Evaluation and Improvement of SMOS and SMAP Soil Moisture Products for Soils with High Organic Matter over a Forested Area in Northeast China

Mengjie Jin 1,2, Xingming Zheng 1,*, Tao Jiang 1, Xiaofeng Li 1, Xiaojie Li 1 and Kai Zhao 1

1 Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China; jinmengjie@126.com (M.J.); jiangtao@iga.ac.cn (T.J.); lixiaofeng@iga.ac.cn (X.L.); lixiaojie@iga.ac.cn (X.L.); zhaokai@iga.ac.cn (K.Z.)

2 University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: zhengxingming@iga.ac.cn; Tel.: +86-431-8554-2224

Academic Editors: Gabriel Senay and Prasad S. Thenkabail
Received: 13 February 2017; Accepted: 15 April 2017; Published: 19 April 2017

Abstract: Soil moisture (SM) retrieval from SMOS (the Soil Moisture and Ocean Salinity mission) and SMAP (the Soil Moisture Active/Passive mission) passive microwave data over forested areas with required accuracy is of great significance and poses some challenges. In this paper, we used Ground Wireless Sensor Network (GWSN) SM measurements from 9 September to 5 November 2015 to validate SMOS and SMAP Level 3 (L3) SM products over forested areas in northeastern China. Our results found that neither SMOS nor SMAP L3 SM products were ideal, with respective RMSE (root mean square error) values of 0.31 cm$^3$/cm$^3$ and 0.17 cm$^3$/cm$^3$. Nevertheless, some improvements in SM retrieval might be achievable through refinements of the soil dielectric model with respect to high percentage of soil organic matter (SOM) in the forested area. To that end, the potential of the semi-empirical soil dielectric model proposed by Jun Liu (Liu’s model) in improving SM retrieval results over forested areas was investigated. Introducing Liu’s model into the retrieval algorithms of both SMOS and SMAP missions produced promising results. For SMAP, the RMSE of L3 SM products improved from 0.16 cm$^3$/cm$^3$ to 0.07 cm$^3$/cm$^3$ for AM (local solar time around 06:00 am) data, and from 0.17 cm$^3$/cm$^3$ to 0.05 cm$^3$/cm$^3$ for PM (local solar time around 06:00 pm) data. For SMOS ascending orbit products, the accuracy was improved by 56%, while descending orbit products improved by 45%.

Keywords: validation; soil moisture; organic matter; soil dielectric model; forest; northeast China

1. Introduction

Soil moisture (SM) is a vital variable in the global land surface water cycle and energy cycle. Accurate estimates of SM can be used to forecast and prevent floods and extreme drought events. In addition, SM estimation accuracy directly influences the reliability of numerical weather predictions and hydrological models. Because forests make up one of the global primary land cover types, SM monitoring under forest cover has important implications for the study of hydrology, ecology, and climate change. Therefore, the acquisition of sufficiently accurate SM data for forest-covered areas has become a pertinent subject. Several studies have reported that L-band passive microwave radiometry is the most promising technique for monitoring SM in vegetated areas, due to its ability to penetrate the canopy cover and sensitivity to SM in all weather conditions [1–3]. The most commonly used satellite-based sources of SM are the Soil Moisture and Ocean Salinity (SMOS) mission, launched in 2009, and the Soil Moisture Active/Passive (SMAP) mission, launched in 2015. Both SMOS and SMAP are equipped with L-band sensors and aim to provide global near-surface SM (0–5 cm depth) products with an accuracy level of 0.04 cm$^3$/cm$^3$. 
The direct measurement of passive microwave sensors is the brightness temperature $T_B$, which is proportional to the emissivity ($e$) and surface physical temperature ($T$): $T_B = e \cdot T$ [4]. Since a forest is a vertically layered system made up of canopy, trunks, understory, litter and soil, the emissions from each layer and their interactions affect the measured $T_B$. Therefore, $SM$ retrieval under forest involves the process of removing the effects of vegetation, litter and roughness to obtain the reflectivity of the smooth bare soil surface. Soil permittivity which is related to the reflectivity is then computed from Fresnel formulations [5]. Finally, taking soil properties into account, the soil dielectric model is applied to translate the soil permittivity into SM content. Therefore, the soil dielectric constant model is an essential component of the $SM$ retrieval algorithms. The reliability of the soil dielectric constant model directly impacts the precision of the retrieved $SM$ values. It is well established that $SM$ retrieval from passive microwave signals over forested areas is difficult [1,6–8] and efforts to obtain sufficient $SM$ accuracy using SMOS/SMAP have not been very successful. For example, Rahmoune et al. compared SMOS level 2 (L2) $SM$ products and ground measurements from the SCAN/SNOTEL network in the U.S. over forests and found that most RMSE (root mean square error) values ranged between 0.07 and 0.12 cm$^3$/cm$^3$ [7], which is far from satisfactory. It is still rare to see validation work for SMAP $SM$ products in forested areas. In order to improve the accuracy of satellite-derived $SM$, numerous studies have been devoted to calibrating the effects of vegetation [9,10], litter [11–13], and roughness [14–17]. Here, we anticipate that some improvement in $SM$ retrieval results may be possible through refinements in the soil dielectric model with respect to the high percentage of soil organic matter (SOM) in forest soil.

