A Five-Year Evaluation of SMOS Level 2 Soil Moisture in the Corn Belt of the United States

Victoria A. Walker, Brian K. Hornbuckle, and Michael H. Cosh

Abstract—The Soil Moisture Ocean Salinity (SMOS) satellite mission is currently retrieving global observations of soil moisture but must be validated before it can be used. We compared SMOS Level 2 Soil Moisture (L2SM) retrievals to the five-year in situ soil moisture dataset available for the predominately agricultural South Fork Iowa River (SFIR) watershed in the U.S. Corn Belt. SMOS L2SM is 0.039 m$^3$ m$^{-3}$ drier than the SFIR network soil moisture and has an unbiased RMSE (ubRMSE) of 0.062 m$^3$ m$^{-3}$ for the period of April 2013–November 2017 (excluding DJF). The bias is 11% of the range of in situ soil moisture. The largest dry biases occur when vegetation and soil surface roughness are changing rapidly. The following potential sources of the dry bias are discussed: bias in auxiliary modeled temperatures; errors in soil texture maps; and non-representative parameterizations of single scattering albedo and soil surface roughness. Auxiliary skin temperature was colder than expected and may explain why SMOS L2SM has a slightly drier bias for evening overpasses but does not explain the overall dry bias. Increasing the parameterized soil surface roughness produces wetter SMOS L2SM retrievals but also decreases the sensitivity to soil moisture. Noise in auxiliary surface temperature and differences in sensing volume between SMOS and the SFIR in situ sensors contribute to the ubRMSE. ubRMSE can potentially be improved in the SFIR by using a non-zero single scattering albedo representative of a corn and soybean canopy at the cost of increasing the dry bias.

Index Terms—Soil Moisture Ocean Salinity (SMOS), soil moisture, U. S. Corn Belt

I. INTRODUCTION

The European Space Agency’s Soil Moisture Ocean Salinity (SMOS) satellite mission is currently using microwave radiometry to map global soil moisture at a spatial resolution of approximately 40 km and a revisit time of no more than 3 days [1]. To be precise, SMOS is sensitive to near-surface soil moisture, the water content of the layer of soil immediately at the surface a few centimeters in depth [2], [3], [4]. This is because SMOS observes the brightness temperature of Earth’s surface within L-band, at a frequency of 1.4 GHz (or wavelength of 21 cm). At this frequency, vegetation is semi-transparent and emission originating from the soil can be detected. Along with its water content, soil surface emissivity depends on the size and chemical composition of its mineral constituents (soil texture), its organic matter content, and the roughness (mm-scale height variations) of the soil surface.

Soil moisture is a small but significant reservoir of the global water cycle. It is important in many contexts, but plays a key role in plant-based agriculture. The quantity of water of most concern is stored in the layer of soil that plant roots can reach. This root-zone soil moisture is typically a meter or more in depth. While SMOS can not directly observe the root zone, frequent observations of near-surface soil moisture (at least every two to three days), in conjunction with models of soil water infiltration, evaporation, and plant transpiration, can be used to infer the total amount of soil water accessible to crops [5], [6]. Water stress resulting from lack of adequate root-zone soil moisture limits photosynthesis and reduces the transport of nutrients from the soil to plants [7]. Even short periods of water stress during important development stages in annual crops, such as tasseling in corn, reduces yield in spite of subsequent irrigation or precipitation [8]. In addition to the immediate effects on agriculture, antecedent soil moisture has been shown to be coupled with precipitation during the growing season in agricultural regions due to its influence on land-atmosphere water and energy exchange [9], [10].

While agricultural productivity is most affected by root-zone soil moisture, near-surface soil moisture, the water directly observed by SMOS, also plays an important role in agricultural ecosystems. Infiltration of precipitation depends on the water content of soil at the very surface. Water that does not infiltrate is either lost as runoff or ponds on the surface. Besides not being captured for use by plants, runoff can erode the soil and remove plant nutrients. High rates of runoff lead to flooding. Ponding of water results in saturated soils which can literally suffocate plant roots. In annual row cropping systems, near-surface soil moisture determines when machinery can be used in the field and thus if planting, harvest, and other types of management can occur.

SMOS Level 2 Soil Moisture (L2SM) retrievals need to be validated before being utilized for agroecosystem and meteorological modeling. We have selected the South Fork Iowa River (SFIR) watershed, a Soil Moisture Active Passive (SMAP) core validation site, for this purpose [11]. This site has a five-year data record (April 2013–present) of in situ soil moisture at a 36 km spatial scale, located within the U.S. Corn Belt, a globally-significant agricultural region. A comparison of the first year of available SFIR data (2013) to SMOS L2SM found that SMOS L2SM retrievals were at least 0.05 m$^3$ m$^{-3}$ drier than the SFIR network average and the root-mean-square-error (RMSE) exceeded the mission accuracy goal of 0.04 m$^3$ m$^{-3}$ [12]. We hypothesize that this dry bias and the unbiased RMSE (ubRMSE) no longer exists as the SMOS L2SM retrieval algorithms have evolved and the in situ data record has lengthened.

V. Walker and B. Hornbuckle are with the Department of Agronomy, Iowa State University of Science and Technology, Ames, IA, 50011 USA.

M. Cosh is with the Hydrology and Remote Sensing Laboratory, United States Department of Agriculture Agricultural Research Service (USDA–ARS), Beltsville, MD 20705 USA.
II. Data

A. SMOS Products

SMOS L2SM retrievals were extracted from the MIR_SMUDP2 (v06.50) product [13]. SMOS L2SM retrievals in the SFIR were filtered following Section III before comparing to the in situ network soil moisture. Surface temperature, soil texture, land surface parameters, and land surface class fractions were obtained from the associated AUX_ECMWF, AUX_DFFSOI, AUX_LANDCL, and AUX_DFFFRA auxiliary datasets [14].

