# An Observing System Simulation Experiment for Soil Moisture Measurements from the SMAP Radiometer

by

Alexandra Georges Konings

Submitted to the Department of Civil and Environmental Engineering
as a supplement to the requirements for the degree of

Bachelor of Science in Environmental Engineering Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2009

Author ..............
Department of Civil and Environmental Engineering
May 8, 2009

Certified by ..............
Dara Entekhabi
Bacardi and Stockholm Water Foundations Professor
Thesis Supervisor

Accepted by ..............
Dennis B. McLaughlin
Chairman, Department Committee on Undergraduate Studies

# An Observing System Simulation Experiment for Soil Moisture Measurements from the SMAP Radiometer

by

Alexandra Georges Konings

Submitted to the Department of Civil and Environmental Engineering
on May 8, 2009, as a supplement to the
requirements for the degree of
Bachelor of Science in Environmental Engineering Science

## Abstract

The Soil Moisture Active Passive (SMAP) satellite, to be launched in 2013, will use both radiometer and radar data to estimate soil moisture. Improved soil moisture knowledge has many applications in hydroclimatology, numerical weather prediction, flood forecasting, and human health. In this thesis, an observing system simulation experiment (OSSE) was used to study the error structure of radiometer measurements using two different retrieval algorithms. In an OSSE, geophysical fields are used to create a model of surface emission, which is coupled to an orbital sampling module and proposed retrieval algorithms. Comparing output from the retrieval algorithm to the starting soil moisture values demonstrates retrieval error. Significant uncertainty remains about the optimal representation of the effect of dielectric mixing, soil roughness, and vegetation opacity on radiometric emissions at a given soil moisture. The effect of this uncertainty on retrieval algorithms is studied by using different representations for each term in the forward and retrieval modules of the OSSE. Uncertainty due to roughness causes less error than errors in dielectric mixing and vegetation opacity treatment. In both algorithms, the retrieval shows a spatially variable bias, which is particularly large when using a single-polarization retrieval algorithm. The spatial and temporal variation of the bias, and the implications for characterization and removal of this bias as a possible error reduction strategy, are discussed.

Thesis Supervisor: Dara Entekhabi
Title: Bacardi and Stockholm Water Foundations Professor

# Acknowledgements

The orbital module used in this thesis was adapted from one provided by Steven Chan of the NASA Jet Propulsion Laboratory (JPL). I thank him for his kindness and patience in explaining the details to me. Indeed, the very foundations of this work were laid during the summer I spent at JPL. I am very grateful for the mentorship and support of Eni Njoku during my time there.

I do not doubt that I will fail to adequately convey the extent of my gratitude towards Prof. Dara Entekhabi for guiding me in this work. His vote of confidence in allowing me a chance to work on SMAP and supervise this thesis has opened a whole new world for me. His encouragement, patience and frequent attention were invaluable in keeping me on the track to understanding and sufficiently confident in my own abilities. The consistent quality and clarity of his remarks and work are an immense source of inspiration, as is his work ethic.

The roots of my growing process as a scientist started with Prof. Dennis McLaughlin. I am exceedingly grateful for his constant generosity of time, advice, and inspiration, and for his willingness to place significant trust in me and give me my first real research responsibility. Few, if any, experiences have given me the chance to learn and grow so much as I did while working with Dennis.

Without Sheila Frankel's constant support and excellent day-to-day management of Parsons Lab, I would have gone insane (or more insane than I am now) a long time ago. I have received much academic and personal advice from many of my friends in the Entekhabi and McLaughlin groups and others - special thanks are due in this context to Mojtaba Bateni, Crystal Ng, Raf Wojcik, Leila Farhadi, Ben Scandella and many more.

Lastly, I must thank my parents and my brothers Herwig and Arno for putting up with such a difficult and negligent family member. Any and all of my positive qualities are due to their influence, patience, and love.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 The Value of Soil Moisture Measurements

Soil moisture is one of the key variables in the earth's hydrologic cycle. It is a strong control of land evaporation, particularly in water-limited regions. As such, soil moisture determines the partitioning of land surface heat fluxes into sensible and latent heat. Additionally, it acts as a limiting factor to infiltration during precipitation events, thereby partitioning the rain into infiltrating and runoff fractions. By controlling not only evaporative fluxes, but also recharge from the surface to the subsurface, soil moisture therefore links the surface and groundwater components of the global hydrologic cycle.

The scientific and engineering applications of improved knowledge of soil moisture are manifold. Among others, the use of soil moisture information can improve forecasting ability of the rainfall-to-runoff ratio [1], allowing for significantly improved flood forecasting [2]. The influence of soil moisture initialization on numerical weather prediction models is well-studied, and ranges across several physical mechanisms (temperature effects to maintain energy balance, height of planetary boundary layer as influenced by buoyancy sensible heat-created turbulence [3]) and precipitation storm types (mesoscale circulations, drylines [4]). Improved soil moisture initialization is therefore a far more promising avenue to increasing weather prediction skill than increasing model resolution, with its associated computational expense. On a larger scale, hindcast experiments have shown that climate systems in several regions have a long-term soil moisture memory, such that accurate soil moisture initialization can significantly improve the skill of seasonal temperature and precipitation predictions [5]. This may lead to improved summertime heat wave prediction, facilitating preparedness to reduce the associated human health costs. Further human health applications for soil moisture information exist in epidemiology. For example, soil moisture variability is the strongest landscape factor in predicting human-biting rates of malaria-bearing mosquitos [6]. The wide range of applications for soil moisture measurements led the National Aeronautics and Space Administration (NASA) to fund the Soil Moisture Active Passive Satellite Mission (SMAP).

## 1.2 The Soil Moisture Active Passive Mission

Because of the high spatial variability of soil moisture, the information that can be provided by *in situ* measurements is limited. Remote sensing measurements, however, can provide a clear image of the overall spatial wetting and drying patterns of the soil. SMAP is intended to provide global soil moisture measurements, and will also describe the freeze/thaw state of the soil, which has a first order impact on the balancing of carbon budgets and our understanding of the carbon cycle.

SMAP has grown out of the Hydros mission [7], which was canceled by NASA in 2005 due to budget concerns. It will have two instruments, a radiometer making passive (i.e. emissivity) measurements at 40 km and a synthetic aperture radar making active (i.e. backscatter) measurements at 3 km. Both instruments will operate at L-band frequencies (centered at 1.41 GHz and 1.26 GHz, respectively). L-band is particularly suited to radiometric soil moisture measurements due to the greater vegetation penetration and reduced atmospheric attenuation and radio frequency interference at this frequency [8]. Although the radar measurements have a higher resolution, this resolution comes at a price of significantly more noisy measurements. Aside from the 40 km and 3 km datasets based on each of the two instruments, a dataset with resolution of 10 km dataset will also be provided, based on combining data from the radar and the radiometer.

SMAP will travel in a low-earth orbit, at an altitude of about 670 m. The radar and radiometer will share a large (6 m diameter) reflector antenna and feedhorn, which scan conically over a 1000 km wide swath. The incidence angle of the beam is constant at 40°. The combination of the conical scan with movement of the satellite along its nadir track allow the satellite to sample a given area by filling in several footprints over time as the conical scan returns multiple times at a slightly offset location. This is illustrated in Figure 1-1. The SAR processing is impossible in the middle of the swath, such that there is a 300 km radar data gap in the middle of the swath (although radiometer observations are still made in these areas). As a result, revisit times (the time between measurements over a given area) will be slightly larger for radar than radiometer data. Although the exact revisit times depend on latitude and the final satellite orbit altitude, they are intended to be about 3 days. This design parameter is determined by the time between precipitation storms and surface wetting/drying timescales [9].

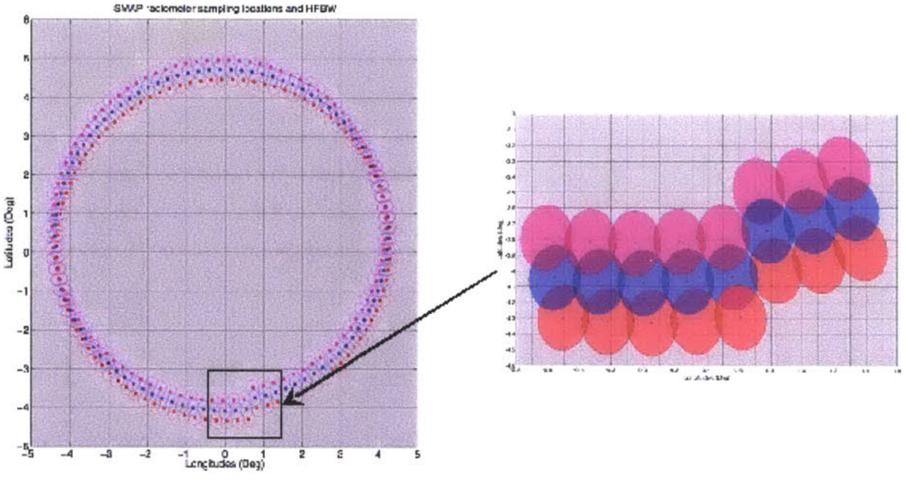


Figure 1-1: Sampling pattern dictated by conical scan. Image by Steven K. Chan, JPL

## 1.3 Motivations

Soil type, topography, and vegetation all have a confounding influence on radiometric soil moisture measurements. It is therefore critical that retrieval algorithms used by SMAP are tested over a wide variety of regions. However, this effort is frustrated by the limited availability of both radiometer measurements and verificiation data. Some data may be obtained by field experiments and measurements from dedicated plane flights (as in Njoku et al. [10], for example), but the expense and resource needs of such campaigns limit their frequency and coverage. An alternative method of testing retrieval algorithms is the use of an observing system simulation experiment (OSSE). In an OSSE, the emission over a given area is simulated using soil moisture and other ancillary fields that are taken to represent true data. After coupling this emission to proposed retrieval algorithms, sometimes via an orbital sampling module, the error associated with the retrieval is given by the difference between the input and output soil moisture values.

Since the emission is explicitly simulated, the starting data merely need to be realistic, rather than exactly equal to the truth. For example, soil moisture fields predicted by land surface models are a valid data source. Due to the influence of soil type, topography, and vegetation cover, ancillary variables are necessary to describe these influences on surface emission. Care must be taken to ensure that errors in these ancillary variables do not break down the realism of the OSSE; the variables must be physically consistent with each other and with soil moisture fields. This can be accomplished by limiting the number of data sources. Using an OSSE also has the advantage that the influence of different factors on retrieval accuracy can be isolated by changing the model for surface emission (the forward component of the OSSE). Furthermore, errors due to subpixel heterogeneity can be studied easily using an OSSE. The ancillary parameters used in retrieval errors often vary significantly within a measurement pixel. Due to the non-linearity of the radiative transfer equation, using a linear average for retrieval leads to error. Explicit control over the amount of variation in the forward model of an OSSE allows for greater study of this nature, including changes to the retrieval procedure to reduce error. Nevertheless, using a synthetic experiment is no all-powerful solution. The accuracy of the inferences from an OSSE is dependent not only on the realism of the datasets used, but also on the realism of the emission model.

Crow et al. [11] carried out an OSSE for Hydros radiometer products in preparation for that mission. That OSSE covered the Red-Arkansas River Basin in the U.S. Southern Great Plains over the course of about a month. As such, it was limited in scope both in space and time, and did not encompass the full range of conditions observed by the satellite. In addition, the forward and retrieval methods were based on the same theoretical description of emission, despite the fact that uncertainty remains about the exact parametric form of the influence of soil and vegetation parameters. Errors in these retrieval models can easily lead to retrieval errors well above the 5% volumetric error requirement for SMAP radiometer estimates. Holmes et al. [12] conducted a global simulation of top of the atmosphere L-band brightness temperature for several different soil effective temperature, dielectric constant, roughness and vegetation representations. However, the impact of the different representations on soil moisture retrievability was not considered.

In this thesis, an OSSE is implemented for SMAP radiometer measurements at the contiguous United States (CONUS) scale, over the course of a full annual cycle. In addition, different model representations are used in the forward and retrieval parts of the OSSE to study the extent of the measurement errors due to imperfections in the emission models on which retrieval algorithms are based. Two different retrieval channels are considered: a single-channel algorithm that uses H-polarization measurements alone to retrieve soil moisture, and a dual-channel algorithm that uses measurements at both H- and V-polarization to retrieve both soil moisture and vegetation water content, the mass of water in above-ground vegetation per area.