The alternative soil dielectric models for the SMOS and SMAP retrieval algorithms are Wang, Dobson, and Mironov [18,19]. These three models all consider soil as a four-phase (free water, bound water, solids, and air) mixture and calculate the permittivity of the soil mixture by summing the weighted permittivity of each component according to its volumetric fraction. The Wang model [20] predicts and illustrates the different impacts of bound water from free water on soil permittivity. However, in this model, the permittivity of bound water is expressed by a linear combination of free water and ice, without consideration of the substantial characteristics of bound water. The Dobson model [21] retains the physical aspects of the dielectric properties of free water in the soil through the Debye equations, while not distinguishing the difference between free and bound water. The Mironov model [22] builds its expressions through a complex refractive index and calculates the dielectric properties of bound water as well as free water using the Debye equations. With clay percentage as the only soil input parameter, the Mironov model derives the needed spectroscopic parameters using a set of regression equations based on large soil databases. These three models have been applied in the SMOS and SMAP retrieval algorithms due to their simple parameterizations and applicability at L-band. The SMOS mission uses the land cover classification to choose the appropriate dielectric model (Wang, Dobson or Mironov), while the SMAP mission contains a switch in the processing steps that selects which dielectric model will be used in the $SM$ retrieval [18,19].

SOM is one of the major components of some soil types, and numerous studies have established its influence on soil dielectric properties. Organic-rich soil is usually characterized by high porosity levels because SOM increases the proportion of air and decreases the bulk density. Furthermore, Malicki et al. suggested that the lower bulk density or higher porosity may lead to a lower permittivity of organic soil than that of mineral soil with the same moisture content [23]. De Jong et al. demonstrated that SOM would increase the fraction of bound water [24]. Also, because the relative permittivity of bound water is smaller than that of free water [25], Bircher et al. attributed the lower permittivity of organic soil than that of mineral soil at a given water content to the higher fraction of bound water in organic soil [26]. Model simulation results also demonstrated the effects of SOM on soil dielectric properties. Mironov et al. used the generalized refractive mixing dielectric model (GRMDM) to compare the dielectric spectroscopic parameters of two agricultural soil samples that contained 6.6% and 0.6% of natural humus respectively [27]. The results of that comparison found that the L-band refractive index of the moist soil ($\sim$0.16 cm$^3$/cm$^3$) in the case of 6.6% humus was lower than that of the soil
(−0.15 cm³/cm³) with 0.6% humus. These above results suggest that in the natural soil surface, SOM is a factor in the soil emissivity as the soil permittivity is indeed reduced by SOM.

However, there are few soil dielectric models that are suitable for organic-rich soil. The temperature-dependent multi-relaxation spectroscopic dielectric model (TD MRSDM) is one example of this kind of model, which was developed for thawed and frozen Arctic organic-rich soil (50% organic matter) by Mironov [28]. TD MRSDM is applicable in a frequency range of 0.05 to 15 GHz and temperature from −30 to 25 °C, and uses bulk density, gravimetric moisture, wave frequency, and temperature as input parameters. The accuracy of the dielectric model was evaluated using RMSE, and the respective values for the soil permittivity and the loss factor were 0.348 and 0.188. The model error is acceptable for practical simulation of the permittivity of thawed and frozen Arctic organic-rich soil. Nevertheless, TD MRSDM, which is entirely based on laboratory dielectric measurements for organic-rich soil (50% organic matter), is not commonly applied in passive microwave SM retrieval algorithms because natural soil may contain a wide variety of organic content. In order to include SOM as an independent variable in the soil dielectric model, Liu Jun et al. measured the dielectric constants of 12 types of soil with organic matter content ranging from 0.03% to 17.84% using the coaxial probe method and network analyzer (0.5 to 40 GHz) at room temperature (approx. 23 °C) [29]. Based on their experimental results, they evaluated the influence of organic matter on soil dielectric constants and developed a semi-empirical dielectric model in which organic matter content was included as an input parameter to describe the dielectric behavior of soil with organic matter. Recently, Bircher et al. developed a simple empirical model for organic substrates [26], which gives out the L-band relative permittivity of organic soil as a function of water content. This empirical model considers a significant variety of organic substrate types and agrees well with the TD MRSDM model. However, this empirical model similar as TD MRSDM is built on the soils that contain at least 50% organic matter.

The objective of this paper was not to continue calibrating the impacts of vegetation, roughness, and other influencing factors, but rather to investigate the potential of Liu’s model [29] in improving SM retrieval results over forested areas. In Section 2, the general characteristics of the selected study area is described. Section 3 provides details on the datasets used, the new semi-empirical soil dielectric model (Liu’s model), and specific methods for this study. Section 4 compares the results of the SMOS/SMAP L3 SM product with in situ sensor network data and the improvements that resulted from implementing the new soil dielectric model in the SM retrieval algorithms. Moreover, the accuracy influencing factors of satellite SM products and other possibilities for improvement are discussed. In Section 5, conclusions are drawn.

