Global Modeling and Assimilation Office

GMAO Office Note No. 12 (Version 1.0)

Soil Moisture Active Passive Mission L4_SM Data Product Assessment (Version 2 Validated Release)

Release Date: 04/29/2016

Global Modeling and Assimilation Office
Earth Sciences Division
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771

This page intentionally left blank.

Soil Moisture Active Passive Mission L4_SM Data Product Assessment (Version 2 Validated Release)

Rolf H. Reichle[1], Gabrielle J. M. De Lannoy[2], Qing Liu[1,3], Joseph V. Ardizzone[1,3], Fan Chen[4], Andreas Colliander[5], Austin Conaty[1,3], Wade Crow[4], Thomas Jackson[4], John Kimball[6], Randal D. Koster[1], and E. Brent Smith[1,3]

1 – NASA Goddard Space Flight Center, Greenbelt, MD, USA
2 – KU Leuven, Leuven, Belgium
3 – SSAI, Lanham, MD, USA
4 – USDA Agricultural Research Service, Beltsville, MD, USA
5 – NASA Jet Propulsion Laboratory, Pasadena, CA, USA
6 – The University of Montana, Missoula, MT, USA

Document maintained by Rolf Reichle (GMAO)

This document should be cited as:
Reichle, R. H., G. J. M. De Lannoy, Q. Liu, J. V. Ardizzone, F. Chen, A. Colliander, A. Conaty, W. Crow, T. Jackson, J. Kimball, R. D. Koster, and E. Brent Smith, 2016: Soil Moisture Active Passive Mission L4_SM Data Product Assessment (Version 2 Validated Release). GMAO Office Note No. 12 (Version 1.0), 55 pp, NASA Goddard Space Flight Center, Greenbelt, MD, USA. Available from http://gmao.gsfc.nasa.gov/pubs/office_notes.

Approved by:

____

Steven Pawson Date

Head, Global Modeling and Assimilation Office
Code 610.1, NASA GSFC

REVISION HISTORY

Revision	Date	Sections Changed	Reason for Change
Initial Version 1.0	4/21/2016	All	Initial Document

TABLE OF CONTENTS

EXECUTIVE SUMMARY

During the post-launch SMAP calibration and validation (Cal/Val) phase there are two objectives for each science data product team: 1) calibrate, verify, and improve the performance of the science algorithm, and 2) validate the accuracy of the science data product as specified in the science requirements and according to the Cal/Val schedule. This report provides an assessment of the SMAP Level 4 Surface and Root Zone Soil Moisture Passive (L4_SM) product specifically for the product's public Version 2 validated release scheduled for 29 April 2016.

The assessment of the Version 2 L4_SM data product includes comparisons of SMAP L4_SM soil moisture estimates with in situ soil moisture observations from core validation sites and sparse networks. The assessment further includes a global evaluation of the internal diagnostics from the ensemble-based data assimilation system that is used to generate the L4_SM product. This evaluation focuses on the statistics of the observation-minus-forecast (O-F) residuals and the analysis increments. Together, the core validation site comparisons and the statistics of the assimilation diagnostics are considered primary validation methodologies for the L4_SM product. Comparisons against in situ measurements from regional-scale sparse networks are considered a secondary validation methodology because such in situ measurements are subject to upscaling errors from the point-scale to the grid cell scale of the data product. Based on the limited set of core validation sites, the wide geographic range of the sparse network sites, and the global assessment of the assimilation diagnostics, the assessment presented here meets the criteria established by the Committee on Earth Observing Satellites for Stage 2 validation and supports the validated release of the data.