## 1.4 Bias Removal

Although numerous field campaigns have demonstrated the feasibility of estimating soil moisture to within 5% volumetric using L-band radiometry data, these campaigns generally rely on area-calibrated values of soil texture and roughness parameters and of the vegetation water content models (e.g. [13, 14]). Such calibration is infeasible on a global scale. Indeed, it may not even be feasible to pin down the optimal functional form of some of the contributing effects before the SMAP launch date. An alternative strategy is to try to characterize and remove the bias of retrieval estimates. Such estimates may be provided by future field campaigns or by using radiometric measurements from the ESA Soil Moisture and Ocean Salinity mission (SMOS), which will provide L-band radiometric measurements at several incidence angles. Bias removal has previously been demonstrated to have the potential for significantly improving the use of remotely sensed soil moisture estimates for land surface modeling, if a sufficiently long climatological history of estimates is available [15]. However, as the authors of this study point out,using a more spatially extensive area might reduce the necessary record length for data assimilation applications if probability matching techniques are used [16]. This thesis also investigates the extent to which errors due to uncertainty in the radiative transfer description, as well as land surface heteorgeneity and measurement errors are removable by defining a static bias.

## 1.5 Outline

Chapter 2 of this thesis describes the radiative transfer model used, including the three terms for which the effect of description uncertainty is investigated in the thesis. Alternative models are presented for each term. It also describes the data sources and orbital module used and the overall configuration of the OSSE. Chapter 3 describes the two retrieval algorithms used (a single-polarization and a dual-polarization one). It also provides results for the full OSSE with each algorithm and several different combinations of forward and retrieval emission descriptions. The analysis of these results includes a study of the potential for bias-removal in both algorithms. Conclusions and some recommendations for future work on L-band radiometric soil moisture retrieval are provided in Chapter 4.

# Chapter 2

# Simulation Design

This chapter describes the design of the OSSE system. The different models used to describe emission are detailed in section 2.1-2.2. Section 2.3 describes the earth-fixed grid and summarizes the data sources used. Section 2.4 describes the orbital module, including its coupling with the rest of the OSSE. Assumptions about expected error are explained in section 2.5.

The overall process is shown in Figure 2-1. The first line of the diagram illustrates the forward model, which uses soil moisture fields with ancillary parameters and an orbital model to create a vector of brightness temperature measurements. As demonstrated in the second line, using knowledge of sensing location and (coarser, aggregated) ancillary fields, the retrieval algorithms generate a vector of soil moisture values. These are processed to a set of estimates on the 40-km earth-fixed grid, as illustrated in the third line. Lastly, comparison of the retrieved values to the original soil moisture field produces an estimate of the retrieval error.

## 2.1 Emission Models

### 2.1.1 $\tau$-$\omega$ Model

Given knowledge of the greybody temperature, radiometer measurements are based on the emissivity $e_p$. The relation between the measured brightness temperature

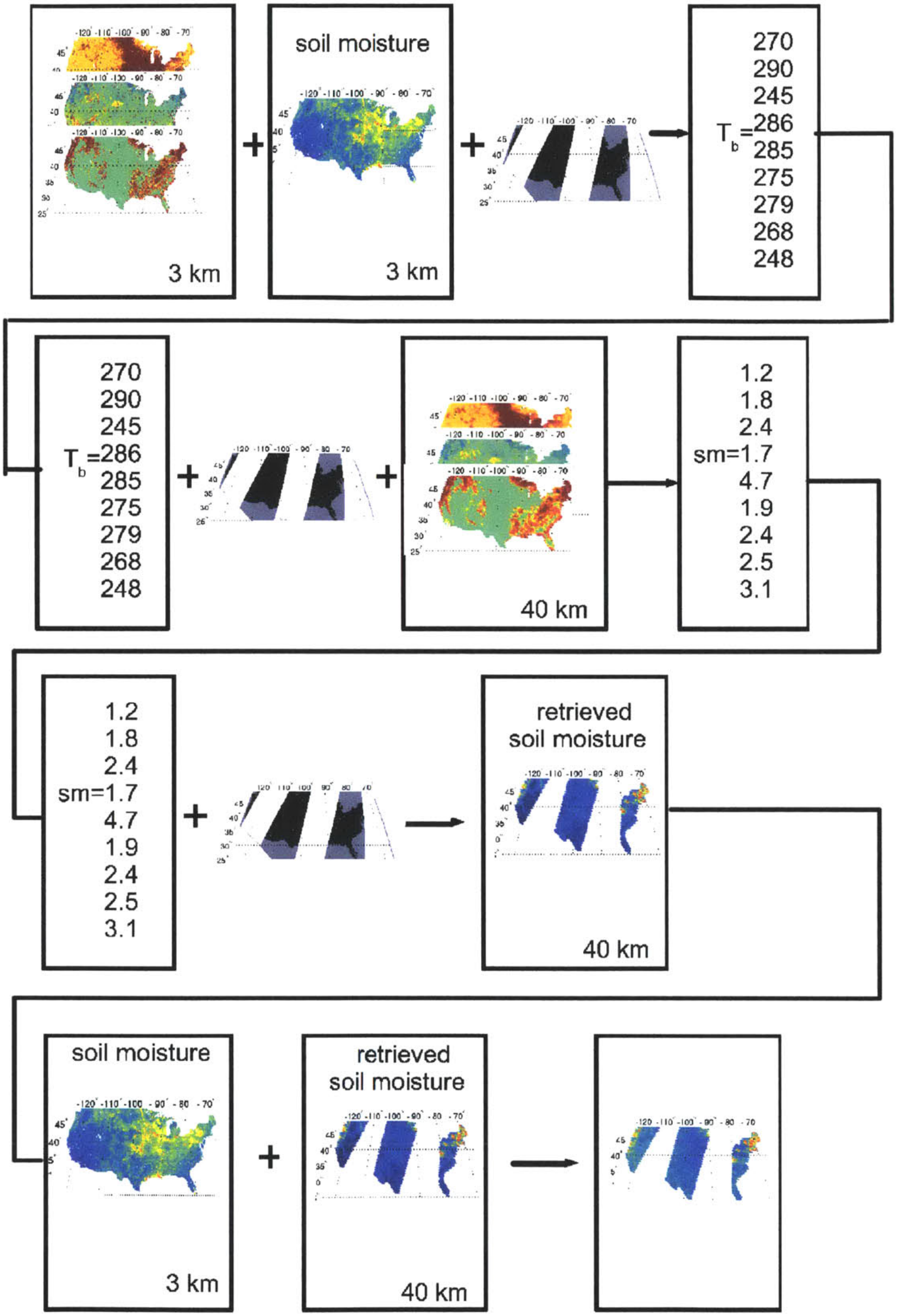


Figure 2-1: Representation of the overall OSSE system. Each line represents the forward model, retrieval, re-processing to the earth-fixed grid, and error comparison, respectively

$T_{bp}$ and $e_p$ can be described using the $\tau$-$\omega$ model, which describes the scattering of emission rays using a two-layer model of the land surface, with a single soil and a single canopy layer.

$$T_{bp} = T_s e_p \exp(\frac{-\tau_p}{\cos\theta}) + T_c(1-\omega_p)(1-\exp(\frac{-\tau_p}{\cos\theta}))(1+R_p\exp(\frac{-\tau_p}{\cos\theta})). \tag{2.1}$$

$T_s$ is the effective soil temperature over the measured depth, $\tau_p$ is the vegetation opacity, $\theta$ is the incidence angle of the measurement, and $\omega_p$ is the single scattering albedo of the vegetation. The subscript p has a value of $V$ or $H$ and describes vertical or horizontal polarization, respectively. $R_p$ is the effective reflectivity of the soil, related to the emissivity (by conservation of energy) by

$$e_p = 1 - R_p. \tag{2.2}$$

Emission from the soil, as attenuated by the vegetation in the canopy layer, is represented by the first term in eq. 2.1, while the second term describes the scattering effect of the canopy layer. Multiple scattering is neglected, as are contributions from downwelling atmospheric radiation. At L-band frequencies, this model is sufficient to adequately describe emission [17].

In this simulation, the dependence of albedo on polarization is neglected. As in the Hydros radiometer study, we parameterize $\omega$ based on land cover type. The land cover data were classified at a spatial resolution of 1 km using a neural network based on infrared remote sensing measurements aboard the MODIS satellite [18]. The class scheme used is the so-called 'University of Maryland' scheme, which has the advantage of having similar classification types to those used to originally derive the VWC parametrization (see also section 2.1.3). Although the land cover categories used are slightly different than those in Crow et al. [11], classes were matched to be as close as possible. The albedo values used for each land cover type are listed in table 2.1.

In order to successfully capture the high temporal variability of $T_s$, $T_c$, and the soil moisture in our simulation, we use model output from the North American Land

| **Land Cover Type** | $b_h$ **[m$^2$/kg]** | $b_v$ **[m$^2$/kg]** | $a_{geo}$ | $\omega$ | $s$ **[m]** | $h$ |
|---|---|---|---|---|---|---|
| Evergreen needleleaf forest | 0.08 | 0.12 | 2/3 | 0.12 | 0.01 | 0.10 |
| Evergreen broadleaf forest | 0.096 | 0.144 | 2/3 | 0.12 | 0.01 | 0.10 |
| Deciduous needleleaf forest | 0.08 | 0.12 | 2/3 | 0.12 | 0.01 | 0.10 |
| Deciduous broadleaf forest | 0.096 | 0.144 | 2/3 | 0.12 | 0.01 | 0.10 |
| Mixed forest | 0.088 | 0.132 | 2/3 | 0.12 | 0.01 | 0.10 |
| Closed shrubland | 0.099 | 0.121 | 1/3 | 0.12 | 0.01 | 0.10 |
| Open shrubland | 0.099 | 0.121 | 1/3 | 0.12 | 0.01 | 0.10 |
| Woody savanna | 0.088 | 0.132 | 1/3 | 0.12 | 0.01 | 0.10 |
| Savanna | 0.094 | 0.127 | 1/3 | 0.08 | 0.01 | 0.10 |
| Grassland | 0.09 | 0.11 | 1/3 | 0.05 | 0.01 | 0.10 |
| Cropland | 0.117 | 0.143 | 1/3 | 0.05 | 0.015 | 0.15 |
| Urban | 0 | 0 | 0 | 0 | 0 | 0 |
| Barren or sparsely vegetated | 0.09 | 0.11 | 0 | 0 | 0.01 | 0.10 |
| Water | 0 | 0 | 0 | 0 | 0 | 0 |

Table 2.1: Soil roughness, albedo, and vegetation model parameters used based on land cover type. Note that the $\tau$-$\omega$ model was not actually applied to areas with cover type water or urban.

Data Assimilation System (NLDAS). NLDAS is a system of four land surface models run with a common set of forcing and ancillary parameters [19]. We use the Noah land surface model, which provides layer output at approximately the SMAP sensing depth. Although NLDAS provides hourly output, all measurements are approximated by assuming they occur within an hour of the 6:00 equator crossing time (see also section 2.4), corresponding to a UMT time of 12:00. The canopy temperature is taken to be equal to the 2 m air temperature (derived from NCEP's Eta-based Data Assimilation System) used to force NLDAS. Radiative transfer theory describes the effective (measured) soil temperature at depth z as

$$T_{eff} = \int_0^\infty T_s z \alpha(z) \exp(\int_0^z \alpha(z')dz')dz \tag{2.3}$$

where

$$\alpha(z) = \frac{4\pi}{\lambda} \frac{\epsilon"(z)}{2\epsilon'(z)^{0.5}} \tag{2.4}$$

where $\epsilon'$ and $\epsilon"$ are the real and imaginary components of the soil dielectric constant [20]. Eq. 2.3 indicates $T_{eff}$ thus depends on the temperature profile of the soil (and, via the large influence of soil moisture on the dielectric constant, the soil moisture profile). To avoid the need for exact knowledge of soil profiles required by equation 2.3, we follow the parametrization of Crow et al. [11]. Thus, $T_s$ is simplified to the average of the skin temperature $T_0$ and the temperature at 5 cm depth $T_5$. $T_0$ is taken to equal the 2 m air temperature, and $T_5$ is the temperature in the top layer of the land surface model (representing a depth from 0 to 10 cm). The measured soil moisture is assumed to equal that in the top Noah layer (although the true sensing depth may be somewhat lower than 10 cm).

$$T_s = \frac{T_0 + T_5}{2} \tag{2.5}$$

where $T_i$ is the temperature at a depth of i cm. The skin temperature $T_0$ is taken to equal the 2 m air temperature, and $T_5$ is the temperature in the top layer of the land surface model (representing a depth from 0 to 10 cm). The exact depth of measured soil moisture sensed is dependent on the moisture and soil roughness profiles. It is generally lower than the temperature sensing depth. In the OSSE, we assume the measured soil moisture equals that in the top Noah layer (although the true sensing depth may be somewhat lower than the 10 cm this layer represents).