2. Study Area

The in situ SM data for this study were obtained from Yichun (latitude 47–49°N, longitude 128–130°E, and altitude 600 m), a 40 × 40 km intensive study area within the Tangwang River catchment of Hei Longjiang Province in northeastern China (Figure 1). Yichun has a cold, temperate and humid climate with an average annual temperature of 1 °C and average annual rainfall of 750 to 820 mm. Yichun is rich in a variety of forested resources, such as birch, larch, oak, and spruce, and 82.2% of the land cover is forest. In this forest area, there is a surface litter layer with a depth of less than 3 cm, which is formed by mainly dead plant materials such as leaves, needles and twigs. Since humus beneath the litter layer of this area is very thin and usually too small to distinguish from soil by eyes, we treated it as a component of soil here. Topsoil in this paper refers to the 0–5 cm soil that calculated starting from the top of the layer without litter. Topsoil in this area is characterized by silt loam with a relatively high content of organic matter (approximately 25%). The detailed characteristics of the soil samples collected from different in-situ SM sites (approximately 6 km apart from each other) within this study area are summarized in Table 1.
Table 1. Topsoil information: soil type, sample numbers (SN) used in each site, textural fractions [%] of clay (<2 µm), silt (2–62.5 µm) and total sand (62.5–2000 µm) according to the FAO (Food and Agriculture Organization of the United Nations) soil texture classification [30].

<table>
<thead>
<tr>
<th>Site</th>
<th>Soil Type</th>
<th>SN</th>
<th>Clay</th>
<th>Silt</th>
<th>Sand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Silt Loam</td>
<td>25</td>
<td>14.21</td>
<td>66.48</td>
<td>19.31</td>
</tr>
<tr>
<td>2</td>
<td>Silt Loam</td>
<td>25</td>
<td>12.90</td>
<td>60.28</td>
<td>26.82</td>
</tr>
<tr>
<td>3</td>
<td>Silt Loam</td>
<td>25</td>
<td>12.37</td>
<td>67.66</td>
<td>19.97</td>
</tr>
<tr>
<td>4</td>
<td>Silt Loam</td>
<td>25</td>
<td>13.59</td>
<td>62.60</td>
<td>23.81</td>
</tr>
<tr>
<td>5</td>
<td>Silt Loam</td>
<td>20</td>
<td>12.32</td>
<td>69.01</td>
<td>18.67</td>
</tr>
<tr>
<td>6</td>
<td>Silt Loam</td>
<td>25</td>
<td>10.81</td>
<td>67.12</td>
<td>22.07</td>
</tr>
</tbody>
</table>

Figure 1. Land use types in the study area and locations of in situ SM stations.

3. Materials and Methods

3.1. Datasets

3.1.1. The Ground Wireless Sensor Network (GWSN) SM Measurements

The GWSN network collected hourly ground SM data from 9 September to 5 November 2015. These data were compared to SMOS and SMAP SM products to assess the performance of the SMOS and SMAP retrieval approaches. Direct comparison was not possible, however, due to the mismatch in spatial scale between the measurements provided by the satellite and in situ sensors [31]. To reduce the uncertainties introduced by scale differences between in situ points and satellite pixels, a high number of distributed sensors were laid out to accurately represent the satellite scale. As Figure 1 shows, the GWSN network was made up of a total of 17 Decagon EC-5 SM sensors that were approximately 10 km apart, distributed across two SMOS grids (Equal Area Scalable Earth [EASE] grid2, 25 km) and one SMAP grid (EASE grid2, 36 km). Of those 17 Decagon EC-5 sensors, 15 of them were located inside the SMAP grid (the grid center point longitude and latitude are 129.025°E and 48.157°N), seven were located inside the upper SMOS grid (the grid center point longitude and latitude are 129.035°E and 48.264°N), and five were inside the lower SMOS grid (the grid center point longitude and latitude are 129.035°E and 47.972°N). Moreover, to remove the influence of spatial variation in SM, the selected SMOS and SMAP grids were all relatively pure. The term “pure” for these grids refers to the following two aspects. On one hand, the soil characteristics for the areas of these grids had...
good spatial consistency as summarized in Table 1. The average SOM for the 17 EC-5 SM sensor-sites was 25% and was related to the soil bulk density in a negative linear fashion (Figure 2). On the other hand, the predominant naturally occurring land cover type is forest, which reaches 95% and above in the selected SMOS and SMAP grids. In this way, the selected SMOS and SMAP grids are considered homogeneous with respect to soil and land cover distribution, meaning that the mean SM values of observations from sensors at different locations are reasonable for representation at the satellite scale.

Another consideration was whether or not the in situ measurement network was reliable. The 17 Decagon EC-5 sensors were installed horizontally at 3 cm (the depth is calculated starting from the top of the layer without litter), with a measurement diameter approximately 5.5 cm. The EC-5 sensor determines volumetric water content by measuring the dielectric constant of the media using capacitance/frequency (70 MHz) domain technology. The EC-5 sensor automatically applies factory calibration functions [32] to convert voltage (mV) output to volumetric water content (\( \theta \), cm\(^3\)/cm\(^3\)). Factory calibration functions are applicable for most mineral soils with accuracy at least 0.03 cm\(^3\)/cm\(^3\). However, site-specific calibration could not be neglected because the accuracy of those sensors may be influenced by the organic-rich soil. Bircher et al. considered this issue and calibrated both Decagon 5TE sensor and Delta-T ThetaProbe ML2x sensor for natural organic soil [33]. Calibration functions in [33] are presented in the form of \( \theta = fct.(\varepsilon) \) for Decagon 5TE and \( \theta = fct.(\varepsilon) \), \( \theta = fct.(\text{voltage}) \) for Delta-T ThetaProbe. In this paper, the EC-5 sensors were also tested and calibrated in soils ranging from 10–55% of organic matter before being installed. The calibration results were obtained by linearly fit ground truth sampling of SM values with EC-5 sensor measurements (Figure 3). The respective overall accuracy of the original and calibrated EC-5 sensor were 0.04 cm\(^3\)/cm\(^3\) and 0.01 cm\(^3\)/cm\(^3\).