B. South Fork Iowa River Network

The SFIR validation network was established by the United States Department of Agriculture Agricultural Research Service (USDA–ARS) in 2013 and is supported by scientists at the University of Iowa and Iowa State University. There are twenty permanent in situ soil stations distributed throughout the watershed such that Thiessen-polygon weighted average soil moisture (WASM) can be produced for the 36 km, 9 km, and 3 km scales required by the SMAP mission. Each station consists of four soil moisture and temperature probes (Stevens Water HydraProbe and Campbell Scientific CS655) inserted horizontally at 5, 10, 20, and 50 cm depths and two precipitation gauges (MetOne 380 tipping bucket and Campbell Scientific TE525).

Figure 1 shows the location of the SFIR in situ sites (denoted by ●) in central Iowa alongside the SMOS Discrete Global Grid points (DGGs) centered over the SFIR. DGGs are illustrated in Figure 1 as 50 km circles with centroids denoted by ▲. While DGGs are typically assumed to have a spatial resolution of ≈ 40 km, realistically it varies according to the proximity of the DGG to the center of the overpass swath and can approach 50 km. Approximately 85% of the land cover in the SFIR watershed is croplands; 60% corn and 40% soybean. [15]. The SFIR is roughly homogeneous at the spatial scale relative to SMOS (≈ 40 km); the AUX_DFFFRA dataset parameterizes the land surface as 94.1% “nominal” (i.e., croplands, grasslands, and bare soil), 5.4% forest, 0.4% open water, and 0.3% urban for the area outlining the three DGGs in Figure 1. Tile drainage is common in the area as the land is flat (grade of less than 2% [16]) and poorly drained.

The SFIR network soil moisture used in this analysis, hereafter referred to as “SFIR WASM,” is the 36 km Thiessen-polygon weighted average of the in situ probes located at a depth of 5 cm. The data record extends over five growing seasons from April 2013–November 2017. The winter months (DJF) are excluded from the comparison to remove retrievals and in situ observations occurring in potentially frozen soil.

The community practice for validating satellite soil moisture products is to compare to in situ networks which have a large distribution over the domain of interest [17]. The SFIR has an ideal landscape for a validation site as it is relatively homogeneous for both the ≈ 40 km retrieval grid and the larger 123 km working area. However, conflicts with seasonal farm management in the SFIR result in the permanent in situ stations being located on the edges of fields. An in–field soil moisture observation campaign was conducted during the summer of 2014 to investigate the differences between the edge–of–field sites and the soil moisture within corn and soybean fields in the SFIR. We compared the SFIR WASM to the in–field WASM calculated from 44 sampling sites for approximately 40 dates (Figure 2). The soil moisture for each site was calculated by averaging in–row (in line with the plants) observations with those sampled 1/4, 1/2, and 3/4 across the row. The SFIR WASM averages 0.005 m$^3$m$^{-3}$ wetter than the in–field WASM; the RMSE was 0.023 m$^3$m$^{-3}$.

C. Metrics

The first metric used to quantitatively assess SMOS L2SM retrievals in the SFIR is the bias as defined by (1). SMOS L2SM consistently underestimating the in situ obser-
viation (too dry) would result in a negative bias; overestimation (too wet) results in a positive bias.

$$\text{bias} = (\text{SMOS L2SM} - \text{SFIR WASM})$$  \hspace{1cm} (1)

The second metric used is the unbiased root–mean–squared–error (ubRMSE) as defined by (2) and (3). RMSE is inherently dependent on bias as defined by (2) and (3): a large magnitude bias will result in a high RMSE even if the random error between SMOS L2SM and in situ observations is relatively low. This effect is removed by using the ubRMSE as the measure of “noisiness” in the relationship instead. The mission accuracy goal for SMOS L2SM retrievals is that random error should not exceed 0.04 m$^3$ m$^{-3}$ [1].

$$\text{ubRMSE} = \sqrt{\text{RMSE}^2 - \left(\frac{\text{SMOS L2SM} - \text{SFIR WASM}}{\text{MAVA0}}\right)^2}$$  \hspace{1cm} (2)

$$\text{RMSE} = \sqrt{(\text{SMOS L2SM} - \text{SFIR WASM})^2}$$  \hspace{1cm} (3)

### III. SMOS L2SM Retrieval Filtering

#### A. Radio Frequency Interference

Anthropogenic sources of radiation within a frequency band protected for passive remote sensing is called radio frequency interference (RFI). Significant RFI has been observed in several areas of the world, particularly in Eastern Europe, the Middle East, and East Asia. When RFI is present in the main lobe of the antenna pattern SMOS measures a brightness temperature, $T_B$, that is higher than what is naturally produced by the terrain in its field of view. Increased $T_B$ is interpreted as a higher soil emissivity and consequently SMOS L2SM will retrieve a drier soil moisture value in the presence of RFI. While it may be possible for RFI in secondary lobes to have an opposite (wetting) affect, contributions from these lobes are orders of magnitude smaller than those of the main lobe.

