The Evaluation of SMAP Enhanced Soil Moisture Products Using High-Resolution Model Simulations and In-Situ Observations on the Tibetan Plateau

Chengwei Li 1,2, Hui Lu 1,2,*, Kun Yang 1,2,3,4, Menglei Han 1,2, Jonathon S. Wright 1,2, Yingying Chen 3,4, Le Yu 1,2, Shiming Xu 1,2, Xiaomeng Huang 1,2 and Wei Gong 5

1 Ministry of Education Key Laboratory for Earth System modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China; licw13@mails.tsinghua.edu.cn (C.L.); yangk@itpcas.ac.cn (K.Y.); nnhx614@163.com (M.H.); jswright@tsinghua.edu.cn (J.S.W.); leyu@tsinghua.edu.cn (L.Y.); xusm@tsinghua.edu.cn (S.X.);hxm@tsinghua.edu.cn (X.H.)
2 Joint Center for Global Change Studies, Beijing 100875, China
3 Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China; chenyy@itpcas.ac.cn
4 CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing 100101, China
5 Institute of Land Surface System and Sustainable Development, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; gongwei2012@bnu.edu.cn

* Correspondence: luhui@tsinghua.edu.cn; Tel.: +86-10-6277-2565

Received: 4 February 2018; Accepted: 28 March 2018; Published: 31 March 2018

Abstract: The Soil Moisture Active Passive (SMAP) mission was designed to provide a global mapping of soil moisture (SM) measured by L-band passive and active microwave sensors. In this study, we evaluate the newly released SMAP enhanced SM products over the Tibetan Plateau by performing comparisons among SMAP standard products, in-situ observations and Community Land Model (CLM) simulations driven by high-resolution meteorological forcing. At local scales, the enhanced SMAP products, the standard products and CLM simulations all generally compare well with the in-situ observations. The SMAP products show stronger correlations (0.64–0.88) but slightly larger unbiased root mean square errors (ubRMSE, ~0.06) relative to the CLM simulations (0.58–0.79 and 0.037–0.047, for correlation and ubRMSE, respectively). At the regional scale, both SMAP products show similar spatial distributions of SM on the TP (Tibetan Plateau), although, as expected, the enhanced product provides more fine details. The SMAP enhanced product is in good agreement with model simulations with respect to temporal and spatial variations in SM over most of the TP. Regions with low correlation between SMAP enhanced products and model simulations are mainly located in the northwestern TP and regions of complex topography, where meteorological stations are sparse and non-existent or elevation is highly variable. In such remote regions, CLM simulations may be problematic due to inaccurate land cover maps and/or uncertainties in meteorological forcing. The independent, high-resolution observations provided by SMAP could help to constrain the model simulation and, ultimately, improve the skill of models in these problematic regions.

Keywords: Soil Moisture Active Passive; evaluation; in situ observations; CLM simulations

1. Introduction

Soil moisture (SM) is an essential variable for the understanding, modeling and forecasting of weather and climate [1–3], the monitoring and early warning of floods and droughts [4,5], and the estimation of crop yield [6,7]. Estimates of SM with high accuracy and fine spatiotemporal resolution are necessary to meet these and other needs. However, in-situ observations are too sparse to
adequately represent spatial variations in SM. Microwave remote sensing provides a promising approach to provide SM at large spatial scales and high temporal resolution under a wide range of weather conditions.