![Figure 2](image-url) Soil organic matter content (SOM) and bulk density (g/cm\(^3\)) of the soil samples from the 17 EC-5 SM sensor-sites.

![Figure 3](image-url) The calibration results of EC-5 SM sensor.
3.1.2. SMOS Data

The SMOS mission is in sun-synchronous orbit that ascends crossing the equator at 6 am and descends crossing the equator at 6 pm. SMOS measurements are made over a range of incidence angles (0 to 55°) across a swath that is approximately 1000 km wide and with a spatial resolution of 35 to 50 km. It takes three days to complete global measurements.

Since June 2011, SMOS has provided what are known as L3 data products in NetCDF format and projected on the EASE grid version 2 with a spatial resolution of 25 km × 25 km [34]. The algorithmic core of the L3 SM product is separate from the L2 SM processor, which is composed of a forward model (L-MEB) and an iterative inversion procedure [35]. The L-MEB model simulates multi-angular and dual-polarized T_B by quantifying the contributions of atmospheric interference, vegetation, and surface roughness, as well as the contribution of the soil itself. SM is then iteratively calculated by minimizing the cost function, which is built on simulated and measured T_B values. Note that the L-MEB model simulates the soil emissivity based on a dielectric mixing model (Dobson or Mironov) and the Fresnel equation. The difference between the daily L3 SM product and the L2 data is that the L3 SM product is processed based on the ESA L1B product. Moreover, the L3 processing algorithm uses a multi-orbit approach, in which calculations are done using three successive orbits within a seven-day moving window. When several values are available for a given day, the best estimation of SM is selected for each grid point. For the SMOS L3 SM product, ascending and descending orbits are processed separately. In this paper, both the ascending and descending SMOS L3 SM daily products for the study area were used. The used L3 data were the last set of reprocessed data using the version of the processor V3.00 at the time this work was conducted.

3.1.3. SMAP Data

Different from SMOS, SMAP only collects observations at an incidence angle of 40°. Based on a sun-synchronous orbit that ascends and descends the local overpass time at 6 pm and 6 am, respectively, SMAP produces global maps of passive soil moisture (0 to 5 cm) every two to three days under nested 9 and 36-km Earth grids.

The typical passive SM derivation approach proposed for SMAP is different from that proposed for SMOS. Five algorithms (including SCA-H, SCA-V, DCA, E-DCA and LPRM) were suggested at the initial release of the Algorithm Theoretical Basis Document (ATBD) [18]. Of these, the Single Channel Algorithm-H polarization (SCA-H), Single Channel Algorithm-V polarization (SCA-V), and Dual Channel Algorithm (DCA) are the most mature. Furthermore, SCA-V is the current baseline algorithm for SM calculation using the SMAP radiometer. In the SCA-V approach, five steps are used to extract SM from the measured T_B: (1) T_B in V polarization is proportionally converted to emissivity; (2) emissivity is corrected by removing the effects of vegetation; (3) emissivity is further corrected by accounting for the effects of soil surface roughness; (4) the Fresnel equation is used to relate emissivity to the soil permittivity; and (5) a dielectric mixing model (Mironov) is used to relate the dielectric properties to SM. This paper used the Version 4 SMAP L3 radiometer surface SM product, which is in HDF5 format and has the 36-km EASE-Grid 2.0 global projection.

3.1.4. Additional Data

For the purpose of validating the SMOS/SMAP L3 SM products and Liu’s model, we downloaded additional data for the study area within the study time period from 9 September to 5 November 2015: (1) precipitation data from Yichun station (ID 50774) provided by NOAA’s NCEI (NOAA is short for National Oceanic and Atmospheric Administration and NCEI is short for National Centers for Environmental Information); and (2) a 30-m Global Land Cover Dataset. This dataset is supported by the National High Technology Research and Development Program of China. In this paper, the Global Land Cover Dataset provides a reference for the selection of the study area.
3.2. Methodology

3.2.1. Methodology for Comparison

A robust validation of retrieved SM products can be useful for reevaluating and adjusting the algorithms used in the data processing steps. With the expectation of the fact that the two missions (SMOS and SMAP) can produce more accurate data products over forested areas, the retrieved SMOS and SMAP SM products were first compared directly with the in situ SM data for forested areas in northeastern China. Because the selected SMOS and SMAP grids in which the in situ SM sensor sites were distributed are homogeneous (as described in Section 3.1.1), a simple drop-in-bucket type of approach was used for the SMOS and SMAP grids. In a first step, all the observations from SMOS and SMAP that are RFI (radio frequencies interferences) and open water flagged are filtered out. The daily 06:00 am and 06:00 pm variations (including minimum, maximum, and mean values) of the in situ SM observations from the 15 ground sensors that were located in the SMAP grid were compared with L3 SMAP AM (local solar time around 06:00 am) and PM (local solar time around 06:00 pm) data respectively for the time period between 9 September and 5 November 2015. For SMOS validation, only the upper SMOS grid was selected because it had better sensor coverage (seven in total), as seen in Figure 1. Meanwhile, the five sensors located within the lower SMOS grid lacked spatial variation, as they were clustered in the top half of that grid. The daily 6 am and 6 pm SM values (including minimum, maximum, and mean values) from the seven sensors distributed within the upper SMOS grid were compared with the L3 SMOS AM and PM data respectively to evaluate the SMOS algorithm.