An analysis of the time average surface and root zone soil moisture shows that the global pattern of arid and humid regions are captured by the L4_SM estimates. Results from the core validation site comparisons indicate that "Version 2" of the L4_SM data product meets the self-imposed L4_SM accuracy requirement, which is formulated in terms of the ubRMSE: the RMSE after removal of the long-term mean difference. The overall ubRMSE of the 3-hourly L4_SM surface soil moisture at the 9 km scale is 0.035 m^3/m^3. The corresponding ubRMSE for L4_SM root zone soil moisture is 0.024 m^3/m^3. Both of these metrics are comfortably below the 0.04 m^3/m^3 requirement. The L4_SM estimates are an improvement over estimates from a model-only SMAP Nature Run version 4 (NRv4), which demonstrates the beneficial impact of the SMAP brightness temperature data. L4_SM surface soil moisture estimates are consistently more skillful than NRv4 estimates, although not by a statistically significant margin. The lack of statistical significance is not surprising given the limited data record available to date. Root zone soil moisture estimates from L4_SM and NRv4 have similar skill. Results from comparisons of the L4_SM product to in situ measurements from nearly 400 sparse network sites corroborate the core validation site results.

The instantaneous soil moisture and soil temperature analysis increments are within a reasonable range and result in spatially smooth soil moisture analyses. The O-F residuals exhibit only small biases on the order of 1-3 K between the (rescaled) SMAP brightness temperature observations and the L4_SM model forecast, which indicates that the assimilation system is largely unbiased. The spatially averaged time series standard deviation of the O-F residuals is 5.9 K, which reduces to 4.0 K for the observation-minus-analysis (O-A) residuals, reflecting the impact of the SMAP observations on the L4_SM system. Averaged globally, the time series standard deviation of the normalized O-F residuals is close to unity, which would suggest that the magnitude of the modeled errors approximately reflects that of the actual errors.

The assessment report also notes several limitations of the "Version 2" L4_SM data product and science algorithm calibration that will be addressed in future releases. Regionally, the time series standard deviation of the normalized O-F residuals deviates considerably from unity, which indicates that the L4_SM assimilation algorithm either over- or underestimates the actual errors that are present in the

system. Planned improvements include revised land model parameters, revised error parameters for the land model and the assimilated SMAP observations, and revised surface meteorological forcing data for the operational period and underlying climatological data. Moreover, a refined analysis of the impact of SMAP observations will be facilitated by the construction of additional variants of the model-only reference data. Nevertheless, the “Version 2” validated release of the L4_SM product is sufficiently mature and of adequate quality for distribution to and use by the larger science and application communities.

1 INTRODUCTION

The NASA Soil Moisture Active Passive (SMAP) mission provides global measurements of soil moisture from a 685-km, near-polar, sun-synchronous orbit. SMAP data are used to enhance understanding of processes that link the water, energy, and carbon cycles, and to extend the capabilities of weather and climate prediction models (Entekhabi et al. 2014).

The suite of SMAP data products includes the Level 4 Surface and Root Zone Soil Moisture (L4_SM) product, which provides deeper layer soil moisture estimates that are not available in the Level 2-3 products. The L4_SM product is based on the assimilation of SMAP brightness temperatures into the NASA Catchment land surface model (Koster et al. 2000) using a customized version of the Goddard Earth Observing System, version 5 (GEOS-5) land data assimilation system (Figure 1; Reichle et al. 2014a). This system propagates the surface information from the SMAP instrument data to the deeper soil. The latency of the L4_SM product is about 2.5 days and is driven by the availability of the gauge-based global precipitation product that is used to force the land surface model (Reichle et al. 2014b; Reichle and Liu 2014).

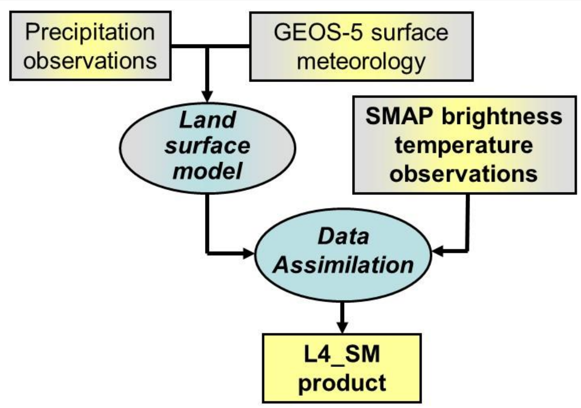

Figure 1. Schematic of the L4_SM algorithm.