### 2.1.2 Bare Soil Emission

Several empirical and semi-empirical models exist to relate the soil dielectric constant $\epsilon$ to the soil moisture content. We use the model of Wang and Schmugge [21] in the forward model, and the model of Dobson et al. [22] for retrieval. This model has several empirical parameters based on regressions from the dataset of Hallikainen et al. [23]. We use the slightly differing fitted parameters, based on the same dataset, derived by Peplinski et al. [24]. Each model contains a different consideration of both the mixing properties of the dielectric contributions from rock, ice, water, organic matter, and air. Furthermore, the dielectric constants of each of the above are not modeled in the same manner. To derive dielectric contributions from soil rock, each model relies on (different) empirical relations based on bulk density ($\rho_b$), sand fraction (sf) and clay fraction (cf) of the soil. The Wang and Schmugge [21] model further considers porosity and the soil wilting point (also as an empirical function of the above properties). The pure water component of the Dobson et al. [22] model, using the full Debye equations, also has a small dependence on temperature. At L-band, some limited experiments have shown a closer agreement with measurements of the Wang and Schmugge model than the Dobson et al. model. The difference between the two model predictions depends on moisture as well as soil type, and is generally more pronounced for the (relatively larger) real than the imaginary component [25].

The soil properties used are from the CONUS-SOIL database, a reprocessing of the State Soil Geographic Database (STATSGO) by the USDA. This reprocessing dataset is intended specifically for hydrological modeling applications [26], and is also used in the NLDAS land surface models. It has a spatial resolution of 30 arcseconds, or a little less than 1 km. CONUS-SOIL discretizes each soil property in 11 depth layers. The values used are averages of those in the top two layers of 0-5 cm and 5-10 cm, respectively. Missing values were filled by taking the mode of the surrounding 3 by 3 cells (or, if missing values occurred in a large block, the mode of an n by n cell, where n is the smallest odd number usable).

For perfectly smooth soil, $\epsilon$ can be related to the effective reflectivity (and by eq. 2.2 to emissivity) by the Fresnel equations

$$r_h = \left| \frac{\cos\theta - \sqrt{\epsilon - \sin^2\theta}}{\cos\theta + \sqrt{\epsilon - \sin^2\theta}} \right|^2 \tag{2.6}$$

$$r_v = \left| \frac{\epsilon\cos\theta - \sqrt{\epsilon - \sin^2\theta}}{\epsilon\cos\theta + \sqrt{\epsilon - \sin^2\theta}} \right|^2 \tag{2.7}$$

This formulation, however, does not account for scattering from soil particles themselves. Such soil roughness effects are particularly important over agricultural fields.

Models of soil roughness tend to take one of two approaches. Several theory-based models start from the basic electromagnetic formulations and perform ray tracing to calculate overall emission. Such models require assumptions about the statistical distribution of soil roughness, including parameter values that are impossible to obtain without field measurements. This virtually eliminates them from operational use or use in synthetic experiment such as this one without making strong assumptions about soil parameters. They also suffer the drawback of being very computationally expensive. Furthermore, they are not invertible, preventing their use in soil moisture retrieval algorithms. This approach is used in the forward component of the OSSE. $R_p$ is first expanded into a noncoherent and coherent component

$$R_p(\theta) = R_p^{non}(\theta) + R_P^{coh}(\theta) \tag{2.8}$$

Using the Fresnel reflectivities of eqs. 2.6 and 2.7, the coherent term is defined as in Shi et al. [27],

$$R_p^{coh}(\theta) = r_p \exp(-4k^2 s^2 \cos^2\theta), \tag{2.9}$$

where k is the angular wave number in free space and s is the standard deviation of the surface height. The noncoherent term can be found by integrating the bistatic scattering component over the hemisphere above the soil,

$$R_p^{non} = \frac{1}{4\pi\cos\theta}\int_0^{2\pi}\int_0^{pi/2}[\sigma_{pp}(\theta,\theta_j,\phi_j) + \sigma_{pq}(\theta,\theta_j,\phi_j)] sin(\theta_j) \mathrm{d}\theta_j \mathrm{d}\phi_j. \tag{2.10}$$

In the above, $\theta_j$ and $\phi_j$ are spherical coordinates of the scattering direction. $R_p^{coh}$ can calculated by the integral equation model (IEM) of Fung [28]. IEM assumes a given random distribution of soil particles and perfectly conducting surface, and uses Maxwell's equations to derive the scattering of a soil. An exponential correlation function is assumed to describe the soil. Due to the difficulty of obtaining an approximation for the correlation length $l$, no wide-scale datasets exist. We therefore assume, as in Holmes [29], that $l$ is linearly related to the surface root-mean-square height $s$, $l = \frac{6}{2.2}s$. Although $s$ is no more straightforward to obtain reliably, an attempt to parametrize based on land cover has been made [11]. The parametrization used is based on those values, and is listed in table 2.1. Since the dielectric mixing model is used within the IEM, information about soil bulk density, sand fraction and clay fraction is also required. The double integral in equation 2.10 is relatively expensive to calculate. Using IEM for each footprint in this OSSE would require prohibitively large computational resources. Furthermore, discretization efforts are hindered by the sensitivity of the model to each of the 4 main input parameters ($\rho_b$, sf, cf, and mv). Instead, the parameterization of IEM developed by Shi et al. [27] is used. It fits IEM to equations of the form

$$R_p^{non} = A_p r_p^{B_p}$$

where $A_p$ and $B_p$ are of the form

$$exp(a) + b\log(ks) + c(ks) + dW$$

where W is the power spectrum and a, b, c, and d, are quadratic functions of $\theta$ depending on polarization.

Alternatively, soil roughness can be described by semi-empirical parameterizations. The h-Q model is the predominant model of this type. In its original formulation, the h-Q model is given by

$$r_p = [(1-Q)r_p + Qr_q]\exp(-h_p \cos^N \theta), \tag{2.11}$$

where Q, h, and N are model parameters [30]. q is the polarization orthogonal to p; when p is $H$, q is $V$ and vice versa. The Q parameter increases with frequency. Although not all studies predict a Q value of exactly 0 at L-band, an extensive review of the literature has nevertheless found that Q can generally be taken to equal 0 for L-band studies [31]. Similarly, N is generally taken to equal 0 at L-band. Using those values, eq. 2.11 reduces to

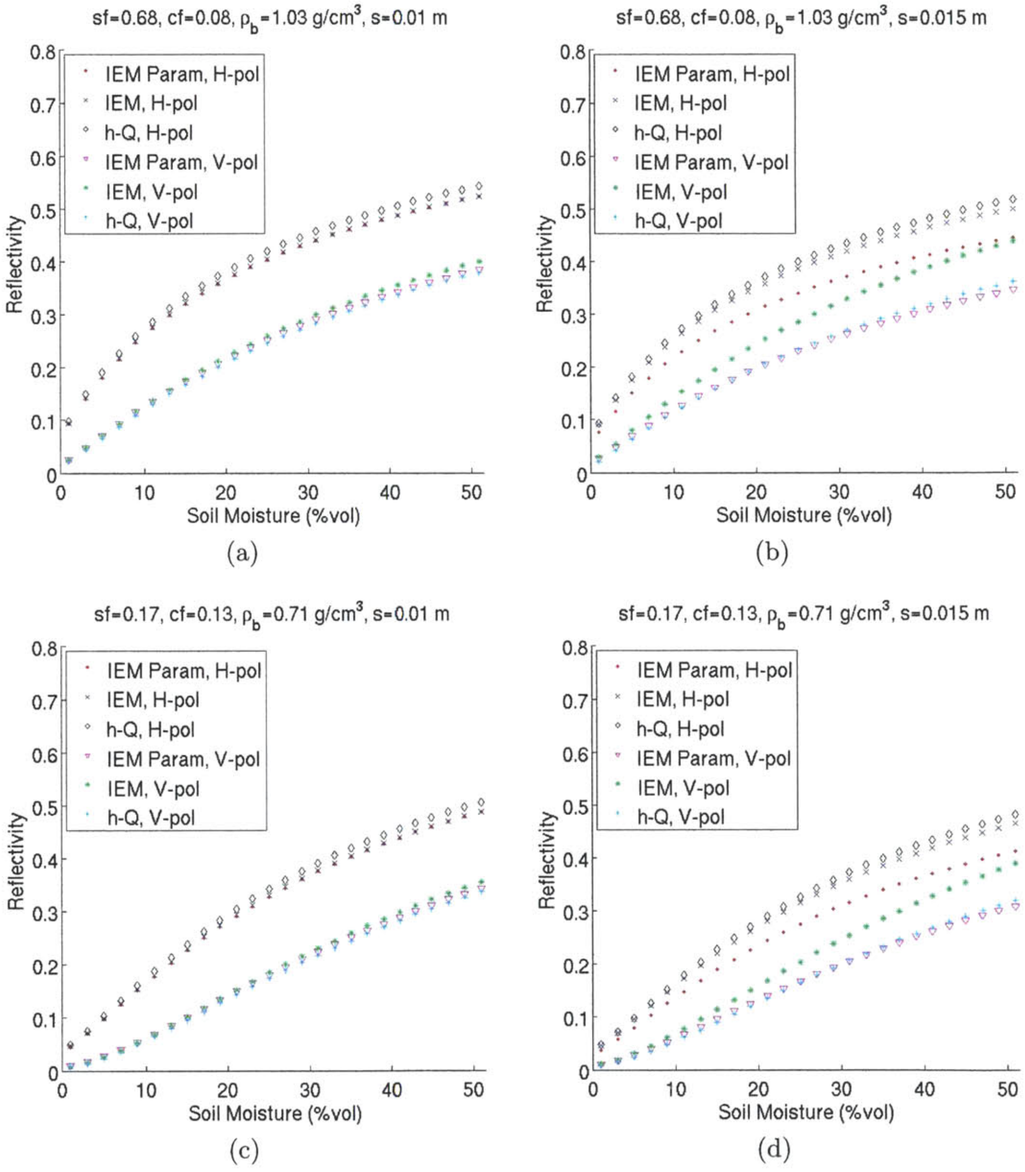
$$r_p = r_p \exp(-h) \tag{2.12}$$

Several parameterizations for h exist, including ones with dependence on s and soil moisture [32, 33] and the ratio s/l and soil moisture [34]. In this simulation, we follow the Hydros radiometer OSSE and assume h=10s (where s is in meters and h is dimensionless). Since s has been parametrized by land cover type, h is also a function of land cover type. The very abundance of different parameterizations shows the behavior of h is not fully understood. In addition, the h-Q model assumes the surface roughness effect is independent of polarization, despite the fact that it significantly influences the ratio $R_v/R_h$ [27].

Neither model is able to perfectly imitate true soil behavior. Nevertheless, a model of soil roughness effects is a necessary component of a soil moisture retrieval algorithm. We can obtain a glimpse of the effect of assuming an imperfect roughness model by comparing the predictions of the different models for typical soil conditions. Figure 2-2 compares the reflectivity variation with soil moisture as calculated for two sets of typical soil physical properties (sf, cf, $\rho_b$). Since s is parametrized as 0.01 m over the vast majority of land cover types, figures 2-2a and 2-2c assume s=0.01m. Figures 2-2c and 2-2c, using the same physical properties as 2-2a and 2-2c, respectively, take s = 0.015 m, the value assigned over cropland. It is immediately obvious that for s = 0.015 m, the different models diverge more. The IEM parameterization underestimates the full IEM at both polarizations. At V-pol, however, the IEM predictions are significantly lower than those from the h-Q model, such that the parametrization and h-Q still match up particularly well. The error between various models, for a given set of soil conditions and measurement polarization, also shows moisture dependence, such that the bias in soil moisture retrievals will not be time-independent. Figures 2-2a and 2-2c indicate that although predicted reflectivity values change for different soil conditions, the difference between the two models used is roughly the same, indicating that the error induced by the differences in roughness model may not vary significantly space. Nevertheless, the moisture dependence mentioned above, as well as the change in slope (and thus, different relative size of the measurement noise), will cause spatial variations. Note that vegetation cover will significantly affect these predictions.