To assess the performance of SMOS/SMAP SM retrievals and Liu’s model, statistics such as bias (B), root mean square error (RMSE), unbiased root mean square error (ubRMSE), and the Pearson correlation coefficient (R) with 95% confidence intervals were used to quantify the differences and relationships between SMOS/SMAP SM products and SM ground measurements from the GWSN. The ubRMSE was calculated as a function of the bias and the RMSE [36]:

\[ B = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{SM_{SMOS/SMAP} - SM_{GWSN}}{n} \right) \]  

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (SM_{SMOS/SMAP} - SM_{GWSN})^2}{n-1}} \]  

\[ ubRMSE = \sqrt{RMSE^2 - B^2} \]

where \( n \) is the total number of SM retrievals considered in the comparison.

3.2.2. Methodology for Obtaining Revised Satellite SM

In order to investigate the potential of Liu’s model in improving SM retrieving accuracy, it was added to the processing steps of both SMOS and SMAP in place of their original alternative soil permittivity models (Dobson or Mironov). The methods and parameters that dealt with surface roughness, vegetation effects, and other influencing factors in the SMOS and SMAP inversion algorithms remained unchanged. Revised SM values were again compared to the ground measurements to see whether refinement of the soil dielectric model for organic matter rich soils improved the outcome. The general approach that was used to obtain revised satellite-derived SM values was the following:

1. Obtain the key parameters of the L-MEB model, such as canopy parameters optical depth (\( \tau_{NAD} \)) and albedo (\( \omega \)), land surface temperature parameter \( T_S \), and the four soil roughness parameters \( Q_R \), \( H_R \), and \( N_{RP} \) (\( p = H \) or \( V \)) of the model employed by Marie Parrens et al. [15]. For SMOS, \( \tau_{NAD} \) and \( T_S \) were obtained from the SMOS L3 daily SM products, and \( \omega \) and \( H_R \) were assumed to be constant as described in [6], with \( \omega = 0.07 \) and \( H_R = 1.2 \). For SMAP, the values of \( \tau_{NAD}, \omega, \)
Remote Sens. 2017, 9, 387 8 of 17

Tₜ, and Hᵣ were obtained from the SMAP L3 daily SM products. Here, we assumed that Nᵣᵣ = 0 (p = H or V) for both SMOS and SMAP. Qᵣ was also approximated to zero as had been done in the SMOS and SMAP retrieval algorithms, since Qᵣ was low at L-band [17,37,38].

(2) Run Liu’s model instead of the Mironov model within the L-MEB model to obtain new expressions for Tₑ Sim for SMOS and SMAP, based on their own L-MEB model parameters.

(3) Minimize the cost function CF (Equation (4)) by a generalized least squares iterative algorithm to achieve the revised SM values. For both SMOS and SMAP, the initial value for SM from the inversion process was the corresponding satellite SM product value. Tₑ Obs was the corresponding satellite observed brightness temperature value, which was obtained from the SMOS (incident angle approximately 42.5°) and SMAP (incident angle approximately 40°) L3 daily SM products.

$$CF = \sum_{p=H,V} (T_{B\_Sim}(p) - T_{B\_Obs}(p))^2$$

3.2.3. Liu’s Model

Liu’s model [29] is built on the basis of the Mironov model and has two variables (the bound water content $m_{vt}$ and the dielectric properties of solid soil particles $\varepsilon_d$) improved by organic matter content. The detailed expressions of this semi-empirical model are as follows [29]:

$$\sqrt{\varepsilon_m} = \begin{cases} \sqrt{\varepsilon_d} + (\sqrt{\varepsilon_d} - 1) \cdot m_v & m_v \leq m_{vt} \\ \sqrt{\varepsilon_d} + (\sqrt{\varepsilon_b} - 1) \cdot m_v + (\sqrt{\varepsilon_u} - 1) \cdot (m_v - m_{vt}) & m_v \geq m_{vt} \end{cases}$$

where $\varepsilon$ is the relative complex permittivity, $m_{vt}$ (cm³/cm³) is the bound water content and $m_v$ is the volumetric water content. The subscripts $m$, $d$, $b$ and $u$ denote “soil mixture”, “dry soil”, “bound water” and “free water”, respectively.

$m_{vt}$ is the point at which water switches from bound to free state. Humus, as the main component of SOM, is a hydrophilic substance and has a strong ability to absorb water. Therefore, soil with higher organic matters also contains a high proportion of bound water. The impact of SOM on the bound water content is expressed by:

$$m_{vt} = 0.02982 + 0.01068 \times SOM + 0.00089 \times CLAY$$

where SOM and CLAY are the mass percentages of soil organic matter and clay, respectively. SOM will also change the soil structure, reduce soil bulk density, and increase porosity, thereby increasing the proportion of air and reducing the soil permittivity when under a certain content of water. Thus, the raw density of the solid soil particles $\rho_s$ (g/cm³) is different than that of mineral soil and is determined by SOM:

$$\rho_s = \frac{m_{\text{mineral}} + m_{\text{organic}}}{v_{\text{mineral}} + v_{\text{organic}}} = \frac{\rho_{\text{mineral}} + \rho_{\text{organic}}}{(1 - \text{SOM}) \cdot \rho_{\text{organic}} + \text{SOM} \cdot \rho_{\text{mineral}}}$$

where $m_{\text{mineral}}$ and $m_{\text{organic}}$ are the quality of mineral soil particle and organic matter respectively. $v_{\text{mineral}}$ and $v_{\text{organic}}$ are the volume of mineral soil particle and organic matter respectively. $\rho_{\text{mineral}} = 2.65$ g/cm³ is the density of a mineral soil particle and $\rho_{\text{organic}} = 1.5$ g/cm³ is the density of organic matter. The subscript $s$ denotes “solid soil particles”.

