td>
</tr>
</tbody>
</table>
3.1 Overview of study area showing location of all the SCAN sites. The dark blue circles indicate sites with most consistent data availability and are being used for comparison and validation in this study. The right figure shows a land cover map (MODIS-2008) for the study area.

3.2 Map of bias, RMSE, unbiased RMSE and Correlation over multiple years (2006-2010) at different layer depths: top panel: 0-10cm; middle panel: 10-40cm and bottom panel: 40-100cm.

3.3 Comparison of Noah and POME SM profiles at multiple layer depths by Land Cover across Southeast U.S.

3.4 Statistics at SCAN sites showing bias, Correlation, RMSE and ubRMSE between in-situ observations and POME SM profiles at multiple depths.

3.5 Time series of soil moisture condition at 10 NRCS SCAN sites from the POME model (Blue); Noah LSM (green) and in-situ observations (red) at three layer depths (2006-2010).

3.6 POME/ALEXI profiles and Noah statistics at all SCAN sites compared against observations averaged across layer depths.

3.7 Triple collocation analyses of SM profiles from Noah (green), POME (blue) and in-situ observations (red) at SCAN site locations at the depths of 0-10; 10-40 and 40-100 cm.

3.8 POME model sensitivity to (a) boundary condition; (b) sensitivity to profile mean input towards profile Bias and RMSE in terms of effective SM.

3.9 Map of Southeast U.S. demonstrating temporal statics in bias, RMSE, ubRMSE and correlation between coarse (top panel) and fine (bottom panel) resolution AMSR-E (MW) and Noah LSM surface SM (2006-2010).


4.1 Continental U.S. ecological domain with active NRCS SCAN site locations.

4.2 SM estimates from SMAP at coarse resolution Passive (36-km); Active (3-km); Active/Passive (9 km); and Enhanced product (9-km) compared with TIR-Downscaled SM data (3 and 9-km) on 8 June, 2015. The white spaces indicate no data availability.
4.3 A daily time series of spatial correlation (top); bias (middle) and coefficient of ubRMSE (bottom) at 9-km and 3-km spatial scales over CONUS between SMAP and TIR-Downscaled SSM products. 104

4.4 Map of CONUS displaying statistics between SMAP-AP, E and TIR-downscaled SM at 9-km scale: correlation coefficient (top); Bias (middle) and ubRMSE (bottom) distribution across CONUS for the period of Apr-June 2015 (left two panels) Apr 2015 - Nov-2016 (right panel) 106

4.5 Statistics between SMAP-A and TIR-Downscaled SM at 3-km scale over CONUS. 107

4.6 Overall gain statistics at 3-km spatial resolution SSM products: SMAP-A (Left) and TIR-Down (Right) 110

4.7 Map displaying overall gain statistics NRCS SCAN observations relative to SMAP SM (AP and E) and downscaled SM at 9-km scale. 112

4.8 Plot displaying percent of SCAN sites with positive gain in moving from coarse to finer resolution against SCAN in-situ observations. 112

4.9 Boxplot displaying the statistics distribution between SCAN observations with SMAP-A/AP and TIR-Downscaled SSM at 3 and 9-km spatial scale. (3 month data period). 114

4.10 Boxplot of statistics between SMAP-E and TIR-Down (3 and 9-km) resolution SSM products against SCAN observations as a function of ecological domain. (19 month data period) 116

4.11 Percent of SCAN sites showing positive gains by ecological region. 117

4.12 Mean annual fraction of vegetative cover over CONUS based on MODIS LAI (2016). 119

4.13 Comparison of SMAP-E and TIR-Downscaled SSM statistics as a function of fractional vegetation cover (9-km). 119

4.14 Results of triple collocation error analysis between NLDAS2, SMAP-E and TIR-Downscaled SSM. 121

5.1 Overview of study area showing landcover type and locations of the selected regions for detailed analysis. 134
5.2 SM profile comparison statistics against SCAN observations for the period of 2006-2010. SP* - SCAN vs POME (earlier version of Mishra et al. (2017)); SN - SCAN vs Noah LSM; and SP - SCAN vs POME (current version). ........................................... 144

5.3 Boxplot showing yield at different regions between GriDSSAT simulation in rainfed (green) and data assimilation (blue) mode with NASS yields (red dot) from 2006-2010. *years NASS yield data was unavailable 146

5.4 Relative errors (%) of open-loop (rainfed-green) and remotely sensed profile (blue) assimilated annual yields compared to NASS (2006-2007). *years NASS yield data was unavailable .......................... 147

5.5 Relative errors in yields by county for rainfed and data assimilation simulations against NASS reported yields (2006-2007). .............................. 148

5.6 Boxplot showing yield at different regions between GriDSSAT simulation with degraded climate inputs in rainfed (green) and data assimilation (blue) mode with NASS yields (red dot) from 2006-2010. *years NASS yield data was unavailable .......................... 149

5.7 Relative errors (%) of rainfed (green) and MW-TIR SM assimilation (blue) yields with a single GHCND station climate runs compared to NASS (2006-2007). *years NASS yield data was unavailable .......................... 151

5.8 Relative errors in yields by County for rainfed and DA simulations: (a) with high resolution gridded precipitation data; and (b) with single GHCND station climate data against NASS reported yields (2006-2007) 152

5.9 Overall absolute errors (%) for gridded climate (left) and single GHCND climate data simulations with and without DA against NASS yields. 155

5.10 Mean SM difference (mm) time series at non-irrigated (left panel) and irrigated (right panel) grid point for years: 2006, 2007 and 2009. .......................... 158
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Soil characteristics at SCAN site as well as the study field</td>
<td>17</td>
</tr>
<tr>
<td>2.2 Summary of POME based soil moisture profile results for all possible three cases as compared to the observed values at SCAN site.</td>
<td>29</td>
</tr>
<tr>
<td>2.3 Summary of POME based soil moisture profile results for dynamic case against SCAN site observed data with derived inflection point as well as inflection moisture content.</td>
<td>32</td>
</tr>
<tr>
<td>2.4 Summary of wetting front simulations</td>
<td>33</td>
</tr>
<tr>
<td>2.5 Summary of drying phase simulations: IP is the inflection point depth (cm); A is surface effective SM; B is effective soil moisture content at IP; C is bottom (100 cm) effective soil moisture content; and D is total amount of water (cm) contained in profile. E is the actual evapotranspiration in cm of water per day.</td>
<td>36</td>
</tr>
<tr>
<td>2.6 Inflection point development results from physical model run and moisture redistribution technique</td>
<td>42</td>
</tr>
<tr>
<td>2.7 Summary of irrigation efficiency observed by the POME (in cm of water)</td>
<td>42</td>
</tr>
<tr>
<td>3.1 SCAN site 5 km dominant land cover (MODIS 2008) and soil characteristics (SCAN) at the surface and depth of 100 cm [S-sand; L-loam; Si-silt]</td>
<td>55</td>
</tr>
<tr>
<td>3.2 Results of temporal comparisons in absolute bias, RMSE, ubRMSE and correlation at 10 sites between the developed profile and Noah SM profiles against SCAN observations at 0-10; 10-40 and 40-100cm depths [NP - Noah vs POME; SP - SCAN vs POME; and NS Noah vs SCAN]</td>
<td>76</td>
</tr>
</tbody>
</table>
3.3 Statistical comparison before and after disaggregation of coarse resolution MW SM against SCAN observations \( [A = \text{Bias}; B = \text{RMSE}; C = \text{ubRMSE}; D = \text{correlation coefficient}; N = \text{number of days data points was available}; \text{maximum possible } N = 1825]; ^\text{non-significant correlation using two-tailed t-test at 99\% CI} \) 82

4.1 Summary statistics between remotely sensed SSM and SCAN observations across CONUS with average number of SCAN sites and days available for comparison. 108

5.1 Study area location and cropland information with mean rainfall and NASS yields for the period of 2006-2010 \( [{growing season, N \text{ refers to number of grid points (5x5 km}^2) \text{ after masking out water and forested regions}}] \) 135

5.2 Summary yield results of rainfed and DA simulation results compared with reported NASS yields (2006-2010): (a) with gridded climate data (b) with single GHCND station climate data. 154
CHAPTER 1

INTRODUCTION

Global food security is one of the most pressing issues of the 21st century. Growing populations, competing uses for land and natural resources (e.g. water), and climate change will result in significant stresses on world food supplies. Consequently, methods have been developed to increase crop yields including the development of new cultivars, more effective fertilizers, and more efficient irrigation equipment and strategies. Agricultural simulation models are a key component in testing advances in agricultural technology and predicting crop responses to current and future climate forcings. However, crop models depend on accurate specifications of time-dependent SM conditions to precisely capture their impacts on crop conditions and at-harvest yield estimates (Ritchie, 1972). Typically, SM is estimated using measurements of precipitation (water input) in conjunction with a soil water transport model, which distributes the moisture over the root-zone profile (Mishra et al., 2013). Obtaining accurate precipitation estimates is challenging, especially in data-limited regions of the world.

Satellite derived SM estimations on the other hand are in more advanced stages as compared to precipitation estimates and are available with considerable accuracy,
specifically from passive MW based instruments such as the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E), the Soil Moisture and Ocean Salinity (SMOS), Soil Moisture Active Passive (SMAP) etc. MW radiometers typically have coarse (24-40 km$^2$) spatial resolution, additional disadvantage associated with MW remote sensing is that it provides SM estimates only near the surface, usually in the top-most 2-5 cm of the soil column (Ulaby et al., 1996). Since, hydrological or biophysical models require SM content in the entire root-zone (surface to $\sim$2 m depth), MW remote sensing estimates alone are not sufficient to drive these models. Therefore it is required to: (a) downscale the coarse resolution MW SM estimates and (b) to extend the SM information available at the surface to represent root-zone SM content as well. Those two issues are at the core of the proposed research.

In this study, a physically based semi-empirical disaggregation approach based on soil evaporation efficiency and its relation to the surface SM condition known as Disaggregation based on Physical and Theoretical scale Change (DisPATCh) model (Merlin et al., 2013, 2008) has been used to downscale MW surface SM estimates. Thermal infrared (TIR) techniques, unlike MW instruments, can be used to retrieve root-zone moisture based on energy fluxes at relatively higher spatial resolutions. The Atmospheric Land Exchange Inverse (ALEXI) model (Anderson et al., 2011b, 1997, 2007a; Hain et al., 2011), based on the TSEB approach, is a well-documented model for estimating surface energy fluxes. The ALEXI model is used in this study as the data source for TIR based ET and root zone SM estimates. In order to run a crop model using satellite derived SM data, the development of a vertical SM profile is
necessary. The Principle of Maximum Entropy based SM profile development model by Al-Hamdan and Cruise (2010) requires minimum initial datasets making it ideal for use with satellite-derived inputs has been used in this study to develop a vertical SM profile.

Lastly, the satellite derived SM profile will be assimilated with the ground based Decision Support System for Agro-Technology Transfer (DSSAT) (Hoogenboom et al., 2010; Jones et al., 2003) crop model. In this study, the standard DSSAT source code was modified to assimilate satellite derived daily SM profile values (available on cloud free days) via Ensemble Kalman filter (EnKF).

1.1 Aims and Purposes of the Study

This study is focused on developing a technique that utilizes the best of both MW and TIR based SM estimations to develop a vertical SM profile that can be input to the gridded crop model. The proposed study is broadly divided into four sections; each section was/is being developed into a peer reviewed scientific publication:

- Article 1 - Application and extensive validation of the principle of maximum entropy based vertical soil moisture profiles (Published in *Entropy* (Mishra et al., 2015)).

- Article 2 - Downscaling of MW SM estimates utilizing SEE based disaggregation approach and coupling with TIR based root zone estimates for optimum profile development via the POME model. (*Under Review* in *Hydrology and Earth System Sciences*)
• Article 3 - An Initial Assessment of SMAP disaggregation using Fine Resolution Thermal Infrared Surface Evaporation data over CONUS (Submitted to *International Journal of Applied Earth Observation and Geoinformation*).

• Article 4 - Assimilate the developed profiles into the DSSAT model through an ensemble Kalman filter and apply to the modified agricultural model under diverse climatic conditions. (Submitted to *Agriculture and Forest Meteorology*).
CHAPTER 2

MODELING SOIL MOISTURE PROFILES IN IRRIGATED FIELDS

BY PRINCIPLE OF MAXIMUM ENTROPY


Abstract

Vertical soil moisture profiles based on the principle of maximum entropy (POME) were validated using field and model data and applied to guide an irrigation cycle over a maize field in north central Alabama (USA). The results demonstrate that a simple two-constraint entropy model under the assumption of a uniform initial soil moisture distribution can simulate most soil moisture profiles that occur in the particular soil and climate regime that prevails in the study area. The results of the irrigation simulation demonstrated that the POME model produced a very efficient irrigation strategy with minimal losses (about 1.9% of total applied water). However, the results for finely-textured (silty clay) soils were problematic in that some plant stress did develop due to insufficient applied water. Soil moisture states in these soils fell to around 31% of available moisture content, but only on the last day of the drying side of the irrigation cycle. Overall, the POME approach showed promise as a general strategy to guide irrigation in humid environments, such as the Southeastern U.S.

2.1 Introduction

Irrigation efficiency is a primary concern of irrigating farmers around the world. It is defined as the ratio of the amount of water beneficially used by the crop or its ecosystem to the volume of water applied minus the change in soil moisture storage
Beneficial water use is the amount of water that is utilized for crop evapotranspiration (ET), soil evaporation for climate control, leaching to reduce soil salinity, and water storage incorporated in the biomass (Burt et al., 1997; FAO, 2009). Thus, to maximize irrigation efficiency, the goal is to only apply the amount of water that the crop can beneficially use.

Irrigated agriculture contributes more than its share in terms of productivity and profit. Fereres and Connor (2004) pointed out that only 17% of the agricultural land is irrigated across the globe, however, it accounts for approximately 40% of the total agricultural products. In the United States, only 5-7% of the total agricultural land is regularly irrigated (Ozdogan and Gutman, 2008), and yet it is responsible for over 50% of total crop income (Schaible and Aillery, 2012). Though only a fraction of the total agricultural land area, the water use for irrigation accounts for over 60% of the total fresh water withdrawals world-wide (FAO, 2009). In the United States alone, 484.5 billion liters of water per day are used for irrigation (Kenny et al., 2009). Therefore irrigation efficiency is extremely important, particularly in water scarce regions of the world. Over irrigation can not only waste valuable water, but also harm soil productivity by leaving residual chemicals or toxins (solutions for example) in the soil when the excess water evaporates (Fereres and Soriano, 2007). On the other hand, under irrigation can lead to sub-optimal crop yields (Zhang et al., 1998). Therefore for advanced irrigation efficiency and beneficial water use, optimal irrigation scheduling is required. Traditionally, irrigation scheduling is accomplished by measuring plant-water stress or soil moisture deficits (Sammis et al., 2012). Knowledge of the soil moisture profile during the irrigation cycle is essential to achieving
maximum irrigation efficiency. Soil moisture at shallow depths are known to be extremely variable temporally (Starks et al., 2003) and can show significant variability with depth \((z)\) (Scott et al., 2003). Therefore frequent *in-situ* soil moisture monitoring is required. In technologically advanced operations, the irrigation equipment can be linked to soil moisture probes in the field that can regulate the amount of water applied. However, it is often the case that rules of thumb are followed based on known or assumed evaporation rates for the area. At best, these rules are crude since they are not based on real time information (weather, soil moisture states, etc.) and often lead to poor irrigation efficiencies. Therefore, there is a need for a low cost method to accurately monitor soil moisture during the irrigation cycle.

During the irrigation cycle, the available information includes the application rate and the total applied water at any time. The distribution of this water within the root zone will not be known unless multi-layer soil moisture probes are employed. Yet, this distribution can be of vital importance to the crop sustainability due to the distribution of plant roots within the zone. The uptake of water by the plants is a function of root mass and suction pressure, both of which vary vertically within the soil profile. Thus, an accurate estimate of the vertical soil moisture distribution is critical to controlling irrigation rates in real time.

There have been multiple studies in the past addressing the development of a vertical soil moisture profile from limited available information. Broadly the approach can be categorized as: regression (Arya and Richter, 1983; Kondratyev et al., 1977; Srivastava et al., 1997); inversion (Kostov and Jackson, 1993); knowledge based (Jackson, 1980; Reutov and Shutko, 1986); water balance approach (De Troch et al.,
1996; Singh, 1988); probabilistic (Al-Hamdan and Cruise, 2010; Assouline et al., 1998; Pachepsky et al., 2006); and intelligence (Koekkoek and Booltink, 1999; Kornelsen and Coulibaly, 2014). Most of the recent studies are hybrid in nature and focused more on total soil moisture estimation rather than soil moisture as a function of depth (Al-Hamdan and Cruise, 2010).

Data assimilation techniques can be utilized to improve the soil moisture state in both hydrological and atmospheric models (Huang et al., 2008). There have been many previous studies where soil moisture states have been treated as random variables in one way or another (Al-Hamdan and Cruise, 2010; Mays et al., 2002; Moore, 1985; Pachepsky et al., 2006; Pan et al., 2011; Singh, 2010b). Pachepsky et al. (2006) provided a comprehensive review of how information theory could be applied to soil moisture transition states. Following this theory and building on recent work in the field, the objective of this study was to develop an entropy based model that will allow for easy and cost effective monitoring of irrigated soil moisture states.

2.2 Soil Moisture States and Irrigation

Irrigation is employed to enhance crop yields in many places in the world, even those that have sufficient rainfall on an annual basis. In many cases the precipitation is not distributed uniformly throughout the year so that growing season rainfall may not be sufficient to supply crop requirements on a sustainable basis (Mishra et al., 2013). The method chosen for delivery of irrigation water is an important element in determining irrigation efficiency. Low cost methods include simply flooding the field, either entirely (border irrigation) or by rows (furrow), while more technologically
advanced methods include spraying (pivot, or linear systems) or via perforated lines buried in the root zone (drip).

Surface irrigation methods rely on a complex infiltration process that is related not only to soil properties, but also on processes by which the plants uptake the water from the soil. The soil moisture infiltration/transpiration process is complex and a number of models have been proposed to simulate it (Bras, 1990). Infiltration of water into a porous media proceeds in three stages (Singh, 1992): on first contact the soil particles absorb water depending on the organic content of the soil; next water will begin to wet the surface of the particles, and finally, the water will begin to fill the pore spaces between the particles until the soil matrix becomes saturated (pore spaces are filled to capacity). Water will tend to move in the soil from a wetter area to a drier one through a combination of diffusion and advection processes. Under these forces, water will drain vertically from the soil until it reaches field capacity, which is the water volume held in the pores against gravitational forces. However, plant root uptake can continue until the wilting point is reached, i.e., the point at which negative pore pressures in the soil counteract the root uptake forces. At this point, movement of moisture within the soil column is theoretically ended.

The one dimensional (vertical) infiltration process is described by the Richards Equation (Al-Hamdan and Cruise, 2010):

$$\frac{\partial \theta}{\partial t} - \frac{\partial}{\partial z} \left( \frac{K(\theta)}{\partial \psi} \right) + \frac{\partial K(\theta)}{\partial z} = 0$$  \hspace{1cm} (2.1)
Where \( \theta \) = soil moisture content, \( K(\theta) \) = hydraulic conductivity at \( \theta \) (mm/hr); and \( \psi \) = suction head (mm). There is no analytical solution of Equation (1) for general initial and boundary conditions but a number of commonly employed analytical infiltration models are designed to simulate it to some degree (Bras, 1990). There are many uncertainties associated with any method with estimated soil properties being the source of perhaps the largest uncertainty in the process. Properties estimated in laboratory settings rarely equate to actual field values, thus characteristic tables usually give a mean (or median) value for a particular soil with a large spread around that value, sometimes encompassing orders of magnitude (Bras, 1990). In addition, there is not a clear relationship between soil properties and actual behavior (\( \theta \) vs \( k(\theta) \) for example) and many different models have been proposed to explain these relationships (Clapp and Hornberger, 1978). There have been multiple infiltration models commonly used in hydrology, agricultural applications and watershed management such as (Green and Ampt, 1911; Holtan; Horton, 1939; Kostiakov, 1932) etc. The uptake of moisture by plant roots depends on many plant characteristics (root distribution, for example) which are difficult to predict.

It is well known that soil moisture movement and it’s associated distributive processes are inherently complex. These complexities pose significant limitations on the ability to physically model the system. Several authors have argued that soil moisture’s uncertainties and complexities can be best described through description of it’s entropy (Mays et al., 2002; Pachepsky et al., 2006; Singh, 1997, 2010a).
2.3 Entropy Applied to Soil Moisture States

In the classic Shannon Entropy formulation, the information associated with a system in state $i$, is given as (Shannon, 1948):

$$I_i = \ln \left( \frac{1}{p_i} \right) \tag{2.2}$$

Where $p_i$ is the probability that the system is in state $i$, and $\ln$ is the naperian logarithm.

Then, on average, the information content of the data related to the state is given by the Shannon Entropy Function which is just the negative expected value of the log probability:

$$H = -\sum_{i=1}^{n} p_i \ln(p_i) \tag{2.3}$$

or, in the continuous case: $H=\int_{0}^{\infty} f(x) \ln (f(x)) \, dx$ : where $f(x)$ is the continuous pdf.

As pointed out by Pachepsky et al. (2006), the fluctuation of entropy as the system transitions between states is given by:

$$\sigma_{i,j}^2 = \sum_{i,j} p_{i,j} \left( \ln \left( \frac{p_i}{p_j} \right) \right)^2 \tag{2.4}$$

Thus, the information content of the data associated with the system can be defined by a probability distribution. According to Mays et al. (2002), the information entropy is a measure of the correspondence between a probability distribution function
(pdf) constructed from a data set associated with a system and the pdf associated with minimum information about the system. For a situation where minimum information is known a priori, or the system is unpredictable, the pdf will be uniform and the entropy will be high (Mays et al., 2002; Pachepsky et al., 2006). Incidentally, in the case of maximum entropy, $\sigma_{i,j}^2$ will be zero (Pachepsky et al., 2006).

The Maximum Entropy formulation was developed by (Jaynes, 1957). The Principle of Maximum Entropy (POME) proposes that if inferences are to be made based on incomplete information, they should be drawn from the probability distribution that has maximum entropy permitted by the information that is already available (Barbé et al., 1991). Based on this concept, Al-Hamdan and Cruise (2010) developed an entropy-based model for simulating soil moisture profiles corresponding to all phases of the physical process. The model was then incorporated into a more general hydrologic formulation given by Singh (2010a) and further evaluated by Singh (2011).

2.4 Soil Moisture profile Development Models

2.4.1 Development of Soil Moisture Profiles from POME

Al-Hamdan and Cruise (2010) developed a set of soil moisture profiles assuming a uniform distribution throughout the soil column. The method was based on an approach originally employed by Chiu (1987) to compute vertical velocity distributions in open channels. Development of three distinct soil moisture profile possibilities exist: first, as water is applied to an initially dry soil, the column will be wetter near
the surface and decrease thereafter as shown in Figure 2.1 below (solid line). Then, after a wetting event, the profile will be dry in the upper layers, but retain moisture in the bottom of the column thus exhibiting one of the profiles shown in the dashed or short dotted lines in the Figure 2.1, depending on the time since the wetting. The case of the parabolic shape (dashed line) is referred to as the dynamic case Al-Hamdan and Cruise (2010) and may develop in the hours immediately after the event.

![Figure 2.1: Possible soil moisture profiles [after Al-Hamdan and Cruise (2010)]](image)

The application of POME to develop a one-dimensional soil moisture profile requires two constraints: the total probability constraint:

$$\int_{\Theta_L}^{\Theta_0} f(\Theta) d\Theta = 1 \quad (2.5)$$

and the constraint to satisfy mass balance:

$$\int_{\Theta_L}^{\Theta_0} \Theta f(\Theta) d\Theta = \Theta \quad (2.6)$$
where $\Theta$ is the effective saturation and computed as $\frac{\theta - \theta_r}{\eta - \theta_r}$ whereas $\overline{\Theta}$ is the mean value of the soil column, $\eta$ is soil porosity and $\theta_r$ is the irreducible water content of the soil; $\Theta_0$ and $\Theta_L$ are the surface and bottom effective saturations. The second constraint serves to connect the first moment in probability space to the mean water content of the soil column in physical space. The Shannon (1948) entropy is given by:

$$I = -\int_0^\infty f(x) \ln(f(x)) \, dx$$  \hspace{1cm} (2.7)

Where $f(x)$ is the probability density function of the variable. Maximizing $I$ in Equation (2.7) for the uniform pdf subject to the constraints above, Chiu (1987) developed the 1-D profile of a variable decreasing monotonically from the surface down (wet case) using the method of Lagrange multipliers (put into soil moisture terms by Al-Hamdan and Cruise (2010):

$$\Theta(z) = \frac{\ln[exp(\lambda_2 \Theta_0) - \lambda_2 exp(1 - \lambda_1) (\frac{z}{L})]}{\lambda_2}$$  \hspace{1cm} (2.8)

The Lagrange multipliers ($\lambda$’s) can be solved from application of the constraints and boundary conditions: surface effective saturation, $\Theta_0$, effective saturation at the bottom of the column, and mean effective saturation value of the soil column ($\overline{\Theta}$), $z$ is calculation depth, and $L$ is total depth of the column.

Following the same procedure, Al-Hamdan and Cruise (2010) developed the profile for the second case (monotonically increasing vertically i.e., dry case):
\[
\Theta(z) = \frac{ln \left[ \exp (\lambda z_0) + \lambda_2 \exp (\lambda_2 (1 - \lambda_1) \left( \frac{z}{L} \right)) \right]}{\lambda_2}
\]  

\hspace{1cm} (2.9)

Relationships between the Lagrange multipliers and the constraints and boundary conditions have also been developed, as in the first case, with the resulting system of nonlinear equations solved for the multipliers by a technique given by Barbé et al. (1991). Finally, Al-Hamdan and Cruise (2010) demonstrated that the third case (dynamic) can be handled by a combination of the wet and dry profiles.

2.4.2 Physical Soil Moisture Profile Model

In order to further evaluate the entropy procedure against profiles that are possible but may not be present at the validation side, a mathematical model simulating the physics of infiltration, drainage and plant uptake was constructed for the test field. To simulate soil moisture processes in the root zone, a one dimensional model accounting for the root distribution throughout the root zone (Yadav et al., 2009) as well as the soil water drainage and availability (Neitsch et al., 2011) was developed. The total vegetative water extraction is determined using a semi-empirical root water extraction term S that can be expressed as a function of \( S_{max} \), i.e. maximum potential transpiration and a soil water availability factor (\( \alpha \)):

\[
S(t) = \alpha S_{max}(t); \quad \alpha \leq 1
\]  

\hspace{1cm} (2.10)

The soil water availability factor accounts for a restriction of transpiration in either a soil moisture deficit in which the suction capacity of the soil exceeds that of
the root or in an excess of soil moisture where the plants are unable to uptake water due to insufficient aeration (Feddes et al., 1978)

To simulate drainage, water is allowed to move vertically in each soil layer modeled if the water content is above field capacity and the layer below is below saturation. The amount of water that moves from one layer to the next is calculated on the storage routing methodology of the Soil Water and Assessment Tool [SWAT, (Neitsch et al., 2011)]. A full description of the physically-based mathematical model used in this study is provided in the Appendix.

2.5 Study Area

The study area is an irrigated corn (maize) field located in Toney (Madison County), Alabama (Lat: 34°54'; Long: 86°36'). The field is under center pivot irrigation. Alabama is a subtropical, humid region in the Southeastern U.S. and receives over 1400 mm of precipitation annually (Mishra et al., 2013). However, less than 300 mm of precipitation occurs during the growing season, on an average, and thus water intensive crops such as corn can benefit greatly from supplemental irrigation (Paudel et al., 2005).

The field covers 30.5 ha (75.5 ac) comprised of 6 different soil types. The soil properties as weighted averages in depths of 100cm are shown in Table 2.1. The study area was selected due to the fact that a US Department of Agriculture Soil Climate Analysis Network (SCAN) monitoring station (2078) is located adjacent to the field. This site provided hourly soil moisture and climate observations used in the verification of the methodology. Soil moisture is measured with a dielectric probe and
climate data include precipitation, temperature, relative humidity, and wind speed and direction.