The National Aeronautics and Space Administration’s (NASA) Soil Moisture Active Passive (SMAP) mission [8] was launched on 31 January 2015, and it has been providing high resolution global maps of SM and freeze–thaw states since 31 March 2015 (passive sensor) and 13 April 2015 (active sensor). The SMAP satellite carries an L-band radar (centered at 1.26 GHz) and an L-band radiometer (centered at 1.41 GHz) that provide backscatter information at a 3-km spatial resolution and brightness temperature observations at 36-km resolution. The low frequency of the operating active and passive channels, and the use of a large antenna (6 m diameter) help SMAP to reach a higher sensitivity to SM relative to previous sensors [9]. Unfortunately, the SMAP radar stopped working on 7 July 2015 due to a mechanical failure. As a result, the radar observations and SM products at 3-km resolution and the combined (active–passive) SM products at 9-km resolution are only available for a short period of about three months. Only the passive SM products (e.g., L2/3_SM_P [10]) are available beyond those three months. Despite the loss of the radar, SMAP continues to provide high-resolution SM observations to the extent possible by using two approaches [11]. The first approach combines the current SMAP coarse-resolution passive observations with high-resolution radar observations from other satellites in orbit [12,13]. The second approach applies the Backus–Gilbert (BG) optimal interpolation technique [14,15] to antenna temperature (T_A) measurements in the original SMAP Level 1B Brightness Temperature Product (L1B_TB) [16]. The SMAP Level 1C Enhanced Brightness Temperature Product (L1C_TB_E) on 9 km grid is then derived from the interpolated T_A by using standard correction/calibration procedures [17]. The L1C_TB_E product is then used as the primary input to subsequent passive geophysical inversions to produce the SMAP Level 2 Enhanced Passive Soil Moisture Product (L2_SM_P_E) [18]. In addition, a contributing domain of 33 km on a side is chosen to approximate the spatial extent of the SMAP radiometer in L2_SM_P_E generation process [11]. Comparison with the L2_SM_P product indicates that the higher spatial resolution of L2_SM_P_E product does not introduce large errors, while allowing greater acuity in spatial details [11,19]. The enhanced SMAP L3_SM_P_E product, which is the focus of this paper, is a daily global composite of the enhanced SMAP L2_SM_P_E product.

The validation of SMAP SM products for different climatic conditions and land covers types is of great utility, not only in view of possible applications (e.g., assimilation in land surface models, drought monitoring, and/or land–atmosphere interaction studies) [20–22], but also to help guide further algorithm improvements. Previous validations have been based on in situ observations [23,24] or other satellite-based products [25]. However, most validation of SMAP SM products has relied on in-situ measurements collected in temperate climate regions [26]. Additional efforts to evaluate SMAP SM products at larger spatial scales and in other climatic zones, are expected. Land surface model simulations provide means of evaluating SMAP SM products at larger spatial scales, as these models can capture dynamic changes in SM well when forced by high quality atmospheric forcing data [27,28].

The Tibetan Plateau (TP) is the highest and largest plateau in the world, with an average elevation of over 4000 m and an area of approximately $2.5 \times 10^6$ km$^2$. Owing to strong solar heating and complex topography, the TP experiences strong land–atmosphere interactions and plays an important role in the development of the Asian monsoon [29], in many ways dominating the regional energy and water cycles in Asia. However, key variables related to these interactions, such as SM, are poorly understood due to a lack of observations. Remote sensing data can be used to supplement to the existing in-situ observations if they can reliably capture temporal dynamics and spatial variations in these key variables. Incorporation of reliable high-resolution (~10 km) SM data could also enhance the understanding and predictability of regional weather systems worldwide [30]. In these aspects, the SMAP enhanced SM products could be extremely useful for research on land processes and land–atmosphere interactions over the TP. Chen et al. [26] evaluated the SMAP standard passive SM product using observations from two networks on the TP. However, the enhanced products have not
yet been validated for this region. Such a validation should ideally apply to large spatial scales. In this study, we evaluate SMAP enhanced SM products in two stages. In the first stage, SMAP enhanced and standard SM products are evaluated against in-situ measurements from two SM and temperature networks on the TP. In the second stage, high-resolution land surface model simulations are used to evaluate the SMAP enhanced SM products at large spatial scales.

Information on the datasets and model simulations is provided in Section 2. The evaluations of SMAP enhanced SM products against both in-situ observations and model simulations are introduced in Section 3. The latter comparison is explored more fully in Section 4, particularly with respect to the factors that influence correlations between SMAP enhanced products and model simulations. The results of this study are summarized in Section 5.

2. Materials and Methods

2.1. SMAP Enhanced Soil Moisture Product

The SMAP is the latest L-band satellite mission that provides global-scale SM and freeze/thaw state measurements [31]. SMAP generates a range of products and SM retrievals. Level 2 refers to half-orbit products, Level 3 to the daily gridded composites and Level 4 [32] to model-assimilated products. In this study, SMAP Level 3 Radiometer Global Daily 9 km EASE-GRID enhanced passive SM product (version 3) is evaluated against in-situ observations, the coarse-resolution (~36 km) SMAP Level 3 passive (radiometer) SM product (L3_SM_P), and model simulations over the TP.