The L4_SM product provides surface and root zone soil moisture (along with other geophysical fields) as 3-hourly, time-average fields on the global, cylindrical, 9 km Equal-Area Scalable Earth, version 2 (EASEv2) grid in the "geophysical" (or "gph") output Collection (Reichle et al. 2015a). Moreover, instantaneous soil moisture and soil temperature fields before and after the assimilation update are provided every three hours on the 9 km global EASEv2 grid in the "analysis update" (or "aup") output Collection, along with other assimilation diagnostics and error estimates. Time-invariant land model parameters, such as soil porosity, wilting point, and microwave radiative transfer parameters, are provided in the "land-model-constants" (or "lmc") Collection (Reichle et al. 2015a).

For geophysical data products that are based on the assimilation of satellite observations into numerical process models, validation is critical and must be based on quantitative estimates of uncertainty. Direct comparison with independent observations, including ground-based measurements, is a key part of the validation. This Assessment Report provides a detailed description of the status of the L4_SM data quality prior to the validated release of the L4_SM data product. The L4_SM validation process and data quality of the beta-release data product are discussed in a publicly available NASA GMAO Technical Memorandum (Reichle et al. 2015b).

2 SMAP CALIBRATION AND VALIDATION OBJECTIVES

During the post-launch SMAP calibration and validation (Cal/Val) phase each science product team pursues two objectives:

1. Calibrate, verify, and improve the performance of the science algorithm.
2. Validate the accuracy of the science data product as specified in the science requirements and according to the Cal/Val schedule.

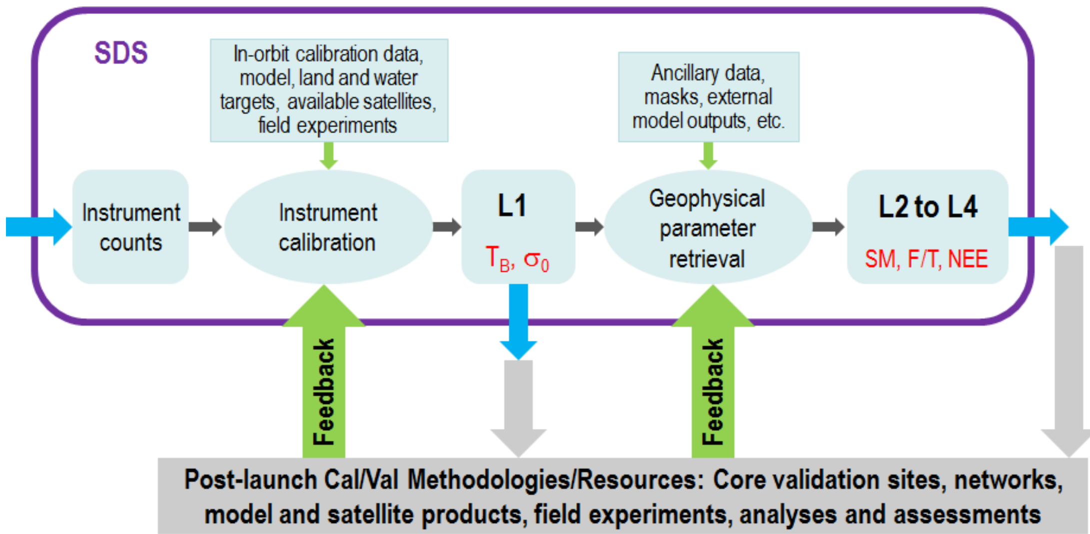

Figure 2. Overview of the SMAP calibration and validation process.

The overall SMAP Cal/Val process is illustrated in Figure 2. This Assessment Report describes how the L4_SM team addressed the above objectives prior to the validated release. The validation approach and procedures follow those described in the SMAP Science Data Cal/Val Plan (Jackson et al. 2014), the SMAP L2-L4 Data Products Cal/Val Plan (Colliander et al. 2014), and the Algorithm Theoretical Basis Document for the L4_SM data product (Reichle et al. 2014b).

SMAP established unified definitions to address the mission requirements. These are documented in the SMAP Handbook (Entekhabi et al. 2014), where calibration and validation are defined as follows:

- *Calibration*: The set of operations that establish, under specified conditions, the relationship between sets of values or quantities indicated by a measuring instrument or measuring system and the corresponding values realized by standards.
- *Validation:* The process of assessing by independent means the quality of the data products derived from the system outputs.