**Table 2.1**: Soil characteristics at SCAN site as well as the study field

<table>
<thead>
<tr>
<th>Soil Type</th>
<th>Sand%</th>
<th>Clay%</th>
<th>$\theta_{wp}$ ($m^3 m^{-3}$)</th>
<th>$\theta_{fc}$ ($m^3 m^{-3}$)</th>
<th>$\theta_{sat}$ ($m^3 m^{-3}$)</th>
<th>$K_{sat}$ (mm/hr)</th>
<th>Bulk Density (gm/cm$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad (SIL)</td>
<td>08.5</td>
<td>33.3</td>
<td>0.204</td>
<td>0.379</td>
<td>0.512</td>
<td>6.30</td>
<td>1.29</td>
</tr>
<tr>
<td>Co (SIL)</td>
<td>17.9</td>
<td>35.1</td>
<td>0.216</td>
<td>0.372</td>
<td>0.498</td>
<td>5.11</td>
<td>1.33</td>
</tr>
<tr>
<td>Dt (SIL)</td>
<td>18.4</td>
<td>26.7</td>
<td>0.174</td>
<td>0.347</td>
<td>0.490</td>
<td>8.11</td>
<td>1.35</td>
</tr>
<tr>
<td>Gs (SIL)</td>
<td>40.8</td>
<td>20.6</td>
<td>0.137</td>
<td>0.288</td>
<td>0.459</td>
<td>15.49</td>
<td>4.43</td>
</tr>
<tr>
<td>Df (SICL)</td>
<td>09.5</td>
<td>53.3</td>
<td>0.310</td>
<td>0.432</td>
<td>0.534</td>
<td>2.52</td>
<td>1.24</td>
</tr>
<tr>
<td>Dc (SIC)</td>
<td>18.4</td>
<td>47.6</td>
<td>0.286</td>
<td>0.414</td>
<td>0.507</td>
<td>1.92</td>
<td>1.31</td>
</tr>
<tr>
<td>SCAN</td>
<td>08.7</td>
<td>48.8</td>
<td>0.287</td>
<td>0.426</td>
<td>0.508</td>
<td>1.41</td>
<td>1.30</td>
</tr>
</tbody>
</table>

**Figure 2.2**: Study area and SCAN site (2078) location.

**2.6 Methodology**

The study consisted of the following elements:
• Verification of the entropy soil moisture profiles with all known input parameters for all cases shown in figure 2.1 using observed data from the USDA SCAN site located beside the field.

• Verification of the entropy soil moisture profiles with derived input parameters at the SCAN site.

• Application of the entropy method to simulate a complete irrigation cycle of the center pivot on the field and comparison with a physical model.

The initial task of the entropy validation takes a two-step approach. Initially errors within the POME model are quantified against observed data with all known input parameters. This is equivalent to the laboratory experiments given by Al-Hamdan and Cruise (2010) and Singh (2010a) and can be compared to those results. Then further validations of the model results are done using parameters derived only from weather records. The second stage of validation includes the incorporation of the ET and drainage algorithms to adjust the total mass of water in the soil column. This simulates the information that would be available to the user in near real time. This type of field validation of the method has never been done before. The model performance was evaluated using mean absolute error (MAE) as well as Nash-Sutcliffe \((R^2_{NS})\) efficiency statistics (Nash and Sutcliffe, 1970). The \(R^2_{NS}\) is recommended for hydrologic error analysis by the American Society of Civil Engineers (ASCE) (ASCE, 1993).

\[
MAE(\%) = \frac{[abs(x_{obs} - x_m) \times 100]}{x_{obs}} \quad (2.11)
\]
\[ R^2_{NS} = 1 - \frac{\sum_{1}^{n}(x_{\text{obs}_i} - x_{m_i})^2}{\sum_{1}^{n}(x_{\text{obs}_i} - \bar{x}_{\text{obs}})^2} \]  \hspace{1cm} (2.12)

here \( x_{\text{obs}} \) and \( x_{m} \) are the observed and modeled values. The \( R^2_{NS} \) values ranges from \(-\infty \) to 1, where 1 represents a perfect fit between model and observed values. Any negative result indicates that the mean would have been the better representation for the observed data whereas a positive result indicates some model skill. The more positive the value, the more accurate the model predictions.

Once the POME model was validated, it was applied to evaluate the continual soil moisture states throughout a specific irrigation strategy (center pivot in this study). Center pivots operate by rotating a boom about a hub located near the center of the field (Figure 2.2 background). Water enters the boom through the hub and flow rates and pressures can be regulated via a computerized control system. Most fields will contain multiple soil types (as shown for our test field in Figure 2.2) and the boom must stay in place until the soil with the maximum water deficit is irrigated to the desired degree.

The level at which the soil should be irrigated is the vital question to be answered. On the one hand, irrigation should not be applied in excess of field capacity lest the water drain out of the soil column and be lost, while on the other hand sufficient water should be applied to supply plant requirements. The plants use water not only in the photosynthetic process, but they also sequester water in their biomass. Traditionally, irrigation strategies depend on estimated water loss due to climatological estimations of ET or simply examining the surface soil moisture by
The optimum irrigation strategy would be to bring the root zone to its field capacity without exceeding it.

One way to accomplish this would be to slightly exceed field capacity in the preceding layers of the soil profile while approaching moisture levels slightly below field capacity in the subsequent layers. This allows the excess water in the preceding layers to drain into the subsequent layers thus raising their water content to near field capacity while minimizing the possibility that water will drain out of the soil altogether. This does, however, increase the potential that the plant could become stressed briefly in the upper layers due to over saturation before drainage becomes effective. Evaluation of this strategy would be impractical using soil moisture probes. A physical model would need to be calibrated for each soil type and horizon encountered. However, computation of the complete vertical soil moisture profiles by POME offers an ideal guide for monitoring the moisture in each soil layer.

2.6.1 Irrigation Phase Simulation

The main purpose of this study is to assess the applicability of the entropy approach to evaluate soil moisture profiles during an irrigation cycle (start of irrigation through the drying phase). The POME model was run to simulate soil moisture states during an irrigation event as the sprinkler boom cycled across the field. The irrigation phase is represented by the POME wetting front (Equation 2.8), whereas the succeeding drying phase (through ET) is represented by the dynamic case runs (Equations 2.8 and 2.9 applied successively). Irrigation is applied to each soil until the moisture content at the bottom of the root-zone shows an increment from its initial
state indicating that the water has drained to that level. Initial surface moisture content for each soil is assumed at 50% of the available water capacity (field capacity minus wilting point). Further, the vertical soil column is assumed to be approximately uniform with each layer at similar soil moisture contents initially. The initial surface values are presented in Table 2.4 in the results section.

During irrigation, the sprinkler spraying rate is variable and is kept consistent with the infiltration capacity of each soil type to being sprayed to limit any possible water loss as surface runoff. Infiltration capacity is calculated using the Green and Ampt (1911) infiltration equation. The wet case POME model was applied initially at the top layer (15 cm) assuming the surface at saturation and the moisture content based on the initial profile. The mean moisture content is calculated as: 

\[
\text{mean moisture content} = \frac{W_D - W_1}{T_{\text{depth}} - L_{\text{depth}}}.
\]

Where, \(W_D\) and \(W_1\) are the mass of water content (in units of length) in current soil column depth (D) and surface, respectively; \(T_{\text{depth}}\) and \(L_{\text{depth}}\) are total and layer depths. If the total mass of water applied is within 1% of the mass of water obtained through the POME generated soil moisture profile, then time is incremented (1 minute). Else, if mass balance cannot be achieved within 1%, then this is indicative of irrigation water reaching the next vertical layer and therefore this subsequent layer is added to the analysis and the POME model is reapplied until the total mass balance error falls within 1% for both layers. This process is repeated, adding each layer successively, until the moisture content at 100 cm demonstrates an increment. The complete process of wetting front development using the POME model is shown through the flow chart in figure 2.3.
2.6.2 Drying Phase Simulation

As soon as irrigation ceases, the boom changes position and the drying phase begins. Since the processes associated with this phase (ET and drainage) occur more slowly, the time step is changed to daily, particularly since there is no effective means of computing hourly ET estimates. As shown in Figure 2.1, immediately after a rainfall/irrigation event, soil moisture profiles usually are not monotonic and tend to develop an inflection point (dotted line in Figure 2.1).

As mentioned previously, the POME model requires three inputs: the upper and lower boundary conditions (surface and bottom layer soil moisture contents) and the total water in the soil column for calculation of the mean constraint. As in the wetting case, mass balance was allowed to determine the lower boundary value. This also allowed the calculation of the volume of water that would reach the boundary and
thus possibly be lost to the system (non-beneficial use of water). The surface moisture value is adjusted daily via a drainage calculation based on an exponential decaying function suggested by Neitsch et al. (2011) where the decay factor is a function of the average water content of the particular layer.

Plant transpiration and soil evaporation were calculated to decrease the daily total mass of water in the soil column. The standard United Nations Food and Agricultural Organization Drainage Paper 56 method which employs a standardized version of the Penman-Monteith approach was used and adjusted to correct for actual ET due to limiting soil moisture. The correction was via an exponential function proposed by Poulovassilis et al. (2001) where the decay rate is a function of average moisture content and PET. Wu et al. (1999) showed for mature corn crops nearly 90% of the roots are distributed within top 60 cm depth. For simplification, the ET is uniformly distributed among the top 60 cm of the soil column.

In past work, no method of determining the inflection point on the dynamic profile has been presented. Therefore, a method needed to be developed for this exercise. In order to find the inflection point on the POME moisture profiles, an approach based on redistribution of water above field capacity was adopted. It was reasoned that the inflection point would most likely initially occur at field capacity since vertical drainage would cease at that point although ET would still be withdrawn during subsequent time periods. Figure 2.4 demonstrates this point at the USDA SCAN site where soil moisture profiles are shown based on several different amounts of applied water. In each case, moisture must be redistributed between the upper and lower soil layers until the inflection point (e.g., field capacity) is achieved. The
dots on the curves mark the estimated inflection points with the caveat that in some cases a single inflection point was not obvious so a range is indicated. The method was tested against observed data and the results given later in this presentation.

The same methodology as described above was used to derive the inflection points for the field soils. For each soil extant in the test field, the ET and drainage algorithms described above were employed for various amounts of applied water until field capacity was obtained for each soil. Then, in order to simulate the drying phase, the wet model (equation 2.8) was applied initially from the surface to the point of inflection, and then Equation-2.9 was applied from the inflection point to the bottom of the root-zone at the daily time step. The overall water content was kept consistent with mass balance and the process was repeated in a circular fashion throughout the field.

**Figure 2.4:** Results of derived inflection points at multiple applied water amounts at SCAN site the using moisture redistribution technique
2.7 Results and Discussions

2.7.1 Varification of Entropy Profiles

2.7.1.1 SCAN site varification

The first step in the study was to verify the entropy profiles against the SCAN site observations. The SCAN site provides hourly data at depths of approximately 5, 10, 20, 50 and 100 cm. Although Al-Hamdan and Cruise (2010) and Singh (2010a, 2011) provided extensive verification of entropy-based profiles in a laboratory setting whereas Mishra et al. (2013) applied POME generated profiles for crop yield estimations, yet no verification has been done up to this time using actual field soil moisture data. The verifications were done for all possible cases.

The first case represents a wetting event beginning at 16:00 hours on Julian day 101 in the year 2013 when 26.4 mm of precipitation fell over a period of 12 hours. The simulations continued for 11 hours during the day. Typical resultant profiles are shown in Figure 2.5. The results show that a wetting front developed at a depth of 10 cm in the third hour of the event and remained at that depth thereafter. The entropy simulations were run as the simple wetting case, i.e., (Equation-2.8) with no wetting front development. Even so, the average error in effective soil moisture over the 11 simulations was only 2.21% indicating that even in this simplified representation the entropy method gave very accurate results. The $R_{NS}^2$ was computed to be 0.86 indicating that the POME model profile results represents the observed values far better than would have been represented by just assuming one layer value equivalent to the observed mean.
The next SCAN site simulation represents the drying case beginning Julian day 87 (2012). Daily simulations were run in this case over a 9 day period when no precipitation occurred. Incidentally, the daily time step precludes the development of the dynamic moisture profile discussed above. Typical results are shown in Figure 2.6. Again, the entropy simulations appear to be very accurate, resulting in an average error over the 9 days of only 0.75% in effective soil moisture. The mean $R^2_{NS}$ efficiency value for the dry case simulations were found to be 0.95 again indicating an extremely good representation of dry case observed values by the POME model. Relatively fewer errors in dry case simulation can be attributed to the fact that dry case profiles did
not develop any noticeable inflection point during the simulation period and remained monotonic.

Figure 2.6: Typical dry case POME based profiles against SCAN site observed data

The last simulations of the SCAN site was run to verify the dynamic case profiles. Dynamic profiles are characterized by the presence of at least one prominent inflection point. The SCAN site developed such profiles between Julian days 29-40 (2013). The observed data showed an inflection point was consistently present at 20 cm. In the first simulation, the POME model was applied assuming this was known, the case corresponding to the laboratory experiments reported by Al-Hamdan and Cruise (2010) and Singh (2010a). A dry case POME was run initially through 0-20 cm and then a wet case subsequently through the remaining 80 cm (20-100 cm). The
dynamic case profiles resulted in mean absolute error of 2.89% in effective soil moisture as compared to the SCAN site observed values. This compares very favorably to the errors reported earlier by Al-Hamdan and Cruise (2010) and Singh (2010a) which averaged 2.1%. The $R^2_{NS}$ efficiency statistic available was not calculated for this case since, three out of five values available for comparison were provided as input. Typical profiles from the dynamic case are presented in such simulations are shown in Figure 2.7. A summary of the entire validation procedure can be found in Table 2.2.

![Figure 2.7: Dynamic case POME based profiles against SCAN site observed data](image-url)
Table 2.2: Summary of POME based soil moisture profile results for all possible three cases as compared to the observed values at SCAN site.

<table>
<thead>
<tr>
<th>Time/day</th>
<th>10 cm</th>
<th>20 cm</th>
<th>50 cm</th>
<th>R²</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Xₓobs</td>
<td>Xₓm</td>
<td>MAE</td>
<td>Xₓobs</td>
<td>Xₓm</td>
</tr>
<tr>
<td>Wet</td>
<td>16:00</td>
<td>0.722</td>
<td>0.757</td>
<td>4.75</td>
<td>0.730</td>
</tr>
<tr>
<td>case</td>
<td>17:00</td>
<td>0.826</td>
<td>0.817</td>
<td>1.14</td>
<td>0.730</td>
</tr>
<tr>
<td>(50cm</td>
<td>18:00</td>
<td>0.896</td>
<td>0.838</td>
<td>6.47</td>
<td>0.735</td>
</tr>
<tr>
<td>Depth</td>
<td>19:00</td>
<td>0.898</td>
<td>0.842</td>
<td>6.19</td>
<td>0.764</td>
</tr>
<tr>
<td></td>
<td>20:00</td>
<td>0.896</td>
<td>0.853</td>
<td>4.74</td>
<td>0.787</td>
</tr>
<tr>
<td></td>
<td>21:00</td>
<td>0.935</td>
<td>0.910</td>
<td>2.72</td>
<td>0.863</td>
</tr>
<tr>
<td></td>
<td>22:00</td>
<td>0.935</td>
<td>0.897</td>
<td>4.08</td>
<td>0.891</td>
</tr>
<tr>
<td></td>
<td>23:00</td>
<td>0.928</td>
<td>0.886</td>
<td>4.48</td>
<td>0.886</td>
</tr>
<tr>
<td></td>
<td>00:00</td>
<td>0.923</td>
<td>0.875</td>
<td>5.13</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td>01:00</td>
<td>0.923</td>
<td>0.867</td>
<td>6.02</td>
<td>0.878</td>
</tr>
<tr>
<td></td>
<td>02:00</td>
<td>0.920</td>
<td>0.861</td>
<td>6.48</td>
<td>0.876</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.582</td>
<td>0.601</td>
<td>3.26</td>
<td>0.653</td>
</tr>
<tr>
<td>Dry</td>
<td>88</td>
<td>0.562</td>
<td>0.583</td>
<td>3.88</td>
<td>0.636</td>
</tr>
<tr>
<td>Case</td>
<td>89</td>
<td>0.549</td>
<td>0.558</td>
<td>1.58</td>
<td>0.616</td>
</tr>
<tr>
<td>(100cm</td>
<td>90</td>
<td>0.549</td>
<td>0.566</td>
<td>3.05</td>
<td>0.621</td>
</tr>
<tr>
<td>Depth</td>
<td>91</td>
<td>0.547</td>
<td>0.555</td>
<td>1.48</td>
<td>0.616</td>
</tr>
<tr>
<td></td>
<td>92</td>
<td>0.539</td>
<td>0.542</td>
<td>0.47</td>
<td>0.609</td>
</tr>
<tr>
<td></td>
<td>93</td>
<td>0.537</td>
<td>0.535</td>
<td>0.41</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>0.532</td>
<td>0.524</td>
<td>1.55</td>
<td>0.591</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0.524</td>
<td>0.512</td>
<td>2.66</td>
<td>0.586</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.582</td>
<td>0.601</td>
<td>3.26</td>
<td>0.653</td>
</tr>
<tr>
<td>Dynamic</td>
<td>29</td>
<td>0.757</td>
<td>0.727</td>
<td>3.89</td>
<td>–</td>
</tr>
<tr>
<td>Case</td>
<td>30</td>
<td>0.835</td>
<td>0.823</td>
<td>1.38</td>
<td>–</td>
</tr>
<tr>
<td>(100cm</td>
<td>31</td>
<td>0.797</td>
<td>0.778</td>
<td>2.34</td>
<td>–</td>
</tr>
<tr>
<td>Depth</td>
<td>32</td>
<td>0.775</td>
<td>0.748</td>
<td>3.53</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>0.759</td>
<td>0.733</td>
<td>3.67</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>0.759</td>
<td>0.734</td>
<td>3.41</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.747</td>
<td>0.715</td>
<td>4.29</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>0.752</td>
<td>0.719</td>
<td>4.35</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>0.747</td>
<td>0.711</td>
<td>4.82</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>0.738</td>
<td>0.757</td>
<td>2.63</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.738</td>
<td>0.754</td>
<td>2.22</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>0.735</td>
<td>0.750</td>
<td>2.01</td>
<td>–</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.757</td>
<td>0.727</td>
<td>3.89</td>
<td>–</td>
</tr>
</tbody>
</table>

2.7.1.2 Entropy Validation with Derived Parameters

In the previous validation of the dynamic case, the necessary boundary conditions (surface and bottom soil moisture content) as well as the additional input
parameter for the dynamic case (the location and moisture content of the inflection point) were assumed known. In a more realistic scenario, the validation was done by estimating these parameters by the methods described above. This simulates the real world case where these values would necessarily be unknown to the user. These profiles were then compared with observed SCAN site moisture values to test the robustness of the inputs.

The first step is to select the inflection point on the profiles. Since the assumption is that each soil will have an inflection point located at the point where field capacity first occurs, then the calculations that led to Figure 2.4 were consulted. The figure shows that for the soil at the SCAN site, the inflection point can vary from 15 to 35 cm depth depending on the amount of applied water. So, following the premise that the user would have minimum prior information, a mean value of 25 cm was selected as the inflection point depth for these simulations. As discussed, the moisture content at the inflection point was assumed to be field capacity initially. The ET and drainage computations previously discussed were used to compute the surface and mean moisture values. While the lower boundary is adjusted if required to keep the mean water content within the upper and lower boundary conditions and the location of the inflection point was kept constant.

Typical results of the simulated profiles are shown in figure 2.8 and a complete error analysis is given in Table 2.3. During the 10 days of simulation, the individual layer mean absolute errors ranged from 0.22 to 8.6% with mean error of about 3.8%. It can be noted that the actual observed inflection point was 20 cm (compared to the computed of 25) but the difference in inflection point depths still results in acceptable
errors across the distribution as a whole. Further, the surface value MAE was observed to be 3.8% with the highest errors observed at 6.16% and 5.03% on days 37 and 40 respectively. The higher errors are, in part, due to extremely low ET estimations due to slight precipitation observed on 36 (0.15 cm) and 39 (0.4 cm). Due to this, the surface moisture content observed by moisture probes showed an increase on days 39 and 40 however, the POME model did not catch this since it was applied with no additional precipitation/irrigation information. The $R^2_{NS}$ efficiency statistic was computed to be 0.93, verifying that the derived input parameters of the POME model are sufficient in capturing the distribution of the observed soil moisture profile.

![Figure 2.8](image)

**Figure 2.8**: Representative dynamic case POME based soil moisture profiles with derived inputs parameters against SCAN site observed data
Table 2.3: Summary of POME based soil moisture profile results for dynamic case against SCAN site observed data with derived inflection point as well as inflection moisture content.

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>SCAN SITE - 2078</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>0.733</td>
</tr>
<tr>
<td>10</td>
<td>0.797</td>
</tr>
<tr>
<td>20</td>
<td>0.823</td>
</tr>
<tr>
<td>50</td>
<td>0.584</td>
</tr>
<tr>
<td>100</td>
<td>0.548</td>
</tr>
<tr>
<td>5</td>
<td>0.747</td>
</tr>
<tr>
<td>10</td>
<td>0.803</td>
</tr>
<tr>
<td>20</td>
<td>0.839</td>
</tr>
<tr>
<td>50</td>
<td>0.618</td>
</tr>
<tr>
<td>100</td>
<td>0.554</td>
</tr>
</tbody>
</table>

| 5         | 1.95 | 2.98 | 2.96 | 3.25 | 3.90 | 5.10 | 6.16 | 4.34 | -2.15 | -5.03 |
| 10        | 0.79 | 2.18 | 3.49 | 2.71 | 3.68 | 2.64 | 2.00 | 1.95 | 1.95  | 2.28  |
| 20        | 1.98 | 4.10 | 4.34 | 4.45 | 4.38 | 3.61 | 3.12 | 2.87 | 3.24  | 3.56  |
| 50        | 5.82 | 5.19 | 7.13 | 7.67 | 7.44 | 8.05 | 6.06 | 7.76 | 7.76  | 8.63  |
| 100       | 1.11 | 0.22 | -1.03| -2.29| -1.87| -2.77| -2.35| -3.26| -2.60 | -3.45 |
| Mean      | 2.33 | 2.93 | 3.38 | 3.16 | 3.51 | 3.32 | 3.00 | 2.73 | 1.63  | 1.20  |

2.7.2 Application of the POME model to an irrigation cycle

The POME model was used to simulate a complete irrigation cycle for the test field shown in Figure 2.3. As shown in the figure, at any position, the boom will apply water to multiple soil types. The spray rate was variable across the boom and was set at the infiltration capacity of each soil to avoid producing surface runoff. The boom remained in this position until all the soils received the optimal amount of water measured as when the moisture profile reached the bottom layer. Not only was surface runoff avoided, but another requirement was to not let excess water drain from the soil to pollute groundwater. Thus, the algorithm was run, allowing each vertical layer to become irrigated in turn, until the mass balance operation required the lower
boundary to be increased, thus indicating that water had reached the bottom of the soil. At that point, the irrigation of the particular soil ceases, i.e. the spray stops; however, the boom remains in position until the last soil is irrigated. Table 2.4 shows the rate of irrigation applied as well as the total time required for irrigation to reach the 100 cm depth for each soil type. The 75 ac field is fully irrigated with 120 radial boom movements. Sprinkler boom movement is governed by the soil with the slowest infiltration rate within the boom range. Out of total 120 boom movements, 50 times the Dc soil with maximum irrigation time (141 min) was covered and the other 70 times the second slowest infiltrating soil Df (115 min) was covered. This results in a total of 10.5 days (24x7) to irrigate the entire field.

The simulation assumes a corn (maize) crop at or near maturity during the growing season in North Alabama. A total of nearly 14.7 million liters of water was required for a complete simulated irrigation cycle. Figure 2.9 shows typical soil moisture profile development during irrigation phase simulations.

### Table 2.4: Summary of wetting front simulations

<table>
<thead>
<tr>
<th>Soil Type</th>
<th>Irrigation Rate (cm/min)</th>
<th>Total time to Saturate (min)</th>
<th>Effective SM</th>
<th>Total Water (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Surface Initial/Final</td>
<td>Bottom Initial/Final</td>
</tr>
<tr>
<td>Ad (SIL)</td>
<td>0.104</td>
<td>56 (10)</td>
<td>0.55/1.0</td>
<td>0.57/0.625</td>
</tr>
<tr>
<td>Co (SIL)</td>
<td>0.104</td>
<td>58 (10)</td>
<td>0.57/1.0</td>
<td>0.55/0.632</td>
</tr>
<tr>
<td>Dt (SIL)</td>
<td>0.102</td>
<td>58 (11)</td>
<td>0.52/1.0</td>
<td>0.55/0.584</td>
</tr>
<tr>
<td>Gs (SIL)</td>
<td>0.101</td>
<td>59 (13)</td>
<td>0.44/1.0</td>
<td>0.47/0.513</td>
</tr>
<tr>
<td>Dc (SIC)</td>
<td>0.019</td>
<td>141 (40)</td>
<td>0.65/1.0</td>
<td>0.66/0.665</td>
</tr>
<tr>
<td>Df (SICL)</td>
<td>0.035</td>
<td>115 (20)</td>
<td>0.71/1.0</td>
<td>0.68/0.720</td>
</tr>
</tbody>
</table>

Once irrigation ceases for a particular area, the drying phase begins at a daily time-step. The drying phase is simulated through the dynamic POME model. For
soils that develop an inflection point, the depth and moisture content at that point are additional derived input requirements. By performing the simple mass distribution exercise demonstrated earlier for the SCAN site, inflections points were easily identified. Inflection points were identified for 4 out the 6 simulated soil types. The four soils that developed inflection points tended to have higher hydraulic conductivities and lower field capacities than those that did not, further reinforcing the derivation technique.

The field is divided into 120 radial segments each representing a sprinkler boom spatial extent. Given the 10.5 days required to irrigate the field, then about 11 days of drying will occur for the initially irrigated segments before the boom would return to its initial position. Thus, the drying phase simulations (dynamic case POME) were run for 11 days. For example, Figure 2.10 shows the fraction of available water content ($\theta_{AWC}$) for day’s 3, 6, 9, and 11 at multiple depths. The portion of the field

---

**Figure 2.9:** Typical wet case soil moisture profile results during irrigation phase simulation a) Ad- Silt loam soil; b) Dc - Silty clay soil
currently under irrigation usually had moisture content above field capacity (> 100% of AWC) at the surface. As clear from the figure 2.10 with the 9 days of drying none of soils at any depth showed sign of water stress. Soil water stress impact on plants is extremely specific to plant type and it’s growth stage (Çakir, 2004). In this study a rather simple and conservative definition of water stress is used based on Luo et al. (2008) as moisture content less than 20% of AWC. While only 0.73 ac (0.97 % approx.) of the total field showed signs of stress with 9 days of irrigation by the end of 11 days of irrigation nearly 4 ac (∼5.16%) of the field begins to experience soil moisture stress at the surface. Out of which, Dc soil with maximum 11 days of drying at surface had the maximum stress of 1.17% of AWC. While actually only 3.9 ac of the field area had water content less than 20% of AWC there is another 7.8 ac of the field just outside of water stress limit (20-30% of AWC) by the end of irrigation cycle. However at depths 20 cm and below, no soil at any field portion showed water stress condition. Table 2.5 shows the summary of the drying phase results at the end of each day. The above results depend on the initial position of the boom itself. If the boom was started from the opposite side of the field, a different portion of the field would be under drying for 11 days.

2.7.3 Comparision to Physical Model

The previous validation of the POME model employed fairly simple observed profiles at the USDA SCAN site. It is known that in some cases, field soils can develop soil moisture profiles that are much more complex than those observed at the SCAN site. In addition, it is desirable to have some basis of comparison of the proposed
Table 2.5: Summary of drying phase simulations: IP is the inflection point depth (cm); A is surface effective SM; B is effective soil moisture content at IP; C is bottom (100 cm) effective soil moisture content; and D is total amount of water (cm) contained in profile. E is the actual evapotranspiration in cm of water per day.