### 2.1.3 Vegetation Opacity

Vegetation attenuates the soil emission. The amount of attenuation, represented by $\tau$ in Eq. 2.1, varies based on both the type of vegetation and its water content. Effects due to vegetation can be parametrized based on land cover. Accounting for



(a) (b)

(c) (d)

Figure 2-2: Reflectivity dependence on soil moisture derived using different roughness models, for typical soil conditions.

the influence of vegetation water content (VWC) requires knowledge of the amount of VWC, defined as mass of water per area. Unfortunately, no methods for remotely sensing VWC directly exist. Although empirical relationships between VWC and NDVI have been found, these have only been tested over very limited cover type conditions [35, 36]. An alternate approach which separates VWC into contributions from the woody stems of the vegetation and contributions from the foliage, was used here [37].

$$VWC = \frac{C_f W_f}{0.45} + \frac{C_w W_w}{0.45} \tag{2.13}$$

$C_x$ denotes the mass of carbon per area for the foliage component f or the woody component w and $W_x$ denotes the fraction of the biomass component x composed of water. The amount of biomass and the mass of carbon are related by assuming the biomass is composed of 45% carbon. Variations in biomass over the growing season are tracked by leaf area index (LAI) measurements. These are related to the above-ground carbon mass by two empirical relations

$$C_f = \frac{(LAI)(LAI_f)}{SLA} \tag{2.14}$$

$$C_w = f_w(1.25 LAI_{mx}) \tag{2.15}$$

In equation 2.14, $LAI_f$ is the ratio of all-side LAI to 1-sided area and SLA is the canopy specific leaf area. In equation 2.15, $LAI_{mx}$ is the annual maximum LAI, and $f_w$ is the fraction of the vegetation composed of woody biomass. Values for $LAI_f$, $SLA$, $f_w$, $W_f$, and $W_w$ are based on land cover type, and are derived from values in the literature (see [38] for references). The values used are listed in table 2.2. LAI measurements are derived, like the land cover dataset, from the MODIS satellite [39]. The dataset used is a mosaic of MODIS data at 1 km spatial resolution and monthly temporal resolution, covering a full annual cycle from January 2002 to December 2002.

The resulting dataset ranges from a minimum of 0 to 21 kg/m$^2$. The annual average VWC is shown in figure 2-3. The discretization in land cover classes causes

| **Land Cover Type (Source Land Cover)** | $LAI_f$ | $SLA$ | $f_w$ | $W_w$ | $W_f$ |
|---|---|---|---|---|---|
| Evergreen needleleaf forest | 2.6 | 8.2 | 1.0 | 1.0 | 0.681 |
| Evergreen broadleaf forest | 2.0 | 32.0 | 1.0 | 1.0 | 2.34 |
| Deciduous needleleaf forest | 2.0 | 22.0 | 1.0 | 1.0 | 1.95 |
| Deciduous broadleaf forest | 2.0 | 32.0 | 1.0 | 1.0 | 1.95 |
| Mixed forest | 2.3 | 20.0 | 1.0 | 1.31 | 1.0 |
| Closed shrubland | 2.3 | 12.0 | 0.12 | 0.54 | 0.82 |
| Open shrubland | 2.3 | 12.0 | 0.06 | 0.54 | 0.82 |
| Woody savanna (woodland) | 2.0 | 40.0 | 0.5 | 1.0 | 1.86 |
| Savanna (wooded grassland) | 2.1 | 30.0 | 0.25 | 0.54 | 1.30 |
| Grassland (grassland & cereal crops) | 2.0 | 49.0 | 0 | N/A | 1.78 |
| Cropland (broadleaf crops) | 2.0 | 12.0 | 0 | N/A | 2.0 |
| Barren or sparsely vegetated | 2.3 | 12.0 | 0 | N/A | 0.82 |

Table 2.2: Parameters used to calculate VWC, based on land cover type

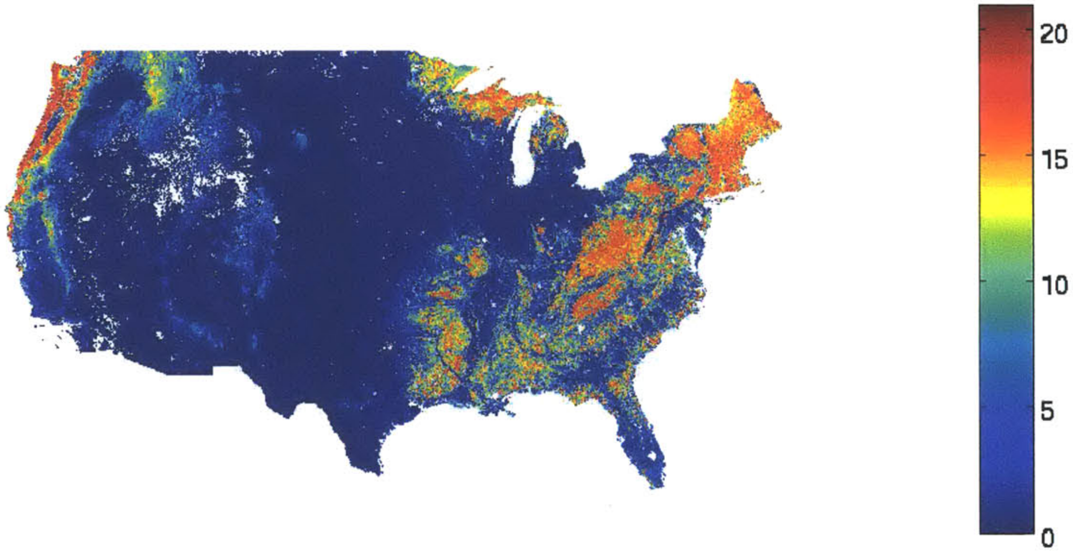


Figure 2-3: Annual mean derived vegetation water content, in kg/m$^2$

some areas to move from high to low VWC more abruptly than is likely to be realistic. Nevertheless, the dataset adequately capture the larger-scale spatial variation of VWC. In particular, it is preferable to models intended for a smaller variety of land cover type. Figure 2-4 shows the annual range of VWC. Note that range and typical amount are not correlated.

Again, we use different models in the forward and retrieval modules to reflect uncertainty about these parameterizations. The forward vegetation opacity model is

Figure 2-4: Annual range of derived vegetation water content, in kg/m$^2$

based on a modification to the low-frequency model of Kirdyashev et al. [40].

$$\tau = a_{geo} \frac{2\pi}{\lambda} \frac{W}{\rho_W} \epsilon''_{sw}. \tag{2.16}$$

$\lambda$ is the wavelength, $\rho_W$ is the density of water (with a minor dependence on the canopy temperature $T_c$), $\epsilon''_{sw}$ is the imaginary component of the dielectric constant of saline water, and $W$ is the vegetation water content (VWC). $a_{geo}$ was originally a theoretical parameter. However, it is far too difficult to calculate exactly for actual surface covers. Some values for different vegetation types are given in Wegmuller et al. [41] and Holmes [29]. The complete parametrization for each land cover type is based on these values; $a_{geo}$ is 2/3 for forests and 1/3 otherwise (see also table 2.1). To calculate $\epsilon''_{sw}$, the model by Swift and Klein [42] is used, assuming a salinity of 6 parts per thousand, as in Holmes [29]. Although this salinity is itself likely to vary with vegetation type and soil conditions, the dependence of $\tau$ on the salinity is small enough that using a single representative value is acceptable.

Rather than starting from a theoretical formulation such as that in eq. 2.16, a simple linear relationship between $\tau$ and VWC is also commonly used [43].

$$\tau = b_p \tag{2.17}$$

The b-parameter at polarization p depends on land cover type. Its parametrization is based on that in Crow et al. [11]. The values are outlined in table 2.1.

In preparation for SMOS (a forthcoming radiometer-only soil moisture and ocean salinity satellite by the European Space Agency), several studies were undertaken to characterize the dependence of the b-parameter on incidence angle, polarization, and time. Wigneron et al. [44] reviewed the findings of these studies, and found that taking b to be constant in time is a fair assumption during all but the early stages of the growing season. This review also suggests that at the SMAP incidence angle of $40^o$, the vertical structures in vegetation canopy tend to cause $b_v$ to be greater than $b_h$. However, this observation is dependent on crop type. Since most studies were over agricultural or grass land, it is not clear whether such a polarization dependence exists for all land cover types. Note that $b_v > b_h$ in the parameters in table 2.1, but that the Kirdyashev model does not include any polarization dependence.

## 2.2 Areas Without Retrieval

Obviously, soil moisture measurements cannot be made over the ocean. Three additional area types exist where no retrieval is attempted: snow-covered areas, urban areas, and inland water bodies such as lakes or rivers. Snow itself has a large and highly variable emission at L-band that dominates the soil moisture signal and makes retrieval impossible. Any pixel with a fractional snow cover (as predicted by NLDAS) greater than zero is removed from consideration. Indeed, as soon as any part of a footprint contains snow cover, the entire footprint is considered flagged and removed from the OSSE.

Unlike snow, the contribution of inland water bodies to a given sample's brightness temperature can be simulated reasonably well. Rather than throwing away possible data as soon as any part of a footprint has water cover, we merely remove it from the observation. After calculating the fraction of the footprint with water cover $f_w$ and the brightness temperature over the water areas $T_{bw}$, the brightness temperature over land areas $T_{bl}$ can be calculated as

$$T_{bl} = \frac{T_b - f_w T_{bw}}{1 - f_w}. \tag{2.18}$$

The dielectric model of Swift and Klein [42] is used to calculate $T_{bw}$. A constant salinity of 0.5 parts per thousand, a typical value for freshwater, is assumed. Since water temperatures are not available from NLDAS, we assume the temperature of the observed water is equal to the 2 m air temperature. This model is used for both the forward and retrieval models.

Although soil moisture cannot be retrieved over urban areas, it is necessary to model the contribution of sub-pixel scale urban areas to the pixel-wide brightness temperature. No models for L-band emission from urban areas exist. Here, the $\tau$-$\omega$ model is adapted to represent emission from covered areas and modulated by scattering from buildings, such that

$$T_{bu} = r_u T_s \exp(-\tau) + T_c(1 - \omega_u)(1 - \exp(-\tau))(1 + r_u \exp(-\tau)), \tag{2.19}$$

where $\exp(-\tau) = 0.95$ and $r_u$ is the reflectivity assuming a dielectric constant $\epsilon = 12.5$ for concrete [45]. $\omega_u = 0.22$, based on surface albedo data [46, 47]. Although this model almost certainly does not provide an accurate description of urban emission, it provides an adequate guess. Precise modeling of urban areas is not necessary for this study, as the brightness contributions from these areas are removed in retrieval much like they are removed for inland water bodies. Nevertheless, since the relative influence of instrument error is dependent on the magnitude of removable urban emission in the signal, it is useful to include an approximate estimate of emissions. Knowing $T_{bu}$, (2.18) is extended for urban areas:

$$T_{bl} = \frac{T_b - f_w T_{bw} - f_u T_{bu}}{1 - f_w - f_u} \tag{2.20}$$

where $f_u$ is the fraction of the pixel covered by urban areas. Pixels covered entirely by water, urban areas, or a combination of both are also removed from the OSSE.

## 2.3 OSSE Grid

Each of the datasets used are converted to a single 3 km earth-fixed grid by repeating values or averaging, depending on the initial resolution. The data used, including their spatial and temporal resolution, are summarized in table 2.3. The low number of data sources reduces the chance of modelling physically unrealistic conditions due to combination of error.

| **Source** | **Variables** | **Spatial Res. [km]** | **Temporal Res.** |
|---|---|---|---|
| NLDAS output | sm, $T_s$, snow flag | ~11 | hourly |
| NLDAS forcing | sm, $T_c$ | ~11 | hourly |
| MODIS LAI | VWC | 1 | monthly |
| MODIS land cover | VWC,$\omega$,h,s,l,$b_p$, | 1 | monthly |
| CONUS-SOIL | $\rho_b$,sf,cf | ~0.7 | static |
| GTOPO30 | topography | ~0.7 | static |

Table 2.3: Summary of datasets used

## 2.4 Orbital Model

To simulate measurements, the CONUS-wide models need to be linked to a model of sampling locations with time. We use an orbital sampling model adapted from code written at the NASA Jet Propulsion Laboratory. It uses standard atmospheric flight equations to track the boresight sampling of a single orbit around the earth. This orbit is then moved according to the earth's rotational velocity to follow the nadir track as it shifts longitudinally around the globe. Trying to define a closed form equation for the areal extent of each footprint and interpolating within would be impossibly expensive computationally. Instead, each footprint is defined by a finite number of points at a constant latitude-longitude spacing, each within a 3 dB power contour. The precise incidence angle for each point is also tracked.