In order to insure the public's timely access to SMAP data, the mission is required to release validated data products within one year of the beginning of mission science operations. The objectives and maturity of the SMAP validated release products follow the guidance provided by the Committee on Earth Observation Satellites (CEOS) Working Group on Calibration and Validation (CEOS 2015):

- Stage 1: Product accuracy is assessed from a small (typically < 30) set of locations and time periods by comparison with in situ or other suitable reference data.

- Stage 2: Product accuracy is estimated over a significant set of locations and time periods by comparison with reference in situ or other suitable reference data. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and time periods. Results are published in the peer-reviewed literature.
- Stage 3: Uncertainties in the product and its associated structure are well quantified from comparison with reference in situ or other suitable reference data. Uncertainties are characterized in a statistically robust way over multiple locations and time periods representing global conditions. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and periods. Results are published in the peer-reviewed literature.
- Stage 4: Validation results for Stage 3 are systematically updated when new product versions are released and as the time-series expands.

For the Version 2 validated release the L4_SM team has completed Stage 1 and Stage 2. (Publication of the results in a peer-reviewed journal is pending.) The Cal/Val program will continue through the above stages over the SMAP mission life span. Incremental improvements are ongoing as more measurements become available from the SMAP observatory. Version 2 data will be replaced in the archive when upgraded product versions become available.

3 L4_SM CALIBRATION AND VALIDATION APPROACH

During the mission definition and development phase, the SMAP Science Team and Cal/Val Working Group identified the metrics and methodologies that would be used for L2-L4 product assessment. These metrics and methodologies were vetted in community Cal/Val Workshops and tested in SMAP pre-launch Cal/Val rehearsal campaigns. The following validation methodologies and their general roles in the SMAP Cal/Val process were identified:

- *Core Validation Sites*: Accurate estimates at matching scales for a limited set of conditions.
- *Sparse Networks*: One point in the grid cell for a wide range of conditions.
- *Satellite Products*: Estimates over a very wide range of conditions at matching scales.
- *Model Products*: Estimates over a very wide range of conditions at matching scales.
- *Field Campaigns*: Detailed estimates for a very limited set of conditions.

With regard to the CEOS Cal/Val stages (section 2), core validation sites address Stage 1, and satellite and model products are used for Stage 2 and beyond. Sparse networks fall between these two stages.

For the L4_SM data product, all of the above methodologies can contribute to product assessment and refinement, but there are differences in terms of the importance of each approach for the validation of the L4_SM product.

The assessment of the L4_SM data product includes comparisons of SMAP L4_SM soil moisture estimates with in situ soil moisture observations from core validation sites and sparse networks. The assessment further includes a global evaluation of the internal diagnostics from the ensemble-based data assimilation system that is used to generate the L4_SM product. This evaluation focuses on the statistics of the observation-minus-forecast (O-F) residuals and the analysis increments. Together, the core site comparisons and the statistics of the assimilation diagnostics are considered primary validation methodologies for the L4_SM product.

Comparisons against in situ measurements from regional-scale sparse networks are considered a secondary validation methodology because such in situ measurements are subject to upscaling errors from the point-scale to the grid cell scale of the data product.

Due to their very limited spatial and temporal extent, data from field campaigns play only a tertiary role in the validation of the L4_SM data product. Note, however, that field campaigns are instrumental tools in the provision of high-quality, automated observations from the core validation sites and thus play an important indirect role in the validation of the L4_SM data product.

4 L4_SM ACCURACY REQUIREMENT

There is no formal Level 1 mission requirement for the validation of the L4_SM product, but the L4_SM team self-imposed an accuracy requirement mirroring the one that applies to the L2_SM_AP product. Specifically, the L4_SM surface and root zone soil moisture estimates are required to meet the following criterion:

> **ubRMSE ≤ 0.04 m^3 m^{-3}** within the data masks specified in the *SMAP Level 2 Science Requirements* (that is, excluding regions of snow and ice, frozen ground, mountainous topography, open water, urban areas, and vegetation with water content greater than 5 kg m^{-2}),

where ubRMSE is the RMSE computed after removing long-term mean bias from the data (Entekhabi et al. 2010; Reichle et al. 2015b, their Appendix A). (The ubRMSE is also referred to as the standard deviation of the error.) This criterion applies to the L4_SM instantaneous surface and root zone soil moisture estimates at the 9 km grid-cell scale from the “aup” Collection. It is verified by comparing the L4_SM product to the grid-cell scale in situ measurements from the core validation sites (section 6.2).