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>200</th>
<th>201</th>
<th>202</th>
<th>203</th>
<th>204</th>
<th>205</th>
<th>206</th>
<th>207</th>
<th>208</th>
<th>209</th>
<th>210</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0.80</td>
<td>0.72</td>
<td>0.69</td>
<td>0.66</td>
<td>0.65</td>
<td>0.63</td>
<td>0.62</td>
<td>0.59</td>
<td>0.56</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>B</td>
<td>0.70</td>
<td>0.69</td>
<td>0.67</td>
<td>0.65</td>
<td>0.64</td>
<td>0.63</td>
<td>0.62</td>
<td>0.62</td>
<td>0.64</td>
<td>0.62</td>
<td>0.61</td>
<td>0.60</td>
</tr>
<tr>
<td>C</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.57</td>
</tr>
<tr>
<td>D</td>
<td>35.3</td>
<td>34.3</td>
<td>33.8</td>
<td>33.5</td>
<td>33.0</td>
<td>32.8</td>
<td>32.5</td>
<td>32.2</td>
<td>31.7</td>
<td>31.2</td>
<td>30.7</td>
<td>30.3</td>
</tr>
<tr>
<td>E</td>
<td>–</td>
<td>0.56</td>
<td>0.60</td>
<td>0.53</td>
<td>0.42</td>
<td>0.25</td>
<td>0.27</td>
<td>0.32</td>
<td>0.52</td>
<td>0.48</td>
<td>0.45</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Method on the actual test field. To that end, a complex physically based mathematical model of soil moisture movement was constructed for the test field. The details of this model are provided in the Appendix. In the model, the mass distribution of water contained in the soil is a function of drainage, ET and its partition as well as its distribution through the column. Evaporative losses are partitioned between plant
transpiration and bare soil evaporation. A root distribution for maize commonly used in crop models was employed in the mathematical model. Inflection points develop naturally as a function of plant root uptake and vertical drainage.

**Figure 2.10**: Field map with soil moisture conditions on day 3; 6; 9 and 12 at surface; 20cm; 45cm and 60cm depths
The first goal of any model is to conserve mass. The overall mean MAE for all soils over the complete drying phase simulation was -1.9 to 1.7% \((R^2 > 0.98\) with physically-based mathematical model). The Mass balance closure criteria of 1% and round-off errors can aggregate to slightly increased total error. However, it appears that the simplified method of estimating \(T_p\) in the entropy model did not unduly bias the method in comparison to the model. Figure 2.11 shows the range of mean mass errors observed by each soils over the 11 days of simulation against the model.

![Figure 2.11: Total mass error of the POME model against the model](image)

After ensuring overall mass balance, the next comparison is by individual soil layers. This would enable quantification of errors associated with derived inflection points; assumption of field capacity at inflection points and distribution of \(T_p\) and drainage through the soil column. Figure 2.12 shows some typical profile comparisons of the entropy model vs the model for each soil type at various stages of an irrigation cycle.
The figure demonstrates how more complex profiles are possible for these soils than were exhibited by the SCAN site data. The combination of the bare soil evaporation routine and the transpiration associated with the root zone distribution resulted in numerous profiles with multiple inflection points that were not observed at the SCAN site and not assumed by the entropy model. Still, the figure shows that, although only one, or in some cases no inflection points were assumed by the entropy model, the model still came close to resolving the profile in each case even though it did not know the profile in advance.

An overall error analysis of the profile comparisons is given in the Box and Whisker plot shown in Figure 2.13 which summarizes all errors across all layers of the model. Overall mean absolute error for all soils together at each layer was about 2.95% (max = 14.87% by Gs soil on day 11 at surface). Maximum errors were observed at the surface and around inflection points. Mean error at the surface was found to be 5.8% whereas errors at the inflection point and 5 cms around it averaged 4.11%. These errors can be mainly attributed to the differences in moisture content due to the development of additional inflection points (see Table 2.6). Another source of error was associated with the surface and bottom moisture content. While discrepancy at the surface moisture content is again due to the complexity of the physical model which included crop leaf abstractions/shadow effects on soil evaporation and a complex root distribution when, of course, these issues were not a part of the entropy model. Whereas, the entropy model’s deviation of the bottom moisture content from the mathematical model results from error at the surface and inflection points being
Figure 2.12: Typical soil moisture profile comparisons between physical model and entropy generated profiles for days 1, 5 and 11 of an irrigation cycle. Vertical dashed lines represents the soil specific effective wilting point (left) and field capacity (right) positions.
carried over to the bottom in order to guarantee conservation of mass balance over the entire profile.

Figure 2.13: Box Plot of errors of the POME model against the mathematical model at layer depths.

2.8 Conclusion

The purpose of the exercise described here was to validate the POME soil moisture model against observed field data, compare POME to the results of a com-
Table 2.6: Inflection point development results from physical model run and moisture redistribution technique

<table>
<thead>
<tr>
<th>Soil Type</th>
<th>Physical model Result</th>
<th>Redistribution Technique Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First 2 prominent inflection point depths (cm) (initial)</td>
<td>First 2 prominent inflection point depths (cm) (final)</td>
</tr>
<tr>
<td>Ad (SIL)</td>
<td>20, 00</td>
<td>15, 25</td>
</tr>
<tr>
<td>Co (SIL)</td>
<td>25, 00</td>
<td>15, 25</td>
</tr>
<tr>
<td>Dt (SIL)</td>
<td>30, 00</td>
<td>15, 35</td>
</tr>
<tr>
<td>Gs (SIL)</td>
<td>40, 00</td>
<td>15, 40</td>
</tr>
<tr>
<td>Dc (SIC)</td>
<td>00, 00</td>
<td>15, 00</td>
</tr>
<tr>
<td>Df (SICL)</td>
<td>00, 00</td>
<td>15, 00</td>
</tr>
</tbody>
</table>

plex physically-based mathematical model, and to demonstrate how POME can be used to monitor and guide an irrigation system with relative accuracy and ease. The comparison of POME to field and model data was discussed above and the errors enumerated. The third purpose of the study will be discussed in terms of irrigation efficiency measures and development of potential plant stress or lack thereof. This discussion revolves around the amount of water added to the field and how that water was used, or wasted, during the irrigation cycle. This information is summarized in Table 2.7 below.

Table 2.7: Summary of irrigation efficiency observed by the POME (in cm of water)

<table>
<thead>
<tr>
<th>Soil Type</th>
<th>Water added</th>
<th>Total ET</th>
<th>ΔS</th>
<th>Δ(S + ET)</th>
<th>ΔSM</th>
<th>ΔSM100cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad (SIL)</td>
<td>5.8</td>
<td>4.81</td>
<td>0.8</td>
<td>5.61</td>
<td>0.19</td>
<td>0.137</td>
</tr>
<tr>
<td>Co (SIL)</td>
<td>6.0</td>
<td>4.75</td>
<td>0.9</td>
<td>5.65</td>
<td>0.35</td>
<td>0.198</td>
</tr>
<tr>
<td>Dt (SIL)</td>
<td>5.6</td>
<td>5.22</td>
<td>0.3</td>
<td>5.52</td>
<td>0.08</td>
<td>0.081</td>
</tr>
<tr>
<td>Gs (SIL)</td>
<td>5.6</td>
<td>5.25</td>
<td>0.2</td>
<td>5.45</td>
<td>0.15</td>
<td>0.096</td>
</tr>
<tr>
<td>Dc (SIC)</td>
<td>2.7</td>
<td>4.99</td>
<td>-2.1</td>
<td>2.89</td>
<td>-0.19</td>
<td>0.011</td>
</tr>
<tr>
<td>Df (SICL)</td>
<td>3.9</td>
<td>4.65</td>
<td>-0.9</td>
<td>3.75</td>
<td>0.15</td>
<td>0.095</td>
</tr>
</tbody>
</table>
The Table shows the amount of water added to each soil class, the actual evapotranspiration for that soil, the change of soil moisture from initial (beginning of irrigation) state, and the volume of water that drained to the bottom of the soil profile. It should be noted that these data relate to the soil moisture column as a whole and not to any particular layer. As discussed in the introductory material, ET demand and enhanced soil moisture states are considered beneficial uses of irrigation water, while surface runoff or vertical drainage to groundwater are not. In this demonstration, the irrigation was applied in a manner to preclude surface runoff, so the principle non-beneficial use of the water would be drainage. In this regard the table shows that a relatively small percentage of the applied water even reached the bottom of the profile (average of 1.9%). Otherwise, the added water was used to supply ET demands and enhance soil moisture status.

The Table demonstrates how the finer texture, silty clay soils (Dc and Df) responded to the irrigation in a fundamentally different manner than did their coarser brethren. Since these soils have a smaller water holding capacity, they absorbed considerably less water than the other soils present in the field. However, since the ET demand is set primarily by weather conditions, ET did not vary much over the field so that the available soil moisture was drawn down in those two cases. The ET and SM demands were met fairly well in the case of the Df soil, with even a slight overall surplus at the end, while the crops residing on the Dc soil would not fare as well. The overall negative water balance in this case indicates that the soil moisture fell below 50% of the available water capacity (the initial soil moisture condition) considering the column as a whole. In fact, the results indicate that for this particular soil class,
the moisture state overall reached about 31% of the available water capacity of the soil.

The next question is, would this condition stress the maize plants? Figure 2.10 can be used to help answer that question. The figure shows that, neglecting the surface which of course dries fairly quickly in the Southern U.S. in the absence of rain, the moisture state at the 20 cm depth drops to about 20-30% of available capacity by day 11 for both of the soils with finer textures. In fact, this level is considered to be a rather conservative threshold under which maize may start to experience stress by some authors (Luo et al., 2008; Taylor and Ravet). However, the moisture level never drops to that state for any other depths during the irrigation cycle.

Taking all evidence into account, it appears that the POME model trod a fine line between efficiency and stress, particularly for the finer soil classes. However, even these soils did not become stressed until the very end of the cycle, and it is not known what the impacts of this small stress would be on the final yield. A good percentage of the root mass for mature maize plants resides around the 20 cm depth, though, so if this stress level was to be maintained for some time, it would undoubtedly affect yields. How long this stress would be maintained would depend on conditions that prevailed subsequent to this irrigation cycle. In conclusion, this study has demonstrated the utility of the POME model to aid in irrigation scheduling. The POME generated profiles were shown to be good approximations of those actually observed at a site very near the test field and to represent profiles generated from a complex mathematical model with overall errors less than 10%. When applied to irrigation, it led to a very efficient strategy, although perhaps overly efficient for
finer soils with a high percentage of clay content. This is an issue that should be investigated further.
CHAPTER 3

DEVELOPMENT OF SOIL MOISTURE PROFILES THROUGH COUPLED MICROWAVE-THERMAL INFRARED OBSERVATIONS IN THE SOUTHEASTERN UNITED STATES


Abstract

The principle of maximum entropy (POME) can be used to develop vertical soil moisture profiles. The minimal inputs required by the POME model make it an excellent choice for remote sensing applications. Two of the major input requirements of the POME model are the surface boundary condition and profile-mean moisture content. Microwave-based soil moisture estimates from Advanced Microwave Scanning Radiometer (AMSR-E) can supply the surface boundary condition whereas thermal infrared-based moisture estimated from the Atmosphere Land Exchange Inverse (ALEXI) surface energy balance model can provide the mean moisture condition. A disaggregation approach was followed to downscale coarse resolution (∼ 25 km) microwave soil moisture estimates to match the finer resolution (∼ 5 km) thermal data. The study was conducted over multiple years (2006-2010) in the southeastern U.S. Disaggregated soil moisture estimates along with the developed profiles were compared with the Noah land surface model (Noah LSM) within the framework of NASA Land Information System (LIS), as well as in-situ measurements from 10 Natural Resource Conservation Services (NRCS) Soil Climate Analysis Network (SCAN) sites spatially distributed within the study region. The overall disaggregation results at the SCAN sites indicated that in most cases disaggregation improved the temporal correlations with unbiased root mean square errors in the range of 0.01-0.09 in volumetric soil moisture. The profile results at SCAN sites showed a mean bias of 0.03 and 0.05; unbiased RMSE of 0.05 and 0.06; and correlation coefficient of 0.44.
and 0.48 against SCAN observations and Noah LSM, respectively.

### 3.1 Introduction

Although soil moisture (SM) represents a relatively small part of the overall hydrologic cycle, it is perhaps the most important part to human survival. SM is the source of water for all vegetation on Earth. It also plays an important role in water and energy exchanges between the land surface and atmosphere. Hydrologically, SM is an indicator of drought or lack thereof, and antecedent moisture conditions are important determinants of runoff response to rainfall events. Thus, SM is a vital part of any terrestrial ecosystem analysis as well as land surface and climate models.

Much of the recent efforts particularly in remote sensing of SM estimation have been focused on surface or near surface observations (0-5 cm); however, moisture throughout the root zone can be just as prevalent. The moisture within the root zone exerts a controlling influence on land-atmospheric fluxes of energy and water under vegetated condition. The actual distribution of root zone moisture is a function of vegetation canopy root density and distribution (Mishra et al., 2013). For this reason, SM at shallow depths (< 100 cm) is known to be extremely variable both as functions of time (Starks et al., 2003) and depth (Scott et al., 2003).

Although several approaches have been proposed for determining SM profiles, most require either observed profile data so that a regression or inversion model can be developed (Arya and Richter, 1983; Kondratyev et al., 1977; Kostov and Jackson, 1993; Singh, 1988). A common approach is to estimate surface or total
root zone moisture using remote sensing and then assimilate those observations into a land surface model (LSM) to determine root zone SM distributions. The NASA Land Information System (LIS) contains a suite of land surface models and data assimilation tools for this purpose and are commonly utilized as a source of SM data.

However, LSMs have their own issues (e.g., bias, ancillary data requirements, computational expense) so it would be advantageous if SM profiles could be deduced directly from satellite observations without the use of a LSM or the availability of in-situ profile data. In-situ SM profile data are only available generally at a few locations over the continental U.S. for any given period of time. In addition, a number of field campaigns over the years have produced high-density observations, but only for very short time periods. In-situ data suffer from the fact that they are site specific and may not be representative of wider surrounding regions. Thus, they are of limited value for modeling or operational purposes. This deficiency has led to the increased reliance on remote sensing to retrieve SM. However, remotely sensing SM estimates alone cannot deduce the distribution of moisture within a soil column.

Due to the inherent complexities involved with the movement of SM in the column, several studies have argued that SM uncertainties and complexities can be best described through the description of its entropy (Mays et al., 2002; Pachepsky et al., 2006; Singh, 2010a). The maximization of entropy characterizes the diffusion of moisture through the soil column over a period of time. The principle of maximum entropy (POME) states that if the inferences had to be drawn from incomplete information then they should be based on the probability distribution with maximum entropy allowed by the a-priori information. Al-Hamdan and Cruise (2010) used
the maximum entropy formulation of Jaynes (1957) based on the Shannon entropy (Shannon, 1948) to formulate the POME-based SM profile development algorithm. Subsequent to its introduction the POME method has been adopted and extended by several authors [e.g., (Mishra et al., 2013, 2015; Pan et al., 2011; Singh, 2010a)]. Initial studies by Al-Hamdan and Cruise (2010) and Singh (2010a) compared their results against experimental data under laboratory settings. However studies by Pan et al. (2011) and Mishra et al. (2013) involved application and validation of the POME model outside laboratory environment. More recently, Mishra et al. (2015) provided extensive validation of the profiles developed using the POME approach against a U.S. Department of Agriculture Soil Climate Analysis Network (SCAN) site located in northern Alabama, as well as with a detailed mathematical model of moisture movement in the soil profile.

The objective of this study is to develop SM profiles from remotely sensed data over the southeastern U.S without the aid of observed profile data or the use of a LSM. The approach utilizes both microwave (MW) data (to supply surface estimates) and thermal infrared (TIR) estimates (for total root zone moisture) within the POME profile methodology. The POME model requires only the upper and lower boundary conditions, as well as the mean moisture content, as input. The surface and mean moisture contents can be supplied by satellite estimates, whereas the lower boundary condition (∼100-200 cm) is often fairly stable and can be parameterized. This makes the POME modeling approach quite feasible when working with remotely sensed SM datasets.
Within this study, before the SM profiles can be calculated, the disparity in spatial resolution between the MW and TIR data must be resolved. MW data are available at much coarser spatial resolutions (25-40 km) than are TIR data (1-10 km). The approach selected here is to downscale (or disaggregate) the coarse MW data to the resolution of the TIR data. This is accomplished via an evaporative efficiency method proposed by Merlin et al. (2013, 2015, 2012). The spatial resolution selected is 4.7 km (∼5 km hereafter) that corresponds to the operational scale of the NWS Multisensor Stage IV precipitation product (Lin and Mitchell, 2005). This facilitates the future integration of the profiles into operational land surface, hydrologic, or agricultural models. It is quite possible that these models could be improved through assimilation of observed SM profiles, especially in regions of the world where climate information is sparse.

As stated earlier, the overall objective of the study is to determine the efficacy of SM profiles developed directly from remotely sensed data only, without the use of a LSM or ancillary data. The study consists of three parts: (a) a multiyear disaggregation of the coarse resolution MW surface SM to the 5-km spatial resolution; (b) calculation of SM profiles for each 5-km grid using the POME approach with the downscaled MW data serving as the surface boundary condition and TIR estimates providing mean SM; (c) validation of the SM profiles against a gridded LSM and in-situ data; and (d) error analyses including evaluation of downscaled MW surface SM estimates against LSM and in-situ data. Two independent data sources are used for comparison and validation purposes, using ground observations from 10 available
NRCS SCAN sites and gridded 3-km Noah LSM SM data aggregated to the 5-km spatial resolution.

3.2 Study Area and Data Sources

3.2.1 Study Area

The study area for this research is the southeastern U.S. consisting of four states including Alabama, Georgia, Florida and South Carolina (Figure 3.1). The southeastern U.S. represents a subtropical humid climate that typically has relatively hot and humid summers and precipitation that is generally evenly distributed throughout the year. The mean annual precipitation is 1250-1500 mm based on the 1981-2010 period. Mean annual temperature ranges from 14°C in Northern Alabama to nearly 24°C in southern Florida. The region is roughly 31% forested; 54% shrubs; 12% agricultural land and rest of the area is covered by urban (1.9%), savanna (1.8%), water etc. according to Moderate Resolution Infrared Spectroradiometer (MODIS) 2008 land cover data aggregated to 5-km spatial resolution. The majority of the soils (nearly 80%) at the surface are classified as sand with loamy sand and sandy loam, as determined from the Soil Information for Environmental Modeling and Ecosystem Management (Miller and White, 1998). These soils are known to have relatively low water holding capacity that can lead to great temporal variation in upper level (1-10 cm) SM conditions and relatively frequent short-term droughts (1-4 week period) during growing seasons in various parts of the region (McNider et al., 2015). The
Figure 3.1: Overview of study area showing location of all the SCAN sites. The dark blue circles indicate sites with most consistent data availability and are being used for comparison and validation in this study. The right figure shows a land cover map (MODIS-2008) for the study area.

Southeastern U.S. is one of the more data rich regions of the world (climate and soils data) providing ample opportunity for calibration as well as validation of results.

3.2.2 Data Sources

3.2.2.1 Microwave Surface SM

Over the past several years, much attention has been given to the use of MW sensors to measure surface SM remotely. The use of the MW band is the only remote sensing technique that is physically based as well as quantitative (Kondratyev et al., 1977; Schmugge et al., 1992). Furthermore, due to their all-weather and day/night capabilities, MW sensors are widely used globally and offer high temporal data availability. This study employs one of the more extensively used and validated MW-based
SM data sets from the Advanced Microwave Scanning Radiometer (AMSR-E) mission operating in the X-band frequency from 2002-2011. The data were obtained from the National Snow and Ice Data Center (NSIDC) and were generated using the so-called standard NASA retrieval algorithm - an iterative multichannel inversion process to deduce surface moisture conditions through comparison of observed and computed brightness temperatures (Njoku et al., 2003). It is primarily impacted by vegetation cover and water content, as well as soil temperature and moisture (Cho et al., 2015). The daily Level-3 AMSR-E SM X-band product (AELand3) (Njoku, 2004) from the ascending (1:30 pm local time) overpass was collected for this study. The ascending overpass was selected to be consistent with the ALEXI retrievals, which are forced with morning and local noon skin temperatures obtained from the Geostationary Operational Environmental Satellite (GOES) Imager instrument. The Level-3 AMSR-E SM estimate is a 25-km gridded data product.

3.2.2.2 Thermal Infrared - ALEXI

Techniques to retrieve root-zone moisture that rely upon TIR data are inferred from surface energy fluxes typically retrieved at relatively high spatial resolutions. TIR-based evapotranspiration (ET) estimates are generally related to LST and vegetation cover fraction. Models such as the Surface Energy Balance System [SEBS: (Su, 2002)]; the Surface Energy Balance Algorithm for Land [SEBAL: (Bastiaanssen et al., 1998)]; and the Two Source Energy Balance [TSEB: (Norman et al., 1995)] exploit this relationship with varying degree of complexities. A two-source based Atmospheric Land EXchange Inverse (ALEXI) (Anderson et al., 1997, 2007a; Hain et al.,
2011) model has been implemented over the continental U.S. and used as a source of surface energy fluxes (Anderson et al., 1997; Norman et al., 2003); evapotranspiration (ET) (Anderson et al., 2011b, 2007a); SM (Hain et al., 2011; Mishra et al., 2013); and an Evaporative Stress Index (Anderson et al., 2013, 2011a). A continental-scale implementation of the ALEXI model was used in this study to estimate instantaneous energy fluxes. ALEXI fluxes are available at approximately 4.7 km (0.04°) spatial resolution on a daily time-step since the year 2000 over the continental U.S., generated using 15-min resolution GOES 10.7 m channel TIR data. ALEXI estimates of actual ET and SM are used in this study. A known drawback of TIR-based methods is that they are limited to cloud-free conditions.

3.2.2.3 In-situ Observations

The study area contains 25 operational U.S. Department of Agriculture SCAN (Bell et al., 2013; Schaefer et al., 2007) monitoring stations. In addition to meteorological observations such as precipitation, air temperature, relative humidity etc. these monitoring stations measure soil temperature and moisture content primarily at depths of 5, 10, 20, 50 and 100 cm at hourly and daily time steps. The SCAN sites use Hydra Probes (Stevens) to observe SM conditions (Schaefer et al., 2007). Most of these 25 SCAN sites are located in northern and central Alabama. Ten sites with the most consistent data availability and with good geographical distribution across the study area were employed for the comparison. The SM data were obtained from http://www.wcc.nrcs.usda.gov/scan/. Table 3.1 lists the major land cover type (at 5 km scale) along with soil characteristics at these ten sites.
Table 3.1: SCAN site 5 km dominant land cover (MODIS 2008) and soil characteristics (SCAN) at the surface and depth of 100 cm [S-sand; L-loam; Si-silt]

| SCAN Site | Lat/Lon     | Landcover         | Soil Type
|-----------|-------------|-------------------|------------
|           |             | Surface 100cm     |
| 2009      | 30.3/-84.4  | Savannas/Mix Forest S S |
| 2013      | 33.8/-83.4  | Crop/Savannas SL C |
| 2027      | 31.5/-83.5  | Cropland S SL     |
| 2037      | 34.3/-79.7  | Crop/Shrubland - - |
| 2038      | 32.6/-81.2  | Crop/Savannas - - |
| 2053      | 34.9/86.5   | Cropland SiCL SiC |
| 2078      | 34.9/-86.6  | Cropland SiCL C   |
| 2113      | 34.2/-86.8  | Crop/Savannas L SCL |
| 2114      | 32.6/-88.2  | Savannas SCL CL   |
| 2115      | 32.4/-85.7  | Savannas LS SC    |

3.2.2.4 Noah Soil Moisture

The Noah SM product generated with the NASA LIS (Kumar et al., 2006) framework was selected as a comparison dataset. The Noah model is driven by actual meteorological forcing, and thus serves as a valuable comparison dataset by which to measure the MW downscaling and profile results. While Noah SM also has biases and uncertainties, the comparisons reveal regional patterns of agreement (disagreement) with the remote sensing estimates. In the event that the POME profiles prove to be superior to the LSM in certain instances, this would indicate that the LSM (or other hydrologic or agricultural models) might be improved through assimilation of the remotely sensed SM profiles. The comparison assumes that errors in the Noah model are independent from the errors associated with MW and TIR based estimates. Noah SM estimates are available in four layers: 0-10; 10-40; 40-100 and 100-200 cm depths. It should be noted that there are inconsistencies in the surface layer depths.
between Noah and MW data: The surface layer in the Noah model is the top 10 cm of the soil column, while the downscaled MW represents the top 2-2.5cm. The Noah 3-km SM products were aggregated to 5-km to be product consistent with the downscaled MW product.

Additionally, the NLDAS2 gridded temperature forcing data (0.125° resolution) were also utilized for computing of potential evapotranspiration (PET). The air temperature forcing data was available from NASA Land Data Assimilation System (NLDAS2) (https://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php). The GTOPO30 digital elevation model (DEM) was used as source of elevation information for the study area. The GTOPO30 product was made available by the U.S. Geological Surveys EROS Data Center (https://lta.cr.usgs.gov/GTOP30). The 1-km gridded soil characteristic data for the study area was available from the Soil Information for Environmental Modeling and Ecosystem Management (Miller and White, 1998).

3.3 Methodology

3.3.1 ALEXI Retrievals

3.3.1.1 Surface Evaporation

A time differential application of ALEXI to monitor 10.7 m brightness temperatures that constitute the land surface temperature (LST) rise, specifically from morning to local noon which are used to diagnose the partitioning of net radiation into sensible; latent and soil heat fluxes. The rise in LST from morning to near-noon is known to be correlated with the moisture content of the soil: compared to a dry
land surface, wetter surfaces warm slowly, thus requiring more energy for evaporating surface moisture (Hain et al., 2011; Kustas et al., 2001). The soil heat conduction flux is parameterized as a function of net radiation following (Santanello and Friedl, 2003); latent heat from the canopy (transpiration) is estimated assuming a non-stressed modified Priestley-Taylor (Priestley and Taylor, 1972) approach. Finally, the soil (surface) latent heat is the residual of the canopy latent heat and latent heat of the soil and canopy system: \( LE_s = LE_{sys} - LE_c \). Here \( LE_s, LE_{sys} \) and \( LE_c \) represent the latent energy of surface, system and canopy, respectively. Detailed model description and derivation is provided in earlier studies (Anderson et al., 2007a; Hain et al., 2011). If the residual is negative [an indicator of condensation, an unlikely process during daytime (Hain et al., 2011)] then the canopy transpiration is relaxed iteratively until it reaches zero. The surface evaporation from ALEXI is used to compute the soil evaporative efficiency (SEE) function required for the disaggregation (described in section 3.3.2).

3.3.1.2 Mean Rootzone Moisture Retrieval

The ratio of actual to potential ET \( (f_{PET}) \) is functionally related to the fraction of available water \( (f_{AW}) \). Multiple relationships between the ratios of PET and available water have been proposed with varying degrees of success including linear; non-linear; piecewise linear or threshold (Hain et al., 2009). Large-scale applications prefer simpler linear functions as sensitivity to SM is constant and thus relatively less detailed soil characteristics are required (Song et al., 2000). In this study a linear
relationship proposed by Wetzel and Chang (1987) is employed: \( f_{PET} = 0.85 \times f_{AW} \).

The resulting ALEXI SM estimation is given as:

\[
\theta_{ALEXI} = (\theta_{fc} - \theta_{wp})(0.85 \times f_{AW}) + \theta_{wp}
\]  

(3.1)

Here \( \theta_{fc} \) and \( \theta_{wp} \) represent the field capacity and wilting point of the soil, respectively. It is argued that the SM retrieval from diagnosed evaporative fluxes is reasonable when the SM content is within the limits of wilting point and field capacity (Hain et al., 2011). ALEXI retrievals can be interpreted based on fraction of vegetation cover \( (fc) \) as either surface moisture content \( (fc < 0.3) \); predominantly root-zone moisture \( (fc > 0.75) \) or a composite of both surface and root-zone moisture for \( fc \) between these limits. In this study Priestly-Taylor PET was used with ALEXI actual ET to compute \( f_{AW} \).