The enhanced SMAP L3_SM_P_E product is a daily gridded global composite based on the enhanced SMAP L2_SM_P_E product. The development of L2_SM_P_E largely parallels that of the SMAP Level 2 passive SM product (L2_SM_P) [33,34]. Both products share the same processing flow, ancillary data, and retrieval algorithms. First, fore- and aft-look brightness temperature observations from L1C_TB_E are combined to provide the primary input to the L2_SM_P_E processor. The retrieval is then evaluated against preprocessed finer resolution ancillary data (for example, freeze/thaw fraction and soil temperature). The processor will then further evaluate the quality of the retrieval if the retrieval is considered feasible at a given location. If the surface conditions are deemed favorable to SM retrieval, corrections are then applied for surface roughness, effective soil temperature, vegetation water content, and the radiometric contributions by water bodies. Once all steps are complete, the brightness temperature observations and ancillary data are used as inputs to the baseline SM retrieval algorithm, producing L2_SM_P_E on a 9 km EASE-Grid 2.0 global projection. Further details on the development of L2_SM_P_E can be found in the Product Specification Document [33].

2.2. In-Situ Observations

In this study, two SM and temperature monitoring networks are used to evaluate the SMAP SM products. Figure 1 shows the locations of these two networks.

The first network is the Naqu network, which is located in a cold semiarid environment. The Naqu basin consists of largely flat terrain with rolling hills. This area has dry winters and rainy summers. Annual precipitation is less than 500 mm, ~75% of which occurs during the monsoon season (June–August). A total of 58 SM stations were deployed since July 2010 within a 100 km by 100 km area. Further details on the Naqu network can be found in Reference [35].

The second network is the Maqu network, which is located in a cold semi-humid environment. The Maqu network was installed in July 2008 near the head waters of the Yellow River, south of Maqu County in Gansu province, China. The network consists of 20 stations in an area of approximately 40 km (north to south) by 80 km (west to east). Further details on the Maqu network can be found in Reference [36].
Figure 1. The locations of the two soil moisture and temperature monitoring networks used in this study (denoted as two small black rectangles on the Tibetan Plateau). The distribution of land cover on the Tibetan Plateau is shown for context. Land cover is based on MODIS (Moderate Resolution Imaging Spectroradiometer), with classifications drawn from IGBP (International Geosphere-Biosphere Programme) land cover types.

2.3. High-Resolution Land Surface Modeling

To assess the spatial variability of the SMAP SM product, we use a state-of-the-art land surface model, the Community Land Model (CLM) [37] developed by the National Center for Atmospheric Research (NCAR). The latest version of the CLM model series (CLM4.5) is used. In CLM4.5, land surface spatial heterogeneity is represented as a nested subgrid hierarchy in which grid cells are composed of multiple land units, snow/soil columns, and plant functional types [37]. CLM4.5 contains 15 soil layers. The average SM in the top two soil layers (0–4.51 cm) is selected to match the approximate depth of the SMAP SM product.

The China Meteorological Forcing dataset (CMFD) [38] is used to drive the CLM model. The CMFD is a hybrid combination of data obtained from other meteorological forcing datasets and observations from 740 operational stations of the China Meteorological Administration, provided at 3-h temporal resolution on a 0.1° × 0.1° grid. The meteorological forcing data include the Tropical Rainfall Measuring Mission (TRMM) 3B42 precipitation analysis [39], the Asian Precipitation–Highly Resolution Observational Data Integration Towards Evaluation of the Water Resources (APHRODITE) precipitation analysis [40], the Global Energy and Water Cycle Experiment—Surface Radiation Budget (GEWEX-SRB) shortwave radiation data [41], and other fields derived from the Princeton meteorological forcing dataset [42]. The dataset is recognized as one of the best forcing datasets in China [43,44], and has been used in multiple land surface and hydrological modeling studies in China [43,45,46]. Other required data, such as soil texture and leaf area index, are derived from surface data pool. Details of raw data used in CLM4.5 can be found in the CLM4.5 technical documentation [37]. The simulation is conducted starting from 1980 to provide sufficient model spin-up.