Every point is related to one of the OSSE's 3 km earth-fixed grid cells. In order to sample each of the square OSSE grid cells covered by the footprint, we use a latitude spacing of $0.018^o$, or 2 km. This, in turn, results in 3 km grid cells being sampled multiple times. All points that do not sample a unique grid cell are discarded. As a result, the number of point-based calculations per footprint is variable. Since all processing is done on a daily basis, a given grid cell may still be visited multiple times by overlapping footprints in consecutive scans. However, since the non-overlapping part of the footprints are different, the two may not predict the exact same soil moisture. Predictions of multiple passes are linearly averaged. SMAP's low-earth orbit is such that the satellite track crosses the equator on its ascending (Northward) pass at the same local time during each orbit. Although the exact time of this crossing depends on the details of the final mission, we assume here that it is 18:00. This implies that the satellite will cross the equator at 6:00 local time on its descending passes. Morning passes are preferable for soil moisture retrieval, both because the soil temperature and moisture profiles are most uniform, and because the atmospheric ion content is at a minimum, so that the associated Faraday rotation of the signal is also minimized. Therefore, SMAP will take measurements only on ascending passes. Accordingly, samples taken while SMAP is moving southward are not included in the OSSE.

## 2.5 Error Introduction

This OSSE neglects to seperately model individual expected sources of error to the SMAP instrument measurements. These errors are included by adding a combined Gaussian error with standard deviation of 1.5 K to the forward simulated brightness temperatures at the footprint level. The magnitude of this noise is based on system estimates by the SMAP science team and accounts for both expected instrument noise (including radio frequency interference and precision and calibration errors) and uncertainty in the ancillary parameters.

# Chapter 3

# Retrieval Performance

## 3.1 Retrieval Algorithms

This OSSE design was tested with two different retrieval algorithms. The first algorithm, referred to as single-polarization since it only uses measurements at one polarization, is based on Jackson [48]. It assumes $T_s = T_c$ and uses aggregated values of all the ancillary parameters to invert equation 2.1. The inversion is not directly solvable due to the non-lineary of the dielectric model. However, by taking advantage of the monotonicity of the soil moisture-reflectivity relationship, it is possible to use an iterative bisection search algorithm, in which a soil moisture guess is repeatedly refined until it predicts a reflectivity close to (within a certain tolerance) that predicted by the inversion of equation 2.1 [49].

The non-linearities in the $\tau$-$\omega$model complicate determination of the optimal aggregation method. The aggregation scheme of $\tau$ of Zhan et al. [50] were used. It neglects the influence of vegetation albedo, and attempts to invert the $\tau$-$\omega$ model such that (2.1) reduces to,

$$T_b = T_s(1 - r\exp(\frac{-2bW}{\cos\theta})). \tag{3.1}$$

Assuming brightness temperatures average linearly and neglecting heterogeneity in

$T_s$ and r, (3.1) leads to

$$\exp(\frac{-2b_{\text{eff}}W_{\text{eff}}}{\cos\theta}) = \frac{1}{n}\sum_{i=1}^{n}\exp(\frac{-2b_iW_i}{\cos\theta}) \tag{3.2}$$

Even though the baseline error was actually somewhat lower for vertical polarization than horizontal polarization, the single-polarization algorithm is tested at H-polarization. The OSSE is run for 1 year to reflect variability over the course of a growing season. For each day, every 3 km earth-fixed pixel, if visited, is assigned a soil moisture value by averaging the estimates of each of the covering footprints. The estimates are then spatially averaged to the 40 km earth-fixed grid to produce a daily soil moisture estimate. Only 40 km pixels for which at least 50% of the associated 3 km pixels have a retrieval are assigned a soil moisture value. For each 40 km pixel, the exact soil moisture knowledge is used to calculate an annually averaged bias. The reported root mean square error (RMSE) values are calculated with bias-removed estimates.

The dual-polarization algorithm, using both horizontally and vertically polarized measurements, is that described in Crow et al. [11]. It assume no prior knowledge of VWC is available, and attempts to retrieve both soil moisture and VWC by adjusting both parameters to minimize the least-squares-sum difference between predicted and measured brightness temperatures at both polarizations using a gradient descent algorithm.

## 3.2 Single-polarization Retrieval

Without bias-correction, only 0.5% of the pixels meet the required SMAP error criterion of 4% volumetric error. With perfect bias removal, 88% of all pixels have an error smaller than the threshold. If areas with average annual VWC greater than 5 kg/m$^2$ are not considered (these pixels will be flagged in the SMAP product), this number increases to more than 97%. To achieve this, however, bias characterization is crucial. Figure 3-1 shows maps of annually averaged bias and RMSE. The bias is

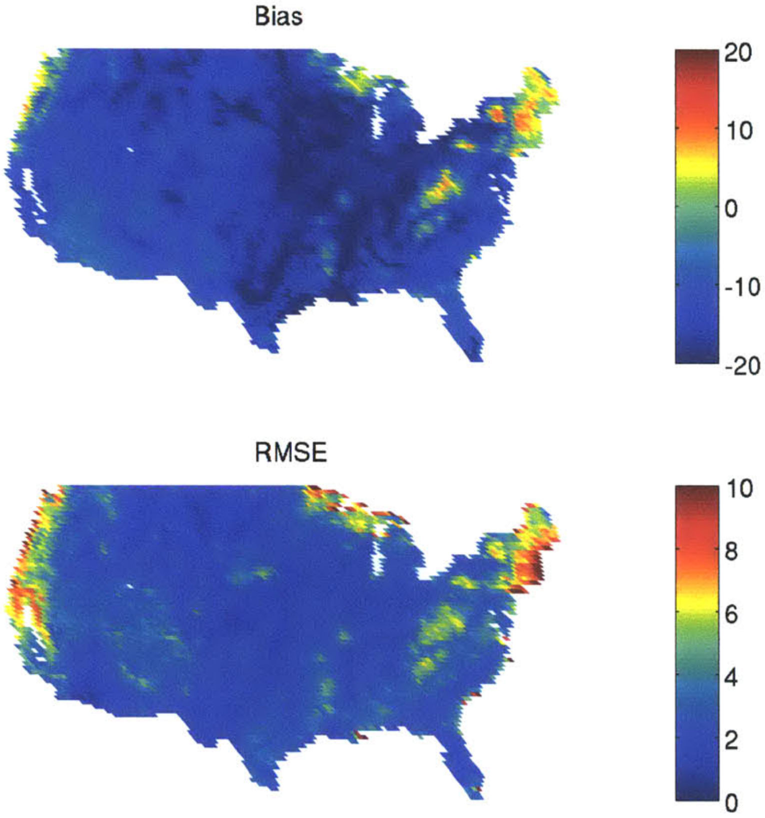


Figure 3-1: Maps of a) annual average bias (%vol), and b) RMSE (%vol) of the single-polarization retrieval based on baseline model settings

generally higher than the random error; the spatial average of the absolute value of the bias is 13%, while the CONUS average RMSE is only 5%.

### 3.2.1 Bias Identification and Removal

The success of bias removal is partially limited by the extent to which the errors themselves, at any given location, are stable over time and do not have a seasonal cycle or, more generally, is not dependent on any variable. Such bias is identifiable during the mission calibration and validation phase. It is unrealistic to imagine that biases can be estimated on a finer temporal scale than the annual case considered above. Nevertheless, seasonal changes in vegetation, temperature and soil moisture regimes cause significant seasonality in the bias. Figure 3-2 maps, for the baseline case, the range of seasonal biases (bias calculated over MAM, JJA, SON, and DJF). Over areas of lower vegetation (i.e. Montana plains), the temporal variation bias depends largely on the variability of soil moisture across the year, since the difference between both the different dielectric and the different roughness models increase with soil moisture.

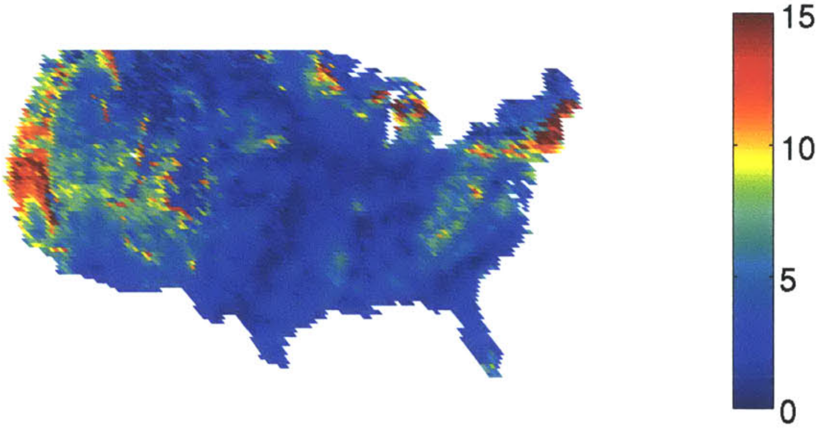


Figure 3-2: Range of soil moisture bias calculated per season (% vol).

Although errors are generally introduced at the brightness temperature level, this study is limited to studying a bias at the soil moisture level since the true SMAP measurements are not known. Future work, whether measurement or OSSE-based, will have a similar constraint. Using bias removal to improve radiometric soil moisture measurements requires that errors in brightness temperature can be expressed as an equivalent error in soil moisture. Figure 3-3 displays the distribution of the removability (difference between RMSE with and without bias removal, normalized by the original RMSE) of the bias. Over 88% of pixels, more than 80% of the bias is removable.

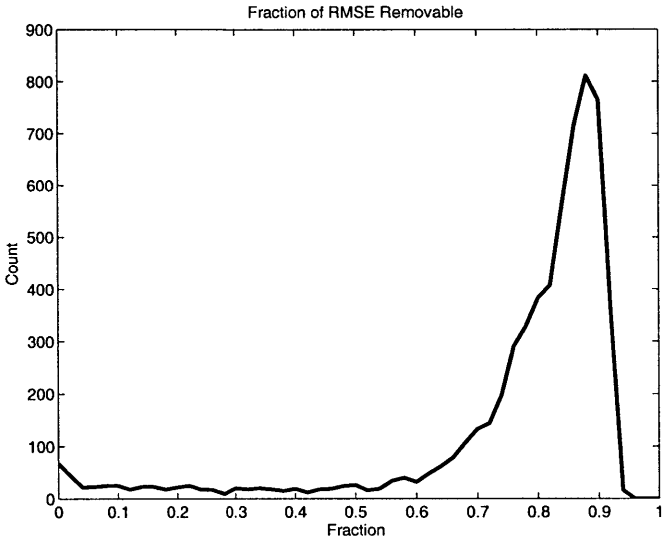


Figure 3-3: Distribution of fraction of single-polarization RMSE removed by bias-removal

The limited spatial extent of available field data requires some understanding of the behavior of the bias' spatial variation will be necessary. The aggregate annual bias varies predominantly with a region's annual mean soil moisture, annual mean VWC, and soil sand fraction. Figure 3-4 plots the discretized mean bias against each of these static variables. The trends are largely monotonic. Although the overall varation of bias varies most over the dynamic range of VWC, it does not vary significantly with VWC within the $<5$ kg/m$^2$ regime (where radiometric soil moisture retrieval is more feasible). Similarly, once a threshold level of soil sand fraction has been reached (about 0.6), the average bias stops increasing.