The SMOS L2SM retrieval algorithm flags individual measurements of $T_B$ as contaminated with RFI if their value is too high to be geophysically plausible. Each SMOS L2SM retrieval is obtained from a number of $T_B$ observed at several incidence angles, $\theta$. The probability of instantaneous RFI (IRRFI) is defined as:

$$\text{IRRFI} = \frac{\text{NRFIX} + \text{NRFIY}}{\text{MAVA0}}$$  \hspace{1cm} (4)

The IRRFI is used to determine the likelihood of RFI contamination negatively impacting the soil moisture retrieval for a given DGG and overpass as a function of the $T_B$ flagged for RFI in the X–direction (NRFIX), $T_B$ flagged in the Y–direction (NRFIY), and the total number of $T_B$ measured (MAVA0). The X– and Y– directions are in the antenna reference frame [14].

The upper limit for allowable IRRFI is dependent on how often the region of interest experiences high levels of RFI. In areas of the world where sustained RFI is prevalent, a more relaxed filter ($\text{IRRFI} \leq 0.30$) may be desirable in order for the amount of valid retrievals to not be overly reduced, assuming that the amount of RFI is low enough that useful soil moisture retrievals can still be obtained (P. Richaume, personal communication). We implemented a stricter filter of $\text{IRRFI} \leq 0.05$ due to the absence of strong RFI sources in the Corn Belt; this discards 3.2% of SMOS L2SM retrievals in the SFIR. Retrievals that were removed by this filter are less due to strong RFI signatures and more to the retrieval being performed with a very small sampling of $T_B$. When the DGG is on the edge of the SMOS swath and few $T_B$ are measured (small MAVA0), any $T_B$ flagged for RFI may result in the retrieval being removed by the IRRFI filter.

#### B. Retrieval Validity

SMOS L2SM retrievals are considered “successful” when the processor is able to retrieve a soil moisture that minimizes the difference between observed and simulated $T_B$. Soil moisture is retrieved in the SMOS L2SM processor by modeling $T_B$ for incremental soil moisture and optical depth values and then minimizing the difference between the modeled and measured $T_B$ [13]. However, even when the processor is “successful” there is a possibility that the fit between the modeled and measured $T_B$ is poor. The reported $\chi^2$ probability is a measure of the cumulative $\chi^2$ distribution, normalized by the degrees of freedom (number of retrieved parameters), where $\chi^2$ is a measure of the “goodness” of fit. The use of a $\chi^2$ probability threshold of 0.05 limits the chance of rejecting valid retrievals, and in turn potentially accepting those with a poor fit, to less than 5%. When this $\chi^2$ threshold was applied after the RFI filter, the total number of valid retrievals was reduced from approximately 31 retrievals per DGG per month to 24. The removal of statistically poor retrievals ($\chi^2$ probability < 5%) increases the overall confidence in the data.

### IV. Comparison of SMOS L2SM and SFIR WASM

Figure 3 presents a comparison of SMOS L2SM retrievals and SFIR WASM for the three DGGs that intersect the SFIR from April 2013–November 2017 (excluding DJF). The data used to create this comparison, as well as a description of the in situ network, are provided in Section II. Over the entire time period, SMOS L2SM is biased -0.039 m$^3$ m$^{-3}$ as compared to the SFIR WASM (SMOS L2SM is drier than the network). The bias is 11% of the 0.08 to 0.44 m$^3$ m$^{-3}$ range of SFIR WASM. In addition, the ubRMSE of SMOS L2SM observations is 0.062 m$^3$ m$^{-3}$. This performance does not meet the mission accuracy goal of an ubRMSE less than 0.04 m$^3$ m$^{-3}$.

A month–by–month breakdown of the bias and ubRMSE, as well as ascending (6 AM) versus descending (6 PM) overpasses, is presented in Tables I and II. There are strong seasonal patterns. The largest multi–year average bias occurs in July (bias = -0.072 m$^3$ m$^{-3}$). The multi–year bias is greater than -0.05 m$^3$ m$^{-3}$ in April, June, July, and November. This pattern is consistent with seasonal changes that occur on the landscape in this region. The vast majority of the SFIR is used to grow two annual crops, corn and soybean, which are typically planted in May and rapidly accumulate biomass in June, July, and August. In November and April farmers often perform tillage which roughens the soil surface and changes its emissivity. Hence the SMOS L2SM dry bias is largest when vegetation and soil roughness, two other factors besides soil moisture that strongly affect terrestrial brightness temperature, are changing the most. Note also that the ubRMSE is above
that has shown that vegetation and soil roughness affect atypical late–spring snowfall.

Fig. 3. Comparison of SMOS L2SM to SFIR weighted average soil moisture (WASM) averaged over three Discrete Global Grid points (DGGs).

Our observations in the Corn Belt are not unique. Dry biases of approximately 0.10 m$^3$ m$^{-3}$ have been reported in the agricultural regions of the Skjern River Catchment in Western Denmark and the Upper Danube Catchment in Germany [19], [20]. A similar dry bias and ubRMSE in excess of mission goals has been observed in SMAP L2SM retrievals as compared to the SFIR WASM [11], [17].

It is a common hypothesis that biases between L–band satellite soil moisture retrievals and in situ observations exist due to differences in the volume of soil moisture sensed by SMOS L2SM in similar ways [18]. It is also possible the higher values of ubRMSE in the spring could be explained by sharp vertical gradients in near–surface soil moisture caused by rainfall during this period when the soil is either bare or only partially covered by vegetation [12]. However, the exact cause of the dry bias, in terms of both its magnitude and sign, is not yet understood.

Our initial hypothesis was that the SMOS L2SM dry bias and ubRMSE in the SFIR, initially identified using only the first year of SFIR WASM, would no longer exist as the retrieval algorithms evolved and the in situ data record lengthened. This was not supported by our comparison of SMOS L2SM and SFIR WASM. The dry bias still exists, but has improved to -0.039 m$^3$ m$^{-3}$ over the period of April 2013–November 2017 (excluding DJF). The ubRMSE of 0.062 m$^3$ m$^{-3}$ remains in excess of the mission accuracy goal of 0.04 m$^3$ m$^{-3}$. We discuss potential sources of the SMOS L2SM dry bias and ubRMSE in the SFIR in Section V.