2.4. Methods

It is challenging to use in-situ ground measurements at a point location to validate SM in a satellite pixel or model grid cell. To better resolve these inherent issues with spatial mismatch, the arithmetic average of in-situ observations from all stations within a grid cell are often used as ground truth for evaluating retrieved SM. This method has been widely adopted in previous validation studies [47–49] since spatial averaging of station data can also effectively reduce uncertainties in the ground “truth”. Therefore, in-situ surface SM observations from stations in the Naqu (0–5 cm) and Maqu (5 cm) networks are averaged to represent “ground truth” in the corresponding networks. The values of all L3_SM_P_E/ L3_SM_P pixels and CLM simulations grid cells containing SM stations are likewise...
averaged to prepare the SMAP retrievals or model simulations for comparison with network-mean in-situ observations.

SMAP instruments observe the Earth’s surface with a near-polar, Sun-synchronous 6:00 a.m. (descending)/6:00 p.m. (ascending) orbit. In this study, SMAP radiometer SM products derived from the observations acquired from the 6:00 a.m. (local solar time) descending passes are used. The CLM simulations used in this study have a 30-min time step with forcing data interpolated to the model time step by CLM. The output interval for the simulated SM is also 30 min. Observations of Naqu network are provided at 30 min interval, while observations from the Maqu network are provided every 15 min. The SM comparison is conducted using SMAP observations, CLM simulations and in-situ observations from 6:00 a.m. local solar time. In addition, since the SMAP L3_SM_P_E and L3_SM_P products are not available during the frozen season, our evaluations of these products are conducted only for months when the areas containing the networks are not in the frozen season (i.e., May–September for 2015 and 2016). Performance was evaluated based on four statistical metrics, including the mean bias (Bias) and the time series correlation (R). Due to the mismatches in spatial representativeness between in-situ observations and SMAP products, the root mean square error (RMSE) may be inflated by the bias. The ubRMSE (unbiased RMSE) [50] is therefore introduced to evaluate dynamic temporal variability in the SMAP SM products and CLM simulations [31,48,49].

Before comparison with CLM simulations, differences between SMAP standard and enhanced Level3 SM products are checked to assess the advantages of the enhanced product (i.e., its capability to represent spatial information in a more detailed way). Pixels with vegetation water content (VWC, defined as the mass of vegetation water per unit area) larger than 5 kg m\(^{-3}\) in both the standard and enhanced SMAP SM products are masked before comparison. VWC is the primary determinant of the optical depth of the canopy layer in the microwave frequency. For comparison with the model-simulated SM, the SMAP L3_SM_P_E product is first gridded to 0.1° resolution to match the model simulation. The spatial distribution of temporal correlations between the SMAP L3_SM_P_E products and the CLM simulations for each pixel is then calculated. The spatial pattern correlations of SM during the unfrozen seasons of 2015 and 2016 are also compared between the CLM simulation and the SMAP L3_SM_P_E product.

3. Results and Discussions

3.1. Comparison with In Situ Observations

Figure 2a,b shows time series and scatterplot of SM based on the in-situ observations (black solid line), the SMAP L3_SM_P_E product (green circle), the SMAP L3_SM_P product (blue triangle) product, and the CLM simulations (orange square) average over the Naqu network. Figure 2c,d shows similar plots for the Maqu network. Quantitative values for the statistical error metrics are listed in Table 1.

In the Naqu network, two SMAP SM products and CLM simulations evidently capture temporal variations in the-in situ observations (R = 0.88 for SMAP L3_SM_P and L3_SM_P_E products, R = 0.79 for CLM). The CLM simulations tend to underestimate SM with a bias of \(-0.022\) m\(^3\) m\(^{-3}\); by contrast, both SMAP products show very small positive biases (0.007 m\(^3\) m\(^{-3}\) for SMAP L3_SM_P and 0.005 m\(^3\) m\(^{-3}\) for SMAP L3_SM_P_E). All three estimates can achieve modest accuracy after removing their respective biases, with ubRMSE values of 0.059 m\(^3\) m\(^{-3}\) for L3_SM_P, 0.055 m\(^3\) m\(^{-3}\) for L3_SM_P_E and 0.037 m\(^3\) m\(^{-3}\) for CLM. Figure 2a,b and the related statistical metrics suggest that the two SMAP products capture temporal variations in SM well, and with a performance similar to that provided by the CLM simulations.