3.3.2 Surface Disaggregation

The spatial resolution of the TIR-based ALEXI SM estimates are roughly 5x5 km². Thus, in order to utilize them in conjunction with the AMSR-E MW data, the coarse resolution MW surface estimates must be downscaled to match the ALEXI spatial scale. A physically based, semi-empirical soil evaporative efficiency (SEE) model in combination with a first order Taylor series expansion around the coarse resolution SM is used to map surface evaporative fluxes to SM content at finer resolutions.
The SEE disaggregation approach has become very popular recently and has been employed by several investigators at varying spatial scales and locations such as: Chen et al. (2017) ($r$: -0.3-0.72, RMSE: 0.06-0.27); Malbêteau et al. (2016) ($r$: 0.70-0.94, RMSE: 0.07-0.09); Merlin et al. (2015) ($r$: -0.22-0.64, RMSD:0.05-0.32); Molero et al. (2016) ($r$: 0.35-0.47, ubRMSE:0.04-0.12). In general, the disaggregation improves agreement with in-situ observations in comparison with coarse-scale estimates.

The disaggregation approach decouples the soil evaporation from the top few centimeters of the soil and the vegetation transpiration through ET partitioning. The disaggregation algorithm used in this study follows the concept of the DISaggregation based on Physical and Theoretical scale CHange [DISPATCH: (Merlin et al., 2013, 2012, 2008)] model. The model accounts for aerodynamic resistance over bare soil in addition to soil parameters such as field capacity via the SEE. Detailed DISPATCH algorithm derivation and description is presented by Merlin et al. (2012). Here we represent the prominent disaggregation equation as:

$$SM_{HR} = SM_{LR} + \left( \frac{\partial SM_{\text{mod}}}{\partial SEE} \right) [SEE_{HR} - \langle SEE_{HR} \rangle_{LR}]$$ (3.2)

HR and LR refer to the high and low-resolution variables, respectively. There have been multiple linear and non-linear relationships proposed between SEE and surface SM in the past (Budyko, 1961; Komatsu, 2003; Lee and Pielke, 1992; Manabe, 1969; Noilhan and Planton, 1989). A nonlinear model suggested by Noilhan and Planton was used in this study to guide the DISPATCH algorithm.
3.3.2.1 Modified SEE Computation

The SEE, which can be defined as the ratio of actual to potential surface soil evaporation (Fang and Lakshmi, 2014; Merlin et al., 2010a), is computed at the high resolution first, and then the SEE results are aggregated to the respective low resolution 25 km MW scale. The studies by Merlin et al. (2010b, 2012) demonstrated the use of MODIS LST, Normalized Difference Vegetation Index (NDVI) and albedo to determine surface and vegetation temperature and evaporation. The SEE was defined as:

\[ \frac{T_{s,max} - T_{s,HR}}{T_{s,max} - T_{s,min}} \]

where \( T_{s,max} \) is the soil temperature at SEE = 0; \( T_{s,min} \) is soil temperature at SEE = 1, and \( T_{s,HR} \) represents soil temperature at the high resolution grid scale. However, in this study we employed the ratio of the estimated surface evaporation from ALEXI to the potential evaporation to compute SEE directly at the 5-km ALEXI resolution. As mentioned earlier, the two-source land surface representation in ALEXI separates surface evaporation and canopy transpiration. The potential surface evaporation is calculated using the Hamon PET (Hamon, 1963). Hamon PET estimates are completely dependent upon atmospheric demand irrespective of soil and vegetation characteristics and can act as a proxy of potential surface evaporation (PE). This represents a subtle change in the definition of SEE from the Merlin formulation in that in our case all land cover/soil matrix combinations are weighted equally as opposed to being weighted by their assumed PE value as in Merlin (approximated as function of surface temperature). Since the Southeastern U.S. is an energy limited, water rich environment (Ellenberg et al., 2016), evaporation is controlled primarily by water availability and atmospheric demand; therefore, the effects of this change are
not expected to be large. Hamon PET estimates have been found to be comparable to radiation based methods (e.g., Priestly-Taylor) to observed ET in the Southeastern U.S. at monthly or longer time scales (Lu et al., 2005), and are computed using air temperatures from the NLDAS2 forcing data subject to terrain adjustment. Terrain adjustment of coarse resolution temperature data was performed using a 30 m digital elevation map of the region and constant lapse rate of \(-6.5 K/km\) (Cosgrove, 2003).

### 3.3.3 Profile Development

A multi-year vertical SM profile was developed for each ALEXI grid cell using the POME model developed by Al-Hamdan and Cruise (2010) over the study area. The application of POME to develop a one-dimensional SM profile requires two constraints; total probability: \(\int_{\Theta_0}^{\Theta_L} f(\Theta) d\Theta = 1\); and the mass balance constraint: \(\int_{\Theta_0}^{\Theta_L} \Theta f(\Theta) d\Theta = \bar{\Theta}\). Here \(\Theta\) is effective saturation and \(\bar{\Theta}\) is the mean moisture of the soil column; whereas \(\Theta_0\) and \(\Theta_L\) are the upper (surface) and lower (bottom) effective saturation. The effective SM is given as: \(\frac{\theta - \theta_s}{\theta_f - \theta_s}\). The second constraint serves to connect the first moment in probability space to the mean water content of the soil column in physical space. The Shannon entropy is given by:

\[
I = -\int_0^\infty f(x) \ln(f(x)) \, dx \tag{3.3}
\]

where \(f(x)\) is the probability density function (pdf) of the variable. Maximizing \(I\) in Eq. (3.3) for the uniform pdf subject to the constraints, Chiu (1987) developed the 1-D profile of a variable decreasing monotonically from the surface down
using the method of Lagrange multipliers. Al-Hamdan and Cruise (2010) applied the same technique to develop vertical SM profiles either increasing or decreasing with depth from the surface:

\[
\Theta(z) = \frac{\ln \left[ \exp (\lambda_2 \Theta_0) \pm \exp (1 - \lambda_1) \left( \frac{z}{L} \right) \right]}{\lambda_2}
\]  

Equation 3.4

The Lagrange multipliers \(\lambda\)'s can be determined from application of the constraints and boundary conditions (surface effective saturation) and mean effective saturation value of the soil column (\(\Theta\)), \(z\) is calculation depth, and \(L\) is total depth of the column. Equation 3.4 is a monotonically increasing (+ sign) or decreasing (- sign) function, representing dry (increasing from the top boundary) and wet (increasing from the bottom boundary) case profiles.

Experience has shown that not all SM profiles are monotonic as given by Equation 3.4. In fact, it is clear that some profiles can be parabolic in shape (i.e., demonstrate an inflection point), especially immediately subsequent to rain events (dynamic case), or due to sharp changes in soil characteristics (Al-Hamdan and Cruise, 2010; Mishra et al., 2015). These cases are identified when mass balance cannot be kept by the monotonic assumption and thus Equation 3.4 has no solution. In these cases, it is assumed that the inflection point is located in the soil layer with the greatest field capacity Mishra et al. (2015). The POME model is then applied twice; from the surface to the inflection point, and then from the inflection point to the bottom boundary. This procedure was only required in 9% of the profiles generated in the study.
3.3.4 Temporal Compositing

The ALEXI data are available from 2000 to present and AMSR-E from 2002-2011. For this study, the years 2006-2010 were selected for analysis as the NRCS SCAN data was most consistently available during this period (nearly 92%). The ascending AMSR-E SM estimates were available 64.5% of the days on an average for all SCAN site locations while ALEXI retrievals were available on only 36% of the days due to cloud cover limitations. Therefore, a three day moving window unweighted mean was used on AMSR-E and ALEXI retrievals to develop a composite dataset that serves as gap filling and also tends to reduce day-to-day noise in satellite retrievals (Anderson et al., 2011a). Compositing of the ALEXI surface ET increased the mean data availability from 36 to nearly 63% over all SCAN sites and in the case of AMSR-E compositing ensured close to 100% data availability. The availability of pixels with intersection of AMSR-E and ALEXI data more than doubled from 22.5% to 58.7% for the study period over all sites.

3.3.5 Evaluation Metrics

The remote sensing derived SM profiles developed using the POME model were compared and validated against in-situ observations from 10 NRCS SCAN sites along with the gridded Noah LSM SM products over the study area. The LSM was used as a basis of comparison since the long term goal of the project is to develop RS SM profiles that can be assimilated into hydrologic and other land surface models. The data gaps in all three datasets restrict the possibility of time series analysis; therefore, pair-wise
temporal statistical comparisons were performed using traditional matrices such as
correlation coefficient ($r$), root mean square error (RMSE) and bias. It has been
argued that in cases with either the model or reference dataset being biased in mean
or amplitude of fluctuations, the traditional RMSE tends to be an overestimation
of true unbiased data (Entekhabi et al., 2010b). Therefore an unbiased RMSE in
addition to traditional RMSE was also computed. The unbiased RMSE can easily be
computed by removing the bias term form the definition as:

$$ RMSE = \sqrt{E[(\theta_{est} - \theta_{obs})^2]} $$ (3.5)

$$ ubRMSE = \sqrt{E\left\{[(\theta_{est} - E[\theta_{est}]) - (\theta_{obs} - E[\theta_{obs}])^2]\right\}} $$ (3.6)

$$ ubRMSE = \sqrt{(RMSE^2 - Bias^2)} $$ (3.7)

where, $E[.]$ is the expectation operator, $\theta_{est}$ and $\theta_{obs}$ are SM values estimated
and observed (or reference), respectively.

To assess the quantitative error between three datasets against an unknown
true observation, the triple collocation (TC) error estimation method was employed
(Stoffelen, 1998). TC has become a very popular technique for simultaneous error
analysis of three data sets since its adaptation to SM states by Scipal et al. (2008).
The procedure is based on the assumption of linear relationships between the three
estimates of the SM at a specific location and the unknown true value. The unknown
truth is eliminated from the linear error equations through subtraction and then
cross multiplied to determine the error variances of the datasets relative to each
other (Gruber et al., 2016). The assumption is that the errors in the three datasets
are independent and random. Multiple recent studies have used the triple collocation
method for error estimation [such as (Crow et al., 2015; McColl et al., 2014; Su
et al., 2014; Yilmaz et al., 2014) etc.]. A detailed review of method derivations and
application to SM error estimation and analysis is presented by Gruber et al. (2016).

3.4 Results and Discussions

3.4.1 Comparison with Noah LSM

For comparing SM profiles, the 5 cm layer depth POME based profiles were
aggregated to the depths consistent with the Noah LSM: 0-10; 10-40; and 40-100 cm.
The analysis can be approached from three perspectives: the surface values represent
the MW downscaling; the bias represents the ALEXI model performance as it is
providing the total SM content in the root zone; and the RMSE is representative of
the entropy model as it measures the moisture distribution within the soil column.
Figure 3.2 shows the statistics of multi-year temporal SM profile comparisons between
the POME and the Noah LSM for the study region. The figure shows the mean
RMSE and ubRMSE tends to be relatively stable with depth over the entire region,
an indication of relative stability for the profile developed using the POME model. As
depth increased, pixel bias from 0.05-13 indicating that the mean SM data from the
ALEXI model is positively biased compared to the Noah LSM, although the mean bias
was $\leq 0.05$ for all layers. The overall RMSE at all layers was found to be under 0.085 in volumetric SM. Moreover $< 97\%$ pixels across the study area showed ubRMSE of less than 0.06 across all layers, indicating good agreement between the POME model and the Noah SM estimates. Comparing Figure 3.2 with the landcover map (Figure 3.1), it seems that the higher correlations ($r > 0.6$) occur more prominently in the agricultural dominant portions of the study area for the top two layers (0-40 cm). The overall correlations in the range of 0.46-0.54 across layer depths suggest that the temporal variabilities from remotely sensed driven POME model compared fairly well against Noah SM.

Comparison between POME and Noah SM profiles by land cover type (Figure 3.3) indicate that the absolute bias tends to increase with depth in the savannah, shrub, and forest land covers while the reverse is evident for the urban, grass and crop coverages. It appears that overall bias is lowest in the savannah, forest, and agricultural land classes and since those classes (particularly forest) dominate the region, this naturally leads the relatively low overall region-wide bias shown in Figure 3.2.

The RMSE (and ubRMSE) present an opportunity to judge the overall profile development process. It is clear from Figure 3.3 that the RMSE improves from the surface to the middle layer and then increases again in the bottom layer in every land cover class except shrub. The top and bottom layer RMSE is being impacted by the boundary conditions placed on the POME integral by the MW and the parameterized lower boundary. Clearly, the POME process tends to improve the imprecise
Figure 3.2: Map of bias, RMSE, unbiased RMSE and Correlation over multiple years (2006-2010) at different layer depths: top panel: 0-10cm; middle panel: 10-40cm and bottom panel: 40-100cm.

In terms of correlation, the mid layer (10-40 cm) has the highest correlation (overall mean $r = 0.54$) for all land cover types with the highest mean correlation of 0.7 for crop dominated landcover. This further demonstrates the capabilities of the ALEXI model to estimate root-zone mean SM content in comparison to the Noah LSM. Incidentally, for most crops, the majority of the root mass is distributed in the
top 60 cm of the soils column (Wu et al., 1999). The higher root density ensures the strong coupling of the land-plant-atmosphere system which tends to improve the accuracy of ALEXI in that zone. Increased correlations in the 10-40 cm layer indicate the ability of ALEXI to mimic the temporal patterns in the root-zone consistently relative to Noah. As depth increases, the root density is reduced and thus the coupling between land and atmosphere is also reduced. This fact, along with the relatively coarse parameterization of the lower boundary on the POME profile, leads to a relative decrease in correlation at layer 3 (40-100 cm) at all land covers except for trees (forest). The cropland showed the highest correlations with the Noah profile while keeping the RMSE and bias consistent with other land types. Agricultural areas demonstrated correlations ranging from 0.5 to 0.7 with a mean correlation of 0.62.

![Figure 3.3](image.png)

**Figure 3.3:** Comparison of Noah and POME SM profiles at multiple layer depths by Land Cover across Southeast U.S.
The overall analysis by layer depths appear to indicate that the profiles developed through the POME model using the disaggregated MW and the ALEXI derived mean SM content is in good agreement with the Noah LSM in the Southeastern U.S. and in very good agreement in agricultural areas of the region.

### 3.4.2 Comparison with in-situ Observations

The comparison against Noah LSM SM estimates provided useful insights towards the performance of TIR-based SM profiles developed through the POME model. The comparisons against the LSM specifically adds to the analysis of results as a function of land cover, yet as mentioned earlier, the analysis does not assume that Noah is a perfect model - it may have its own errors. Therefore multiple NRCS SCAN site in-situ observations are used for further validations. When comparing remotely sensed data to site specific in-situ observations, disparities in spatial scale and sensing depth must be considered. Although some authors prefer to remove bias due to the differing scales before comparisons are made (Brocca et al., 2011), it is also quite common to do the comparisons without adjusting for scale, even when only one in-situ site is available (McCabe et al., 2005; Sahoo et al., 2008). In this study no bias corrections were performed.

Figure 3.1 shows the location of each of the sites used for validation along with the underlying land cover map. Table 3.1 summarizes the SCAN site characteristics, dominant land cover types and soil characteristics at surface and 100 cm depth. Dominant land cover for sites 2009, 2114 and 2115 are predominantly savannas and forest type (hereafter referred as forest sites), whereas sites 2013, 2037, 2038 and 2113
are a mix of cropland either with savannas or shrubs (hereafter referred as mixed cropland sites). Only sites 2027, 2078 and 2053 (hereafter referred as cropland sites) are predominantly cropland at the 5-km spatial resolution footprint. The crop and mixed crop sites are shown in bold in the following text. The SCAN sites monitored SM at depths of 5, 10, 20, 50 and 100 cm. The POME based profiles are developed at 5 cm layer depth increments down to 100 cm depth.

The results of the developed profiles in comparison to the SCAN site observations alone are shown in Figure 3.4. First, it is evident in all the statistics except the correlation that the pattern demonstrated in the previous comparisons persists in that the statistics often tend to improve with depth with occasional deterioration when the lower boundary is encountered. Considering the performance of ALEXI initially, the bias appears reasonable in most cases where the majority of instances the absolute bias is less than 0.1, but it appears to be best in the mixed cropland areas (mean absolute bias of 0.07 across all depths) and worse in forested sites (mean absolute of 0.13). In fact, at seven of the ten total sites the overall bias is considerably less than the average moisture content at the SCAN site itself. At the two sites with the highest bias (2009 and 2027), the mean moisture content from ALEXI was about twice the observations at all layers, indicating that the satellite estimates showed considerable positive bias (mean bias 0.17 and 0.13 respectively). Hain et al. (2011) pointed out that sensitivity of the ALEXI model decreases as moisture content nears either the wilting point or the field capacity. Both sites 200 and 2027 had sandy soils at the SCAN site and exhibited the lowest mean moisture content of all sites. At site 2009 with sandy soil through the column, the mean SM content was
0.05 cm³cm⁻³ against the wilting point of 0.033 cm³cm⁻³ while 2027 site had sand at the surface and sandy loam (wilting point = 0.095 cm³cm⁻³) at the 100 cm depth and the mean SM content was 0.12 cm³cm⁻³. Moreover, the site 2009 is located in a forest-dominated region. Whereas for site 2027 (located in southwest Georgia), the higher bias in remotely sensed observations can be attributed to additional SM content due to irrigation. Southwest Georgia is one of the most irrigated regions of the study area. In contrast, the SCAN site observations are primarily governed by precipitation alone.

In the case of RMSE, half the sites showed an average RMSE of 0.1 or less. RMSE tends to be better at the mixed land use sites, while poor performances at sites 2009, 2115 and 2027 skewed the forest and cropland results respectively. As in the bias case, these sites demonstrated the highest mean RMSE values (Figure 3.4). However, with the exception of these sites, the average RMSE was less than the SCAN average moisture content in all cases. The ubRMSE, on the other hand, at all sites was better with the overall ubRMSE for all layer depths and land cover types exhibiting an average ubRMSE of 0.07. The ubRMSE tended to improve with depth for all cases (Figure 3.5) up to the depth of 50 cm, but showed a rise at the 100 cm depth as discussed previously. Improvements in ubRMSE with depth indicate the ability of the POME model to converge and correct itself from the effects of the noisy surface boundary condition until the assumed lower boundary affected the performance in that layer.

The correlation coefficient (r) results are interesting and do not necessarily track the other two indices. It is clear from Figure 3.4 that POME tended to perform
better in agricultural land use areas than in other environments. Similar to the bias results, correlation was poorest at forested locations. In all, three sites showed average correlation above 0.5 with four other sites showing a correlation above 0.4. Two sites (2009, 2113) produced average correlations of 0.16 and 0.32 across all depths. As discussed earlier, site 2009 is forested while 2113 is located near a water body (Lake Catoma). Overall, the crop sites showed the highest correlations (0.51) followed by mixed crop sites (0.42), an indication of the ability of the satellite derived surface and mean moisture content estimates to mimic wetting and drying patterns over time across depths.

Figure 3.4: Statistics at SCAN sites showing bias, Correlation, RMSE and ubRMSE between in-situ observations and POME SM profiles at multiple depths.

However, the correlation consistently declined with depth at most of the agriculture and mixed agriculture sites. The decline most often became more pronounced
after the second (or sometimes third) layer indicating that the influence of the parameterized lower boundary extends through the lower 50 cm of the profile, at least to some extent. This phenomenon was not evident in the forest areas where the SM was not as variable in the lower layers.

3.4.3 Intercomparison of Noah, POME with In-situ Observations

The POME profiles have been compared with Noah LSM across the study region against in-situ observations at ten locations. However, as mentioned earlier, both analyses have some limitations either in terms of proxy ground truth (in case of LSM) and spatial representation (in-situ observations). Therefore, in this section an intercomparison between the three datasets is performed to assess the relative strength of each SM dataset. Figure 3.5 shows the time series of the SM state from Noah LSM, SCAN observations and the POME model. Consistent with the layer depths of the Noah, the POME profile and the SCAN observations were aggregated to 0-10; 10-40; and 40-100 cm layer depths.

Table 3.2 shows the detailed statistics of comparison between Noah LSM SM, in-situ observations and POME profiles at each SCAN site location. The results are further summarized across all sites in Figure 3.6. The overall results show that the satellite-based and LSM SM estimates are reasonably comparable based on error statistics of ubRMSE (0.05 vs 0.04) and absolute bias (0.08 vs 0.07). For the surface layer (0-10 cm) comparisons, the Noah correlations are superior to the POME model ($r = 0.75$ vs $0.54$), although in several cases the Noah correlations decrease vertically through the soil column to the point that the two approaches are much more compa-
Figure 3.5: Time series of soil moisture condition at 10 NRCS SCAN sites from the POME model (Blue); Noah LSM (green) and in-situ observations (red) at three layer depths (2006-2010)
rable (Figure 3.6). This case does not show the steep decline in correlation through the POME profiles as before, indicating that amalgamation of the lower layers into one 60 cm layer has dampened that effect. In terms of mean bias across layers, the POME model is superior in four cases, Noah is superior in four cases and in the other two cases (2115 and 2053) the two models perform the same. In terms of ubRMSE, the POME is superior to Noah at three locations while at other six locations the difference is within 0.01 (in $cm^3cm^{-3}$). Overall, the average statistics across all depths and all sites, the Noah/SCAN average RMSE was 0.09 in comparison to the POME RMSE of 0.10 against ground based SCAN observations. The unbiased RMSE between Noah and SCAN was 0.04, and for the POME it was 0.05 in volumetric SM. Figure 3.6 shows that the Noah LSM tended to become less accurate with depth while the POME generally showed the reverse.

The three data sets can be further compared through TC analysis. TC has the advantage that the SCAN observations are treated equally with the LSM and POME as just another estimate of the true SM state. The analysis is performed for three layers to be consistent with the LSM model configuration (Figure 3.7). The surface results (0-10 cm) showed that in most instances the SCAN observations are closer to the true SM compared to the Noah and POME data; however, the latter two data sets also show high coefficient of determination ($R^2$) values at several sites. The middle and bottom layer results appear to indicate that the Noah LSM is superior (with 5 and 9 instances of $R^2 > 0.8$, respectively), while the SCAN observations and the POME model track each other fairly well with 6 and 5 instances, respectively, of $R^2 > 0.4$ for the POME and 5 and 4 such instances for SCAN observations. The Noah results
Table 3.2: Results of temporal comparisons in absolute bias, RMSE, ubRMSE and correlation at 10 sites between the developed profile and Noah SM profiles against SCAN observations at 0-10; 10-40 and 40-100cm depths [NP - Noah vs POME; SP - SCAN vs POME; and NS  Noah vs SCAN]

<table>
<thead>
<tr>
<th>Site</th>
<th>Bias</th>
<th>RMSE</th>
<th>ubRMSE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NP</td>
<td>SP</td>
<td>NS</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0-10 cm</td>
</tr>
<tr>
<td>2009</td>
<td>0.02</td>
<td>0.13</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>2014</td>
<td>0.00</td>
<td>-0.10</td>
<td>-0.10</td>
<td>0.06</td>
</tr>
<tr>
<td>2015</td>
<td>0.01</td>
<td>0.15</td>
<td>0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>2017</td>
<td>0.03</td>
<td>0.19</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>2013</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>2015</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>2013</td>
<td>0.06</td>
<td>0.02</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>2013</td>
<td>0.06</td>
<td>0.08</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>2013</td>
<td>0.04</td>
<td>0.07</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>2013</td>
<td>0.06</td>
<td>0.18</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>2013</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>2013</td>
<td>0.06</td>
<td>0.14</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>2013</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>2013</td>
<td>0.01</td>
<td>-0.06</td>
<td>-0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>2013</td>
<td>0.14</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>2013</td>
<td>0.07</td>
<td>0.10</td>
<td>0.02</td>
<td>0.08</td>
</tr>
<tr>
<td>2013</td>
<td>0.08</td>
<td>-0.01</td>
<td>-0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>2013</td>
<td>0.08</td>
<td>0.06</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>2013</td>
<td>0.08</td>
<td>0.20</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>2013</td>
<td>0.04</td>
<td>-0.17</td>
<td>-0.21</td>
<td>0.08</td>
</tr>
<tr>
<td>2013</td>
<td>0.00</td>
<td>-0.02</td>
<td>-0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>2013</td>
<td>0.08</td>
<td>0.05</td>
<td>-0.02</td>
<td>0.10</td>
</tr>
<tr>
<td>2013</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.04</td>
<td>0.08</td>
</tr>
<tr>
<td>2013</td>
<td>0.03</td>
<td>-0.08</td>
<td>-0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>2013</td>
<td>0.07</td>
<td>-0.03</td>
<td>-0.10</td>
<td>0.11</td>
</tr>
<tr>
<td>2013</td>
<td>0.09</td>
<td>0.04</td>
<td>-0.04</td>
<td>0.10</td>
</tr>
<tr>
<td>2013</td>
<td>0.13</td>
<td>-0.10</td>
<td>-0.22</td>
<td>0.14</td>
</tr>
<tr>
<td>2013</td>
<td>0.08</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.15</td>
</tr>
</tbody>
</table>

May be problematic in that the basic assumption of TC analysis is that the errors are random and unrelated. In the case of a LSM such as Noah, the deterministic SM equation (e.g., Richards Equation) governs the movement of moisture through the column and some of the random errors are eliminated. This would not affect the
surface layer, which is governed by precipitation and surface evaporation. Thus, the errors in the LSM at the deeper layers may be dampened. The conclusion may be that the LSM cannot be fairly evaluated through a purely stochastic analysis such as TC.

### 3.4.4 Error Characterization

The developed profile results are impacted by the boundary conditions applied to the POME as the integral serves to transition the profile between the upper and lower boundary conditions. The upper boundary is associated with the MW surface SM estimates while the lower boundary was assumed for this study and potentially could be parameterized or used as a calibration parameter. In addition, the mean

![Figure 3.6: POME/ALEXI profiles and Noah statistics at all SCAN sites compared against observations averaged across layer depths](image-url)
SM estimated from ALEXI determines the total mass to be distributed. Earlier studies by Al-Hamdan and Cruise (2010) and Mishra et al. (2015) showed that the POME model is capable of producing profiles with significant accuracy with mean absolute errors in the range of 0.5-3.0% for known input conditions. However, in this study inputs to the POME model are derived from remotely sensed measurements, in addition to a parameterized bottom boundary condition. Hence, profile errors may be characterized in terms of errors in input parameters.

Figures 3.8a and b shows the sensitivity of the profile in terms of bias and RMSE to variations in the mean and surface constraints. From Figure 3.8a it is clear that, even if the surface boundary condition is off by 50% (in effective SM), the overall profile RMSE and bias is less than 0.35 (in effective SM), and the maximum possible deviation in the surface boundary results in bias and RMSE of 0.62 and 0.67.
respectively. The sensitivity study of the mean moisture content (Figure 3.8b) shows that the bias and RMSE of the profile (in terms of effective SM) are linearly related to the deviations in the assumed mean. Further, Figure 3.8 indicates that the profile is more sensitive to errors in the mean than it is to deviations in the surface boundary condition.

Figure 3.8: POME model sensitivity to (a) boundary condition; (b) sensitivity to profile mean input towards profile Bias and RMSE in terms of effective SM.

3.4.4.1 Effect of Disaggregation of AMSR-E MW Data

Figure 3.8 shows that the POME profile is sensitive to the surface boundary conditions. In this study these conditions are provided by AMSR-E; therefore, it is instructive to examine the relative accuracy of the downscaled MW data. To that end, the AMSR-E surface SM before and after disaggregation is compared to both the Noah LSM and the in-situ SCAN data to quantify the effect of the SEE downscaling algorithm. The results from a temporal analysis between coarse and downscaled (fine) resolution MW surface SM with the Noah LSM surface is shown in
Figure 3.9 for the study domain. The figure shows that the generally negative bias of the original AMSR-E data (overall mean = -0.08) when compared to the Noah LSM was transformed by the disaggregation to a positive bias in the eastern half of the study area although the overall bias remained slightly negative. The positive bias in the eastern zone was largely in the 0.04 to 0.13 range. It is also apparent that this same area exhibited a substantial increase in correlation between the downscaled MW and Noah data. Comparing Figure 3.9 to the land cover image in Figure 3.1, it can be see that the increase in correlation was largely in the agricultural bands in the southwestern Georgia leading into southeastern Alabama. However, a few areas, such as extreme southwestern and east-central parts of Alabama, showed degradation in correlation on downscaling. The land cover map shows that these areas are generally forested. Overall the temporal correlation ($r$) showed a modest increase from 0.21 to 0.25 with downscaling for the study area indicating that downscaled AMSR-E is slightly more comparable to Noah LSM surface SM. Perusal of the figure shows that the poor results in Florida and along the eastern seaboard are primarily responsible for the low correlations. It also demonstrates the fundamental property that the downscaling process will be compromised in areas where the original MW data was of exceptionally poor quality to begin with.