V. Discussion

Our observations in the Corn Belt are not unique. Dry biases of approximately 0.10 m$^3$ m$^{-3}$ have been reported in the agricultural regions of the Skjern River Catchment in Western Denmark and the Upper Danube Catchment in Germany [19], [20]. A similar dry bias and ubRMSE in excess of mission goals has been observed in SMAP L2SM retrievals as compared to the SFIR WASM [11], [17].

It is a common hypothesis that biases between L–band satellite soil moisture retrievals and in situ observations exist due to differences in the volume of soil moisture sensed by
in situ probes and satellites [21, 22]. The volume observed by SMOS, approximately the top 3–5 cm of soil, is different than the volume observed by sensors inserted at 5 cm in the SFIR. However, this difference in sensing volume does not account for the SFIR dry bias [12]. Exponential fits were used to determine rates of soil drying after rainfall events for 0–3 cm and 4–6 cm layers as simulated by an agroecosystem model. The same analysis was performed for in situ sensors inserted at 1.5 cm and 4.5 cm in a 1 km × 1 km field. It was determined that while the shallower soil layer observed by SMOS does initially wet and dry faster than in the in situ layer, the effect is negligible after a few days for soils in the SFIR. While the differences in sensing volume did not impact the long–term bias they do contribute to the ubRMSE due to individual SMOS L2SM retrievals occurring directly after rainfall and long periods of dry weather.

We have identified potential sources of the SMOS L2SM dry bias in the SFIR by considering radiometric theory and analyzing the SMOS L2SM retrieval procedure [13]. Earth’s surface emits radiation in proportion to its emissivity, ϵ, and its temperature, T, as given by (5). Increases in soil moisture reduce the emissivity at microwave frequencies and therefore result in a lower TB observed by SMOS.

\[ T_B = \epsilon \ T \]  

(5)

SMOS L2SM is retrieved from observations of the 1.4 GHz TB using the τ–ω model (6), a zero–order radiative transfer model that simulates TB as a function of: soil temperature, \( T_g \); soil emissivity, \( \epsilon_s \); vegetation optical depth, \( \tau \); canopy temperature, \( T_c \); observation incidence angle, \( \theta \); and the single scattering albedo, \( \omega \) [23, 24].

\[ T_B = T_g \ \epsilon_s \ e^{-\tau/\cos \theta} + (1 - e^{-\tau/\cos \theta})(1 - \omega)T_c 
+ (1 - e^{-\tau/\cos \theta})(1 - \omega)T_c \ (1 - \epsilon_s) e^{-\tau/\cos \theta} \]  

(6)

SMOS L2SM uses \( T_B \) at multiple \( \theta \) and estimates of \( T_g \), \( T_c \), and \( \omega \) to simultaneously retrieve soil moisture and \( \tau \) for each DGG. A soil dielectric model and soil roughness model is used to determine soil moisture from \( \epsilon_s \) [25].

We have identified the following as likely causes of the dry bias:

1) bias in modeled auxiliary temperatures \( T_g \) and \( T_c \);
2) errors in auxiliary soil texture maps used to find soil moisture from \( \epsilon_s \);
3) non–representative parameterization of \( \omega \); and
4) an incorrect estimation of soil surface roughness which will affect \( \epsilon_s \).

A. Surface Temperature

SMOS does not have the capability to directly observe surface temperature so modeled temperature data from European Centre for Medium–range Weather Forecasts (ECMWF) is used in SMOS L2SM retrievals. According to (5), \( \epsilon \) and T are inversely proportional. If the auxiliary temperature used in the SMOS L2SM retrieval algorithm is too cold, \( \epsilon_s \) will be increased to simulate the same observed \( T_B \). Therefore \( \Delta T < 0 \), where \( \Delta T \) is defined by (7), will result in a SMOS L2SM dry bias.

\[ \Delta T = \text{SMOS } T - \text{SFIR } T \]  

(7)

The τ–ω model (6) portions observed \( T_B \) into contributions from the soil and the canopy. This requires using separate \( T_g \) and \( T_c \) instead of a single \( T \) for the surface. The SMOS L2SM retrieval algorithm calculates \( T_g \) following (8) and (9), where \( T_{surf} \) is the temperature of ECMWF soil layer 1 (0–7 cm), \( T_{depth} \) is the temperature of ECMWF soil layer 3 (28–100 cm), \( SM \) is the soil moisture of ECMWF soil layer 1, and \( w_0 \) and \( b_0 \) are static parameters that define the thermal profile of soil temperature.

\[ T_g = T_{depth} + C \ (T_{surf} - T_{depth}), \]  

(8)

\[ C_t = \min \left\{ \frac{SM}{w_0}, 1 \right\}. \]  

(9)

To compare \( T_g \) between SMOS L2SM retrievals and the SFIR, the SMOS \( T_g \) was calculated from (8) and (9) and the auxiliary ECMWF data associated with each retrieval. SFIR \( T_g \) was calculated with the 5 cm (\( T_{surf} \)) and 50 cm (\( T_{depth} \)) soil temperatures. The \( T_g \) bias, \( \Delta T_g \), is defined by (7) where \( T = T_g \).