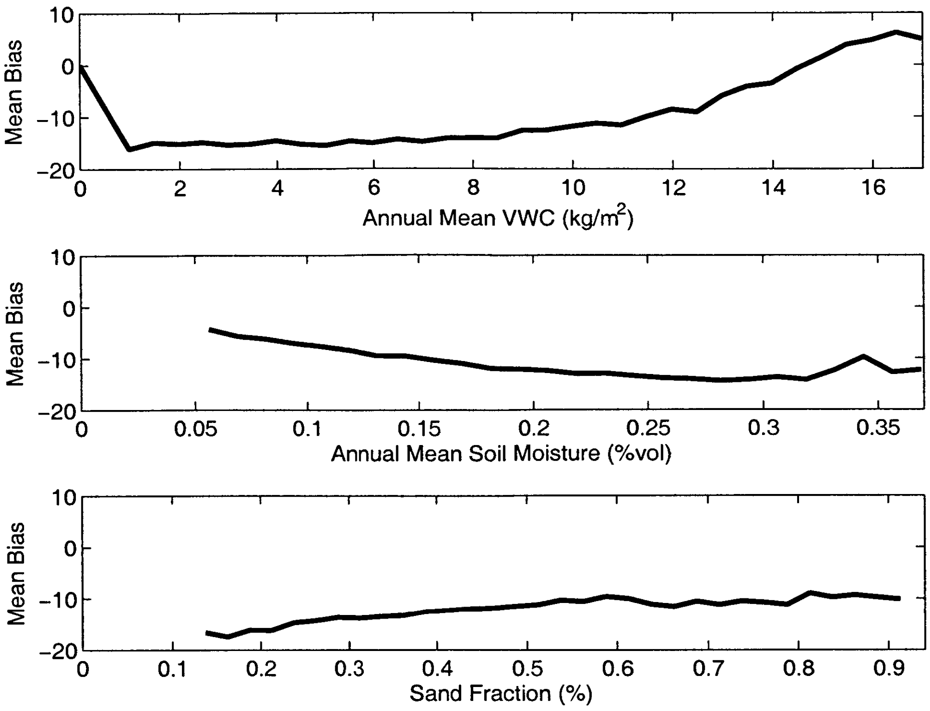


Figure 3-4: Single-polarization bias mean (% vol) discretized against pixel annual mean soil moisture (% vol), mean annual VWC (kg/m$^2$), and sand fraction(%)

### 3.2.2 Spatial Variation of Contributing Effects

To explore the changes in retrieval error structure if one or more of the sources of parametric uncertainty is removed, the different sources of uncertainty were removed one by one. In each case, the retrieval algorithm was modified to use the same model as the forward simulation for the term considered. The models used in each run are summarized in Table 3.1.

| | **Forward** | **Baseline** | **Different Dielectric Only** | **Different Roughness Only** | **Different Opacity Only** |
|---|---|---|---|---|---|
| $\epsilon$ | Wang & Schmugge | Dobson | Wang & Schmugge | Dobson | Dobson |
| $r_{sp}$ | IEM | h-Q | h-Q | IEM | h-Q |
| $\tau$ | Kirdyashev | $b_p$ | $b_p$ | $b_p$ | Kirdyashev |

Table 3.1: Modular configuration of different OSSE runs

Figure 3-5 depicts the difference between the baseline RMSE and the relevant retrieval case RMSE. Figure 3-6 depicts the difference between the bias in each case and the baseline bias. Generally, the greatest change in RMSE is found by removing uncertainty about vegetation opacity. Note that a positive difference in Figure 3-5 implies that RMSE is actually higher if the radiative transfer knowledge is increased. This is possible due to compensating errors. Fluctuations in brightness temperature lead to different magnitude changes in soil moisture depending on both the size of the fluctuation and the model structure. As a result, if errors from different sources cancel, changing the bias-removal for each run cannot fully remove those cancellation effects and removing an error source (such as a specific radiative transfer term) may therefore actually increase estimate RMSE.

When equalizing only the dielectric mixing model, the largest increase in bias and decrease in relative error occur in pixels whose soils have a high clay content. The largest decrease in bias occurs over areas with relatively high mean soil moisture. Since the Wang and Schmugge dielectric model tends to estimate lower dielectric constants, and therefore lower reflectivities, the difference in dielectric model will ordinarily cause an underestimation of soil moisture. Therefore removing the dielectric model difference increases soil moisture estimates. As a result, over Southwestern (dry) pixels with a relatively low negative bias, the variability-caused error (RMSE) actually increases in the 'different dielectric only' case.

The results of removing roughness treatment uncertainty leads to bias changes (and, to a lesser extent, changes in error) that are broadly similar to those of removing the dielectric mixing model uncertainty. This occurs because both result in a change in $r_{sp}$ in (2.1). However, the contribution is generally lower than the analogous one for the dielectric mixing model and is not as sensitive to soil properties (except insofar as they dictate the soil moisture behavior). This suggests further investigation of roughness models may be relatively unfruitful. Indeed, for most of the Western United States, removing roughness uncertainty would change the baseline error by less than 5%.

The effect of vegetation model on retrieval bias and error is particularly intricate. As in the dielectric case, reducing uncertainty in the vegetation model actually increases both bias and RMSE in some locations. Single-polarization soil moisture re-

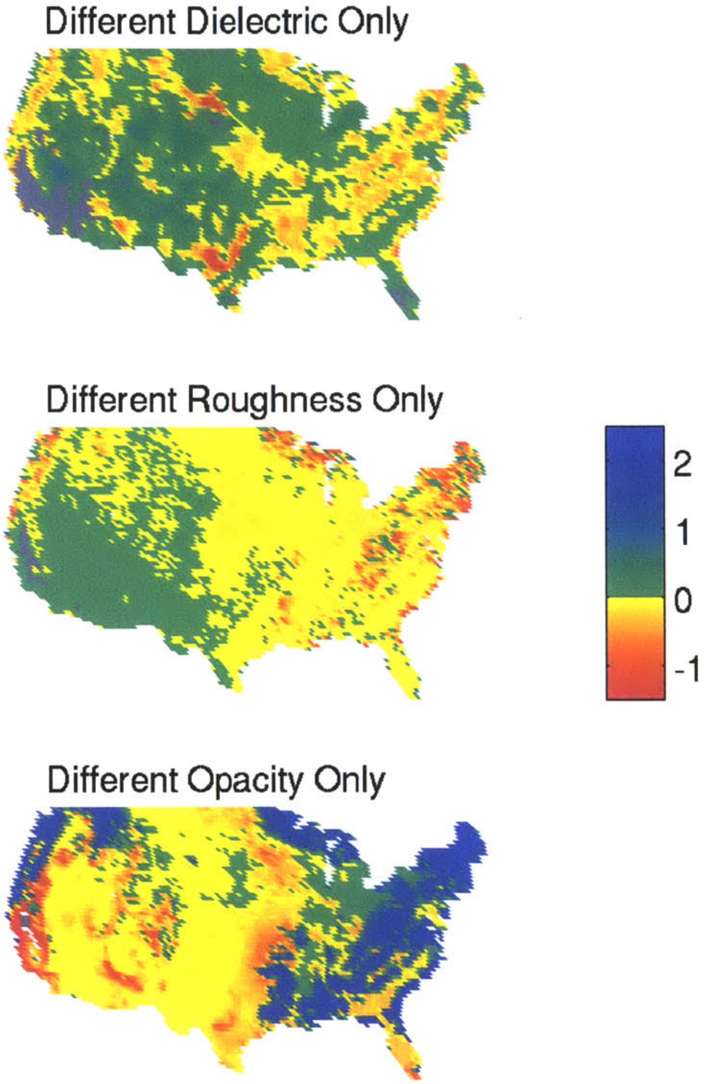


Figure 3-5: Differences in single-polarization RMSE between alternative configuration runs and baseline for the a)different dielectric only, b)different roughness only and c)different opacity only cases

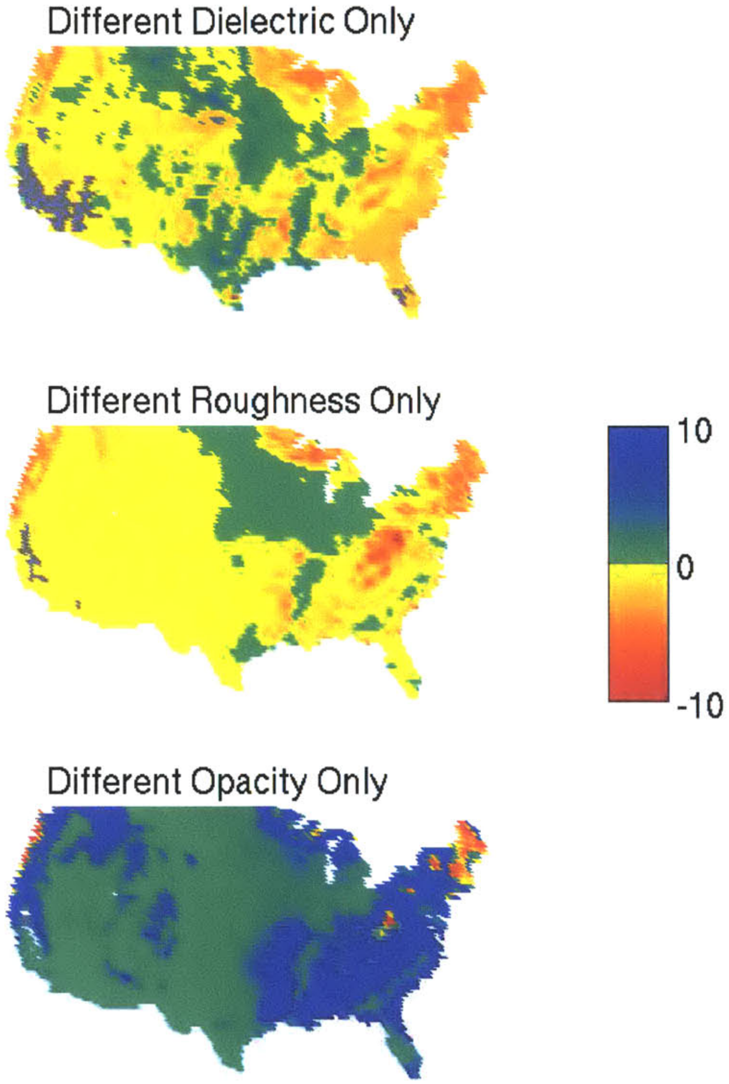


Figure 3-6: Differences in single-polarization retrieval bias between alternative configuration runs and baseline (%vol) for the a)different dielectric only b)different roughness only and c)different opacity only cases

trievals are positively biased over high vegetation cover. Since the Kirdyashev model generally provides higher estimates of $\tau$ than the b-parameterization, retrieval algorithms using the latter model will tend to underestimate the effect of vegetation, and therefore overestimate soil moisture. In doing so, they cancel the general underestimation caused by using an 'incorrect' roughness and dielectric model. This occurs even over certain areas of low vegetation. It should be noted, however, that the vegetation water content over these regions is far too high for soil moisture retrieval to be possible operationally, and probably well beyond the regime of validity of the $\tau$-$\omega$ model employed in the OSSE. More relevantly to SMAP measurements, the contribution to error can also be very high for areas such as the cropland near Iowa, with high temporal variability of opacity relative to the mean opacity.

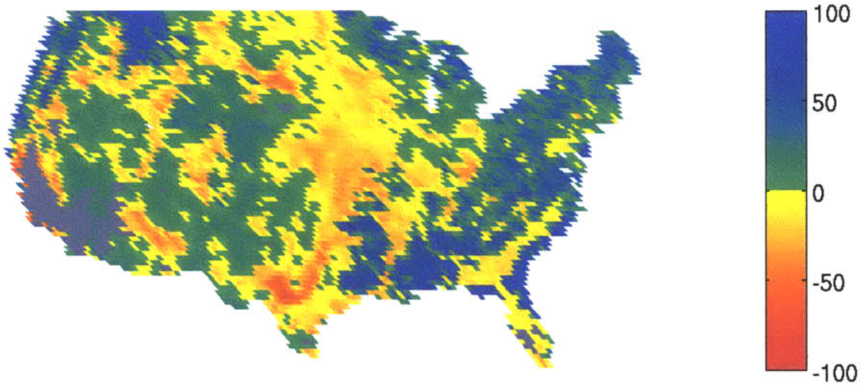


Figure 3-7: Sum of differences in single-polarization RMSE between each of the alternative configuration cases and the baseline, normalized by baseline error (%).