For the Maqu network, both SMAP products overestimate SM relative to in-situ observations. The Maqu network is located in a semi-humid region with more extensive vegetation (leaf area index (LAI) \(~0.1–3.1\) based on MODIS observations during 2000–2009) than the Naqu network (LAI \(~0.13–1.27\)). These positive biases may thus arise due to the influence of the more extensive canopy. When the bias is removed, CLM and SMAP products show modest accuracy, with ubRMSE values...
of 0.058 m$^3$ m$^{-3}$ for L3_SM_P, 0.059 m$^3$ m$^{-3}$ for L3_SM_P_E and 0.047 m$^3$ m$^{-3}$ for CLM. Both SMAP products and the CLM simulations can reasonably capture time variations in SM in the Maqu network (R = 0.64 for L3_SM_P, R = 0.65 for L3_SM_P_E and R = 0.59 for CLM).

In summary, both SMAP products are reasonably accurate relative to in-situ measurements of SM in the Naqu and Maqu networks.

**Figure 2.** Comparison of the Soil Moisture Active Passive (SMAP) L3 soil moisture product against in situ soil moisture data for the Naqu and Maqu network: (a) the time series for the Naqu network; (b) the scatterplot for the Naqu network; (c) the time series for the Maqu network; and (d) the scatterplot for the Maqu network. Blue triangles represent the standard SMAP soil moisture product, green circles represent the enhanced SMAP soil moisture product, orange squares represent the CLM simulations, and black solid lines represent the in-situ observations.

**Table 1.** Performance of the Soil Moisture Active Passive (SMAP) L3_SM_P and L3_SM_P_E products as well as Community Land Model (CLM) simulations for the Naqu and Maqu network, respectively.\(^1\)

<table>
<thead>
<tr>
<th>Networks</th>
<th>Product</th>
<th>ubRMSE</th>
<th>RMSE</th>
<th>BIAS</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naqu</td>
<td>L3_SM_P</td>
<td>0.059</td>
<td>0.060</td>
<td>0.007</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>L3_SM_P_E</td>
<td>0.055</td>
<td>0.055</td>
<td>0.005</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>CLM</td>
<td>0.037</td>
<td>0.043</td>
<td>−0.022</td>
<td>0.79</td>
</tr>
<tr>
<td>Maqu</td>
<td>L3_SM_P</td>
<td>0.058</td>
<td>0.133</td>
<td>0.120</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>L3_SM_P_E</td>
<td>0.059</td>
<td>0.127</td>
<td>0.113</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>CLM</td>
<td>0.047</td>
<td>0.057</td>
<td>0.030</td>
<td>0.58</td>
</tr>
</tbody>
</table>

\(^1\) RMSE, ubRMSE, and BIAS are the root mean square error (unit: m$^3$ m$^{-3}$), the unbiased RMSE (unit: m$^3$ m$^{-3}$), and the mean bias (unit: m$^3$ m$^{-3}$), respectively. R is the temporal correlation.
3.2. Comparison between the SMAP L3_SM_P and L3_SM_P_E Products

Figure 3a,b shows the spatial variations of summer (June–August) SM on the TP based on the SMAP L3_SM_P and L3_SM_P_E products for 2015. Both SMAP products capture the expected spatial patterns of SM on the TP; with SM decreasing from the southeast to the northwest. With L1C_TB_E as the input to the baseline SM retrieval algorithm, the 9-km L3_SM_P_E product shows additional fine-scale structure relative to the 36-km L3_SM_P product. This enhancement of spatial details is further illustrated in Figure 3c, which shows summer-mean SM based on L3_SM_P (black dashed line) and L3_SM_P_E (red solid line) along identical east–west transects across the TP (black lines in Figure 3a,b). The enhanced and standard SMAP SM products generally show similar variations along the longitudinal direction without obvious biases or unusual artifacts. However, fine-scale structure is evidently enhanced in the L3_SM_P_E product relative to the L3_SM_P product (e.g., near 80°E and 86°E–88°E).

Figure 3. Maps of summer-mean (June–August) soil moisture in 2015 (unit: m³ m⁻³) based on: (a) L3_SM_P; and (b) L3_SM_P_E; and (c) longitudinal distributions of soil moisture based on L3_SM_P (red solid line) and L3_SM_P_E (black dashed line) along the transect, marked by black horizontal lines in (a, b).
The enhanced 9 km L1C_TB_E product (the primary input for L2_SM_P_E) contains additional spatial information beyond that available in the standard 36-km L1C_TB product (the primary input for L2_SM_P). Consequently, the L2_SM_P_E product contains finer spatial detail beyond that available in L2_SM_P. The observed enhanced spatial detail identified in L2_SM_P_E is primarily contributed by the additional spatial information in L1C_TB_E [11].