Over the five–year SFIR record, the average bias \( \Delta T_g \) is 0.32 K. Monthly biases are shown in Table III. While the overall \( \Delta T_g \) is small, individual months vary and have both positive and negative \( T_g \) biases. There is a seasonal pattern in the sign of \( \Delta T_g \): July and August have \( \Delta T_g < 0 \) while the spring and fall months have \( \Delta T_g > 0 \). The monthly biases are less than 1.5 K with the exception of April 2014 and March 2015. Evening overpasses have a stronger \( \Delta T_g \) than morning overpasses (1.1 K vs -0.15 K).

The ubRMSE between SFIR and ECMWF–derived SMOS \( T_g \) are presented in Table IV. The \( T_g \) ubRMSE is 1.5 K for the five–year period. Evening overpasses have a higher \( T_g \) ubRMSE than morning overpasses. Previous analysis of ECMWF soil layer 1 temperature found that the modeled temperature was noisier in the evenings as compared to the single–point SCAN and SNOTEL in situ networks [26].

A realistic soil surface is not perfectly smooth. The sensitivity of soil moisture retrieval to surface temperature increases as both soil surface roughness increases. We therefore modified \( \epsilon_s \) in (6) to approximate the emissivity of a fairly rough soil surface instead of a quasi–specular surface by using a standard deviation of soil surface height \( \sigma = 25 \) mm and a correlation length \( l_c = 60 \) mm [27]. The modified \( \epsilon_s \) is given by:

\[ \epsilon_s = 1 - (1 - \epsilon_{\text{smooth}}) e^{-h_s}, \]  

(10)

where

\[ h_s = A \theta_v B \left( \frac{\sigma}{l_c} \right)^C, \]  

(11)

where \( \epsilon_{\text{smooth}} \) is the emissivity of a quasi–specular soil surface, \( \theta_v \) is the volumetric soil moisture, and \( A, B, \) and \( C \) are empirical parameters equal to 0.5761, -0.3475, and 0.4230, respectively [28].
TABLE III
SMOS – SFIR $T_g$ bias averaged over three Discrete Global Grid points (DGGs).

<table>
<thead>
<tr>
<th>bias</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Mar–Nov</th>
<th>6 AM</th>
<th>6 PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013, K</td>
<td>*</td>
<td>-0.29</td>
<td>0.78</td>
<td>0.04</td>
<td>-0.61</td>
<td>0.01</td>
<td>1.4</td>
<td>0.78</td>
<td>-0.38</td>
<td>0.30</td>
<td>0.09</td>
<td>0.66</td>
</tr>
<tr>
<td>2014, K</td>
<td>*</td>
<td>2.1</td>
<td>0.67</td>
<td>0.71</td>
<td>-0.53</td>
<td>-0.50</td>
<td>-0.44</td>
<td>-0.11</td>
<td>-0.43</td>
<td>0.13</td>
<td>-0.40</td>
<td>1.0</td>
</tr>
<tr>
<td>2015, K</td>
<td>3.1</td>
<td>1.0</td>
<td>0.37</td>
<td>-0.52</td>
<td>-0.86</td>
<td>-0.26</td>
<td>0.34</td>
<td>0.77</td>
<td>0.16</td>
<td>0.29</td>
<td>-0.15</td>
<td>1.1</td>
</tr>
<tr>
<td>2016, K</td>
<td>0.64</td>
<td>0.81</td>
<td>0.68</td>
<td>1.2</td>
<td>-0.43</td>
<td>-0.08</td>
<td>0.67</td>
<td>1.5</td>
<td>-0.41</td>
<td>0.53</td>
<td>-0.15</td>
<td>1.7</td>
</tr>
<tr>
<td>2017, K</td>
<td>0.65</td>
<td>0.75</td>
<td>0.41</td>
<td>0.10</td>
<td>-0.24</td>
<td>-0.34</td>
<td>1.1</td>
<td>0.70</td>
<td>-0.54</td>
<td>0.33</td>
<td>-0.10</td>
<td>1.0</td>
</tr>
<tr>
<td>All Years, K</td>
<td>2.7</td>
<td>0.88</td>
<td>0.58</td>
<td>0.31</td>
<td>-0.53</td>
<td>-0.24</td>
<td>0.62</td>
<td>0.73</td>
<td>-0.32</td>
<td>0.32</td>
<td>-0.15</td>
<td>1.1</td>
</tr>
</tbody>
</table>

* The SFIR in situ network was installed in April 2013.
* March 2014 was significantly affected by snow and had few successful retrievals. The $T_g$ bias was 6.3 K.

TABLE IV
ubRMSE between SMOS and SFIR $T_g$ averaged over three Discrete Global Grid points (DGGs).

<table>
<thead>
<tr>
<th>ubRMSE</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Mar–Nov</th>
<th>6 AM</th>
<th>6 PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013, K</td>
<td>*</td>
<td>1.2</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
<td>1.9</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>2014, K</td>
<td>*</td>
<td>1.4</td>
<td>1.4</td>
<td>1.2</td>
<td>1.1</td>
<td>0.84</td>
<td>1.6</td>
<td>0.96</td>
<td>1.7</td>
<td>1.5</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>2015, K</td>
<td>1.4</td>
<td>1.6</td>
<td>1.2</td>
<td>0.88</td>
<td>0.64</td>
<td>1.1</td>
<td>1.2</td>
<td>1.4</td>
<td>1.5</td>
<td>1.5</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>2016, K</td>
<td>1.8</td>
<td>0.76</td>
<td>1.4</td>
<td>1.7</td>
<td>0.99</td>
<td>1.1</td>
<td>1.3</td>
<td>1.4</td>
<td>2.7</td>
<td>1.6</td>
<td>0.90</td>
<td>1.9</td>
</tr>
<tr>
<td>2017, K</td>
<td>1.4</td>
<td>1.2</td>
<td>1.3</td>
<td>1.3</td>
<td>0.84</td>
<td>0.87</td>
<td>1.4</td>
<td>1.3</td>
<td>1.3</td>
<td>1.2</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>All Years, K</td>
<td>1.4</td>
<td>1.2</td>
<td>1.3</td>
<td>1.2</td>
<td>0.93</td>
<td>1.1</td>
<td>1.4</td>
<td>1.2</td>
<td>1.8</td>
<td>1.5</td>
<td>1.3</td>
<td>1.5</td>
</tr>
</tbody>
</table>