It is difficult to determine the impact of the disaggregated MW surface SM estimates on the profiles compared to the LSM. First, the statistics shown in Figure 3.9 are for the sensing depth of the raw AMSR-E data (0-5 cm) while the relatively better statistics shown in Figure 3.2 are for the top layer corresponding to the Noah LSM (0-10 cm). This disparity in depth is undoubtedly affecting the results. The
introduction of the mean SM from ALEXI also affects the near surface layer in the POME profile since mass balance must be maintained throughout the soil column.

In any case, comparison of Figure 3.2 and 3.9 shows that the profile statistics are considerably improved compared to the MW surface values and thus the noise in the MW data has a minimal effect when compared to the Noah LSM.

The results of the comparison with the SCAN sites are perhaps more instructive and are given in Table 3.3 below. The table shows that in terms of correlation, the disaggregated data were better related to the in-situ data than were the original coarse scale MW data ($r = 0.53$ vs $r = 0.31$). This result was particularly evident at the agricultural SCAN sites ($r = 0.64$ vs $r = 0.42$). These results were obtained at a slight cost in the bias (bias=0.07 vs bias= -0.02) and RMSE (RMSE=0.1 vs
RMSE=0.12), although the difference was not as great in unbiased RMSE. In the case of Table 3.3, the SCAN depth is the same as the MW so comparisons are apt. In cases of relatively high bias in the MW data (e.g., sites 2009, 2114, 2053, 2078) this error is introduced into the POME profile. Figure 3.8 shows that errors in the surface boundary of about 0.1 translate to bias and RMSE in the profile of about 0.05. It appears from Table 3.3 that at the sites demonstrating the consistently higher bias and RMSE, the error in the surface boundary could be responsible for one third to one half of that total.

**Table 3.3:** Statistical comparison before and after disaggregation of coarse resolution MW SM against SCAN observations [A = Bias; B = RMSE; C = ubRMSE; D = correlation coefficient; N = number of days data points was available; maximum possible N = 1825]; *non-significant correlation using two-tailed t-test at 99% CI

<table>
<thead>
<tr>
<th>Site</th>
<th>SCAN</th>
<th>MW (25k)</th>
<th>MW (5k)</th>
<th>N</th>
<th>SCAN/MW(25km)</th>
<th>SCAN/MW(5km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>2009</td>
<td>0.06</td>
<td>0.18</td>
<td>0.19</td>
<td>841</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
<td>-0.12*</td>
</tr>
<tr>
<td>2014</td>
<td>0.26</td>
<td>0.15</td>
<td>0.19</td>
<td>1055</td>
<td>-0.11</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.09</td>
<td>0.30</td>
</tr>
<tr>
<td>2015</td>
<td>0.08</td>
<td>0.15</td>
<td>0.25</td>
<td>1103</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.42</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.03</td>
<td>0.12</td>
<td>0.05</td>
<td>0.22</td>
<td>0.08</td>
</tr>
<tr>
<td>2027</td>
<td>0.08</td>
<td>0.13</td>
<td>0.29</td>
<td>1241</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
<td>0.48</td>
</tr>
<tr>
<td>2053</td>
<td>0.24</td>
<td>0.14</td>
<td>0.32</td>
<td>1160</td>
<td>-0.10</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.08</td>
<td>0.43</td>
</tr>
<tr>
<td>2078</td>
<td>0.25</td>
<td>0.14</td>
<td>0.31</td>
<td>1080</td>
<td>-0.12</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.05</td>
<td>0.34</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>-0.06</td>
<td>0.10</td>
<td>0.05</td>
<td>0.42</td>
<td>0.12</td>
</tr>
<tr>
<td>2113</td>
<td>0.20</td>
<td>0.14</td>
<td>0.18</td>
<td>1014</td>
<td>-0.06</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
<td>0.44</td>
</tr>
<tr>
<td>2037</td>
<td>0.16</td>
<td>0.14</td>
<td>0.21</td>
<td>1157</td>
<td>-0.02</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td>2038</td>
<td>0.14</td>
<td>0.16</td>
<td>0.16</td>
<td>1067</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.45</td>
</tr>
<tr>
<td>2013</td>
<td>0.21</td>
<td>0.15</td>
<td>0.23</td>
<td>1218</td>
<td>-0.06</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.21</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>-0.03</td>
<td>0.08</td>
<td>0.06</td>
<td>0.26</td>
<td>0.02</td>
</tr>
<tr>
<td>Overall Mean</td>
<td></td>
<td>-0.02</td>
<td>0.10</td>
<td>0.06</td>
<td>0.31</td>
<td>0.07</td>
</tr>
</tbody>
</table>

### 3.4.4.2 Effect of Mean SM Inputs

The mean SM content within soil column in this study obtained from TIR based ALEXI model served as one of the two remotely sensed input parameters for
the POME model. Therefore the mean SM content retrieved from the ALEXI model is compared with the Noah LSM. The results of the temporal analysis between the two datasets are shown in Figure 3.10. The overall bias between the two datasets is 0.04. The overall RMSE is 0.08 with ubRMSE of 0.04 indicating that the mean SM content of the two datasets is similar. In terms of correlation coefficient, the root zone correlation nearly doubled ($r = 0.49$) compared to the surface correlations (Figure 3.9). Further, comparison of Figure 3.10 with Figure 3.1 reveals that, similar to the surface SM analysis, the mean SM content with the highest correlations ($r \geq 0.5$) are observed mostly in agriculture-dominated areas.

Figure 3.8b shows that the translation of the error in the mean SM content to errors in the POME profile is linear, so an error of 0.04 in the ALEXI mean compared to the LSM would translate into a similar error in the computed profile. Examination of column 2 (NP) in Table 3.2 above shows that this error represents the majority of the errors in the computed POME profiles compared to the LSM.

### 3.5 conclusions

This study evaluated the feasibility of linking downscaled MW surface SM with TIR root zone estimates to develop entropy-based vertical SM profiles. The SM profiles (including surface values) were compared to in-situ data at the Southeastern U.S. as well as the Noah LSM within the NASA LIS. Initial results are encouraging. The SEE disaggregation method of Merlin et al. (2012), guided by high resolution TIR estimates from the ALEXI model, showed promise when compared to the in-situ and modeled estimates in a humid semi-tropical region of the U.S. The POME
generated SM profiles generally compared favorably with the SCAN site profiles and the Noah LSM. In summary:

- When the Noah LSM and the POME profiles were compared to the in-situ data in terms of bias, the POME-generated profiles were clearly superior in at four sites, the LSM was superior at four sites and the two methods were the same at the other sites. The maximum correlation in the range of 0.4-0.65 was observed in agriculturally dominant areas. Further the highest correlations were found at the depth of 10-40 cm, coinciding with the maximum root density for crops and thus offering a better coupling between land and atmosphere. The
ALEXI model was able to pick the wetting and drying trends in the root-zone consistently.

- Compared to *in-situ* observations, the bias and RMSE of the Noah model often tended to degrade vertically with depth while the reverse was evident in most of the POME profiles. This characteristic of the remote sensing driven POME method seems to imply that profiles from land surface models could be improved in terms of bias and RMSE through the assimilation of the remotely sensed profiles.

- TC analysis revealed that the POME and observed SCAN site observations tracked well, while the LSM appeared to show less variability, possibly due to the use of the deterministic Richards Equation to model SM movement through the soil column.

Error analyses revealed that the majority of the error in the POME generated profiles was due to error in the mean SM deduced from the ALEXI retrievals and the parameterized lower boundary condition. The SEE downscaling procedure increased the correlation of the surface SM compared to both the LSM and the SCAN sites, especially in agricultural areas where correlations in the range of 0.5-0.8 were achieved. In the meantime, the overall bias was reduced by a factor of 4 and the RMSE was only slightly increased (0.09 to 0.10). Downscaling generally was less effective in locations where the AMSR-E demonstrated positive bias and appeared to lose effectiveness as the bias increased. MW surface observations can be contaminated when a high percentage of the pixel is dominated by water, as near large streams or lakes or in
the near coastal region. Dense vegetation also tends to degrade the MW results. Overall, analysis revealed that the surface SM estimates accounted for, at most, for one third to one half of the error in the SM profiles and for most cases, the mean SM and the parameterized lower boundary accounted for the majority of the error. Recent advances such as the L-band sensor aboard the SMAP mission, offers the potential for even better correlated MW data. In addition, further analysis of the lower boundary condition parameterization could improve the profiles, particularly in the lower layers. For example, Mishra et al. (2013) used POME generated profiles to update SM within a crop model using the lower boundary condition from the model itself. If sufficient ground truth data are available, calibration could be accomplished, or the lower boundary could be set as a function of soil properties in the bottom layer of the profile.

The relatively sparse (5-10 day recurrence interval) availability of the ALEXI TIR-based SM retrieval is the major weakness of the procedure and necessitated compositing of the data into three day running means. However, the issue is a function of the semi-tropical humid climate of the Southeastern U.S. Drier regions of the world would not suffer as much from this issue. Thus it is possible that the proposed method could be employed to deduce vertical SM profiles in regions of the world where observed climate data are scarce or insufficient to drive ecological models. These profiles could be assimilated into the models to help correct for model bias due to the poor climate inputs.
CHAPTER 4

AN INITIAL ASSESSMENT OF SMAP SOIL MOISTURE DISAGGREGATION SCHEME USING TIR SURFACE EVAPORATION DATA OVER THE CONTINENTAL UNITED STATES

Submitted to International Journal of Applied Earth Observation and Geoinformation

Abstract

Soil Moisture Active Passive (SMAP), is a mission dedicated toward global soil moisture mapping. Typically the L-band microwave radiometer spatial resolution is on the order of 36-40 km, which is known to be too coarse for many specific hydro-meteorological and agricultural applications. With the failure of SMAP radar within three months of becoming operational, and this, an intermediate (9-km) and finer (3-km) scale soil moisture product solely from the SMAP mission is no longer possible. As an alternative, and the focus of this study, a disaggregation of the 36-km resolution SMAP surface soil moisture (SSM) was performed using the Soil Evaporative Efficiency (SEE) approach to 3 and 9-km spatial scales was performed. The SEE was computed using thermal-infrared (TIR) estimation of surface evaporation over CONUS. Once available the disaggregation results were compared with the 3 months of SMAP-Active (SMAP-A) and Active/Passive (AP) products, while comparisons with SMAP-Enhanced (SMAP-E), SMAP-Passive (SMAP-P), as well as with more than 180 Soil Climate Analysis Network (SCAN) stations across Continental U.S., were performed for 19 months period. At the 9-km spatial scale, the TIR-Downscaled data correlated strongly with the SMAP-E SSM both spatially ($r = 0.90$) and temporally ($r = 0.87$). In comparison with SCAN observations, overall correlations of 0.49 and 0.47; bias of -0.022 and -0.019 and unbiased RMSE of 0.105 and 0.100 were found for SMAP-E and TIR-Downscaled SSM across the Continental U.S., respectively. At 3-km scale, TIR-Downscaled and SMAP-A had mean temporal r of only 0.27. In terms of gain statistics, the highest percentage of SCAN sites with positive gains (<55%) was observed with the TIR-Downscaled SSM at 9-km. Over-
all, the TIR-based downscaled SSM had a strong relation with SMAP-E; compared to SCAN, and overall both SMAP-E and TIR-Downscaled performed similarly, however, gain statistics shows that TIR-Downscaled SSM slightly outperformed SMAP-E.

4.1 Introduction

Soil moisture quantity, quality, and state are essential components of both the hydrologic and energy budgets. The amount of moisture in the soil drives a wide variety of hydrological, geotechnical, agricultural, and meteorological processes (Romano, 2014). Soil moisture (SM) can be estimated through ground based *in-situ* measurements; biophysical and land surface models (LSMs) or through remote sensing techniques. Unfortunately, ground based stations are sparse so that currently, LSMs offer the most common source for SM retrievals. However, models can be subject to considerable error and bias and for this reason, other sources of SM data have been developed to correct for model inaccuracies. In particular, remote sensing technologies have come to the forefront in this issue.

The characteristics of remotely sensed SM estimates are highly dependent on the wavelengths of the instruments being utilized to retrieve them. Longer wavelengths penetrate deeper into the soil column and are normally available at higher spatial resolutions but do not penetrate through cloud cover. In contrast, shorter wavelengths do not penetrate as deep in the soil and exhibit lower spatial resolutions, but can penetrate through cloud cover and some vegetation canopies, and thus are available more frequently than are the longer wavelength data.
The goal of LSMs (including hydrologic and agricultural models) is to simulate relevant processes at operational space and time scales. Thus, there is a dichotomy in this matter between the long and short wavelength sources. In recent times, the emphasis has been on the development and application of passive shortwave (microwave) sensors. Microwave (MW) sensors, since their inception in late 1970s, have been used to estimate large scale SM, typically from higher frequency C-band [\(\sim 6 \text{ GHz}\)] and X-band [\(\sim 10 \text{ GHz}\)] sensors such as the Scanning Multichannel Microwave Radiometer (SMMR) (Owe et al., 2001); Special Sensor Microwave/Imager (SSM/I) (Paloscia et al., 2001); and the Advanced Microwave Scanning Radiometer (AMSR-E) (Njoku et al., 2003). Sensors such as the Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2010) instrument and the recently launched Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010a) are the only missions dedicated toward global SM mapping operating at L-band [\(\sim 1 \text{ GHz}\)] frequencies. Low frequency L-band radiometers have penetration depths of 3-5 cm (approx.). Further, the measurements are sensitive to soil moisture through moderately thick vegetation water content (5 \(kg.m^{-2}\)) (Entekhabi et al., 2010a). Although exhibiting relatively higher accuracy and attenuated atmospheric absorption compared to the C- and X-bands, the L-band MW radiometer spatial resolution is on the order of 36-40 km (Merlin et al., 2015). Such spatial resolutions are acceptable for hydro-climatological studies but are known to be too coarse for many hydro-meteorological and agricultural applications (Brown et al., 2013).

Due to the relatively low spatial resolution of passive MW sensors, a great deal of effort has been expended on the development of algorithms to disaggregate (or
downscale) the coarse resolution data to finer operational scales (e.g., 10-km or less). The SMAP mission aimed to approach the problem by including active microwave L-band radar onboard in addition to the passive L-band sensor. As demonstrated in this study, integration of active with passive MW observations has been used as a disaggregation scheme to reduce the spatial footprint of coarse resolution radiometers with some success (Bindlish et al., 2009; Das et al., 2011; Narayan and Lakshmi, 2008; Rudiger et al., 2016). The drawbacks associated with an active-passive based disaggregation scheme including inconsistent retrieval times, low temporal revisits of the radar (∼12-16 days), and the difference in operational bands limits the applicability of such methods. However, in SMAP both radiometer and radar were developed to operate at L-band, hence avoiding the issues pertaining to inconsistent retrieval timing and bands. The primary objective of the SMAP mission was to develop an integrated active-passive SSM product at fine resolution of 3-km and an intermediate resolution of 9-km with radiometer-like accuracy of $0.04 \, m^3 m^{-3}$ (Das et al., 2011; Entekhabi et al., 2010a). However the radar malfunctioned within three months of it becoming operational and thus the possibility of a finer (3-km) resolution integrated product solely from the SMAP platform is unlikely, although the intermediate 9-km data stream continues.

SMAP offers an excellent opportunity to test disaggregation schemes that have been developed to downscale coarse resolution MW SM data and thus to evaluate the SMAP products themselves. The objectives of this study are twofold: (a) to compare a recently developed disaggregation scheme to the SMAP SM products at the 9-km resolution and to a lesser extent, the 3-km resolution; and (b) to compare
the SMAP products themselves, along with the disaggregation scheme, to in-situ data from all available U.S. Department of Agriculture SM sites in the Continental U.S. (CONUS). The SMAP active radar was fully operational between mid-April to early-June 2015 and provided SMAP SM products at mid (9-km) and high (3-km) resolutions along with a gridded 36-km radiometer product. In this study, the disaggregation of the SMAP radiometer SM estimates was performed over CONUS for the periods coincident with the products; the 3-km SMAP product was compared to the downscaled SM for the three months coinciding with the period the SMAP radar was operational, however, the 3-km downscaled product was continued for the entire period from April 2015-November 2016 (19 months) for further evaluation against in-situ data. The 9-km products (SMAP enhanced and downscaled) were evaluated against each other, as well as to the downscaled 9-km data for the entire period up to November 2016. The fine resolution (3-km) SMAP and downscaled SM were compared with active Natural Resource Conservation Service (NRCS) Soil Climate Analysis Network (SCAN) sites across CONUS during the study period. Comparison with SCAN sites allows evaluation of the 3-km SMAP radar SM product before its failure along with the performance of the disaggregation scheme. However, the disaggregated 3-km data were compared to the SCAN data for the entire period to November 2016.

4.2 Background

To overcome the coarse resolution issue associated with MW SSM retrievals while leveraging its high temporal (2-3 days) global coverage at acceptable accuracy,
several studies have focused on assimilating MW soil moisture observation within LSMs (Crosson et al., 2002; Ines et al., 2013; Lievens et al., 2015; Reichle et al., 2002; Sahoo et al., 2013) and in-situ observations (McCabe et al., 2005). Another technique to disaggregate coarse scale MW data have been developed involving the use of finer resolution visible and thermal infrared (TIR) imagery. Such approaches are based on the apparent triangle/trapezoidal pattern relationship between land surface temperatures (LST) and vegetation indices (VI) linked with underlying moisture content (Carlson, 2007; Carlson et al., 1981). Multiple variants of the triangle approach have been studied and applied either directly as polynomial fitting or indirectly as evaporative fraction. Chauhan et al. (2003) used normalized Difference Vegetation Index (NDVI) and LST along with surface albedo from Advanced Very High Resolution Radiometer (AVHRR) instrument to disaggregate SSM/I for the Southern Great Plains (SGP-97) experiment. Later, a multitude of studies attempted similar downscaling approach using finer resolution data form Moderate Resolution Imaging Spectroradiometer (MODIS), Meteosat Second Generation Spinning Enhanced Visible and Infrared Imager (MSG-SEVIRI) etc. under different combination setups (such as: inclusion of albedo, brightness temperature etc.) e.g. Knipper et al. (2017); Piles et al. (2011, 2016); Sanchez-Ruiz et al. (2014) etc. The results of these studies were decidedly mixed, although there were some very good results, even at fine (500 m) resolutions [e.g., Sanchez-Ruiz et al. (2014)].

A variant of the triangle approach that is relatively more theoretically and physically-based than polynomial fitting was proposed by Merlin et al. (2010a, 2012) which relates the soil evaporative efficiency (SEE) to surface moisture content These
authors used finer resolution MODIS VI, LST and surface albedo to compute SEE based on the triangle approach to generate a downscaled SMOS SSM product up to 1-km resolution in southern Australia (Merlin et al., 2012). Multiple recent studies have used the SEE-based algorithm to downscale SSM from AMSR-E, SMOS, SMAP etc. with some success includes Chan et al.; Djamai et al. (2015); Malbèteau et al. (2016); Mishra et al. (2017); Molero et al. (2016). A comparative study of multiple disaggregation schemes by Kim and Hogue (2012) in the semi-arid climatic conditions of the Western U.S. indicated that the evaporation efficiency based disaggregation technique performed better than the empirical polynomial fitting approach. Although, the SEE-based approach has shown some promising results, the accuracy of the algorithm is limited by the accuracy in SEE computation as well as the underlying SEE/SM model being used. One of the limitations of the visible (VIS)/infrared (IR) based disaggregation is the lower cloud penetration capabilities of such bands, resulting in data gaps under cloudy conditions. Multiple other downscaling algorithms exists and an excellent review of SSM downscaling approaches is presented by Peng et al.

In this study, the evaporative efficiency based algorithm from Merlin et al. (2012) was used to disaggregate the SMAP radiometer SSM product over CONUS. The SEE-based approach was selected due to its relatively more theoretical and physical basis than polynomial schemes and relatively better results reported by earlier studies. SEE can be defined as a ratio of actual to potential soil evaporation (Fang and Lakshmi, 2014; Merlin et al., 2010b). As opposed to the indirect approach of (Merlin et al., 2010a, 2012), the SEE in this study is computed directly from the
surface actual evaporation and potential surface evaporation data. The Atmospheric Land Exchange Inverse (ALEXI) model (Anderson et al., 2011b, 1997) was used to obtain surface soil evaporation. The ALEXI model is an energy balance model that utilizes time differential rise in LST data from Geostationary Operational Environmental Satellites (GOES) to retrieve actual evapotranspiration (ET) (Anderson et al., 2007a,b; Hain et al., 2012). The ALEXI model is a two-source model that estimates the partition of surface evaporation and plant transpiration from the total system ET. Currently, the ALEXI model is operational over CONUS at 0.04° and daily resolution; however, although the model is operational at a daily time step, the retrievals are only available on substantial cloud free locations relative to a GOES IR pixel (Hain et al., 2011; Mishra et al., 2013). Potential surface evaporation, defined here as the atmospheric demand, is computed using Hamon PET (Hamon, 1963) and is independent of the underlying soil and plant characteristics and therefore, acts as a proxy for potential surface evaporation.

4.3 Data Description

4.3.1 SMAP Soil Moisture Data

The coarse resolution L-band MW SSM product from SMAP-Passive (SMAP-P) was used as an input to the disaggregation algorithm. Whereas the intermediate (9-km SMAP-Active/Passive and SMAP-Enhanced) and fine (3-km SMAP-Active) SM products from the SMAP mission were used for comparison and validations purposes. The Active radar (SMAP-A at 3-km) and Active/Passive combined (SMAP-
AP at 9-km) products are available from April 2015 to July 2015, while the SMAP-P and SMAP-E SSM products are available from March 2015 to present. The Level-3 daily SMAP products are projected over fixed ease-grid at 36-km (Passive), 9-km (Active/Passive and Enhanced) and 3-km (Active) resolutions. The 1,000-km wide swath allows SMAP 2-3 day global revisit, however this creates data gaps at a daily time step. Therefore, a 3-day composite SM dataset was created at all three resolutions during the study period.

4.3.2 ALEXI Surface Evaporation

A continental scale implementation of the TIR-based ALEXI model was used in this study. The 5-km ALEXI product is ideal for this study since its resolution falls neatly between the 3-km and 9-km SMAP products. The gridded surface evaporation from ALEXI was resampled to 3 and 9-km consistent with the SMAP resolution using the nearest neighbor technique. Due to the cloud free constraint, despite being a daily product, ALEXI data are available only on substantially cloud-free locations within a GOES satellites field-of-view. To fill data gaps and maintain consistency with SMAP SM products, a three day composite of ALEXI surface evaporation was computed over CONUS consistent with that done for the SMAP data.

4.3.3 NRCS SCAN Observations

Ground-based observations of surface volumetric SM were available from NRCS SCAN sites. A total of 228 active SCAN sites are present in the study area, however not all stations reported surface SM data over the study period. SCAN stations pe-
periodically monitor multiple meteorological parameters such as precipitation, air temperature, relative humidity, etc. along with SM and temperature at various depths at near real time with hourly and/or daily sampled time steps. This study utilizes the SM measurement from the top 2 inches (∼5 cm) using a Hydra Probe instrument (Bell et al., 2013; Schaefer et al., 2007). The SCAN sites, despite having low density compared to the gridded 3 to 36 km footprints of satellite-derived SM datasets, cover a wide range of soil and climatic conditions across the CONUS. Figure 4.1 shows the location of all the active sites used in this study within the CONUS domain (Baily, 1995).

In addition to above mentioned datasets, daily air temperature data were also used in this study. The North America Land Data Assimilation System (NLDAS2) air temperature forcing data at 0.125° resolution was used to compute Hamon PET. The forcing data were obtained from the NASA Land Data Assimilation System. Terrain adjustment of coarse resolution temperature data was performed using a 30-m digital elevation map (GTOPO30 digital elevation model) with a constant lapse rate for the study region. The GTOPO30 elevation map for the CONUS was obtained from the U.S. Geological Surveys EROS Data Center.

4.4 Methodology

4.4.1 Surface SM Disaggregation

As the aim of this study is to test the effectiveness of an alternate scheme to disaggregate the SMAP radiometer SM product, with the early mission malfunc-
tioning of the SMAP radar, the search for effective alternatives is of high priority within the agricultural and hydro-meteorological communities (Chen et al., 2017). A semi-empirical physically based disaggregation scheme introduced by Merlin and his associates (Merlin et al., 2013, 2012, 2008), called DISaggregation based on Physical And Theoretical scale CHange (DISPATCH), was used in this study. The disaggregation approach is based on underlying SEE, which is a model used to map surface evaporative fluxes to the moisture content at finer scales. Its basic premise is that the ratio of actual to potential surface evaporation is scale invariant and related to surface SM. As pertinent to this study we re-present the equation of the scheme that reflects the fundamental theoretical basis of the algorithm:

\[
SM_{HR} = SM_{LR} + \left( \frac{\partial SM_{mod}}{\partial SEE} \right) \left[ SEE_{HR} - \langle SEE_{HR} \rangle_{LR} \right]
\] 

(4.1)
Here, HR and LR refer to high and low resolution variables, respectively. The SEE is computed initially at the native ALEXI higher resolution \(0.04^\circ\) and then resampled to lower resolutions. \(M\) is the partial derivative function that relates SEE to the underlying SM content. \(\langle SEE_{HR}\rangle_{LR}\) is high resolution SEE aggregated to low resolution MW scale.

Multiple models have been proposed that describe the relationship between SEE and surface moisture content in the past. In earlier studies, (Merlin et al., 2012, 2008) employed variants of non-linear relationships by (Komatsu, 2003; Lee and Pielke, 1992; Noilhan and Planton, 1989). A comparative study by Merlin et al. (2010b) suggests that the non-linear model suggested by Noilhan and Planton (1989) was superior to the other non-linear models. Recent studies by Merlin et al. (2013, 2015) indicated that a linear model performed better than earlier proposed non-linear methods over relatively dry climatic conditions of South Australia and Spain. This study employs the non-linear model suggested by Noilhan and Planton (1989) to assess the effectiveness of each across the CONUS as well as at regional scale, and given as:

\[
M_{LR} = \frac{SM_{LR}}{\cos(-1)(1 - 2SEE_{LR}\sqrt{SEE_{LR}(1 - SEE_{LR})}}
\]

### 4.4.1.1 Modified SEE computation

The SEE in the original DisPATCH model is computed based on the triangle approach using MODIS LST, VI and surface albedo. However in this study, the actual ratio of surface evaporation to the potential surface evaporation was used to compute
SEE at high resolution. The surface evaporation was obtained from the ALEXI model and the potential ET (PET) was estimated using the Hamon PET model (Hamon, 1963) as a proxy for potential surface evaporation. Hamon PET is solely dependent upon atmospheric demands that are completely decoupled from the underlying soil and canopy characteristics. Therefore, the model can be used as a proxy of potential surface evaporation. The Hamon PET is computed as:

$$H_{PET} = K \cdot (35.755) \cdot N \cdot \frac{e_s}{(T + 273.3)}$$ (4.3)

$K$ is the proportionality constant used as 1. $N$ is the daylight hours in multiples of 12 and $e_s$ is the saturated vapor pressure at the given temperature $T$ (${0^\circ C}$) which is given as: $6.108e^{\frac{17.267}{(234.5+T)}}$, where $T$ is the mean daily temperature. The terrain adjusted daily min/max temperatures from the NLDAS2 forcing data are used to compute daily mean temperatures. Terrain adjustment of the coarse resolution temperature data were performed using a 30 m digital elevation map of the region and a constant lapse rate of $-6.5K km^{-1}$ (Cosgrove, 2003). This represents a subtle change in the definition of SEE from the Merlin formulation, as in our case all land cover/soil matrix combinations are weighted equally as opposed to being weighted by their assumed PET value as in Merlin (approximated as function of surface temperature).
4.4.2 Evaluation Matrices

The aim of this study is to evaluate the applicability of ALEXI-driven modified disaggregation scheme to estimate SM at finer scales. Therefore, TIR-downscaled SM data were compared and validated against remotely sensed SMAP soil moisture products at corresponding resolutions along with in-situ observations from SCAN sites across CONUS. The data gaps in all three datasets restrict time series analysis, hence pair-wise spatial and temporal statistical comparisons were performed using traditional matrices such as: bias, root mean squared error (RMSE) and correlation ($r$). It has been argued that the traditional RMSE can be an overestimated if a bias exists either in model or reference dataset (Entekhabi et al., 2010b). Therefore, an unbiased estimation of RMSE (ubRMSE) is computed by removing the potential impact of bias in the error estimation.