3.3. Correlation between the SMAP L3_SM_P_E Product and CLM Simulations

The comparison with in-situ observations presented in Section 3.1 indicates that CLM simulations can capture much of the temporal variability in SM within the two in-situ measurement networks. Given the limited spatial coverage of in-situ SM observations on the TP, we therefore compare temporal variations in the SMAP L3_SM_P_E product against those produced by the CLM simulations. Figure 4 shows the spatial distribution of the temporal correlations between the CLM simulations and the SMAP L3_SM_P_E product. Only correlations that are statistically significant at the 95% confidence level are shown.

Figure 4. The distribution of the temporal correlation coefficient between SMAP L3_SM_P_E product and Community Land Model (CLM) simulations. Locations where correlations are not significant at the 95% confidence level are masked as white, while locations where soil moisture retrievals were not possible are masked as grey.

Temporal variations in the SMAP L3_SM_P_E product generally agree well with CLM simulations, with 97% of grid cells on the TP producing a significant correlation. The average correlation across these grid cells is 0.62. Correlations between the SMAP L3_SM_P_E product and the CLM simulations are larger in the central part of the TP than in the northwestern part of the TP, as reflected in a southeast-to-northwest decrease in the calculated correlations.

3.4. Spatial Variation of SMAP Product and CLM Simulations

Figure 5a,b shows the spatial distributions of average SM during summer (June–August) from the SMAP L3_SM_P_E product and CLM simulations in 2016. The spatial pattern of summertime SM over the TP is similar between two estimates, as evidence by a spatial pattern correlation of 0.95. However, the CLM simulations are wetter than the SMAP estimate for most (76.3%) regions, with mean bias of 0.032 m$^3$ m$^{-3}$. In addition, the SMAP L3_SM_P_E product shows an obvious wet strip in the far northwestern part of the TP (black squares in Figure 5a,b). Figure 5c,d shows the distributions of
lakes [51] overlaid onto the spatial distribution of SM in this region. The SMAP L3_SM_P_E product estimates wetter soils around these small lakes. The CLM simulations fail to capture this feature, mainly due to errors in the land cover distributions (the model lacks many of the lakes in this regions) and meteorological forcing data.

Figure 5. Spatial variations in summer soil moisture during 2016 based on: (a) the SMAP L3_SM_P_E product; and (b) the CLM simulation. (c, d) Spatial variations in soil moisture along with the distribution of lakes in the regions outline by black squares in (a, b) for the SMAP L3_SM_P_E product and the CLM simulation, respectively.

Figure 6 shows time series of the daily spatial pattern correlation of the SMAP L3_SM_P_E product relative to the CLM simulation for the unfrozen periods (1 May–30 September) of 2015 and 2016. The spatial patterns of SM for the SMAP L3_SM_P_E product and the CLM simulation agree well, as indicated by an average pattern correlation of 0.916. Most (75.5%) days have pattern correlations greater than 0.9, and the minimum pattern correlation is 0.836. These consistently large pattern correlations further affirm the high level of agreement between spatial variations in the SMAP L3_SM_P_E product and those produced by the CLM simulation.

Figure 6. Time series of the daily spatial pattern correlation between the SMAP L3_SM_P_E product and the CLM simulation for: 1 May–30 September 2015 (left); and 1 May–30 September 2016 (right).
4. Discussions

In this paper, SMAP enhanced SM product is evaluated using the ground observations from two networks as well as high-resolution model simulations on the TP, and the corresponding results are shown in Section 3. Regarding the evaluation in two chosen networks, the results are consistent with [11,26], that is, both SMAP enhanced and standard SM product show similar accuracies, as indicated by modest ubRMSE which ranges 0.055–0.059 m$^3$ m$^{-3}$. Temporal correlation is 0.88 for two SMAP products in Naqu networks. Maqu network show lower temporal correlations (0.64–0.65) which may be associated to better vegetation condition in Maqu network. In addition, differences between SMAP enhanced and standard SM products are also checked. SMAP enhanced SM product presents similar spatial variation with the standard one, but shows higher acuity in spatial details, which is consistent with other studies (e.g., [11,19]). Furthermore, SMAP enhanced SM product and CLM simulations agree well in both spatial pattern and temporal dynamics, however, temporal correlations between the CLM simulations and the SMAP enhanced products vary significantly across different regions of the TP. We therefore conduct a simple sensitivity analysis on the SMAP-CLM correlations with respect to soil and topographic characteristics, two important ancillary datasets used in the SMAP production algorithm. The documentation of the SMAP SM production algorithm and some other sensitivity analyses [9,52] identify the soil sand fraction as a key variable representing soil characteristics. Topographic variations are also considerable on the TP. To describe the characteristics of the topography variations, we use the Shuttle Radar Topography Mission (SRTM) 90 m Digital Elevation Database [53]. Topographic variations within each 0.1° pixel are represented as the SD of the more finely resolved (90-m) SRTM elevations. Figure 7a,b shows spatial maps of each characteristic (sand fraction and the sub-grid SD of elevation). Figure 7c,d shows the results of the sensitivity analysis, with error bars on the mean correlation indicating the SD of correlations calculated within each bin.