* The SFIR in situ network was installed in April 2013.
* March 2014 was significantly affected by snow and had few successful retrievals. The $T_g$ ubRMSE was 6.4 K.

We investigated the sensitivity of SMOS L2SM retrievals to errors in surface temperature for the average $T_g$ ubRMSE of 1.5 K. This sensitivity is illustrated for bare soil in Figure 4a. Soil moisture was simulated for horizontally (h-pol) and vertically (v-pol) polarized $T_B$ at $\theta$ of 15°, 40°, and 65° assuming $T_g = 300$ K. The typical range in $\theta$ for measured $T_B$ is 15° to 65°. The simulated sensitivity matches our physical expectations of too–cool temperatures resulting in soil moisture retrievals that are biased dry.

$\Delta T_g = -1.5$ K produces a dry bias of less than 0.005 m$^3$ m$^{-3}$ for most combinations of polarization and incidence angle over a bare soil. This suggests that the average $\Delta T_g = 0.32$ K in the SFIR is too small to have an affect on SMOS L2SM retrievals. The sign of $\Delta T_g$ indicates that biases in $T_g$ are not contributing to the dry bias in the SFIR and may actually produce wetter SMOS L2SM retrievals if the retrieval algorithm is more sensitive than simulated.

Soil moisture retrievals based on v-pol $T_B$ at 65° have a significantly higher sensitivity to errors in surface temperature. This due to $\epsilon_g$ approaching 1 as $\theta$ approaches the Brewster angle, $\theta_B$. When $\theta$ equals $\theta_B$, changes in soil moisture will not have an affect of $\epsilon_g$ and therefore there can be no retrievals of soil moisture. $\theta_B$, which will always be larger than $\theta$ for SMOS L2SM retrievals, becomes closer to $\theta = 65^\circ$ for dry soils and results in an increased sensitivity to surface temperature errors. Consequently there are no soil moisture retrievals for v-pol $T_B$ at 65° in Figure 4a for very dry soils because the retrieved soil moisture would be drier than our lower limit for realistic soil moisture in the simulations (0.05 m$^3$ m$^{-3}$).

While we simulated the effect of $\Delta T$ on soil moisture retrievals for individual polarizations and $\theta$, SMOS L2SM is a product of many $T_B$ at varying $\theta$ for both h-pol and v-pol. Successful SMOS L2SM retrievals are calculated based on 10–100 observations of $T_B$; the average in the SFIR is 71. The amount of $T_B$ and their $\theta$ is dependent on the proximity of a DGG to the center of the overpass swath. The SMOS L2SM retrieval algorithm minimizes the differences between the observed $T_B$ and those simulated by (6), diminishing the impact of $\theta_B$ on SMOS L2SM retrievals as the inclusion of smaller $\theta$ and h-pol $T_B$ weights the retrieval towards lower sensitivity to $\Delta T$.

Observed $\epsilon$ is higher for vegetation canopies ($\tau > 0$ Np) than for bare soil ($\tau = 0$ Np). Unfortunately $T_c$, defined as the ECMWF skin temperature for SMOS L2SM retrievals, is not a common in situ measurement and is not collected in the SFIR. The high water column density, $M_w$, of corn results in increased potential for errors in $T_c$ to propagate into SMOS L2SM retrievals. $\tau$ was set to 0.44 Np using (12) to test the sensitivity of soil moisture retrievals to errors in $T_c$, where the croplands b–parameter is 0.110 and $M_w$ for a typical mixed corn and soybean canopy in late July is 4 kg m$^{-2}$ [23], [11], [29].

$$\tau = b \times M_w$$

(12)

The infrared surface temperature (i.e., skin temperature) averages 1.5 K colder than the 4.5 cm soil temperature at 6 AM solar time and 0.25 K warmer at 6 PM for a corn canopy during July, August, and September [29]. We calculated the difference between SMOS $T_c$ and $T_{surf}$ to determine if the modeled $T_c$ is realistic for a corn canopy. When averaged over July–September, $T_c$ was 2.2 K colder than $T_{surf}$ at 6 AM solar
time and 1.7 K colder at 6 PM for SMOS L2SM retrievals in the SFIR. This suggests that SMOS $T_{\theta}$ is biased colder than $T_{g}$ for both 6 AM and 6 PM overpasses. The colder bias for 6 PM overpasses may be contributing to SMOS L2SM having a slightly drier bias at 6 PM than at 6 AM (Table III).

Figure 4b illustrates the sensitivity to temperature for a late–July corn and soybean canopy assuming that errors in the modeled $T_{g}$, $\Delta T_{g}$, are roughly the same magnitude as $\Delta T_{\theta}$. $\Delta T_{g}$ = -1.5 K produces a dry bias less than 0.03 m$^3$ m$^{-3}$ for most combinations of $\theta$ and polarization. The temperature sensitivity patterns related to $\theta$ and polarization are similar for the vegetated case as in the bare soil case. The decreased sensitivity of $\epsilon$ to changes in soil moisture in the presence of high $M_{\epsilon}$ amplifies the effect of errors in surface temperature. Realistically the sensitivity in the SFIR would vary between the bare soil and vegetated cases as the canopy grows and senesces throughout the year.