Based on Equation (5), the relationship between the dielectric properties of solid soil particles and those of dry soil are described by the following equation:

$$\sqrt{\varepsilon_d} = 1 + (\sqrt{\varepsilon_s} - 1) \cdot \frac{\rho_b}{\rho_s}$$
According to the dielectric properties measurement results of 12 types of dry soil and Equation (8), the permittivity for solid soil particles and its relationship to organic matter content can finally be determined by the following empirical formula:

$$\varepsilon_s = 0.1469 \times SOM + 4.983$$ (9)

Equation (9) only calculates the real part of the permittivity of solid soil particles as a function of SOM and does not improve the imaginary part. The dielectric constants for bound water and free water are determined by the Debye formulas as in [22].

4. Results and Discussion

4.1. Comparison of SMOS/SMAP L3 Data with In-Situ Measurements

SMOS L3 data for the EASE grid2 (129.035°E and 48.264°N) were compared to the SM values that were measured continuously by the seven ground sensors throughout the focus areas (Figure 4a). SM was clearly underestimated by SMOS. From the period of 9 September to 5 November 2015, the daily mean SM value was 0.37 cm³/cm³ for the ground sensors, while SMOS AM results averaged 0.07 cm³/cm³ and PM averaged 0.08 cm³/cm³. These low SMOS SM values (below 0.1 cm³/cm³) seemed unreasonable for the Yichun area, which has a cold, temperate and humid climate. The total rainfall during the study period was 92 mm. Figure 4b shows SMAP L3 SM values for the grid (129.0249°E and 48.1572°N) as compared with the average values from the 15 in-situ sensors, and reveals that the SMAP product also underestimated SM. The SMAP SM values were closer to the ground measurements than SMOS, with a daily AM mean value of 0.28 cm³/cm³ and PM mean value of 0.26 cm³/cm³ for the study period. To quantitatively evaluate the SMOS and SMAP L3 products, statistics (Bias, RMSE, ubRMSE, R) on the comparisons between satellite retrievals and in-situ measurements were summarized in Table 2. This showed that there was a slight difference in the quality of the ascending (RMSE = 0.30 cm³/cm³) and descending (RMSE = 0.31 cm³/cm³) SMOS L3 SM daily products. The RMSE value of AM and PM SMAP SM product was 0.16 cm³/cm³ and 0.17 cm³/cm³ respectively, which were superior to that of SMOS. In general, neither SMOS nor SMAP meet the missions’ accuracy requirements for SM retrievals (0.04 cm³/cm³) over this forested area.

The in-situ SM data corresponded well with rainfall, showing an increasing trend during rainfall events and a decreasing trend after rainfall events (Figure 4). SMOS and SMAP data, on the other hand, did not seem to be able to capture the rainfall events as well as the in-situ data. For example, the top plot of Figure 4b shows an anomalous decreasing trend in the SMAP daily AM data during a rainy time period from DOY (Day Of Year) 281 to 284. This phenomenon is not easy to understand, because emissivity values usually decrease and canopy opacity increases following rainfall events [39], which should result in an overestimation bias in the SM calculations [40]. Figure 4a also shows that the SMOS L3 SM data were not temporally continuous. One of the reasons for this is that the SMOS revisit period is three days at the equator, which means it takes approximately three days to cover the global surface from the SMOS field of view. Another reason is that the SMOS SM data were masked out if the Data Quality Index (DQX) was larger than 0.06, or equal to 0.
we expect that the dielectric model could be improved by including organic matter content as an independent variable. Figure 5 shows the comparison between the permittivity simulated by the Mironov model and Liu’s model for three kinds of soil with the same clay content (13%) and different SOM content (10%, 20% and 30%) at 1.4 GHz. When SOM is 30%, the real part of simulated permittivity by Liu’s model is 15.81. For the Mironov model, which neglects the organic matter impact, the real part of simulated permittivity is much higher, with a value of 21.95. For the imaginary part, the difference was not so dramatic: ε″ = 2.02 for Liu’s model and ε″ = 2.96 for the Mironov model. The simulation results from Liu’s model for soil with SOM ranging from 10% to 30% also revealed that as SOM increases, the permittivity decreases.