As there is a spatial mismatch while comparing gridded SM estimations with in-situ observations, sampling errors can occur (Peng et al.). Multiple upscaling algorithms have been suggested for sparse in-situ monitoring stations to minimize the impact of sampling error; however these methods typically require a dense network of such stations in addition to an independent a-priori error characterization (Crow et al., 2012). One possible alternative is the computation of gain statistics. Merlin et al. (2015) have proposed a performance matrix to compute relative gain in slope, correlation and biases to measure the overall improvement of downscaled SM estimates over coarse resolution data with reference to a given set of point observations. Similar relative gain statistics are used here to assess the quality of SMAP and down-
scaled surface SM with SCAN observations as the reference. The relative gain in slope ($G_{Eff}$: efficiency gain); gain in correlation coefficient ($G_{Prec}$: precision gain); and gain in bias ($G_{Acc}$: accuracy gain) are computed as:

$$G_{Eff} = \frac{(|1 - S_{LR}| - |1 - S_{HR}|)}{(|1 - S_{LR}| + |1 - S_{HR}|)}$$ (4.4)

$$G_{Prec} = \frac{(|1 - R_{LR}| - |1 - R_{HR}|)}{(|1 - R_{LR}| + |1 - R_{HR}|)}$$ (4.5)

$$G_{Acc} = \frac{(|B_{LR}| - |B_{HR}|)}{(|B_{LR}| + |B_{HR}|)}$$ (4.6)

Here $[X]_{LR}$ refer to low resolution SM statistics [S: slope; R: Correlation and B: Bias] against in-situ observations whereas $[X]_{HR}$ refer to the statistics of the high-resolution SM against the in-situ observations. The overall gain ($G_{Down}$) is the simple unweighted mean of the partial independent relative gains (Merlin et al., 2015). Relative gain statistics are advantageous over traditional statistics in that they measure the relative performance of two SSM datasets directly against the target data. Further, the relative nature of the matrix makes it less sensitive to bias in the mean or in the variance. It tends to reduce the uncertainties associated with the mismatch in spatial scales of in-situ and remotely sensed data (Merlin et al., 2015).
4.5 Results

The 2-3 day revisit cycle of the SMAP and cloud constraints on ALEXI makes both datasets prone to data gaps at a daily time-step. Therefore, a 3-day centered moving window compositing scheme was applied to all remotely sensed SSM estimates. In addition to filling data gaps, these composite techniques tend to reduce day-to-day noise associated with satellite products (Anderson et al., 2011a). The disaggregation scheme described in section 4.1 and 4.2 was applied to the coarse resolution SMAP radiometer SSM product over the CONUS. To evaluate the performance of the TIR-Downscaled scheme, the disaggregated SSM estimates were compared spatially and temporally against the available and corresponding SMAP SSM products. To further evaluate the performance, remotely sensed SM estimates from both SMAP and TIR-Downscaled SSM data were further compared and validated against all available SCAN site observations. The following section details the results of comparisons and validation, first among remotely sensed products and then with in-situ observations. Figure 4.2 displays the composited SM conditions from SMAP (P, A, AP, and E), as well as the TIR-downscaled (3- and 9-km scales) for a single day (Julian day 159) during the summer of 2015 over CONUS.

4.5.1 Spatial Analysis

SSM products from SMAP (A, AP and E) and TIR-downscaled data (9- and 3-km resolutions) were compared over the CONUS grids and the average statistics are shown in Figure 4.3. At 9-km resolution, the mean spatial correlation between
Figure 4.2: SM estimates from SMAP at coarse resolution Passive (36-km); Active (3-km); Active/Passive (9 km); and Enhanced product (9-km) compared with TIR-Downscaled SM data (3 and 9-km) on 8 June, 2015. The white spaces indicate no data availability.

SMAP-AP and TIR-downscaled SM was 0.76 with an overall ubRMSE of 0.09 and a negative bias of -0.013. Compared with SMAP-E SSM product, the TIR-Downscaled SSM showed average $r$ of 0.90 with ubRMSE of 0.057 and bias of -0.01. The SMAP-AP and SMAP-E SSM had a correlation coefficient of 0.84, ubRMSE of 0.09 and bias -0.003.
A similar grid analysis of the SSM signals was performed at 3-km resolution between SMAP-A and TIR-downscaled (3-km) SSM estimates and the results are also shown in Figure 4.3. The similarity of the 3-km SSM products (SMAP-A TIR-Downscaled) weakens considerably relative to 9-km products. The average $r$ between the SMAP-A (active radar) SM measurement and TIR-based downscaled SSM was 0.29. The ubRMSE was found to be 0.14 and bias was 0.008. The overall mean bias was close to zero ($= 0.008$) however the daily standard deviation (SD = 0.017) was double of the mean.

**Figure 4.3**: A daily time series of spatial correlation (top); bias (middle) and coefficient of ubRMSE (bottom) at 9-km and 3-km spatial scales over CONUS between SMAP and TIR-Downscaled SSM products.
4.5.2 Temporal Analysis

Temporal analysis at each pixel is limited by the number of days corresponding SSM products coincide. Figure 4.4 shows the map of statistics at 9-km resolution between SMAP-AP, E and TIR-Downscaled SSM products over CONUS. The overall mean temporal correlation between SMAP-E and TIR-downscaled SSM over CONUS (right panel) was found to be 0.87 with ubRMSE of 0.03 and bias at -0.03. Comparison with SMAP-AP the TIR-Downscaled SSM (middle panel) showed an overall $r$ of 0.71, ubRMSE = 0.05 and bias of 0.065 temporally but for a sample size of only 3 months. The SMAP-AP compared with SMAP-E (left panel) showed $r$ of 0.75 and ubRMSE of 0.04 with bias = 0.06 again with the smaller sample size. These results indicate that the 9-km TIR-downscaled SSM most strongly relates to the SMAP-E with high correlation and low ubRMSE values followed by the SMAP-AP SSM product.

In terms of 3-km SSM products (SMAP-A vs TIR-Downscaled), $r$=0.27, with an ubRMSE of 0.097 and bias 0.011. Figure 4.5 shows the map of temporal statistics between the two SSM products. Though it can be seen from Figure 4.5 that both the 3-km SSM products are still most similar in the West-Central U.S. (with $r > 0.6$ and ubRMSE < 0.07), yet the distinction is not as clear as in the 9-km products of similar time frame. The overall bias at the 3-km scale is lower than the 9-km products [0.011 vs 0.065 (with SMAP-AP) and 0.028 (with SMAP-E)], however the variance in bias across CONUS is 0.015 which is 2 and 7 times higher compared to bias in SMAP-AP and SMAP-E, respectively. The higher variance in 3-km indicates a relatively greater spread and instability in results across CONUS despite the low overall mean bias.
Figure 4.4: Map of CONUS displaying statistics between SMAP-AP, E and TIR-downscaled SM at 9-km scale: correlation coefficient (top); Bias (middle) and ubRMSE (bottom) distribution across CONUS for the period of Apr-June 2015 (left two panels) Apr 2015 - Nov-2016 (right panel)

Again, it should be noted that these results are for a sample size of only 3 months while the 9-km results are based on a 19 month sample size.

4.5.3 Comparison with SCAN Observations

The remotely sensed SSM estimates from SMAP (A, AP, E and P) along with TIR-Downscaled (3 and 9-km) were compared with SCAN site in-situ observations across CONUS. While comparing remotely sensed SSM to in-situ observations, disparity at spatial scale as well as the sensing depths must be considered. Some authors prefer to remove the bias due to scale difference before comparisons (Brocca et al.,
2011), however it is common practice to compare in-situ observations without adjusting for scale even when only one observation per pixel is available (McCabe et al., 2005; Sahoo et al., 2008). In this study, remotely sensed SSM estimates are compared directly without bias correction or upscaling of in-situ observations. Although, the absolute value of SM varies spatially at much finer scales (∼ few meters), the temporal dynamics are found to be highly correlated spatially, indicating that the temporal SSM dynamics can be compared between datasets of varied spatial scales (Seneviratne et al., 2010).

A total of more than 180 SCAN sites over CONUS were active and provided daily summary of SM and other meteorological observations (such as, soil temperature, humidity, etc.) during the study period. SSM observations (≤2 inch (∼5cm) depth) were collected from SCAN sites for comparisons with remotely sensed SSM products. Table 4.1 shows the overall statistics of the remotely sensed SSM compared with the SCAN observations over CONUS. The overall correlation between SCAN observations and coarse resolution SMAP-P SSM data was 0.54. Mean bias at all sites
was -0.02 and ubRMSE of 0.06. The intermediate resolution SMAP-E was found to have similar statistics although the correlation was slightly lower (0.49). The finer resolution SSM data from the active radar on the other hand, showed relatively less similarity with SCAN observations ($r = 0.16$, ubRMSE=0.077). Although there is a slight improvement in overall bias compared to coarser resolution SMAP-P and E estimates (0.008 vs -0.022). The SMAP-AP, a combination of passive and active, had statistics relatively better than SMAP-A but poorer than SMAP-P. There is a slight disparity in sample size in case of SMAP-A and AP that should be taken into account while interpreting the results.

**Table 4.1**: Summary statistics between remotely sensed SSM and SCAN observations across CONUS with average number of SCAN sites and days available for comparison.

<table>
<thead>
<tr>
<th>SM Product</th>
<th>No. of sites</th>
<th>No. of days</th>
<th>$r$</th>
<th>Bias</th>
<th>ubRMSE</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMAP-P</td>
<td>181</td>
<td>563</td>
<td>0.54</td>
<td>-0.021</td>
<td>0.062</td>
<td>0.47</td>
</tr>
<tr>
<td>SMAP-A</td>
<td>156</td>
<td>54</td>
<td>0.16</td>
<td>0.008</td>
<td>0.077</td>
<td>0.19</td>
</tr>
<tr>
<td>SMAP-AP</td>
<td>144</td>
<td>69</td>
<td>0.37</td>
<td>-0.006</td>
<td>0.069</td>
<td>0.49</td>
</tr>
<tr>
<td>SMAP-E</td>
<td>182</td>
<td>570</td>
<td>0.49</td>
<td>-0.022</td>
<td>0.062</td>
<td>0.40</td>
</tr>
<tr>
<td>TIR-Down(3K)</td>
<td>181</td>
<td>306</td>
<td>0.47</td>
<td>-0.019</td>
<td>0.064</td>
<td>0.42</td>
</tr>
<tr>
<td>TIR-Down(9K)</td>
<td>180</td>
<td>300</td>
<td>0.47</td>
<td>-0.019</td>
<td>0.064</td>
<td>0.41</td>
</tr>
</tbody>
</table>

The TIR-Downscaled SSM when compared with SCAN observations, showed statistics similar to SMAP-P and -E products. It can be noticed that the statistics are identical for both the 3-km and 9-km resolutions. The overall ubRMSE increased slightly from 0.062 to 0.064 but there is a gain in bias (-0.022 to -0.019) compared to the SMAP-P SSM estimate. In addition, there was a slight decline in correlation for the downscaled SM to 0.47 compared to 0.49 for the SMAP-E, but better than the 0.37 exhibited by the SMAP-AP (albeit with a much smaller sample size). Inter-
estingly, the correlations of both the SMAP and TIR relatively finer scale products were less than that of the coarser SMAP-P product itself. The overall results indicate that the downscaled SSM products either SMAP-E or TIR-Downscaled showed overall statistics similar to the coarse SMAP-P. In case of SMAP, the brightness temperature from the same source is being used with a similar algorithm to deduce passive and enhanced SSM products. Therefore, similarities between the products are expected.

The TIR-down, on the other hand uses TIR derived evaporative efficiency in addition to passive MW SSM to guide the disaggregation algorithm. Therefore, some similarities can be expected with passive MW when TIR driven SEE at finer and coarse scale are similar i.e. relatively homogenous landscape. But for heterogeneous landscapes, the SEE based algorithm is expected to provide physically based additional details on the underlying SSM state.

4.5.3.1 Gain Statistics

As mentioned earlier in section 4.2, the scale mismatch between \textit{in-situ} observations and gridded remotely sensed SSM data can induce sampling error; therefore gain statistics were computed between coarse and finer resolution SSM data simultaneously against \textit{in-situ} observations. Figures 4.6 and 4.7 display the map of gain statistics across CONUS at 3 and 9-km resolutions, respectively. The figure indicates that a majority of sites with negative overall gain was observed with SMAP-A (3-km) data particularly in the western U.S. as well as the Mississippi River valley. In the case of TIR-Downscaled, most of the overall gains were small i.e., within ±0.1 (<91%). In all cases (efficiency; precision; and accuracy), the percent of sites with positive gains
in TIR-Down (3-km) is higher than the SMAP-A by a factor of nearly 3 (Figure 4.8). Except for gain in bias (accuracy), at least 59% of sites observed positive gains with TIR-Downscaled (3-km) data with more than 65% of sites observing positive gains in efficiency (correlation). Compared to SMAP-A data, the percent of sites with positive gains was more than double in all cases, except for bias, indicating a relatively better representation of finer scale SM condition by TIR-Downscaled than SMAP-A against in-situ observations. Although, the number of sites with positive gains in accuracy (bias) was less than 50% for TIR-Downscaled, it was greater than the positive gains greater than the SMAP-A (37 vs 45%), indicating a relatively modest improvement.

These results clearly indicate that there is a relative improvement in SSM estimate quality in moving from coarse resolution passive MW to TIR-based 3-km resolution with respect to in-situ observations at a majority of the locations. However, the overall improvement is modest at best.

**Figure 4.6**: Overall gain statistics at 3-km spatial resolution SSM products: SMAP-A (Left) and TIR-Down (Right)

At 9-km resolution, the relative overall gains in disaggregating passive MW SSM estimates from SMAP to SMAP-AP, SMAP-E and TIR-Downscaled (9-km)
compared to SCAN observations are shown in Figure 4.7. Similar to SMAP-A, the overall gain in SMAP-AP is noisier with values ranging towards the extreme ends in both directions. Further, less than half of the total SCAN sites observed positive gains in efficiency, precision and accuracy with SMAP-AP data (Figure 4.8). Only 37.8% of SCAN sites had positive overall gains in moving from coarse resolution passive to intermediate resolution SMAP-AP SM estimate. On the other hand, more than 50% of SCAN sites observed positive gains in both SMAP-E and TIR-Downscaled SSM estimates for all the cases. Although at the majority of sites the SSM quality was improved with SMAP-E data, the number of sites with positive gains is even higher with TIR-Downscaled (9-km) compared to SMAP-E in all cases, but most particularly in the precision statistic. Figure 4.8 shows the percent of sites with positive gains with SSM data at 9-km resolution compared to coarse resolution passive MW and SCAN point observations. The results from gain statistics suggest that there is a clear improvement in representation of SSM at the intermediate scale with SMAP-E data compared to SMAP-AP product. More than half of the locations with positive gains, indicate that the intermediate scale SM from SMAP-E is of superior quality than coarse resolution passive MW against in-situ observation. The TIR-based SM at both scales (3 and 9-km resolution) appears to slightly better represent the SM conditions at higher resolution compared to other products with the maximum number of sites having positive gains. Again, however, the difference between the SMAP-E and TIR-Downscaled products is very small in most cases.
4.6 Discussion

4.6.1 Ecological Domain

To further examine the possible spatial patterns and effects of vegetation and climate on the disaggregation, the United States Forest Service (USFS) ecological domains were used (Baily, 1995). The USFS splits the CONUS into three major eco-
logical domains: humid tropical; humid temperate and dry representing contrasting vegetation and varied bioclimatic conditions (Figure 4.1). Out of total 181 available and selected SCAN sites, 91 fall under the humid temperate domain, 90 sites under the dry domain, with only 1 site (not used) under the humid tropical designation. Figure 4.9 shows the boxplots of statistics of the SMAP and TIR-Downscaled SSM data against SCAN observations at each location for the period during which the active radar was operational (3 months). It is understood that the short sample size of these results limits their reliability; however, they do offer an opportunity to judge the efficacy of the active radar against other methods even if in a qualitative, or anecdotal, manner.

In general the SMAP products using the active radar tend to display a little higher ubRMSE and to have poorer correlation with the in-situ data than do the TIR-Downscaled products at both resolutions across all ecological domains. In terms of bias, the radar based data and the downscaled TIR SSM are similar at the 9-km resolution while at the finer (3-km) resolution, the SMAP active radar product shows a slight negative bias in dry regions and positive bias in temperate climates. On the other hand, the TIR based data display a slight positive bias throughout. Overall, based on this relatively small sample size, one would judge the downscaled TIR SSM to be slightly superior to the active radar product.

The results of the comparisons over the longer time period (19 months) are shown in Figure 4.10. This figure compares the SMAP-E and the TIR-Downscaled SSM products at the 9-km resolution as well as the 3-km TIR-Downscaled SSM against the SCAN observations. Taking the 3-km longer term TIR-Downscaled SSM
Figure 4.9: Boxplot displaying the statistics distribution between SCAN observations with SMAP-A/AP and TIR-Downscaled SSM at 3 and 9-km spatial scale. (3 month data period).

first, the statistics are slightly better than the shorter term values shown in Figure 4.8 in some respects. While the correlation remains around 0.5 and the ubRMSE around 0.05 the overall bias has been reduced to near zero. However, the bias in temperate regions has changed from slightly positive to a negative 0.05. Comparing the 9-km SSM products, the SMAP-E and TIR-based downscaled statistics appear very similar across all ecological domains. The most notable attribute seems to be the negative bias in temperate regions displayed by all the SSM products. While there is a slight
improvement in $r$ in the temperate domain compared to the dry region, the ubRMSE tends to be relatively higher in the temperate domain. Some variations do exist, but overall it seems that the remotely sensed SSM products (except for SMAP-A and AP) from SMAP and TIR-derived downscaled are relatively stable across domains for the study period. Although, the mean values of SMAP-E and TIR-Downscaled SSM are close, yet Figure 4.10 indicates a slightly larger spread in TIR-Downscaled data compared to SMAP-E. Involvement of an additional remotely sensed product, TIR derived surface evaporation, seems to be the reason for slight increase in the noise compared to SMAP-E which is derived from single remotely sensed observation. The additional remotely sensed product present in TIR-Downscaled SSM, although noisier, is relatively more physical in nature than the interpolated brightness temperature being used in SMAP-E.

In term of gain statistics ecologically, with respect to SMAP-E, as shown in Figure 4.11 both domains showed nearly the same percentage of sites with positive gains (55 and 53%). Although less than half of the total SCAN sites showed positive gains in SMAP-A and SMAP-AP SSM, a higher percentage of sites in the humid temperate domain showed positive overall gains in both compared to the dry domain. However, the TIR-Downscaled SSM showed a reverse trend, as a higher percentage of sites in the dry domain (>60%) were found to have positive gains compared to the humid temperate domain (<51%). Since nearly the same percentage of sites showed positive overall gains for the SMAP-E product for both domains, this would indicate that the combined effect of bias, correlation, and slope is domain independent for SMAP-E SSM product. Conversely, in the case of TIR-downscaled data (both 9
Figure 4.10: Boxplot of statistics between SMAP-E and TIR-Down (3 and 9-km) resolution SSM products against SCAN observations as a function of ecological domain. (19 month data period)

and 3-km resolution), the disaggregation in the dry domain appears to be superior compared to the humid temperate domain. However, the number of sites showing positive gains in the dry domain was considerably greater for the TIR-Downscaled SSM than for the SMAP products.
4.6.2 Effect of Vegetation Cover

It has been argued that the MW SSM signals are attenuated by thick vegetation cover, especially with higher frequency bands like C- and X- (Albergel et al., 2011; Brocca et al., 2011). With L-band radars, like that of SMAP, the sensitivity to vegetation cover is comparatively reduced, yet errors are still higher over vegetated land surfaces compared to bare soils (Konings et al., 2017). With the ALEXI model, sensitivities decrease as surface moisture content reaches either the wilting point or field capacity (Hain et al., 2011). The partitioning of system (canopy + surface) energy fluxes to surface evaporation in the ALEXI model is limited by the fraction of vegetation cover. The vegetation effects of both the SMAP and ALEXI products could, in part, explain the spatial disparities identified in the east (and far west) and the more central/western states (Figures 4.4 and 4.5). In this section, we analyze the

Figure 4.11: Percent of SCAN sites showing positive gains by ecological region.
effect of vegetation cover on coarse and disaggregated SSM using an independent third SSM source, NLDAS2 (Xia et al., 2012) (Mosaic of Noah and Variable Infiltration Capacity (VIC) LSMs). The analysis does not assume that the LSMs are accurate; models may have their own biases and errors associated with them. The assumption is that the physically-derived SSM from LSM models will not have any vegetative effects associated. The analysis performed using LSM are only to assess the relative dynamics of both remotely sensed SSM products under various vegetative scenarios against a common independent data source. Due to limited data availability resulting in small sample size, as well as their relatively poor performance in the previous analyses, the SMAP-A and -AP products are omitted from this analysis.

Figure 4.12 shows the annual mean fraction of vegetation cover derived using MODIS LAI (Anderson et al., 2011b) over CONUS for the year 2016. In most of the central and western part of CONUS, mean vegetation cover is less than 40%, thus the surface conditions are readily accessible through both MW and TIR based sensing platforms. The frequency distribution of the statistical comparison between SMAP-E and TIR-Downscaled (9-km) SSM as a function of mean fractional vegetation cover is shown in Figure 4.13. The figure clearly indicates the effect of vegetation cover on the statistical relationship between the two soil moisture products. With vegetation cover less than 40%, both SM products seems to be strongly related with \( r > 0.75 \) (bias nearly 0.0 and ubRMSE < 0.03). However, a sharp decline in \( r \) with a simultaneous steep rise in bias and ubRMSE was observed with vegetation cover beyond 70%. For vegetation cover between 40 and 70%, the correlation drops but the fall is relatively less steep compared to vegetation cover of greater than 70%.
Figure 4.12: Mean annual fraction of vegetative cover over CONUS based on MODIS LAI (2016).

Figure 4.13: Comparison of SMAP-E and TIR-Downscaled SSM statistics as a function of fractional vegetation cover (9-km).

Figure 4.13 shows the effects of vegetation cover on remotely sensed SSM products; however, the analysis does not illustrate the effects of vegetation on individual datasets. Therefore, the NLDAS2 SSM product was used as an independent measure to assess the vegetative effect on the individual remotely sensed SSM products.
Triple Colocation (TC) analysis was utilized to compute the error statistics between
the three datasets. TC has become a very popular technique for simultaneous error
analysis of three data sets since its adaptation to SM states by Scipal et al. (2008).
The procedure is based on the assumption of linear relationships between the three
estimates of the SM at a specific location and the unknown true value. The un-
known truth is eliminated from the linear error equations through subtraction and
then cross multiplied to determine the error variances of the datasets relative to each
other (Gruber et al., 2016). The assumption is that the errors in the three datasets
are independent and random. Multiple recent studies have used the TC method for
error estimation [such as (Crow et al., 2015; McColl et al., 2014; Yilmaz et al., 2014;
Zhang et al., 2017) etc.]. A detailed review of method derivations and application to
SM error estimation and analysis is presented by Gruber et al. (2016).

The results of the TC analysis are shown in Figure 4.14 below. The figure
shows that the NLDAS2 errors are fairly stable across vegetation coverages and that
the TIR-Downscaled tracks the NLDAS2 data very well at vegetation fractions above
60%. Conversely, the SMAP-E product does not associate with the NLDAS2 at any
vegetation cover. However, the TIR-based and SMAP-based SM track better at the
lower vegetation fractions, particularly in the range of 20-40%.

Perhaps not coincidentally, it is in the lower vegetation covers that the re-
motely sensed SSM is closer to the true values than are the NLDAS2 estimates. The
TIR based product appears best in the bare soil region (<0.2) but deteriorates more
at the higher vegetation fractions than does the SMAP-E product. It should be noted
that there is a discrepancy in the SSM layer depth definition of the NLDAS2 product.
NLDAS2 had surface SM defined as mean moisture content between 1-10cm depth whereas MW and TIR-based SSM are estimates of typically less than 5 cm depth.

### Error Characterization

The SM content at the surface is the most variable across depth temporally (Starks et al., 2003). However, recent studies by Penna et al. (2013) showed that the SSM dynamics are strongly correlated for a temporal lags less than 5 days. Further, satellite data can be noisier at a daily time step, therefore temporal compositing can be used to reduce daily variability while retaining the temporal dynamics of the SSM (Anderson et al., 2011a). A 3-day centered moving window compositing has been performed to fill in some of the data gaps associated with remotely sensed SSM datasets. The temporal compositing, despite reducing data gaps, can introduce errors into the dataset.
In addition, accuracy of SEE based disaggregation model is dependent upon the accuracy of: (a) SEE estimation and (b) the relationship between SSM and SEE. SEE accuracy can be associated with ALEXI estimation of surface evaporation. As mentioned earlier, ALEXI estimates the total ET and then partitions between the surface and root zone which leads to errors in surface evaporation especially in areas of high vegetation cover (Figure 4.13 and 4.14). Further, the assumption behind using Hamon-PET as a proxy of surface potential evaporation, could further add to the error in SEE estimation. Next, the use of the linear vs non-linear model to relate SEE with SSM is still unclear. Earlier studies [such as Merlin et al. (2012)] used the non-linear approach, while later analyses [such as Merlin et al. (2015)] showed that the linear model performed better than non-linear in dry and arid conditions of Australia. However, recent studies by Djamai et al. (2015) and Mishra et al. (2017) suggested that the non-linear models are better suited for wet and humid climatic conditions than the linear model. This study employed the non-linear model throughout CONUS including the dry domain in the Western U.S.

4.7 Conclusion

This study investigated the effectiveness of the SMAP downscaled products against the soil evaporative based disaggregation scheme over CONUS compared to in-situ data from 180+ USDA observation sites. The study evaluated the performance of the downscaled SM and the SMAP SSM estimates at both 9- and 3-km spatial scales consistent with SMAP SM products.
Based on the limited sample available, the 3-km SMAP products based on the active radar statistics were inferior to the other SSM products with the exception of bias. There was a considerable deterioration in the SMAP 3-km product with the introduction of the active radar as the correlation with SCAN data reduced to 0.16 and RMSE of 0.14. However, the active radar data displayed a low bias of only 0.008. The radar performed most poorly in the Western U.S.

The 9-km SMAP-E and TIR-Downscaled products offered only modest improvements (at best) to the coarse scale SMAP-P (36-km) SSM in terms of overall statistical comparison to the SCAN data. When viewed spatially, there were substantial improvements in some locations across CONUS, particularly in arid climates.

At the 9-km spatial scale, the TIR-Downscaled SSM data correlated strongly with the SMAP-E SSM product both spatially and temporally. The two SSM products also performed very similarly to the SCAN data across all ecological domains with overall correlations of 0.49 (SMAP-E) and 0.47 (TIR-Downscaled) and bias of -0.022 and -0.019, respectively with RMSE of 0.105 and 0.100.

Since both the 9- and 3-km downscaling was based on resampling of the ALEXI TIR data from its native 5-km resolution, perhaps not surprisingly, the statistics of the 3-km downscaled TIR data were similar as in the 9-km case. Clearly, the resampling did not materially affect the results.