Normalizing the three components of Figure 3-5 by the pixel-based baseline RMSE and summing them allows estimation of the total relative contribution of radiative transfer uncertainty to the RMSE (as opposed to sub-pixel heterogeneity or measurement error). The sum is shown in Figure 3-7. Areas for which the total relative effect of radiative transfer uncertainty is the greatest tend to be dominated by a single type of uncertainty (i.e. dielectric effects over high clay content soils, opacity over high vegetation areas). Fifty-eight percent of pixels show a positive relative contribution of radiative transfer treatment. That is, errors due to incorrect descriptions of dielectric mixing and vegetation opacity (the positive roughness contributions are particularly small) compensate for other errors. To the extent that these errors are consistently biased, bias-removal can reduce the effect of error compensation on RMSE. Nevertheless, the non-linear structure of the $\tau$-$\omega$ models prevents full removal of compensating errors. Positive contributions of radiative transfer descriptions imply that a better understanding of the behavior of the soil dielectric constant or vegetation opacity may not actually improve retrieval. Instead, a coupled reduction in error sources is necessary to improve SMAP radiometer-only errors.

## 3.3 Dual-polarization Retrieval

Using the single-polarization retrieval OSSE without explicitly accounting for uncertainty in VWC is somewhat unrealistic, particularly given that current methods of estimating VWC rely on indirect measurements. The dual-polarization retrieval algorithm does not require an exact VWC estimate, as it simultaneously retrieves soil moisture and VWC. Multi-polarization algorithms may also be developed for retrieving ancillary parameters other than VWC, such as $T_s$ (e.g [51]).

An approximate range for VWC is still necessary for the optimization scheme. Since brightness temperature is more sensitive to VWC than soil moisture for several combinations of parameters, the algorithm often retrieves a very high VWC, causing the soil moisture estimate to be too high. This OSSE is run with bounds depending on the vegetation cover. The lower bound is 90% of the minimum VWC estimate of the year, while the upper bound is 110% of the maximum observed VWC. Estimating such bounds is mostly dependent on land cover, and may be done even without the development of exact, well-tested methods for predicting VWC.

| | RMSE | | Bias | |
|---|---|---|---|---|
| | mean | st. dev. | mean | st. dev |
| Single-pol (%vol) | 2.8 | 1.6 | -12.8 | 5.3 |
| Dual-pol (%vol) | 5.5 | 1.5 | 1.8 | 8.3 |
| Dual-pol VWC (kg/m$^2$) | 2.0 | 1.0 | 0.2 | 3.2 |

Table 3.2: Error statistics for each retrieval algorithm

Table 3.2 summarizes the error statistics of baseline runs with the two algorithms. The dual-polarization algorithm minimizes VWC-related error by implicitly finding the optimal aggregation of sub-footprint variations (which may not be the same as that predicted by the single-polarization aggregation scheme) and, since it operates on two measurements, should theoretically be better able to correct for measurement error. Furthermore, it partially compensates for using an 'incorrect' (compared to the synthetic truth) description of dielectric mixing, soil roughness, and vegetation. As a result, the remaining error has a smaller (in absolute value) bias, and the RMSE is actually higher than for the single-polarization algorithm. In addition, the bias is far more spatially variable. As betrayed by the large standard deviation, the vegetation bias statistics are very low because the bias happens to be about equally weighted positively and negatively.

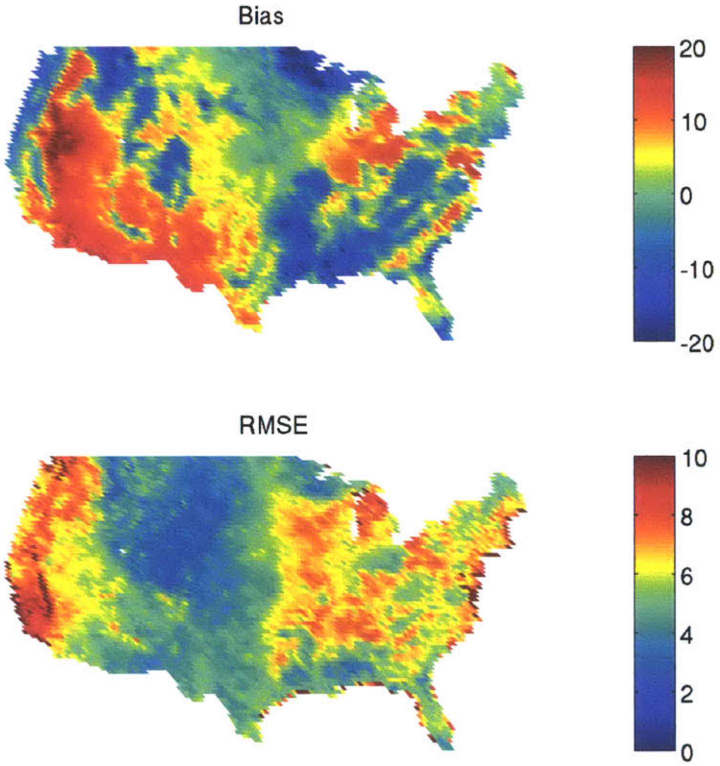


Figure 3-8: Maps of a)annual average bias (%vol) and b)RMSE(%vol) for the soil moisture retrieval of the dual-polarization algorithm. The baseline modular configuration is used.

The bias and RMSE of the soil moisture estimates are displayed Figure 3-8. Fig-

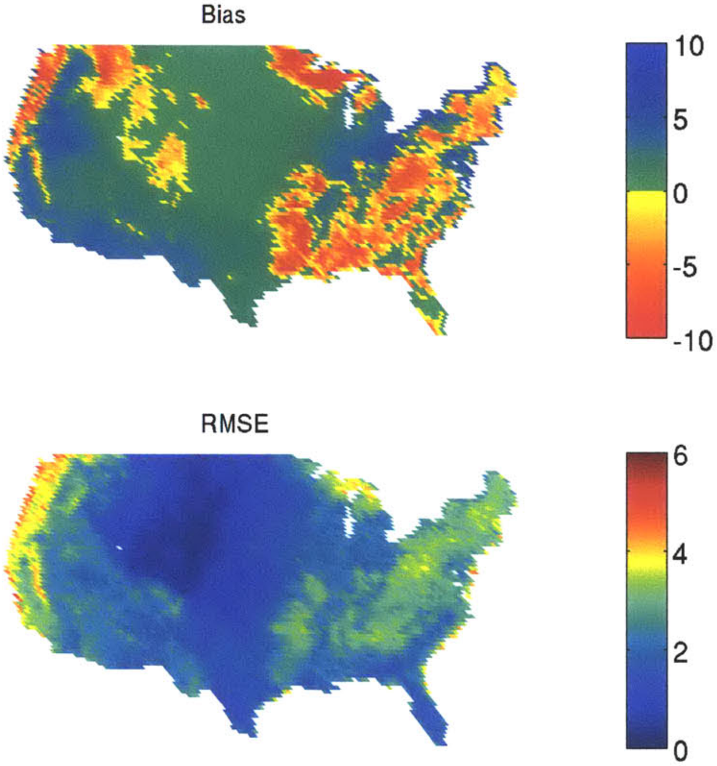


Figure 3-9: Maps of a)annual average bias (kg/m$^2$) and b)RMSE (kg/m$^2$) of VWC estimate using the dual-polarization algorithm. The baseline modular configuration is used.

ure 3-9 displays the bias and RMSE of the VWC estimates. Vegetation retrieval RMSE varies approximately with mean annual VWC. The high errors in VWC estimation, relative to its dynamic range, are notable. As Figure 3-10 demonstrates, the

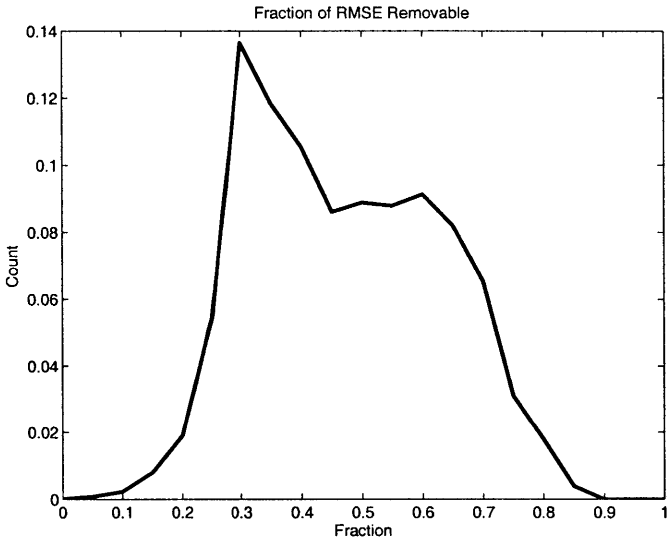


Figure 3-10: Distribution of fraction of dual-polarization RMSE removed by bias-removal

removability (fractional decrease in RMSE upon bias-removal, as in section 3.2.1) is generally far lower for the dual-polarization estimates than for the single-polarization ones. This occurs in part because the errors in VWC estimation include a correction for biases in the retrieval method. The distribution in Figure 3-10 is also much wider than the equivalent single-polarization distribution shown in Figure 3-3; this occurs because the VWC correction effect has considerable spatial variability.

The algorithm is also run using the decomposition configurations of Table 3.1. Unlike with the single-polarization algorithm, removing uncertainty about vegetation opacity decreases both estimate and bias-removed error. However, the contribution of dielectricmodel uncertainty is still negative.

# Chapter 4

# Conclusions and Future Research

Significant uncertainty exists about the representation of dielectric mixing, soil roughness effects and vegetation opacity in L-band radiative transfer. Understanding how the ambiguity of existing models affects soil moisture estimates is particularly important for the development of retrieval algorithms for the forthcoming SMAP satellite. Using an observing system simulation experiment covering a wide temporal and spatial range (one annual cycle over CONUS), the effect of this uncertainty on retrieval error and bias is simulated by using different representations in the forward and retrieval models. Measurement error and error due to land surface heterogeneity are also incorporated in the OSSE. The effects were tested using both a single-polarization algorithm that inverts the radiative transfer model (2.1) with an ancillary parameter aggregation scheme and a dual-polarization method (using data from both H and V polarizations) that retrieves both soil moisture and vegetation water content. Use of the single-polarization algorithm, even without accounting explicitly for uncertainty in ancillary parameters, leads to more than 99% of pixels having an RMSE above the SMAP 5% error criterion.

Using the single-polarization algorithm, retrieval errors are mitigated due to cancellation of errors. For example, whereas uncertainty in roughness models (in this OSSE) tends to overestimate soil moisture, error in the structure of the vegetation opacity treatment will underestimate soil moisture. Errors in the dielectric model and errors in the roughness cause similar retrieval errors, although the dielectric model causes a slightly greater error. The relative importance of each term for the total error is studied by removing uncertainty due to each term one-by-one and comparing the errors to those of a simulation with all terms retrieved with an 'incorrect' model. The largest change in retrieval error occurs when the vegetation opacity and dielectric mixing models are treated as functionally flawless. However, over several areas, mean-removed RMSE actually increases due to error cancellation. It is therefore not sufficient to make a single improvement to one of the sources of error contained in this OSSE (radiative transfer uncertainty, measurement noise, parameter aggregation to account for sub-pixel surface heterogeneity). Similar results occur for the dual-polarization algorithm, although the vegetation opacity effect is more often negative since the algorithm can adjust for errors in opacity by retrieving a higher estimate of vegetation water content.

Due to the high retrieval errors, it is necessary to investigate opportunities for error removal. Meeting the SMAP error requirement will require both reduction of this uncertainty and improved calibration of the spatial variation of ancillary parameters. Doing both to the extent necessary may not be feasible before the SMAP launch date. An alternative strategy is to characterize and remove the bias of the retrieval estimates. Such a scenario may occur in the context of data assimilation for land surface modeling, or if ancillary data are available to predict brightness temperatures (such as from SMOS). The degree to which brightness temperature errors can be represented by a constant bias in soil moisture was investigated. The brightness temperature bias due to dielectric and roughness uncertainty effects is mathematically similar to that induced by a constant bias in soil moisture. With (admittedly unrealistic) perfect bias knowledge, over 88% of the pixels in the study area, the annually averaged RMSE of the single-polarization retrievals can be reduced by more than 80% by removing the bias. It is not as useful a strategy for retrievals for the dual-polarization algorithm, although an error reduction of 30% or more can still be achieved in the vast majority of cases. Annual mean VWC, annual mean soil moisture, and sand fraction each correlate well with single-pol retrieval bias magnitude, suggesting it may be possible to understand and predict the spatial variability of the bias based on exact knowledge of the bias over limited areas. Thus, bias-removal is a potentially usefull strategy for SMAP radiometer retrievals. Indeed, the mean absolute bias is higher than the RMSE (bias deviations) for both algorithms.