4.2. Liu’s Model Performance

In general, the retrieved SM values from both missions were underestimated and were far from the accuracy requirements of 0.04 cm³/cm³. Given the fact that forested areas have high SOM content in the upper layer and that SOM substantially changes soil structure and water holding capacity, we expect that the dielectric model could be improved by including organic matter content as an independent variable. Figure 5 shows the comparison between the permittivity simulated by the Mironov model and Liu’s model for three kinds of soil with the same clay content (13%) and different SOM content (10%, 20% and 30%) at 1.4 GHz. When SM is 0.35 cm³/cm³ and SOM is 30%, the real part of simulated permittivity by Liu’s model is 15.81. For the Mironov model, which neglects the organic matter impact, the real part of simulated permittivity is much higher, with a value of 21.95. For the imaginary part, the difference was not so dramatic: ε″ = 2.02 for Liu’s model and ε″ = 2.96 for the Mironov model. The simulation results from Liu’s model for soil with SOM ranging from 10% to 30% also revealed that as SOM increases, the permittivity decreases.
Figures 6 and 7 and Table 2 display the overall performance of Liu’s model. Our analysis indicated clear improvements through modification of the soil dielectric model. The bias between the in-situ SM and the revised retrieval values from SMOS in ascending orbit was 0.08 cm$^3$/cm$^3$, with a RMSE of 0.13 cm$^3$/cm$^3$. Although these results still do not meet the accuracy requirement of 0.04 cm$^3$/cm$^3$, it is much better compared with the original SMOS SM products, which had a bias of 0.29 cm$^3$/cm$^3$ and a RMSE of 0.30 cm$^3$/cm$^3$. For the descending SMOS orbit, the level of underestimation was also attenuated and the accuracy was improved by approximately 45%. The revised retrieval values from SMAP were much closer to the ground measurements with a satisfactory RMSE of 0.05 cm$^3$/cm$^3$ for PM data. Moreover, the correlation coefficients between the revised retrievals and ground measurements were improved more or less. Consequently, we can conclude that the new semi-empirical soil dielectric model is a reasonable improvement with regard to organic soil and can be applied to forested areas to obtain higher quality estimates of SM.

**Table 2.** The overall statistics of SM retrievals compared with the average values of GWSN SM for the corresponding SMOS and SMAP grids.

<table>
<thead>
<tr>
<th>Statistical Indicators</th>
<th>Satellite SM Retrievals with Mironov Model</th>
<th>Revised SM Retrievals with Liu’s Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SMOS AM</td>
<td>SMOS PM</td>
</tr>
<tr>
<td>Bias</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.30</td>
<td>0.31</td>
</tr>
<tr>
<td>ubRMSE</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>R</td>
<td>0.25</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Figure 6. Comparisons of revised SMOS SM by Liu’s model with SMOS L3 SM products.

Figure 7. Comparisons of revised SMAP SM by Liu’s model with SMAP L3 SM products.
4.3. Discussion

Based on the comparison between the SMOS and SMAP L3 SM products and the in-situ measurement network over the forested study area, it was found that the satellite-derived SM values were underestimated, with RMSE values of 0.31 m$^3$·m$^{-3}$ for SMOS and 0.17 m$^3$·m$^{-3}$ for SMAP. Several aspects may account for the discrepancy between the retrieved SM values and ground measurements. First, the current soil dielectric model, the Mironov model, that is typically applied in the retrieval algorithms neglects the impacts of SOM. Figure 2 shows a strong negative linear correlation between SOM and soil bulk density (BD, g/cm$^3$): $BD = -0.0109 \cdot SOM + 0.695 \ (R^2 = 0.784)$. This means that SOM content greatly decreases the bulk density through changes in the soil structure (such as porosity). Moreover, SOM can increase the fraction of the bound water content of soil. As a result, the dielectric constants of organic soil will be lower than that of mineral soil with the same SM content. Because of this, a higher $T_B$ value observed by the passive sensors would be converted into drier soil conditions if SOM were neglected in the retrieval algorithm. In this paper, we accounted for this influence by introducing Liu’s model into the SM retrieval algorithms, which clearly improved the results (Section 4.2). The differences between SM values calculated using Liu’s model (SM1) and the Mironov model (SM2) were further examined and analyzed in Figure 8. It is clear that as SM1 and SOM increased, so did the discrepancy between SM1 and SM2. This means that the wetter the soil and the higher the SOM content, the more the SOM affects the accuracy of the satellite-derived SM values. However, it must be stressed that Liu’s model has its own limitations and errors. For example, Liu’s model calculates the real part of the permittivity of solid soil particles as a function of SOM based on experimental data; however, the influence of SOM on the imaginary part of the permittivity of solid soil particles is not considered. On the other hand, the Debye equation parameters for free and bound water that are adopted by Liu’s model are also probably affected by SOM and could lead to prediction errors [29]. Additionally, the range of the GWSN measurements in this paper is 0.30–0.47 cm$^3$/cm$^3$ with a standard deviation of 0.03 cm$^3$/cm$^3$. This relatively low dynamics is due to the local climate of the study area. Therefore, the fitness and reliability of Liu’s model in improving satellite-based SM retrievals need to be validated on a larger spatial and temporal scale over the entire wetness range.

![Figure 8](image-url)  
**Figure 8.** The differences between the SM values derived by Liu’s model (SM1) and the Mironov model (SM2). SM1 ranged from 0 to 0.6 cm$^3$/cm$^3$, while SOM ranged from 0 to 60%. SM1 and SOM were input into Liu’s model, and SM2 was then retrieved by the Mironov model based on the $T_B$ that was simulated by Liu’s model. The soil was assumed to be bare and smooth.