Bias in the ECMWF–modeled $T_{g}$ is likely not the source of the SMOS L2SM dry bias in the SFIR. The $T_{g}$ bias of 0.32 K is both too small and the wrong sign to be producing a dry bias in soil moisture retrievals. A portion of the ubRMSE between SMOS L2SM and SFIR WASM may be due to noisiness in the modeled surface temperatures ($T_{g}$ ubRMSE = 1.5 K). While the $T_{g}$ utilized during SMOS L2SM retrievals is a skin temperature representative of the very top of the canopy, the average temperature of the canopy will be somewhere between $T_{g}$ and $T_{surf}$ with the difference between the two being smaller for soybean canopy than in corn. Comparing the difference between $T_{g}$ and $T_{surf}$ to known relationships in a corn canopy indicates that $T_{g}$ is biased colder than anticipated, particularly for 6 PM overpasses, and may contribute to the dry bias at the peak of the growing season when sensitivity to $\Delta T_{g}$ is largest.

B. Soil Texture

There are two main components to soil moisture: free water and bound water. Free water is visible to SMOS as molecules are able to rotate at microwave frequencies. Bound water, which is held immobile against soil particles, is essentially invisible to SMOS. Soils with high clay contents have an increased proportion of bound water due to the larger surface area of clay particles. This results in clay soils having a higher $\epsilon_{s}$ than soils with less clay at the same soil moisture. Therefore in order to retrieve soil moisture using (6), the SMOS L2SM retrieval algorithm needs the fraction of clay, $C$, in the soil. If a too–low $C$ is assumed ($\Delta C < 0$), the retrieval algorithm believes that there is less bound water than actually present in the soil and retrieves a soil moisture that is biased dry.

The SMOS L2SM retrieval algorithm currently uses the Mironov dielectric mixing model [30]. The Dobson and Wang and Schmugge dielectric mixing models are also available [31], [32]. All three react differently to varying $C$. We used (6) and (10) to estimate how errors in $C$ propagate through the dielectric mixing models and into simulated soil moisture.

Figure 5 illustrates the effect of errors in clay content, $\Delta C$, on SMOS L2SM retrievals for h–pol and v–pol $T_{B}$ at $\theta$ of 15°, 40°, and 65°. The simulation shown in Figure 5 assumes the average SFIR soil texture according to the USDA Natural Resources Conservation Service (USDA–NRCS) and a “true” soil moisture of 0.25 m$^3$ m$^{-3}$. The sensitivity of soil moisture retrievals to $\Delta C$ does not differ significantly for wetter soils; drier soils are less sensitive. The Mironov and Wang and Schmugge models, which retrieve a dry–biased soil moisture for $\Delta C < 0$, act as physically expected. The Dobson model retrieves a wet–biased soil moisture for $\Delta C < 0$.

The differences in simulated soil moisture for the three dielectric mixing models are shown in Figure 6. SMOS L2SM retrievals performed using the Wang and Schmugge model would be wetter than those using the Mironov model, however
the Wang and Schmugge has several more required soil textural parameters and thus an increased potential for error to propagate into SMOS L2SM. The Dobson model produces drier soil moistures than the Mironov until the soil nears saturation. This simulation using the average SFIR soil texture is reinforced by the report that global SMOS L2SM became 0.033 m$^3$ m$^{-3}$ wetter after the retrieval algorithm switched from using the Dobson model to the Mironov [33].

The SMOS L2SM auxiliary and USDA–NRCS clay maps of the SFIR are compared in Figure 7. The region depicted with Figure 7 is the outlined region from Figure 1. The general features of the 4 km resolution maps are similar; an area of higher clay along the western boundary that aligns with ancient clay lake deposits is clearly visible. The average clay content, $C$, is 0.25 for the SMOS L2SM auxiliary map and 0.28 for the USDA–NRCS map. It is worth noting that SMOS L2SM versions prior to v06.20 used a significantly different clay map over the SFIR (spatially constant $C = 0.40$) which was creating a wetting effect of $\approx 0.02$ m$^3$ m$^{-3}$. While the use of this map did result in wetter SMOS L2SM retrievals, the parameterization of $C = 0.40$ for the entire SFIR was extremely unrealistic.

The soil texture map currently used for SMOS L2SM retrievals in the SFIR has a $C$ bias of -0.03. This corresponds to a 0.003 m$^3$ m$^{-3}$ dry bias in soil moisture retrievals using the Mironov model and is too small to noticeably impact SMOS L2SM.

C. Single Scattering Albedo

When radiation emitted by the soil passes through a corn canopy, some is scattered away due to the similar magnitudes of the size of the plant components and the wavelength of L–band radiation. This effect, known as scatter darkening, reduces observed $T_B$ [27]. The $\omega$ parameter in (6) is used to simulate the effect of scatter darkening caused by vegetation. SMOS L2SM retrievals currently neglect scatter darkening ($\omega = 0$) in “croplands” land classes such as those over the SFIR. If the retrieval algorithm does not know to account for scatter darkening the low $T_B$ is interpreted as a wetter soil. Therefore increasing $\omega$ to a more realistic (non-zero) value for the SFIR is expected to further dry SMOS L2SM.