![Figure 7](image.png)

**Figure 7.** Spatial distributions (top row) of: (a) soil sand fraction (%); and (b) the elevation variations (m) over the Tibetan Plateau. Average correlations between the SMAP L3_SM_P_E product and the CLM simulations across: (c) different soil sand fractions; and (d) different elevation variations. The area fractions represented by each bin are shown as gray bars and the average correlations are shown as red lines, with error bars representing standard deviation of correlations in each bin.
Mean correlations between the SMAP enhanced product and the CLM simulations show no obvious variation across the range of soil sand fractions. The maximum mean correlation is 0.65 (for sand fractions of 50–52%), and the minimum mean correlation is 0.57 (for sand fractions less than 35%). However, the SD of correlation is larger for low sand contents than for high sand contents. The average SD of correlations for regions with sand fractions greater than 46% are 0.12, while that for regions with sand fractions less than 46% is 0.18. The sub-grid variance of elevation also has little relationship with the mean correlation until the value reaches 300 m (mean correlations for bins below this level range from 0.58 to 0.63). The mean correlation starts to decrease for values larger than 300 m, reaching a minimum of 0.43 in grid cells with sub-grid elevation SDs larger than 500 m. The SD of the correlation also increases across this threshold, with an average SD of 0.25 for bins with elevation variance larger than 300 m against an average SD of 0.15 for bins with elevation variance less than this threshold. In essence, large sub-grid elevation variance appears to reduce the correlation between SMAP and CLM estimates. This may be caused in part by decreases in the accuracy of SMAP retrievals for such cells, or by a reduction in the reliability of the background meteorological fields used in the CMFD owing to the presence of complex terrain.

To further analyze the reason behind the smaller SMAP–CLM correlations in the northwestern TP relative to the southeastern TP, changes in correlations as increasing sub-grid elevation variance increases are examined over both regions. Figure 8a shows the distribution of CMA operational stations that produced observations used in constructing the CMFD. Almost all stations are in the southeastern TP, so that the TP can be divided into a low station-density region (Region 1 in Figure 8a) and a high station-density region (Region 2 in Figure 8a). Figure 8b shows a box-and-whisker plots of the correlation across different sub-grid elevation variance bins in the two different regions.

In general, Region 2 (the high station-density region) has a higher average correlation (0.64) than Region 1 (the low station-density region; 0.50). Differences in the average correlation between the two regions become increasingly pronounced as the sub-grid elevation variance increases, with differences as large as 0.2–0.3 when the elevation SD exceeds 300 m. In Region 2, the median correlation shows no obvious change as the elevation variance increases beyond a slight decrease when sub-grid elevation SD exceeds 500 m. Region 1, by contrast, shows significant decreases in the median correlation for the elevation SD larger than 200 m, decreasing from 0.57 (for cells with elevation SDs within 200–250 m) to 0.27 (for cells with elevation SDs greater than 500 m). A handful of negative correlations, indicating a profound lack of agreement between the temporal variations in the CLM simulation and those in the SMAP enhanced product, are found in the grid cells with the largest sub-grid elevations (SD ≥ 300 m) in Region 1. By contrast, correlations are almost uniformly positive in the high-density station region. The obvious decline of correlations in the low station-density region may be caused by a lack of suitable observational constraints in the CMFD forcing data, therefore increasing the likelihood of large errors in the CMFD dataset. Uncertainties associated with CLM simulations in barren or sparsely-vegetated locations (the main land cover types in the low station-density region) may also contribute.
Figure 8. (a) China Meteorological Administration (CMA) operational stations that contributed observations to the China Meteorological Forcing Dataset (CMFD) dataset on the T. The TP is divided into two regions with high (Region 2) and low (Region 1) station densities. (b) Box-and-whisker plots of correlations between the SMAP L3_SM_P_E product and CLM simulations across different sub-grid elevation variance bins for Region 1 (red) and Region 2 (black).