L4_SM output fields other than instantaneous surface and root zone soil moisture are provided as research products (including surface meteorological forcing variables, soil temperature, evaporative fraction, net radiation, etc.) and will be evaluated against in situ observations to the extent possible given available resources.

As part of the validation process, additional metrics (including bias, RMSE, time series correlation coefficient R, and anomaly R values) are computed for the L4_SM output fields to the fullest extent possible. This includes computation of the metrics outside of the limited geographic area for which the 0.04 m^3 m^{-3} validation criterion is applied.

For the computation of the *anomaly* R metric, the seasonal cycle of the raw data (including the L4_SM product and the in situ measurements) is estimated, separately for each product and each location, by computing, for each day of the year, a climatological value of soil moisture. Anomaly time series are then computed by subtracting the mean seasonal cycle from the raw data. Lastly, the anomaly R metric is derived by computing the time series correlation coefficient of the anomaly time series. Because of the short (11-month) data record of the validated release, anomaly R metrics are not provided in this report.

The validation includes additional metrics that are based on the statistics of the observation-minus-forecast residuals and other data assimilation diagnostics (section 6.4). Reichle et al. (2015b) provide detailed definitions of all the validation metrics and confidence intervals used here.

5 L4_SM VERSION 2 VALIDATED RELEASE

5.1 Process and Criteria

Since the beginning of the science data flow, the team has been conducting frequent assessments of pre-beta and beta-release L4_SM data products and will continue to evaluate the products throughout the intensive Cal/Val phase and beyond. Frequent reviews of performance based upon core validation sites, sparse networks, and assimilation diagnostics were conducted for a period of 11+ months and captured a wide range of geophysical conditions. The assessment presented here is a summary of the latest status of this process.

The validation against in situ measurements includes metrics for a model-only "SMAP Nature Run," version 4 (NRv4). The NRv4 estimates are based on the same land surface model and forcing data as the L4_SM estimates, except that the NRv4 estimates do not benefit from the assimilation of the SMAP brightness temperature observations. Specifically, the NRv4 estimates are the result of a single-member, land model integration within the L4_SM system but without the ensemble perturbations and without the assimilation of the SMAP L1C_TB observations; any accuracy in the NRv4 estimates is thus derived from the imposed meteorological forcing and land model structure and parameter information. The NRv4 estimates are available for the period 1 January 2001 to present and also provide the model climatological information required by the L4_SM assimilation algorithm (Reichle et al. 2014b).

One key finding of this Assessment Report is that the L4_SM accuracy requirements (section 4) have been met, and that the Version 2 L4_SM product is sufficiently mature for the planned validated release on 29 April 2016.

5.2 Processing Options and Science ID Version

The L4_SM product version used to prepare this Assessment Report has Science Version ID **Tv2001**. The Tv2001 data were generated in February 2016 using the L4 Operations System in its "test" (T) configuration (ECS Version ID 777) and are not publicly available. The L4_SM algorithm slated for the validated release on 29 April 2016 and the forthcoming (publicly available) Version 2 L4_SM data product (ECS Version ID 2) are expected to have only very minor differences from Tv2001 algorithm and data product.

The L4_SM Tv2001 algorithm assimilated test data of the Version 3 SMAP L1C_TB brightness temperatures (CRID T12323) that were likewise generated in preparation of the validated release scheduled for 29 April 2016. Otherwise, and except for a few minor bug fixes, the L4_SM Tv2001 algorithm and its ancillary data are the same as those of the L4_SM Version 1 system that was used to generate the public beta-release data product (Science Version IDs Vb1010 for 31 March 2015, 0z-26 October 2015, 0z and Vb1012 thereafter).