We tested the impact of increasing $\omega$ on SMOS L2SM soil moisture by running the SMOS Soil Moisture L2 Prototype Processor v06.20 (SML2PP), developed by Array Systems Computing and the European Space Agency to retrieve soil moisture for custom surface parameterizations, for a high vegetation test case with $\omega=0.05$ for croplands, $\omega=0.05$ was chosen for consistency with SMAP L2SM retrievals over croplands [11]. We then compared soil moisture retrieved by the SML2PP to the operational SMOS L2SM and SFIR network soil moisture for the test case.

During the two–month high vegetation test case, increasing $\omega$ to a more realistic value in the SML2PP increased the magnitude of the dry bias by 41%. Accounting for scattering in the canopy improved both the noisiness and correlation coefficient (R) between SMOS L2SM and the SFIR WASM: the ubRMSE decreased by 22% and R increased by 15% (from 0.67 to 0.77).

D. Soil Surface Roughness

According to (10) and (11), a soil with a high surface roughness (higher $\sigma$) have a larger $\epsilon_t$ than a smooth soil with the same soil moisture. If SMOS L2SM retrieval algorithm uses a value of $\sigma$ that is too low, soil moisture must be decreased to increase $\epsilon_t$. Therefore, the parameterization of a too–smooth soil results in SMOS L2SM retrievals that are biased dry.

We used the SML2PP to simulate SMOS L2SM retrievals assuming smooth ($\sigma \approx 5$ mm), moderately–rough ($\sigma = 15$ mm), and very–rough ($\sigma = 30$ mm) soils with...
\( \omega = 0.05 \). This was done for both bare soil (April–May, 2015) and high vegetation (July–August, 2015) test cases.

Figure 8a compares SMOS L2SM soil moisture to the SFIR WASM for the bare soil SML2PP test case. As \( \sigma \) increases from smooth to very–rough soils, the dry bias improves (from 0.087 m\(^3\) m\(^{-3}\) to 0.013 m\(^3\) m\(^{-3}\)) and the ubRMSE of the retrievals decreases (from 0.052 m\(^3\) m\(^{-3}\) to 0.033 m\(^3\) m\(^{-3}\)). However, these improvements come at the cost of soil moisture retrieval sensitivity to actual soil moisture; the slope of the best-fit line between SMOS L2SM and the SFIR is halved between the moderately–rough and very–rough scenarios. In other words, for the same set of “actual” soil moisture, the range of SMOS L2SM retrieval values is halved between the moderately–rough and rough scenarios. Physically this is due to the increased emissivity of a rough soil reducing the change of observed \( T_b \) with respect to changes in soil moisture.

Figure 8b compares SMOS L2SM to the SFIR WASM for the high vegetation test case. As with the bare soil test case, the dry bias improves (from 0.141 m\(^3\) m\(^{-3}\) to 0.030 m\(^3\) m\(^{-3}\)) and retrieval sensitivity decreases as the roughness parameterization increases. The ubRMSE of the closed canopy case is not affected by increasing \( \sigma \).

VI. Conclusion

SMOS L2SM retrievals are too dry when compared to the SFIR in situ soil moisture network, a satellite soil moisture validation site in the U.S. Corn Belt. The five–year average bias, defined as the mean difference between SMOS L2SM and SFIR weighted average soil moisture (WASM), is -0.039 m\(^3\) m\(^{-3}\). We investigated both auxiliary modeled soil temperature and the soil textural maps utilized by SMOS L2SM and found no evidence to support them as sources of the dry bias. While there are no in situ observations available for direct comparison, the modeled skin temperature was colder than anticipated when compared to soil temperature. The too–cold \( T_c \) for 6 PM overpasses could explain why SMOS L2SM evening retrievals have a slightly drier bias than in the morning but is not the source of the overall dry bias. Increasing assumed soil surface roughness from smooth (\( \sigma = 5 \) mm) to moderately–rough (\( \sigma = 15 \) mm) in the SFIR results in wetter SMOS L2SM retrievals without the extreme loss in retrieval sensitivity to soil moisture that occurs in the very–rough soil (\( \sigma = 30 \) mm) scenario. Utilizing \( \sigma = 15 \) mm during retrievals improves the dry bias but does not fully correct it.

The ubRMSE of 0.062 m\(^3\) m\(^{-3}\) between SMOS L2SM and SFIR WASM exceeds the SMOS mission accuracy goal of 0.04 m\(^3\) m\(^{-3}\). The ubRMSE is highest in the spring months when there is little vegetation. This suggests that differences in the sensing volumes of SMOS and the SFIR network may be contributing to the overall noisiness of SMOS L2SM retrievals for bare soil. The ubRMSE of ECMWF–derived soil temperatures in the SFIR may also be contributing to the noisiness. Increasing the single scattering albedo utilized in SMOS L2SM retrievals has been shown to decrease the ubRMSE of a high–vegetation test case at the cost of worsening the dry bias.

Acknowledgment

Support for this research was provided by the Iowa Water Center, operated in cooperation between Iowa State University and the United States Geological Survey through the Water Resources Institute Act. This research was performed as part of Iowa Agriculture and Home Economics Experiment Station project IOW05387.

References

High vegetation

(a) Bare soil

(b) High vegetation

Fig. 8. SMOS L2S retrievals produced by the Soil Moisture L2 Prototype Processor (SML2PP) for smooth ($\sigma = 5$ mm), moderately–rough ($\sigma = 15$ mm), and very–rough ($\sigma = 30$ mm) soils during low and high vegetation test cases. Operational SMOS L2S retrievals (OPER) are included for comparison.


Victoria A. Walker [Insert biography text for author w/ photo]

Michael H. Cosh [Insert biography text for author w/ photo]

Brian K. Hornbuckle [Insert biography text for author w/ photo]