5. Conclusions

This study provides the first comprehensive assessment of SMAP 9-km enhanced Level 3 passive SM product over the Tibetan Plateau, by comparing it against ground observations in two SM and temperature monitoring networks, the SMAP 36-km Level 3 passive SM product and CLM simulations. Comparison of the standard and enhanced SMAP SM products shows that the enhanced product includes similar spatial variation to those in the standard one, but with a greater ability to produce fine details in the spatial distribution of SM. Compared to station-based measurements, both the standard and enhanced SMAP passive SM products and the CLM simulations can reliably reproduce temporal variations of SM in the two chosen monitoring networks, as indicated by small values of ubRMSE (0.055–0.059 m³ m⁻³) and high temporal correlation coefficients (0.64–0.88). On average, the two SMAP products show better correlations but slightly larger ubRMSE values relative to CLM simulations. Such errors may result from a variety of reasons, including scale mismatches between in-situ observations and gridded products and coarse sampling of landscape heterogeneity.
Spatiotemporal comparison of the enhanced SMAP SM product against the CLM simulations further confirms that the SMAP enhanced SM product can reproduce temporal variations in SM, with a significant positive correlation (at the 95% confidence level) covering 97% of the area of the TP with valid retrievals. Correlations between the CLM simulations and the SMAP enhanced SM product are influenced by the quality of the CMFD dataset used to drive the CLM simulation and by sub-grid variance in land surface elevation. Increases in sub-grid elevation variance tend to reduce correlations between the CLM simulations and the SMAP enhanced SM product, particularly in locations where observational constraints on CMFD are lacking. Larger within-pixel elevation variance may also degrade the accuracy of SMAP products, as well as the reliability of the numerical forecast models, whose simulations are used to generate the background state for CMFD. Further, CLM simulations may be flawed in the northwestern part of the TP due to the low quality of the input forcing data (due to the lack of meteorological observations) and/or the low quality of the base map. In such situations, the SMAP enhanced SM product could provide a potential means of constraining the model simulations, or even be directly assimilated into a land data assimilation system to improve the model's ability to simulate the state of the land surface.

The land surface remains unmonitored or poorly-monitored in many regions around the world, with reasons ranging from economic or sociopolitical challenges (as in some regions of southeast Asia and Africa) to complex topography and harsh natural environments (as in the TP and polar regions). Forcing data and/or raw data (such as land cover maps) covering these regions may be problematic, potentially leading to cause errors in land surface model (LSM) simulations. It is likewise very difficult to evaluate LSM performance in unmonitored regions. High-resolution products from remote sensing observations, such as the SMAP enhanced products evaluated in this paper, can be used to fill some of these gaps: to provide initial conditions, to constraint model parameters, and to validate and evaluate model performance [20,54,55]. Used in tandem with other remote sensing data (such as observations of leaf area index, land surface temperature, and/or land cover type), these data can be used to help identify and correct important sources of errors in LSMs, as illustrated for CLM in this work. In this way, LSM performance and, ultimately, the underlying structures of the LSM itself can be improved even in regions that lack dedicated monitoring sites.

Acknowledgments: This work was jointly supported by the National Basic Research Program of China (2015CB953703), the National Natural Science Foundation of China (91537210 & 91747101), and the National Key Research and Development Program of China (2017YFA0603703). The distribution map of lakes on the TP is provided by the Third Pole Environment Database (http://en.tpedatabase.cn/). The computation for this work is supported by Tsinghua National Laboratory for Information Science and Technology.

Author Contributions: All authors made great contributions to this study. Hui Lu conceived and designed this study. Chengwei Li, Hui Lu, Kun Yang, Menglei Han, Yingying Chen, Le Yu, Shiming Xu, Xiaomeng Huang and Wei Gong performed the experiments. Chengwei Li, Hui Lu and Jonathon S. Wright wrote the manuscript.

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

References


© 2018 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/).