In anticipation of the planned L4_SM validated release on 29 April 2016, the L4_SM team defined the assessment period for this report as **1 April 2015, 0z to 1 March 2016, 0z**. The start date matches the first full day when the radiometer was operating under reasonably stable conditions following instrument start-up operations. The end date was selected to allow sufficient time for analysis and preparation of this Assessment Report as well as other documents (such as the NSIDC User Guide) required for the validated release.

Like the beta-release Version 1, the validated release Version 2 of the L4_SM algorithm ingests only the SMAP L1C_TB radiometer brightness temperatures, contrary to the originally planned use of downscaled brightness temperatures from the L2_SM_AP product and landscape freeze-thaw state retrievals from the L2_SM_A product. The latter two products are based on radar observations and are only available for the period from 13 April to 7 July 2015 because of the failure of the SMAP radar instrument. Neither of these two radar-based products was sufficiently mature in early 2016 to allow for calibration of the L4_SM algorithm with those inputs in time for the Version 2 L4_SM validated release. Therefore, the decision to use only radiometer (L1C_TB) inputs for the validated release was made to ensure homogeneity in the longer-term Version 2 L4_SM data record.

6 L4_SM DATA PRODUCT ASSESSMENT

This section provides a detailed assessment of the L4_SM validated-release data product. First, global patterns, features, and noteworthy events are discussed (section 6.1). Next, we present comparisons and metrics versus in situ measurements from core validation sites (section 6.2) and sparse networks (section 6.3). Thereafter, we evaluate the assimilation diagnostics (section 6.4), which includes a discussion of the observation-minus-forecast residuals, the increments, and the data product uncertainty estimates. Finally, we present a summary of the validation findings (section 6.5).

6.1 Global Patterns and Features

Figure 3 shows global maps of time-averaged L4_SM surface and root zone soil moisture for the validation period (1 April 2015, 0z to 1 March 2016, 0z). The global patterns are as expected – arid regions such as the southwestern US, the Sahara desert, the Arabian Peninsula, the Middle East, southern Africa, and central Australia exhibit generally dry surface and root zone soil moisture conditions, whereas the tropics (Amazon, central Africa, and Indonesia) and high-latitude regions show wetter conditions. One notable exception is that a portion of the Democratic Republic of Congo and adjacent areas appear unexpectedly dry. This is because over Africa, the validated-release version of the L4_SM algorithm uses precipitation forcing directly from the GEOS-5 Forward Processing (FP) system, which has a known dry bias in central Africa.