It should be noted that the results of SMAP-A and SMAP-AP comparisons are based on a low sample size of only 3 months while SMAP-E are based on 19 months.
of data. The results of the present study are similar to those recently reported in the literature at varying spatial scales and locations: Chen et al. (2017) – \( r: -0.3-0.72, \) RMSE: 0.06-0.27; Malbéteau et al. (2016) – \( r: 0.70-0.94, \) RMSE: 0.07-0.09; Merlin et al. (2015) – \( r: -0.22-0.64, \) RMSD:0.05-0.32; Molero et al. (2016) – \( r: 0.35-0.47, \) ubRMSE:0.04-0.12. Most of these earlier studies are site specific with multiple in-situ observations possibly within a single pixel resolution and thus offer better representation of the SSM conditions. However, in this study only single in-situ observations per pixel were available and the approach was applied at the continental scale. Despite these differences, the correlation and RMSE results are comparable to earlier studies.

Due to cloud constraints on the ALEXI retrievals, the TIR-based disaggregation approach has more potential for long-term agricultural and hydrological analysis rather than operational implementation. The relatively short time period of the study (19 months) also limits its reliability. Therefore, a further detailed study is required to completely assess the applicability of the disaggregation scheme over a wide range of seasons and climate. Given the limited scope of this study, a few improvements can be suggested:

- A detailed multi-year assessment of the downscaled SM across seasons is required to fully understand the inter-annual variability and efficiency of the scheme.
• A detailed study is required to quantify the ALEXI surface evaporation and the use of Hamon PET as a proxy of surface potential evaporation towards the computation of SEE.

Here, an ALEXI driven disaggregation scheme was presented as a promising alternative toward obtaining finer scale soil moisture estimates. For hydro-meteorological and agricultural applications an intermediate spatial scale of 9-km is preferred to the coarse radiometer scale, and the disaggregation scheme was found to be efficient at the 9-km scale. Further, the scheme is found to be most efficient under low to moderately thick vegetation cover; therefore can supplement agricultural applications effectively.
CHAPTER 5

ASSIMILATION OF COUPLED MICROWAVE/THERMAL INFRARED SOIL MOISTURE PROFILES INTO A CROP MODEL

To be submitted to Agriculture and Forest Meteorology

Abstract

Global food security is one of the most pressing issues of the current century, particularly for developing nations. Agricultural simulation models can be a key component in testing new technologies, seeds and cultivars etc. However, inaccurate input information in addition to model related errors adds to model uncertainties. Satellite observations of soil moisture, vegetation index etc. can be assimilated into crop models to reduce model uncertainties. This study utilizes a satellite derived microwave and thermal-infrared coupled soil moisture profile assimilated into a crop model via Ensemble Kalman Filter over parts of Southeastern U.S. from 2006-2010. The National Agricultural Statistical Services (NASS) reported yield data at county levels were used for comparison and validation purposes. The rainfed yields in comparison with the reported NASS yields showed an overall absolute error of nearly 38% whereas the error in yields after data assimilation was only 12%. Assimilating remotely sensed SM profiles into the crop model improved the yield estimation considerably in irrigated regions with average errors (< 3%) whereas for relatively non-irrigated regions the mean error with DA was nearly 19%. Overall, over all regions together, by assimilating remotely sensed soil moisture profiles into a rainfed crop model the errors were reduced by a factor of 3 compared to rainfed yield errors against NASS reported yields.

5.1 Introduction

Growing population, competing use of resources, and climate uncertainties have resulted in significant stress on world food supplies. It is estimated that an in-
increase in crop production of nearly 70% will be required by 2050 to meet the demands (FAO, 2009). Crop production can be increased either by expanding agricultural land area or through improved agricultural technology (Tilman et al., 2011). Land resources are finite and only limited intensification is possible. However, technological advances through development of high yielding seeds, drought resistant crops, etc. could supplement production.

Agricultural simulation models can be a key component used to test new technologies and seed cultivars and thus in addressing issues of global food security. Crop models typically depend on accurate estimates of numerous inputs, ranging from human induced management options (e.g., crop type and cultivar, fertilizers amount/type, irrigation method and timing) in addition to field information (e.g., soil type and characteristics) to weather data (e.g., temperature, precipitation, solar radiation). Of these required input types, crop management information can be especially difficult to obtain and hence is often parameterized over a region. Furthermore, many areas of the world lack adequate spatially and temporally consistent monitoring of weather data (at least 1 climate station per 5000 km²) (Aghakouchak et al., 2015). Sparse weather inputs, in combination with inconsistent management options, tend to increase uncertainties within crop models and produce unreliable results and thus substantially limit the applicability of such models as reliable analysis and decision-making tools (de Wit and van Diepen, 2007; Ines et al., 2013; Liu et al., 2016). Moreover, inherent model errors, due to model structure; inaccurate parameterizations; simplification of complex physical processes, etc., can further add to the overall model uncertainties (EL HAJJ et al., 2016; Hansen et al., 2006)].
Uncertainty in cropping system models may be mitigated by assimilating remotely sensed data into the model (Huang et al., 2016; Ines et al., 2013; Jiang et al., 2014). In particular, satellite estimates of soil moisture (SM), leaf area index (LAI), and evapotranspiration (ET) can be assimilated within a crop model. Several earlier studies (such as Dong et al. (2016); Huang et al. (2016); Jiang et al. (2014) etc.) assimilated remotely sensed vegetation information into a crop model which showed success in reducing errors in ET and plant biomass estimations, however little improvements in overall yield estimates was achieved. However, Mishra et al. (2013) updated crop model soil moisture states with thermal infrared (TIR)-based soil moisture profiles in the Southeastern U.S. Their results indicated that the model with the satellite-derived profiles was better able to estimate mean county yields than were rainfed only simulations. Although, the results improved the yields estimates, the study was limited to a single location and did not involve data assimilation as the entire model profile was merely replaced with a profile of remotely sensed SM. In addition, a number of assumptions were necessary in order to derive the TIR based SM profiles. Chief among these was that the surface SM value was assumed to be constant at 50% of saturation.

Despite its limitations, Mishra et al. (2013) demonstrated that the inclusion of remotely sensed root zone SM into the crop model may hold more promise for improving yield simulations than does the use of vegetation information. Hence, as an extension of the Mishra et al. (2013) study, the objective of this study was to improve the previously developed approach by: 1) utilizing microwave remotely sensed data to provide the necessary surface boundary condition on the remotely sensed profiles
and thus integrating the two remote sensing data sources; 2) assimilating the merged passive microwave (MW)-TIR based profiles into the crop model via the ensemble Kalman Filter; and 3) comparing the simulated yields to observed data at multiple locations.

5.2 Background

Integration of remotely sensed data into agricultural crop models has been ongoing for some time (Hansen et al., 2006; Huang et al., 2016; Ines et al., 2013; Jiang et al., 2014). These studies have concentrated on the assimilation of crop vegetation information from remote sensing in order to better characterize the biomass. However, as mentioned previously, although total plant biomass estimates were somewhat improved, there was generally limited improvement in yield predictions.

Ines et al. (2013), however assimilated remotely sensed LAI as well as SM estimates independently, as well as simultaneously to simulate maize yields in Iowa, USA. The results showed that individual assimilation of Moderate Resolution Imaging Spectroradiometer (MODIS) derived LAI and surface SM from the Advanced Microwave Scanning Radiometer Earth Observation system (AMSR-E) had little success in reducing errors in yields while simultaneous assimilation of both parameters had relatively better results. The authors identified two main limitations to their approach, (1) assimilating near-surface SM to update the root-zone wetness increases errors in the root-zone by overestimating drainage (particularly in wet conditions), and (2) improvements as a result of LAI assimilation were limited in dry conditions because the root-zone SM could not meet the increased water demand that result from
improved canopy growth. Addressing these challenges, Ines et al. (2013) found that using a-priori climate information to employ a strategic assimilation strategy worked best. However, in their attempt to characterize the vegetation dynamics using LAI, they were unable to constrain the root zone water demand. However, utilizing a state variable that represents aspects of both the surface process as well as the biophysical plant (root zone) process may provide the way forward.

These earlier studies suggest that the assimilation of LAI might reduce the uncertainties in model biomass estimations, but there will be limited success in improving yield estimations. Satellite derived SM, particularly from MW radiometer sensors (such as AMSR-E, Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP) etc.), on the other hand can be used to better represent regional surface SM dynamics; however the inability to represent root-zone moisture content from MW data, in addition to the coarse spatial resolution (∼25-40 km²), limits their applicability significantly. Therefore, a holistic system representative of high-resolution surface as well as root-zone SM content is required for better yield estimations. It has been shown that relatively higher spatial resolution (∼1-10 km²) root-zone SM can be deduced from TIR brightness temperature data (Crow et al., 2008; Hain et al., 2011). However, the TIR sources only offer a gross estimate of total root zone moisture. The root-zone SM profile can be generally entered into the model by: (a) assimilating the root-zone moisture content (in addition to the surface value) into the land surface model (LSM) and let the model physics handle the distribution [e.g. Kumar et al. (2009)], or (b) using some functional relationship to distribute the root-zone moisture within the model (Mishra et al., 2013, 2017). Kumar et al. (2009)
concluded that the first approach is limited by the linkage between the surface and deeper layers in most LSM and Draper et al. (2011) further noted that the surface SM estimates frequently evaporate prior to their assimilation into the lower model zones.

Mishra et al. (2017) in their study, developed satellite-derived physically based coupled SM profiles. They used the principle of maximum entropy to merge surface SM observations from AMSR-E (MW) with TIR-driven mean moisture content data over the Southeastern U.S. The authors demonstrated how the merged MW and TIR SM product tended to improve the Noah LSM. The study illustrated that the statistics of the Noah SM profiles tended to degrade with depth through the soil column while the RS profiles demonstrated the reverse trend. The profile development techniques as well as results are discussed in detail by Mishra et al. (2017). This study used a modified form of SM profile dataset of Mishra et al. (2017) to assist a standard crop model through data assimilation.

The Decision Support System for AgroTechnology Transfer [DSSAT: (Hoogenboom et al., 2010; Jones et al., 2003)] is a mature, well vetted crop model that contains modules for more than 28 crop types. DSSAT, originally a point model, was modified to run the model in gridded fashion over Southeast US by McNider et al. (2011). This model, here referred to as GriDSSAT, is used for this study to simulate maize crop growth and yield predictions over the study areas.

Sequential data assimilation allows for optimal merging of model estimates and observations while taking their respective errors into account. The data assimilation technique serves to minimize the uncertainties within a model state by enhanced use
of model and observation errors. The ensemble Kalman Filter [EnKF: (Evensen, 2003)] is a Monte-Carlo approximation of sequential Bayesian filtering. The EnKF has received a lot of attention recently in crop model data assimilation (e.g. Huang et al. (2016); Ines et al. (2013); Linker and Ioslovich (2017); Liu et al. (2016); Ramirez-Villegas et al. (2017) etc.) due to its ability to handle non-linear models in addition to ease of implementation, computational efficiency, as well as quick and optimum performance. Given the value of the above assimilation methods, in this study, the satellite derived SM profiles were assimilated in the GriDSSAT crop model using EnKF approach over the study area.

5.3 Study Area and Data Description

5.3.1 Study Area

For this research, 5 areas were selected for intensive study encompassing 19 counties from four states: Alabama (AL); Georgia (GA); South Carolina (SC); and Florida (FL) in the Southeastern U.S. The Southeastern U.S. represents a subtropical humid climate that typically has relatively hot and humid summers and precipitation is generally evenly distributed throughout the year. The mean annual precipitation is 1250-1500 mm with mean annual temperature ranges from 14°C in Northern Alabama to nearly 24°C in southern Florida. Mishra et al. (2017) have already developed merged MW-TIR SM profiles for the entire study area.

The counties selected from the Southeastern U.S. for detailed analysis were grouped into five regions as shown in Figure 5.1. Two of the locations (Southwest GA
and FL) are highly irrigated (55% and 33% approx., respectively) while the Northern AL region is partially irrigated (~10-12%) and the other two have little to no irrigation (~5-8%). Two sites (Southwest GA and North AL) have the highest agricultural crop areas (57.2% and 44.2%, respectively) while the SC location contains about 33% agriculture and the Northeast GA and FL sites contain only 9.2% and 13.2% agricultural land, respectively (see Table 5.1). The primary reasons for selecting these study locations were to get a mixture of irrigated and non-irrigated sites, to offer geographical coverage of the Southeastern U.S., and to select sites ranging from major agricultural producing locations (i.e., Southwest GA) to those with a fair amount of agriculture (SC) to those where agriculture is present but not the major land cover. This last requirement (selection of land uses) was done in order to examine the efficacy of using remotely sensed SM profiles in agricultural models, not only in agriculturally dominant regions, but also in areas where the remote sensing pixels will include a mixture of other related land covers (such as shrubs, grassland etc.). For this study, pixels significantly dominated (>50%) by either water or forests were masked out.

5.3.2 Data Sources

5.3.2.1 Satellite Derived Soil Moisture Profiles

Mishra et al. (2017) employed a principal of maximum entropy model (POME) to merge MW (surface) and TIR (root zone) SM estimates to develop SM profiles across the Southeastern U.S. following the approach of Al-Hamdan and Cruise (2010) for the period of 2006-2010. The POME approach ensures the minimum variance
unbiased profile given the prior information (Barbé et al., 1991). The method is based on the informational entropy concept of Shannon (1948) and the subsequent work on maximization of entropy by Jaynes (1957). Later, Chiu (1987) showed that the POME concept could be used to develop a vertical profile of a random variable assuming an initial uniform distribution of the variable. Al-Hamdan and Cruise (2010) then applied this concept to develop vertical profiles of SM. Subsequent to its introduction, the POME model has been further refined and tested by Mishra et al. (2013, 2015); Pan et al. (2011); Singh (2010a). In particular, Mishra et al. (2015) compared the POME SM profiles to both in-situ SM data and to a detailed mathematical model.
Table 5.1: Study area location and cropland information with mean rainfall and NASS yields for the period of 2006-2010 [*growing season, N refers to number of grid points (5x5 km$^2$) after masking out water and forested regions].

<table>
<thead>
<tr>
<th>Region</th>
<th>State</th>
<th>Counties</th>
<th>Cropland (%)</th>
<th>Irrigation (%)</th>
<th>N</th>
<th>Mean GS* Rainfall (mm)</th>
<th>Avg NASS yields (kg/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alabama</td>
<td>Limestone, Lawrence, Morgan, Madison</td>
<td>44.2</td>
<td>8-12</td>
<td>272</td>
<td>400</td>
<td>6767</td>
</tr>
<tr>
<td>2</td>
<td>Georgia</td>
<td>Hart, Madison, Franklin, Banks, Banks</td>
<td>9.2</td>
<td>5-8</td>
<td>121</td>
<td>340</td>
<td>6336</td>
</tr>
<tr>
<td>3</td>
<td>South Carolina</td>
<td>Lee, Florance, Darlington</td>
<td>32.9</td>
<td>3-5</td>
<td>294</td>
<td>401</td>
<td>5786</td>
</tr>
<tr>
<td>4</td>
<td>Georgia</td>
<td>Baker, Mitchell, Colquitt, Worth, Miller</td>
<td>57.2</td>
<td>55-60</td>
<td>384</td>
<td>340</td>
<td>10034</td>
</tr>
<tr>
<td>5</td>
<td>Florida</td>
<td>Alachua, Suwannee, Gilchrist</td>
<td>13.6</td>
<td>30-35</td>
<td>311</td>
<td>344</td>
<td>6834</td>
</tr>
</tbody>
</table>

of SM movement and found that if the input data are well known, the overall error in the SM profile is less than 3%. The method requires upper and lower boundary conditions on the entropy integral and an estimate of the mean moisture content in the soil column. With the MW surface estimate from AMSR-E providing the upper boundary and the TIR root zone SM (from the two source based Atmospheric Land Exchange Inverse-ALEXI (Anderson et al., 1997, 2007a; Hain et al., 2011) model) leading to the mean moisture content, the POME method is ideal for merging the two remotely sensed SM products. The required lower boundary can be either a parameterized soil property or a calibration parameter.
5.3.2.2 Climate Dataset

The climate dataset developed by Livneh et al. (2013) is used to force the GriDSSAT crop model in its current operational mode (McNider et al., 2015). Therefore, this data set was continued for use in this study. Livneh et al. (2013) expanded the hydrologically based dataset of Maurer et al. (2002) to extend it from 1915 through 2011, and increased the resolution to a 1/16th degree. Using the same methodology as in Maurer et al. (2002), the Livneh dataset includes station-based interpolation of daily precipitation and temperature data, and wind fields from the National Centers for Environmental Prediction (NCEP)National Center for Atmospheric Research (NCAR) reanalysis. Variables that are not directly observed (downward solar and long-wave radiation and humidity) were derived using the algorithms employed in (Maurer et al., 2002), that is, the mountain microclimate simulator [MTCLIM, (Thornton and Running, 1999)]. A spline was applied to daily minimum and maximum temperatures to estimate the diurnal cycle. The Livneh dataset are publically available, free and accessible from the Internet (ftp://ftp.hydro.washington.edu/pub/blivneh/CONUS/).

5.3.2.3 Observed Precipitation Data

One objective of the study is to run the crop model under simulated climate conditions that might mimic those encountered in data scarce regions of the world. The assimilation of satellite SM data might prove most useful in those conditions. For this purpose the Daily Global Historical Climatology Network (GHCND)
station data were employed. GHCND is an integrated database of daily climate summaries from stations located globally as maintained and distributed by National Center for Environmental Information (NCEI, formerly also known as NCDC). Over the United States, the dataset contains daily summaries from a number of networks such as COOP, Automated Surface Observing system (ASOS), Community Collaborative Rain, Hail and Snow (CoCoRaHS) etc. Daily observed precipitation data from multiple gage stations across the study area were obtained from the GHCND. The GHCND dataset contains more than 4300 stations across southeast U.S. Out of these, 5 stations (one from each region) were selected for this study to force the crop model mimicking the data scarce region scenario.

5.3.2.4  *In-situ* Soil moisture data

The Southeast U.S. contains 25 operational U.S. Department of Agriculture Soil Climate Analysis Network [SCAN: (Schaefer et al., 2007)] monitoring stations. These monitoring stations monitor soil temperature and moisture content primarily at depths of 5, 10, 20, 50 and 100 cm at hourly and daily time steps in addition to meteorological observations such as precipitation, air temperature, relative humidity etc. The SCAN sites use Hydra Probes to observe SM conditions. *In-situ* SM data from sites with the most consistent data availability were selected for the comparison and validation of remotely sensed SM profiles. SCAN SM data from are available from [http://www.wcc.nrcs.usda.gov/scan/](http://www.wcc.nrcs.usda.gov/scan/).
5.4 Methodology

The purpose of the study was to determine if assimilation of the merged MW-TIR soil moisture profiles could improve the rain fed crop model, first in a data rich area (Southeastern U.S.) and then in data scarce regions (simulated through degraded climate inputs). The POME-generated profiles developed by Mishra et al. (2017) were slightly modified to better link them to the crop model and then re-compared to the Noah LSM and in-situ data. The profiles were then assimilated into the crop model and compared to rainfed only results for a period of 5 years (2006-2010). The climate inputs were then degraded by selecting one station only to represent each study area and the simulations run again.

5.4.1 Modification of remotely sensed SM Profiles

Using the downscaled AMSR-E MW data as the surface boundary and the ALEXI root zone SM estimates, Mishra et al. (2017) computed SM profiles for the Southeastern U.S. at the 5x5 km spatial resolution. Originally, the necessary lower boundary on the POME integral was set at 50% of available water content. However, this assumption led to an increase in bias and unbiased RMSE (ubRMSE) in the bottommost layer (100-200 cm) and frequently with low correlation in that layer compared to field data. Therefore, a slight change was made to the profiles computed for this study as the SM in the lower layer of the rain fed crop model itself was used as the lower boundary on the POME integral (layer depth 100-200 cm). Past analyses have shown [e.g., Mishra et al. (2013)] that this value is fairly constant over
a growing season and can thus be set as the lower boundary on SM in the model and will serve to better tie the POME profiles to the crop model. In any case, this layer is well below the majority of the root mass in corn plants and thus its moisture level would have minimal effect on the yield estimates. A full description of the profile development methodology and a comprehensive evaluation of results against both the Noah LSM and multiple field sites over the Southeastern U.S. can be found in Mishra et al. (2017).

5.4.1.1 Crop Model Simulation

The DSSAT crop model (Hoogenboom et al., 2010; Jones et al., 2003; Tsuji et al., 1998) is an agriculture modeling system designed to utilize variables such as weather, soil type and profile properties, cultivar-specific inputs and management options including irrigation as well as amount and type of fertilizer for simulating crop growth and yield for more than 28 crop types and fallow systems. DSSAT also has a soil hydrological model that estimates vertical soil water flow and root water uptake as a function of soil type and profile properties and root development (Jones et al., 2003; Ritchie, 1972).

The calibrated DSSAT model is currently run in a gridded format (4.75 x 4.75 km) throughout the Southeastern U.S. (McNider et al., 2015). The model simulates maize production using a cultivar calibrated to field trials and crop management information (planting date, fertilizer application, etc.) obtained from conversations with area farmers. The most dominant agriculture soil type from the Soil Survey Geographic (SSURGO) data base within a grid point was selected as representative
of the entire grid. Daily weather information such as temperature, precipitation and solar radiation were provided as weather inputs from Livneh (Livneh et al., 2013) dataset, as described above. This present study used the well calibrated and validated GriDSSAT framework described by McNider et al. (2011, 2015). For this study, the automatic fertilization (i.e. no nitrogen stress) option within the model was employed, which was done to assess the effects of only the moisture related stress on the crop growth and yield estimates.

5.4.2 Data Assimilation

The remotely sensed SM profiles were assimilated within the DSSAT model via the EnKF. EnKF implementation was based on the works of Evensen (2003). For a given forecast vector and analyzed vectors, $A^f$ and $A^a$, the basic analysis step of the EnKF is given as:

$$A^a = A^f + K[I']$$

Where, $K$ is the Kalman gain and $I'$ is the innovation vector. The Kalman gain ($K$) is given as:

$$K = \frac{P_e}{HP_eH^T + R_e}$$

$H$ is a measurement operator; $P_e$ and $R_e$ are model ensemble and observation error covariance matrices. The model error covariance matrix was obtained from model ensembles:
\[ P_e = \frac{1}{H^T(N - 1)} \sum_{n=1}^{N} (A_f^i - \overline{A_f})(HA_f^i - H\overline{A_f})^T \] 

(5.3)

\( \overline{A_f} \) is the ensemble mean and \( N \) is the number of ensemble members. The observation errors in this study are obtained by computing triple collocation error (Gruber et al., 2016; Scipal et al., 2008) between SCAN observations and POME profiles following Mishra et al. (2017).

The innovation vector is given as: \( I' = (D - HA_f^i) \). In this case, the model observation state i.e. the SM is measured directly, \( H \) is an identity matrix. \( D \) is the observation matrix generated by perturbing the observation around its error vector with zero mean. Ensembles of model simulations can be generated either by perturbing the model forcings or its parameters. In this study, the ensembles were generated by randomly drawing forcing data from the long term climate data set (1915-2011) from Livneh et al. (2013). A recent study by Yin et al. (2015) used a mathematical deduction to determine the optimal ensemble size for a EnKF. The study suggests that for assimilating SM observations through sequential operations, the maximum efficiency can be reached with an ensemble size of 12. For executing the models, years are randomly drawn from this \( \sim 100 \) year Liveneh data set, and since the ensemble only contains 12 members the population will not be oversampled at any given update point.
5.4.2.1 Modeling Strategy

The objective of the study was to determine if the rain-fed DSSAT model can be improved through assimilation of the remotely sensed soil moisture profiles, particularly in data scarce regions of the world. The main objective of this study was accomplished through executing the GriDSSAT model in three modes: (a) first using the best state-of-the art climate inputs available without data assimilation (open-loop); (b) then assimilating the model with the POME based remotely sensed SM profiles where available; (c) and lastly running the model with degraded climate inputs with and without assimilated remotely sensed profiles. The last step is used to simulate conditions in data scarce regions but where the remotely sensed profiles would be available. This was done by selecting one GHCND Station in each study region and applying the data from that station to all model grids within the region. The GriDSSAT model was modified to enable assimilation of remotely sensed soil moisture profile through EnKF, and a fixed ensemble size of 12 was used.

The results of simulations in terms of yields are compared against the county level mean yield data from U.S. Department of Agricultures National Agricultural Statistical Services (NASS). The NASS census contains information such as total yields, yields per acreage, total acreage under agriculture etc. aggregated to the county level. The mean corn yield information from 2006-2010 was obtained from NASS for the selected 19 counties for comparison and validation of simulated runs. The results in terms of yields were compared using statistics such as bias, correlation ($r$) and unbiased root mean square error (ubRMSE).
5.5 Results

5.5.1 Profile Calibration Results

In an earlier study by Mishra et al. (2017), the computed SM profiles were extensively validated against both the Noah LSM within the NASA Land Information System (LIS) and in-situ SM data at SCAN sites around the region. The overall results showed that the satellite based and LSM model generated profiles are reasonably comparable based on error statistics of ubRMSE (0.05 vs 0.04) and absolute bias (0.08 vs 0.07). At the surface (0-10cm) comparisons the Noah correlations were superior to the POME model ($r = 0.75$ vs $0.54$), although in several cases the Noah correlations decreased vertically through the soil column to the point that the two approaches were much more comparable. The Noah LSM tended to become less accurate with increasing depth while the POME generally showed the reverse (Figure 5.2 below) despite having the assumed static lower boundary condition.

Including the lower boundary with the crop model was done to further improve the POME model performance and provide better linkage with the model. The overall error statistics of the updated POME model in comparison with the earlier profile (with static boundary condition) showed an improvement in bias from 0.035 to 0.028 (in absolute values) and slight improvement in ubRMSE from 0.048 to 0.047 against the same in-situ observations, with an increase in correlation from 0.37 to 0.49. As mentioned previously, this lower boundary is at 200 cm which is well below the root mass of corn corps (at or above about 60 cm) so will have minimal effect on the yield results. Figure 5.2 shows the overall profile statistics from 10 SCAN sites of the
POME (earlier version of Mishra et al. (2017) along with calibrated) and Noah LSM SM profiles against *in-situ* observations.

As shown in Figure 5.2, in the top two layers (0-10 and 10-40 cm), the Noah LSM outperforms the POME model with mean bias (absolute) of 0.025 vs 0.044; ubRMSE of 0.041 vs 0.062; and correlation of 0.72 vs 0.51, respectively against *in-situ* observations for the period of 2006-2010. However, the lower layers (40-100 and 100-200 cm) the trend reverses, with POME outperforming the Noah SM profiles with mean bias (absolute) of 0.013 vs 0.085; ubRMSE of 0.035 vs 0.071; and $r$ of 0.46 vs 0.39.

**Figure 5.2:** SM profile comparison statistics against SCAN observations for the period of 2006-2010. SP* - SCAN vs POME (earlier version of Mishra et al. (2017)); SN - SCAN vs Noah LSM; and SP - SCAN vs POME (current version).

### 5.5.2 Data Assimilation Results

Comparison of GriDSSAT model yields between the rainfed runs with the gridded climate rainfall only (open-loop) and the runs with assimilation of MW-TIR profiles are shown in Figure 5.3 below. The figure compares the maize yields ($kg/ha$)
for the models and the actual recorded yields (red dots) for each of the five study regions. The spread of the yields among the 5x5 km grids within each region is also shown in the box plots. The notch area of the boxes represents the 95% confidence intervals on the median. The results show that in nearly all cases the assimilation model improved the simulations compared to the rainfed only yields. In fact, in 18 out of 22 cases the reported yields fall within the 50% spread of the assimilation model compared to only 4 such cases with rainfed only runs. In fact, in 12 out of 22 cases, NASS mean yields fall within the 95% confidence interval of assimilated yields (Figure 5.3). In all but two cases (2007 in Northeast GA and 2008 in SC) the recorded yield is closer to the assimilation median than it is to the rainfed median. Thus, in all but these instances (i.e., in 20 of 22 possible cases) assimilating remotely sensed SM profiles improved the model relative to the observed data. It can also be noted from Figure 5.3 that in all but one case (2007 Northeast GA) the rainfed and data assimilated median yields differ at the 0.05 confidence level (notched areas do not overlap). The variability of the assimilated model is much higher than in the rainfed case due to the remote sensing datas ability to detect differences in land cover and water availability within the pixels that comprise the study regions.