Second, a depth of approximately 3 cm litter is presented in this study area. As previous experimental studies [41,42] found, the sensitivity of the above-canopy $T_B$ to soil moisture may be substantially attenuated by the presence of litter when the soil is wet. Models for correcting the effects of that layer overlaying the soil have been developed, such as by Della Vecchia et al. for needle litter [13], by Schwank et al. for deciduous leaf litter [11], and lately by Bircher et al. for boreal forest
Third, the study area is dominated by forest cover and the attenuation effects of leaves, woody biomass inevitably reduces the sensitivity to SM. Developing reliable vegetation parameterization models with equivalent optical thickness ($\tau_F$) and albedo ($\omega_F$) for standing trees is a crucial step for calibrating these vegetation effects. For $\omega_F$, both SMOS and SMAP consider it to be constant (0.08 and 0.05, respectively) and independent of the angle, polarization, and time. For $\tau_F$, on the other hand, different parameterized expressions are adopted by SMOS and SMAP. In the SMOS algorithm, $\tau_F$ is modeled by $LAI_{\text{max}}$ (the maximum yearly value of the leaf area index [LAI]): $\tau_F = b_F \cdot LAI_{\text{max}}$. For SMAP, $\tau_F$ is dependent upon the vegetation water content (VWC): $\tau_F = b_F \cdot VWC$. For the sake of simplicity, the $b_F$ parameter in both algorithms is treated as constant. However, the values of $b_F$ and $\omega_F$ are not only determined by the mode of observation (polarization, frequency, and incidence angle), but also vegetation properties (structure). Because the structure of the overlying vegetation can be expressed in terms of vegetation type and age, it is likely that $b_F$ will vary with the growth stages of certain types of vegetation under different polarization modes. In the case of forest, there are few studies on the variability of the $b_F$ parameter due to polarization and vegetation growth stages. For crop cover, however, the temporal dynamics of the $b_F$ parameter in the SMAP algorithm have been researched by Wigneron et al. (1996) [44], who found that $b_F = 0.125$ for green vegetation and $b_F = 0.04$ for senescent crops. Similar results were also found by Burke et al. (1999) [45]. The temporal dynamics between forested $b_F$ and $\omega_F$ under different polarizations will be the subject of our future work.

Fourth, and as reported in previous studies [46], the uneven terrain in the study area may cause changes in the overall surface emissions by impacting the scattering mechanism. This topographic impact changes with the viewing angle. SMOS, which uses multi-angled measurements to derive SM, seems to be more influenced by the topography than SMAP. It is possible that topographic impact accounts for the poorer quality of the SMOS SM product than that of SMAP. Furthermore, although the soil properties and land cover type in this study area are relatively uniform, spatial variations in topography and rainfall events may impact the representativeness of the average values of the distributed in-situ SM measurements for the small study area, to some extent.

5. Conclusions

This paper attempted to validate the SMOS and SMAP L3 SM products on a regional scale by comparing them with distributed in-situ SM measurements. In general, the satellite-based SM values were underestimated, with RMSE values of 0.31 cm$^3$/cm$^3$ for SMOS and 0.17 cm$^3$/cm$^3$ for SMAP. Due to the high content of SOM in the top soil layer of the forested study area, a new soil dielectric model (Liu’s model) that considers the impacts of SOM was introduced into the retrieval algorithms of both SMOS and SMAP. These refinements yielded clear improvements in SM retrievals. The RMSE of the revised SM values for SMOS in ascending orbit was reduced from 0.30 cm$^3$/cm$^3$ to 0.13 cm$^3$/cm$^3$. As for SMAP, the RMSE was reduced from 0.16 cm$^3$/cm$^3$ to 0.07 for AM data, and from 0.17 to 0.05 for PM data cm$^3$/cm$^3$. These results indicate that for the SMOS and SMAP mission, it is possible to acquire more accurate SM estimates by considering the effects of SOM in mineral soil. However, there is also room for improvement by considering the effects of litter and topography, as well as employing vegetation calibration methods. In sum, we can conclude that Liu’s model is a reasonable improvement with respect to organic soil and can be applied in forested areas to acquire higher-quality estimates of SM.
Acknowledgments: This work was financially supported by the National Natural Science Foundation of China under grant No. 41301369, 41471289, 41371345 and the National Fundamental Research Funds under grant No. 2014FY210800. The authors would like to acknowledge the Centre Aval de Traitement des Données SMOS (CATDS), and the National Snow and Ice Data Center (NSIDC) for providing the satellite-based soil moisture products. The authors would also like to thank Changchun Jingyuetan Remote Sensing Experiment Station, Chinese Academy of Sciences, for implementing Ground Wireless Sensor Network (GWSN) and providing ground-based soil moisture measurements.

Author Contributions: Mengjie Jin designed the experiment, analyzed the results and wrote this paper. Xingming Zheng proposed the main scientific ideas, considerations and comments. Tao Jiang processed the in situ measurements. Xiaofeng Li, Xiaojie Li and Kai Zhao offered constructive suggestions and comments.

Conflicts of Interest: The authors declare no conflict of interest.

References


34. Brodzik, M.J.; Billingsley, B.; Haran, T.; Raup, B.; Savoie, M.H. EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets. ISPRS Int. J. Geo-Inf. 1990, 1, 32–45. [CrossRef]

43. Putuhena, W.M.; Cordery, I. Estimation of interception capacity of the forest floor. *J. Hydrol.* 1996, **180**, 283–299. [CrossRef]
45. Burke, E.J.; Wigneron, J.P.; Gurney, R.J. The comparison of two models that determine the effects of a vegetation canopy on passive microwave emission. *Hydrol. Earth Syst. Sci.* 1999, **3**, 439–444. [CrossRef]

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).