Future work could include an expansion of this OSSE, such as a study of how error due to uncertainty about how ancillary parameter values interacts with radiative transfer uncertainty errors. Additionally, more sophisticated dual polarization algorithms should be developed to take full advantage of the retrieval estimation. In the current algorithm, multiple views of the same location over a few days are done independently. Since error generally reduces with an increased amount of observations, we expect combining such estimates to help constrain the VWC to a more optimal value, and may allow for also constrain other ancillary variables, such as land surface temperature.

Based on the results of this study, future field campaigns should focus on determining the optimal description of the vegetation opacity and dielectric mixing. The error cancellation effects a renewed investigation of aggregation strategies, including aggregation of soil texture properties. The extent of the polarization dependence of the vegetation model, particularly over areas other than cropland or over measurements measuring multiple crops, needs to be determined. Reducing dielectric mixing uncertainty is particularly important for dry soils. Rather than using field data to perfectly determine a given site's roughness and vegetation parameters, an investigation of the potential for bias characterization and removal should be done with real data.

# Bibliography

[1] W. T. Crow, R. Bindlish, and T. J. Jackson. The added value of spaceborne passive microwave soil moisture retrievals for forecasting rainfall-runoff partitioning. *Geophysical Research Letters*, 32(L18401), 2005.

[2] J. Komma, G. Bloschl, and C. Reszler. Soil moisture updating by ensemble Kalman filtering in real-time flood forecasting. *Journal of Hydrology*, 357:228–242, 2008.

[3] A. R. Desai, K. J. Davis, C. J. Senff, S. Ismail, E. V. Browell, D. R. Stauffer, and B. P. Reen. A case study on the effects of heterogeneous soil moisture on mesoscale boundary-layer structure in the Southern Great Plains, U.S.A.. part i: simple prognostic model. *Boundary-layer meteorology*, 119(2):195–238, 2006.

[4] S. B. Trier, F. Chen, and K. W. Manning. A study of convection initiation in a mesoscale model using high-resolution land surface initial conditions. *Monthly Weather Review*, 132:2954–2976, 2004.

[5] R. D. Koster and M. J. Suarez. Impact of land surface initialization on seasonal precipitation and temperature prediction. *Journal of Hydrometeorology*, 4(2): 408–423, April 2004.

[6] J. A. Patz, K. Strzepek, S. Lele, M. Hedden, S. Greene, and B. Noden. Predicting key malaria transmission factors, biting and entomological inoculation rates, using modeled soil moisture in kenya. *Tropical Medicine and International Health*, 3:818–827, 1998.

[7] D. Entekhabi, E. G. Njoku, P. Houser, M. Spencer, T. Doiron, Y. Kim, J. Smith, R. Girard, S. Belair, W. Crow, T. J. Jackson, Y. H. Kerr, J. S. Kimball, R. Koster, K. C. McDonald, P. E. O'Neill, T. Pultz, S. W. Running, J. Shi, E. Wood, and J. Van Zyl. The Hydrosphere State (Hydros) satellite mission: an Earth System Pathfinder for global mapping of soil moisture and land freeze/thaw. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10):2194–2195, October 2004.

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

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

[10] E. G. Njoku, W. J. Wilson, S. H. Yueh, S. J. Dinardo, F. K. Li, T. J. Jackson, V. Lakshmi, and J. Bolten. Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during sgp99. *IEEE Transactions on Geoscience and Remote Sensing*, 40(12):2659–2673, December 2002.

[11] W. T. Crow, S. T. K. Chan, D. Entekhabi, P. R. Houser, A. Y. Hsu, T. J. Jackson, E. G. Njoku, P. E. O'Neill, J. Shi, and X. Zhan. An observing system simulation experiment for Hydros radiometer-only soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1289–1303, June 2005.

[12] T. R. H. Holmes, M. Drusch, J.-P. Wigneron, and R. A. M. de Jeu. A global simulation of microwave emission: Error structures based on output from ECMWF's operational integrated forecast system. *IEEE Transactions on Geoscience and Remote Sensing*, 46(3):846–856, March 2008.

[13] R. Panciera, J. P. Walker, J. D. Kalma, E. J. Kim, K. Saleh, and J.-P. Wigneron. Evaluation of the SMOS L-MEB passive microwave soil moisture retrieval algorithm. *Remote Sensing of Environment*, 113:435–44, 2009.

[14] U. Narayan, V. Lakshmi, and E. Njoku. Retrieval of soil moisture from passive and active l/s band sensor (pals) observations during the soil moisture experiment in 2002 (smex02). *Remote Sensing of Environment*, 92:483–496, 2004.

[15] W. T. Crow, R. D. Koster, R. H. Reichle, and H. O . Sharif. Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products. *Geophysical Research Letters*, 32(L24405), 2005. doi: 10.1029/2005GL024889.

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

[17] Y. H. Kerr and J.-P. Wigneron. Vegetation models and obervations: a review. In B.J. Choudhury, Y.H. Kerr, E.G. Njoku, and P. Pampaloni, editors, *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*, pages 317–344. VSP, Utrecht, The Netherlands, 1995.

[18] A. Strahler, D. Muchoney, J. Borak, M. Friedl, S. Gopal, E. Lambin, and A. Moody. MODIS Land Cover Product Algorithm Theoretical Basic Document, 5 1999. Version 5.0.

[19] K. E. Mitchell, D. Lohmann, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, B. A. Cosgrove, J. Sheffield, Q. Duan, L. Luo, R. W. Higgins, R. T. Pinker, J. D. Tarpley, D. P. Lettenmaier, C. H. Marshall, J. K. Entin, M. Pan, W. Shi,

V. Koren, J. Meng, B. H. Ramsay, and A. A. Bailey. The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, 109, 2004. doi: 10.1029/2003JD003823.

[20] F. T. Ulaby, R. K. Moore, and A. K. Fung. *From Theory to Applications*, volume 3 of *Microwave Remote Sensing: Active and Passive.* Artech House, Norwood, MA, 1986.

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

[22] M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. A. El-Rayes. Microwave dielectric behavior of wet soil - part ii: Dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, 23(1):35–46, January 1985.

[23] M. T. Hallikainen, F. T. Ulaby, M. C. Dobson, M. A. El-Rayes, and L. Wu. Microwave dielectric behavior of wet soil, Part I: Empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, 23 (1):25–34, January 1985.

[24] N. R. Peplinski, F. T. Ulaby, and M.C. Dobson. Dielectric properties of soils in the 0.3-1.3 GHz range. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):803–807, May 1995.

[25] M. Vall-llossera, M. Cardona, S. Blanch, A. Camps, A. Monerris, I. Corbella, F. Torres, and N. Duffo. L-band dielectric properties of different soil types collected during the MOUSE 2004 field experiment. In *Proceedings of IGARSS 2005*, pages 1109–1112, 2005.

[26] D. A. Miller and R. A. White. A conterminuous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions*, 2(2):1–26, 1998.

[27] J. Shi, K .S. Chen, Q. Li, T. J. Jackson, P. E. O'Neill, and L. Tsang. A parametrized surface reflectivity model and estimation of bare-surface soil moisture with l-band radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, 40(12), December 2002.

[28] A. K. Fung. *Microwave Scattering and Emission Models and Their Applications.* Artech House, Norwood, MA, 1994.

[29] T. R. H. Holmes. *The Radiative Temperature of Earth at Microwave Frequencies.* PhD thesis, Vrije Universiteit Amsterdam, 2008.

[30] J. R. Wang and B. J. Choudhury. Remote sensing of soil moisture content over bare field at 1.4 GHz frequency. *Journal of Geophysical Research*, 86(C6):5277–5282, June 1981.

[31] J.-P. Wigneron, J. Shi, M. J. Escorihuela, and K.-S. Chen. Modelling the soil microwave emission. In C. Matzler, editor, *Thermal Microwave Radiation: Applications for Remote Sensing*, pages 276–287. The Institution of Engineering and Technology, 2006.

[32] T. Mo and T. J. Schmugge. A parametrization of the effect of surface roughness on microwave emission. *IEEE Transactions on Geoscience and Remote Sensing*, 25(4):481–486, July 1987.

[33] M. J. Escorihuela, Y. H. Kerr, P. de Rosnay, J.-P. Wigneron, J.-C. Calvet, and F. Lemaître. A simple model of the bare soil microwave emission at L-band. *IEEE Transactions on Geoscience and Remote Sensing*, 45(7):1978–1987, July 2007.

[34] J.-P. Wigneron, L. Laguerre, and Y. H. Kerr. A simple parameterization of the L-band microwave emission from rough agricultural soils. *IEEE Transactions on Geoscience and Remote Sensing*, 39(8):1697–1707, August 2001.

[35] T. J. Jackson, D. M. Le Vine, A. Y. Hsu, A. Oldak, P.J. Starks, C.T. Swift, J.D. Isham, and M. Haken. Soil moisture mapping at regional scales using microwave radiometry: the Southern Great Plains Hydrology Experiment. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5):2136–2151, September 1999.

[36] T. J. Jackson, D. Chen, M. Cosh, F. Li, M. Anderson, C. Walthall, P. Doriaswamy, and E. R. Hunt. Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment*, 92:475–482, 2004.

[37] K. McDonald and J. S. Kimball, 2008. Personal Communication.

[38] J. S. Kimball and K. McDonald. Global vegetation water content maps for HYDROS ESSP-3 Risk Mitigation. Unpublished report, 2002.

[39] Y. Knyazikhin, J. Glassy, J. L. Privette, Y. Tian, A. Lotsch, Y. Zhang, Y. Wang, J. T. Morisette, P.Votava, R.B. Myneni, R. R. Nemani, and S. W. Running. MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis Document. `http://eospso.gsfc.nasa.gov/atbd/modistables.html`, 1999.

[40] K.P. Kirdyashev, A.A. Chuckhlantsev, and A.M Shutko. Microwave radiation of the earth's surface in the presence of vegetation cover. *Radio Science and Engineering*, 24:256–264, 1979.

[41] U. Wegmuller, C. Matzler, and E. Njoku. Canopy opacity models. In B.J. Choudhury, Y.H. Kerr, E.G. Njoku, and P. Pampaloni, editors, *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*, pages 375–388. VSP, Utrecht, The Netherlands, 1995.

[42] C. T. Swift and L. A. Klein. An improved model for the dielectric constant of sea water at microwave frequencies. *IEEE Transactions on Antennas and Propagation*, 25(1):104–111, January 1977.

[43] A. A. Van de Griend and J.-P. Wigneron. The *b*-Factor as a Function of Frequency and Canopy Type at H-Polarization. *IEEE Transactions on Geoscience and Remote Sensing*, 42(2):786–794, April 2004.

[44] J.P. Wigneron, M. Pardé, P. Wadlteufel, A. Chanzy, Y. Kerr, S. Schmidl, and N. Skou. Characterizing the dependence of vegetation model parameters on crop structure, incidence angle, and polarization atL-band. *IEEE Transactions on Geoscience and Remote Sensing*, 42(2):416–425, February 2004.

[45] C. M. Frey, G. Rigo, and E. Parlow. Urban radiation balance of two coastal cities in a hot and dry environment. *International Journal of Remote Sensing*, 28(12):2695–2712, June 2007.

[46] J. Hafner and S. Q. Kidder. Urban heat island modeling in conjunction with satellite-derived surface/soil parameters. *Journal of Applied Meteorology*, 38: 448–465, April 1999.

[47] R. Z. Schneider, K. P. Papathanassiou, I. Hajnsek, and A. Moreira. Polarimetric and interferometric characterization of coherent scatterers in urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 44(4):971–984, April 2006.

[48] T. J. Jackson. Measuring surface soil moisture using passive microwave remote sensing. *Hydrological Processes*, 7:139–153, 1993.

[49] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C++: The Art of Scientific Computing.* Cambridge University Press, Cambridge, UK, 2002.

[50] X. Zhan, W. T. Crow, T. J. Jackson, and P. E. O'Neill. Improving spaceborne radiometer soil moisture retrievals with alternative aggregation rules for ancillary parameters in highly heterogeneous vegetated areas. *IEEE Transactions on Geoscience and Remote Sensing*, 5(2):261–265, April 2008.

[51] T. Pellarin, J.-P. Wigneron, J.-C. Calvert, and P. Waldteufel. Global soil moisture retrieval from a synthetic l-band brightness temperature data set. *Journal of Geophysical Research*, 108(D12), 2003. doi: 10.1029/2002JD003086.