Relative GriDSSAT model errors for each region are shown in Figure 5.4. Overall, the results illustrate the improvement in the model after assimilation of the remotely sensed profiles. The model degraded slightly in three cases (the two mentioned above plus 2009 in SC), but in most instances the assimilation of remote sensing data profiles resulted in marked improvement. The error results show that in the majority of cases, the relative errors of the assimilation model were no more
Figure 5.3: Boxplot showing yield at different regions between GriDSSAT simulation in rainfed (green) and data assimilation (blue) mode with NASS yields (red dot) from 2006-2010. *years NASS yield data was unavailable

than 10%. In two cases, (2008 in SC and 2006 in FL) errors are in the 40-45% range and in two other cases (2010 in Northeast GA and 2006 in AL) the errors are in the 20-30% range. In all other cases the errors are no greater than 15%. Also of interest, except in the North FL region the assimilation model was relatively unbiased with both positive and negative errors in equal measure.
Figure 5.4: Relative errors (%) of open-loop (rainfed-green) and remotely sensed profile (blue) assimilated annual yields compared to NASS (2006-2007). *years NASS yield data was unavailable

Looking more closely into the simulations, the relative error results by county are shown in Figure 5.5 below. Figure 5.5 shows marked improvement through assimilation of the remotely-sensed SM profiles in most counties with only one county (Banks, Northeast GA) showing slight degradation in performance after assimilation. However, the Figure reinforces the discussion above in that 8 counties show errors
of 10% or less with 3 having errors on the order of 15% in the assimilation model results. Conversely, only one county demonstrates an error of 30% and one other of 25%. All other counties show errors of less than 20% in the remotely assisted model.

![Figure 5.5](image.png)

**Figure 5.5**: Relative errors in yields by county for rainfed and data assimilation simulations against NASS reported yields (2006-2007).

### 5.5.3 Results with Degraded Climate Inputs

The results of the previous runs show that in dry (i.e., non-irrigated regions such as Northeast GA and SC), the rain fed model often performed fairly well without assimilation. This is to be expected since it was driven with a high resolution gridded climate dataset. However, in many parts of the world, such data are not available and those are the areas where the assimilation might be most useful. In order to simulate this situation, the GriDSSAT model was run at all grid points with climate data from a single GHCND station randomly selected from within each of the 5 study regions. The assimilation also uses the same single station climate data for all grids within
a region. The results are shown in Figure 5.6 where the yields (in kg/ha) from the degraded rainfed model (green) are compared to the data assimilation model (blue) under the degraded conditions. As in the previous case, the actual recorded yields are plotted as the red dots.

Figure 5.6: Boxplot showing yield at different regions between GriDSSAT simulation with degraded climate inputs in rainfed (green) and data assimilation (blue) mode with NASS yields (red dot) from 2006-2010. *years NASS yield data was unavailable

Figure 5.6 shows that to a large degree the both model simulations (with assimilation versus the control) performed similarly to the previous case with some notable exceptions. First, the rainfed model did not deteriorate to the degree that
was expected (absolute overall error 49% vs 45%). Still, in 15 of the 22 possible cases the recorded yield fell within the 50% spread of the assimilation model simulations (6 within 95% confidence interval), while only 5 did so for the rain-fed model. In addition, 3 cases demonstrated results where the rainfed model yield estimates were closer to the observed yields than were the assimilation model compared to 2 such instances in the previous run. In addition it can be noted that in all but 2 cases (both in AL) the rainfed and assimilated yields differ at the 0.05 confidence level.

The relative errors in both models are shown in Figure 5.7. The figure shows that assimilation of MW-TIR SM improved the model errors in 19 of the 22 possible cases, the same as in the previous runs. One situation that stands out is the 2008 run for SC where the DA model error is close to 50% while the rainfed simulation error remains at around 10%. Otherwise, assimilation model errors approach 40-60% in three other cases (2010 in northeast GA; 2007 in SC and 2006 in FL) but remain relatively low (<40%) elsewhere. In fact, in 7 cases assimilation model errors are less than 10%. Overall the average absolute error in the assimilation model increased from 11% in the previous run to 16% under the degraded inputs. However, it is still less than half of that of the rainfed only model (40.5%).

Detailed model simulations: (a) forced with gridded high resolution precipitation data (simulation-1); (b) and with degraded precipitation data (simulation-2) for each County in the study area are shown in Figure 5.8. The Figure clearly demonstrates the difficulty in assimilation of remotely sensed SM profiles in the northeast GA (region 2) study area as all four counties show errors in the assimilation model are greater than those in the rainfed model. Looking at Figure 5.7 above, one can
Figure 5.7: Relative errors (%) of rainfed (green) and MW-TIR SM assimilation (blue) yields with a single GHCND station climate runs compared to NASS (2006-2007). *years NASS yield data was unavailable

see that this anomaly is due to the fact that the errors in the assimilation model are positive for all years (positively biased), while the rainfed model demonstrates both positive and negative errors (relatively unbiased) that tend to cancel each other out. Otherwise, with the exception of Lee County in SC (region 3), the assimilation model shows less error than the rainfed in each case. Both models appear to be biased by
region with the study areas showing either all positive or negative errors in particular study areas.

![Figure 5.8](image)

**Figure 5.8**: Relative errors in yields by County for rainfed and DA simulations: (a) with high resolution gridded precipitation data; and (b) with single GHCND station climate data against NASS reported yields (2006-2007)

Table 5.2 summarizes the yield results over the regions for all five years. The Table 5.2a shows the overall yield results of GriDSSAT simulation with gridded Livneh climate datasets, whereas Table 5.2b shows the yields for GriDSSAT simulations with a single GHCND station climate data. In the case of the gridded Livneh driven crop model simulations (Table 5.2a) the highest improvements in yield estimates with assimilation compared to rainfed simulations were achieved for regions 1, 4 and 5. The overall relative errors (except for region 5) of the assimilation model are less than 5% compared to NASS with an average error of 4.3%. The overall mean relative error of the rainfed simulations was -36.8% with a maximum of -64.7% (region 5).
Similarly, looking at GriDSSAT overall results with a single GHCND station climate data simulation (Table 5.2b), the same three regions (1, 4 and 5) showed the highest improvements with DA. Region 2 however showed an increase in error with assimilation compared to the rainfed simulation, whereas region 3 showed only marginal improvement with assimilation. With the single GHCND station data, only 2 regions out of 5 showed errors less than 10% with assimilation compared to 4 with the gridded climate data. Overall, the assimilation or remotely sensed data improved the model yields estimates with both simulations; although the use of a single station climate data did not degrade the rainfed yields substantially to assess the relative gains in assimilating remotely sensed SM profiles.

5.6 Discussion

It must be noted that data assimilation via EnKF being a stochastic process; the results given above represent only one of many possible outcomes. Therefore, the results have to be viewed from a global perspective, i.e., without placing undue significance on a single result. Viewed in this manner it is clear that the assimilation of the remotely sensed SM profiles served to improve the rainfed crop model in terms of bringing the yield estimates closer to the observed yields. However, it is possible that in certain instances the assimilation of noisy remotely sensed data could harm the model if the remotely sensed profile is in error but demonstrates less uncertainty than the model run when the model itself is in fact more accurate. Clearly, this situation prevailed in some years, particularly in the northeast GA and SC regions and most obviously under the degraded climate condition.
Table 5.2: Summary yield results of rainfed and DA simulation results compared with reported NASS yields (2006-2010): (a) with gridded climate data (b) with single GHCND station climate data.

(a)

<table>
<thead>
<tr>
<th>Region</th>
<th>Yields (kg/ha)</th>
<th>error (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NASS</td>
<td>Rainfed</td>
<td>DA</td>
<td>Rainfed</td>
</tr>
<tr>
<td>1</td>
<td>6767</td>
<td>3805</td>
<td>6299</td>
<td>-43.8</td>
</tr>
<tr>
<td>2</td>
<td>6336</td>
<td>5494</td>
<td>6937</td>
<td>-13.3</td>
</tr>
<tr>
<td>3</td>
<td>5786</td>
<td>5715</td>
<td>5798</td>
<td>-1.2</td>
</tr>
<tr>
<td>4</td>
<td>10034</td>
<td>3899</td>
<td>10425</td>
<td>-61.1</td>
</tr>
<tr>
<td>5</td>
<td>6834</td>
<td>2415</td>
<td>7857</td>
<td>-64.7</td>
</tr>
<tr>
<td>Mean</td>
<td>7152</td>
<td>4266</td>
<td>7463</td>
<td>-36.8</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Region</th>
<th>Yields (kg/ha)</th>
<th>error (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NASS</td>
<td>Rainfed</td>
<td>DA</td>
<td>Rainfed</td>
</tr>
<tr>
<td>1</td>
<td>6767</td>
<td>3667</td>
<td>5601</td>
<td>-45.8</td>
</tr>
<tr>
<td>2</td>
<td>6336</td>
<td>5776</td>
<td>7154</td>
<td>-8.8</td>
</tr>
<tr>
<td>3</td>
<td>5786</td>
<td>7254</td>
<td>7186</td>
<td>25.4</td>
</tr>
<tr>
<td>4</td>
<td>10034</td>
<td>3713</td>
<td>9205</td>
<td>-63.0</td>
</tr>
<tr>
<td>5</td>
<td>6834</td>
<td>2390</td>
<td>6901</td>
<td>-65.1</td>
</tr>
<tr>
<td>Mean</td>
<td>7152</td>
<td>4560</td>
<td>7210</td>
<td>-31.5</td>
</tr>
</tbody>
</table>

The results are affected by the manner in which the error covariance matrices are achieved. In terms of the MW and TIR data, SM retrieval methods from both sensors are affected by fraction of vegetation cover. The MW surface SM estimates are more accurate under low vegetation cover particularly from higher frequency C and X-bands (such as from AMSR-E radiometer as in this case) (Albergel et al., 2011; Brocca et al., 2011), while with TIR the reverse is true due to the better connection between the measured ET and root zone SM under moderately thick vegetation (Hain et al., 2011). Since the errors in the two sensor behaved this way, it was decided to use a constant error region-wide with the remote sensing-based profiles. However,
this error may be too small early in the growing season (less vegetation cover) and too large later in the season after the vegetation has emerged. These issues led to the anomalous results seen in the 2008 SC and 2010 Southwest GA degraded climate runs.

The results demonstrate that the assimilation of the MW-TIR developed profiles improved the model best in the regions with a higher percent of cropland (>10%), especially under irrigated conditions such as Southwest GA, FL and AL (Table 5.1). Figure 5.9 shows the mean absolute errors by region for the two scenarios run in this study. The figure illustrates the effectiveness of assimilation in regions 1 (AL) and 4 (Southwest GA) in particular. It is clear that the MW-TIR coupled SM profiles have the ability to detect the greater SM supplied by the irrigation was determinative in these areas as demonstrated in Figure 5.9.

![Figure 5.9](image)

**Figure 5.9:** Overall absolute errors (%) for gridded climate (left) and single GHCND climate data simulations with and without DA against NASS yields.
The results were not quite as good in the other regions for a variety of reasons. First, in the low agricultural areas such as northeast GA (with 9.2% cropland) the remotely sensed SM pixels are contaminated by other land uses. Assimilating these profiles into the crop model would not be expected to improve the model to a great extent, if at all, although Figure 5.9 (region 2) shows that overall the EnKF assimilation performed fairly well. The situation is even more complicated at the SC site (region 3) where surface water is prevalent, particularly in numerous small streams and the floodplain wetlands associated with the Pee Dee River which was not masked as they did not necessarily dominate any particular pixel. The presence of water in the pixels is known to contaminate both the MW and TIR data (Jackson et al., 2010; Norman et al., 1995). This situation is more complex in that small streams and water bodies if present (such as the case with the SC region); pixels might not be masked out and are contaminated by the presence of water. Therefore if assimilated into the crop model these profiles may provide excess water to the root zone and positively bias the crop yield. Such results are evident in SC where the yields are generally overestimated. The relatively poor performance of DA in SC are clearly evident in Figure 5.9 (Region 3) where the assimilation of the remotely sensed profiles even increased the overall average absolute error under the degraded scenario.

As mentioned earlier, the assimilation system performed best in well irrigated regions (4 and 5), and partially irrigated (region 1). The ability of TIR based remotely sensed SM estimates to sense root zone moisture content sets it apart from rainfed model estimates. Remotely sensed TIR profiles are likely to contain signals from irrigated and non-irrigated fields. Irrigated fields can potentially be identified
with elevated root-zone moisture content particularly during the growing season. Figure 5.10 shows the difference in mean moisture content (mm) between rainfed and assimilation simulations at two grid points over multiple years: a) shown in the left panel is a non-irrigated grid from AL and; b) is an irrigated grid (right panel) from Southwest GA (Region 4). The positive difference (remotely sensed assimilated SM > rainfed SM) is shown in blue. It is clear from the Figure 5.10 (right panel) that the remotely sensed SM profile is able to sense the irrigation application consistently through the year (2006, 2007 and 2009) with 2006 and 2007 being relatively dry years (mean total growing season precipitation < 200 mm) whereas 2009 is a wet year (total precipitation > 500 mm). The growing season for both regions is shown as shaded. Irrigated fields showed a consistent positive difference throughout the later part of the growing season clearly indicating irrigation application compared to the period outside of the growing season. The presence of a negative difference towards the period outside of the growing season indicates the relatively unbiased nature of the profile.

For non-irrigated grids, the difference in mean SM content between the rainfed and assimilated simulations is not as significant with differences up to 10 mm. However, the figure shows that the regime of the SM in these areas did change somewhat and the yield results show that these relatively small changes sometimes resulted in a fairly large change in yields (2006 for example). The net result was that the rain-fed model was usually improved even in these areas, but not to the extent exhibited in the irrigated areas. The corn yields from both models for non-irrigated grids were usually closer; however, for the irrigated grids, as expected the differences in yields
Figure 5.10: Mean SM difference (mm) time series at non-irrigated (left panel) and irrigated (right panel) grid point for years: 2006, 2007 and 2009.

were much higher. It seems that the assimilation of remotely sensed SM profile into a crop model can not only improve yields estimates, but can potentially also be used for irrigation mapping.

5.7 Conclusion

This study evaluated the applicability of remotely sensed SM profile assimilation into the GriDSSAT crop model for yield predictions over the Southeast U.S.
The SM profiles developed using remote-sensing data were assimilated via EnKF to the GriDSSAT crop model. The crop model simulations were made in two distinct modes: a) with high resolution gridded climate data; and b) with a single GHCND station precipitation data for all the grids in a particular region. The results in terms of yields were compared against the reported county level NASS yields.

In summary, the rainfed yields in comparison with the reported NASS yields showed an overall absolute error of nearly 38% and 41% for both simulations, respectively. The assimilated yields on the other hand had an error of 12 and 16%, respectively. Assimilating remotely sensed SM profiles into the crop model improved the yield estimation most in irrigated regions (4 and 5) with average errors (< 3%) whereas for the relatively non-irrigated regions (2 and 3) the mean error with data assimilation was nearly 19%. Overall, taking all regions together, the assimilated yield errors were less than half of the rainfed yield errors. The results indicate that the assimilation of remotely sensed SM profiles into the crop model was effective in improving yield estimates compared to rainfed only simulations. The results also highlight the effectiveness of TIR-based SM estimates in sensing irrigation applications on a reginal scale.

The DA assimilation was very effective in areas with significant agricultural land, but was even fairly effective in areas of mixed land use as long as the pixels were not significantly contaminated by the presence of open surface water. It is interesting that in a region such as northern FL with only 13% agricultural land the model performed well, while in SC with substantial open water to contaminate the pixels,
the DA did not perform as well. It appears that even in mixed land use areas the remotely sensed profiles may have some value in guiding agricultural models.

The errors in the remotely sensed SM profiles can originate from the down-scaled surface boundary condition as well as the TIR driven mean moisture content. The MW surface SM seems to be the major source of error for the remotely sensed profiles as evidenced by higher error statistics at the surface. TIR estimates on the other hand, demonstrated relatively high correlation and lower error statistics. Errors in the EnKF application are related to perturbations of both the observation and model SM states. First, a constant region-wide error was applied to all remotely sensed profiles and an ensemble size of only 12 was used to perturb the model as indicated in the literature (Yin et al., 2015). Further, the degradation of the climate data using a single GHCND station data did not inject as much error into the model SM states as expected, and as a result the relative gains with DA cannot be assessed to the desired degree.

Improvements in the study can be made through modifying the application of the EnKF. A variable error term can be applied to the remote sensing SM profiles taking into account the error in the sensors related to vegetation coverage. Finally, refining the method of producing degraded climate inputs to the model would greatly increase the ability to evaluate the robustness of the data assimilation system for reducing errors in yield estimations at regional scale.
CHAPTER 6

CONCLUSION

This study was focused on developing and applying a technique that utilizes the best of both MW and TIR based SM estimations to develop a vertical SM profile and then assimilating that profile into can be input to the gridded crop model. The task was divided into four broad categories: (a) detailed validation of the method to merge MW and TIR SM estimates into a vertical SM profile POME model against in-situ SM observations and a physically based mathematical model as an effective tool towards develop SM profiles; (b) testing and evaluating a disaggregation scheme to downscale coarse resolution MW surface SM estimates to match the spatial scales of TIR-driven root zone SM data; (c) Developing SM profiles from the merged MW and TIR estimates for the Southeastern U.S. and evaluating the results against a standard LSM and in-situ SM data throughout the region; applying the POME model on disaggregated surface, TIR-based root-zone SM content data to develop a multi-year a vertical SM profile for Southeastern U.S., comparison and validation of remotely sensed SM profiles against in-situ and LSM model; and (d) assimilation of the merged SM application of RS SM profiles into a gridded crop model via EnKF and evaluation
against observed county level yields at 5 intensive study areas in the Southeastern U.S. to improve yield estimates over parts of southeast U.S.

The main outcomes from this research are:

- The POME model is capable of generating SM profiles, despite having errors in surface boundary condition as well as inflection point location and values estimate, the overall error was still less 5% compared to SCAN observations. Therefore, low error combined with minimum input requirements two of which can easily be met with satellite observations makes POME a preferred choice for this study.

- The disaggregation had relatively higher impact with higher frequency X-band AMSR-E surface measurement with correlation improving from 0.31 to 0.53. The coupled MW (disaggregated) and TIR SM profiles were comparable to both SCAN and Noah LSM. In particular, the with increased layer depths, the POME profile error statistics tend to improve in contrast with Noah LSM where with depth the profile seems to deteriorate.

- The SEE based disaggregation scheme was generally ineffective in improving the surface SM data quality compared to coarse scale passive measurement particularly with L-band SMAP estimates in terms of correlations and error statistics. However, gain close to 60% of SCAN sites showed positive gains with TIR-Downscaling compared to coarse resolution against SCAN observations.

- Lower frequency L-band radiometer seems to be relatively more accurate than higher frequency X-band AMSR-E surface measurements. It is expected as
SMAP has a dedicated subsystem to detect and mitigate radio frequency interference and is also relatively more advanced in its approach to handle vegetation water content that than AMSR-E or SMOS (Chan et al.).

- The calibration of lower boundary condition for the POME model further improved the model performance in terms of bias and ubRMSE.

- The develop profile were assimilated into crop model via EnKF, the DA yields in comparison with the reported NASS yields showed an overall absolute error of nearly 12%, while the rainfed yields had an error of 44%. Assimilating remotely sensed SM profiles into the crop model improved the yield estimation most in irrigated regions with average errors (< 3%) whereas for the relatively non-irrigated regions the mean error with DA was nearly 19%. Overall, taking all regions together, the DA yield errors were less than half of the rainfed yield errors.

The results indicate that the assimilation of remotely sensed SM profiles into the crop model was effective in improving yield estimates compared to rainfed only simulations. The results also highlight the effectiveness of TIR-based SM estimates in sensing irrigation applications on a regional scale.

Although, the assimilation of remotely sensed SM profile into crop model improved the model yields considerably, yet further work is required particularly in: a) degrading the rainfed climate data to better assess the robustness of DA system; b) a variable error term can be applied to the remotely sensed SM profiles taking into
account the possible errors in MW and TIR retrieval algorithms due to vegetation cover.

6.1 Future Directions

This study presents a modeling framework with capabilities of assimilating remotely sensed SM profiles into a crop model for improved yield estimates, particularly in data limited regions of the world. Although, the current study focused on simulating crop yields in an historical mode (2006-2010), the framework has a potential to be used in nowcast and forecast modes as well. However, as currently constituted, real-time application of the framework is limited by the availability of remotely sensed SM data from MW and TIR sensors as well as processing time. Typically, at present, the remotely sensed SM products are available with 1-3 day latency while processing the profiles and data assimilation into the model can require an additional 4-6 days. The spatial scale of the current operation (5-km) also precludes it from use by most individual producers with the possible exception of large agro-businesses. However, the current framework can be run with forecasting capabilities on larger spatial scales and longer time frames by forcing the crop model with forecast data. During a growing season, the framework would begin with forecast only mode with no ground or satellite forcing data available. Days into the growing season, as rainfall observations and satellite data becomes available, the model would run with the assimilation scheme improving the prediction with each passing day. This has a potential to act as an improved agricultural decision support system particularly for large scale agricultural
agencies and governments. Decisions on plant cultivars and optimal irrigation and fertilization schemes could be aided by the model.

The framework is applicable at regional scale currently. The spatial scale of the framework is limited by the spatial resolutions of MW and TIR based sensors as well as computational resources. However, with advancement in technology such as the introduction of Unmanned Aerial Vehicles (UAVs), the potential for field scale modeling does exist. Flying at much shorter altitudes than satellites, UAVs can be equipped with MW and TIR sensors which can obtain very high resolution SM measurements. As UAVs can fly under cloud cover, the potential data gaps with the TIR instrument due to cloud constraints can also be mitigated. The use of UAV-driven remotely sensed SM data can be helpful at field scale application of the framework. Smaller scale application can also reduce the processing time and may provide helpful information not only at the end but also during growing season. Models run in data assimilation mode always require high computational resources due to the need for multiple runs to define the model error covariance matrix. Some improvements can be made by improving the efficiency of the code, but computer run times will always be long if the model is applied over large spatial domains. However, if run at the field scale with data provided by UAV mounted instruments, then actual real-time decision making may be possible.

Perhaps the greatest use of the tools developed in this project will be in data scarce regions of the world. In remote areas, field scale management information such as: planting date, crop and cultivar type, irrigation date and amount, fertilization type, amount etc. are generally unavailable or often parameterized. There remotely
sensed observations SM as well as vegetation indices (e.g., LAI), ET and even precipitation can be assimilated into the crop model and provide much of the missing information. Remotely sensed observations have the potential to act as forcing as well as correcting factors for crop models; nudging the model towards more realistic states in lieu of missing or inadequate weather and management inputs.
The semi-empirical root water extraction term can be expressed as:

\[ S(t) = \alpha S_{\max}(t); \quad \alpha \leq 1 \]  \hspace{1cm} (A.1)

Where \( S(t) \) is root uptake at time \( t \), \( \alpha \) is soil water availability factor, and \( S_{\max} \) is the maximum potential uptake that is equivalent to the potential transpiration allowed by prevailing atmospheric conditions. The soil water availability factor \( \alpha \) is computed as a function of pressure head following Feddes et al. (1978) formulation (see Yadav et al. (2009) )

\[
\alpha(h) = \begin{cases} 
0, & h \geq h_1 \text{or} \ h \leq h_4 \\
\frac{h-h_1}{h_2-h_1}, & h_2 \leq h \leq h_1 \\
1, & h_3 \leq h \leq h_2 \\
\frac{h_4-h}{h_4-h_3}, & h_4 \leq h \leq h_3 
\end{cases} \hspace{1cm} (A.2)
\]

The values of limiting pressure heads were obtained from Veenhof and McBride (1994) as follows: \( \psi_1 = -15 \) cm, \( \psi_4 = -15000 \) cm. The Van Genuchten (1980) model is used to relate the effective saturation to pressure head:
\[ \theta_e = \begin{cases} \theta_s, & h \geq 0 \\ 1/[1 + |\sigma h|^\omega]^\beta, & h < 0 \end{cases} \]  

(A.3)

Where \( \theta_e \) is the effective saturation ranging from 1 to 0, \( \theta_s \) is the saturated moisture content, \( \theta_r \) is the residual moisture content and \( h \) is the pressure head at the current moisture content. \( \sigma, \omega \) and \( \beta \) are curve fitting parameters. These soil dependent parameters were calibrated to the individual soils used in this study.

Evapotranspiration is estimated using the FAO-Penman two stage approach, partitioning between surface evaporation and crop transpiration (Allen et al., 1998). After the volume of water uptake by the roots is computed, a root distribution function is used to distribute the moisture uptake throughout the root zone. \( S_{\text{max}} \) is calculated assuming the water uptake is proportional to the root density (Perrochet, 1987), and is formulated in a similar fashion to that of Yadav et al. (2009) as:

\[ S_{\text{max}}(Z, t) = \frac{L_{\text{nrd}}(Z_r)T_p(t)}{L_r(t)} \]  

(A.4)

Where \( L_r \) is the total root depth, a function of the growth stage of the crop, \( Z_r \) is the normalized root depth \( (Z_r = Z/L_r) \) and \( L_{\text{nrd}} \) is the normalized root density distribution given by Wu et al. (1999) and expressed as:

\[ L_{\text{nrd}}(Z_r) = R_0 + R_1 Z_r + R_2 (Z_r)^2 + R_3 (Z_r)^3 \]  

(A.5)

where \( R_n \) (n = 0, 1, 2, 3) are crop specific coefficients. The crop specific coefficient were obtained from Wu et al. (1999) from root distribution data attributed
to Upchurch and Ritchie (1984) and Merrill et al. (1987) and are as follows: \( R_0 = 2.15, R_1 = -1.67, R_2 = -2.36, \) and \( R_3 = 1.88. \)

In addition to the limiting soil water availability factor a root compensating index (\( \varphi \)) Li et al. (2001) as a function of both the root density distribution as well as the soil moisture availability is included such that:

\[
\varphi(z, h, t) = \frac{\alpha F}{\sum \alpha F^\lambda}
\]  

(A.6)

for each layer, where \( F \) is a root distribution function and \( \lambda \) is a crop specific coefficient. This additional index allows water stress in one part of the root zone to be compensated for by other moisture root zone layers. Combining both the soil availability factor with the root compensating factor Yadav et al. (2009) presented the following equation using the root density distribution function of Wu et al. (1999):

\[
S(z, h, t) = \frac{\alpha^2(h)[L_{nr}(Z_r)]^\lambda}{\Delta z \sum_0^L \alpha(h)[L_{nr}(Z_r)]^\lambda} T_p(t)
\]  

(A.7)

\( S(z, h, t) \) was used to simulate the root water uptake in this model. The crop specific coefficient (\( \lambda \)) was optimized based on optimal moisture content so that it best fit equation A-4 and hovered around a value of 1.

To simulate drainage, water is allowed to move vertically in each soil layer modeled if the water content is above field capacity and the layer below is below saturation. The amount of water that moves from one layer to the next is calculated on the storage routing methodology of the Soil Water and Assessment Tool [SWAT, Neitsch et al. (2011)]:

170
\[ w = \theta_{\text{excess}} \times \left[ 1 - e^{-\frac{\Delta t}{T_{\text{perc}}}} \right] \] (A.8)

where, for each layer, \( w \) is the amount of water drained, \( \theta_{\text{excess}} \) is the amount of drainable water, \( -\delta t \) is the length of time step in hours, and \( T_{\text{perc}} \) is the travel time for drainage. \( T_{\text{perc}} \) is defined as the difference between the saturated (\( \theta_{\text{SAT}} \)) and field capacity (\( \theta_{\text{FC}} \)) moisture states divided by the hydraulic conductivity (\( K_{\text{SAT}} \)):

\[ T_{\text{perc}} = \frac{\theta_{\text{SAT}} - \theta_{\text{FC}}}{K_{\text{SAT}}} \] (A.9)
REFERENCES


the continental United States based on thermal remote sensing: 2. Surface moisture


FAO (2009). How to feed the world in 2050.


Holtan, H. N.


177


Taylor, T. W. and Ravet, F. W. A meteorologically driven maize stress indicator model.