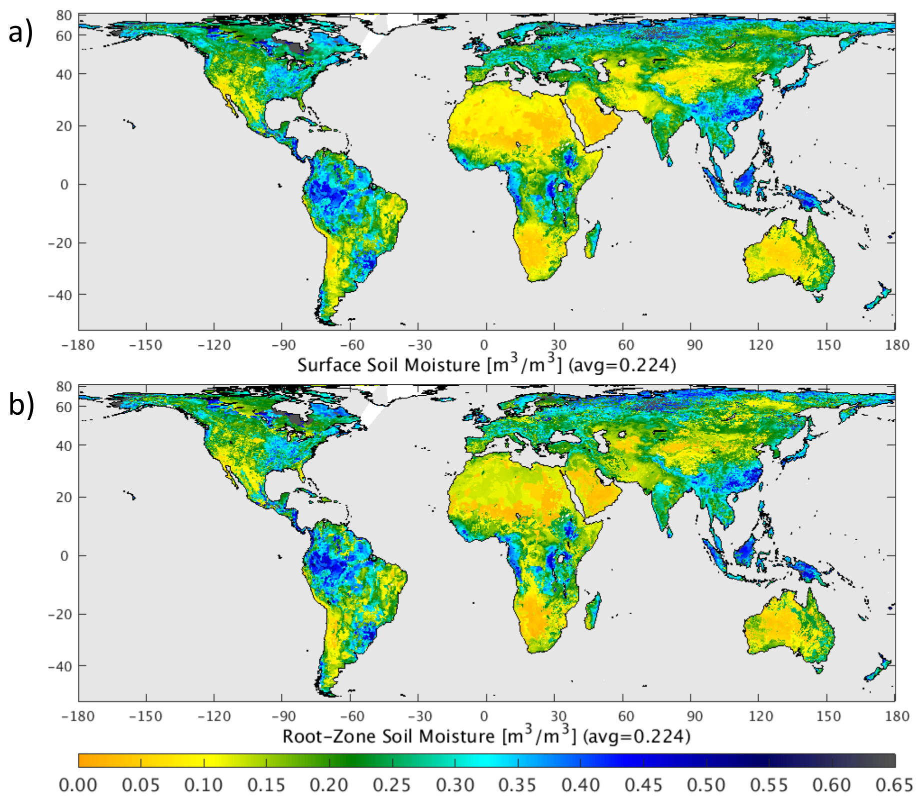

Figure 3. Time-average of L4_SM (a) surface and (b) root zone soil moisture for 1 April 2015, 0z, to 1 March 2016, 0z.

Generally, the global patterns of absolute soil moisture values are dominated by soil parameters and climatological factors. The influence of soil texture is noticeable in the coarse-scale patterns in the Sahara desert, where little is known about the spatial distribution of mineral soil fractions. Areas with peat soil include, for example, the region along the southern edge of Hudson Bay and portions of Alaska. In the land model, the soils in this region are assigned a high porosity value and show persistently wetter conditions than seen in other areas.

The L4_SM product also includes a large number of output fields that are not subject to formal validation requirements. Such "research" output includes the surface meteorological forcing fields, land surface fluxes, soil temperature and snow conditions, runoff, and error estimates (derived from the ensemble).

Global maps of instantaneous L4_SM output fields are very similar between the Version 2 (validated release) and Version 1 (beta release) products and are therefore not repeated here. See (Reichle et al. 2015b; their section 6.1) for more discussion of global patterns and features.

6.2 Core Validation Sites

6.2.1 Overview

This section addresses validation using SMAP core validation sites, which provide in situ measurements of soil moisture conditions at the scale of 9 km and 36 km grid cells. Details about the processing of the data and the validation methodology can be found in (Reichle et al. 2015b; their section 6.2.1). The status of the core validation sites is reviewed periodically. The current set of sites that provide data for the L4_SM validated release assessment are listed in Table 1, along with the details of the 9 km and 36 km reference pixels that are used. The table shows that the present L4_SM validation is based on a total of 33 reference pixels from 13 different core validation sites, representing a small increase relative to the 27 reference pixels from 12 different core validation sites used in the beta-release report (Reichle et al. 2015b). Surface soil moisture measurements are available for all 33 reference pixels, which include 11 reference pixels at the 36 km scale from 11 different sites and 22 reference pixels at the 9 km scale from 13 different sites. For root zone soil moisture, measurements are available for only 17 reference pixels from 7 different core sites, including 6 reference pixels at the 36 km scale from 6 different sites and 10 reference pixels at the 9 km scale from just 5 different sites. The 9 km reference pixels for root zone soil moisture belong to the core validation sites of Little Washita (Oklahoma), Fort Cobb (Oklahoma), South Fork (Iowa), Kenaston (Saskatchewan), and TxSON (Texas). This very limited set obviously lacks the diversity to be representative of global conditions.

Table 1 also lists the depths of the deepest sensors that contribute to the in situ root zone soil moisture measurements. The measurements from the individual sensors are vertically averaged with weights that are proportional to the spacing of the depth of the sensors within the 0-100 cm layer depth corresponding to the L4_SM estimates. At all reference pixels except Little River and Yanco, the deepest sensors are at 45 cm or 50 cm depth. At Little River and Yanco, the deepest sensors are at 30 cm and 75 cm, respectively, with Yanco's second-deepest sensors being installed at 45 cm depth. In all cases, the deepest sensors are therefore weighted most strongly in the computation of the vertical average. To compute the vertically averaged root zone soil moisture at a given time from a given sensor profile, all sensors within the profile must provide measurements that pass the automated quality control.

Across the reference pixels listed in Table 1, the average number of individual sensors that contribute to a given 36 km reference pixel ranges between 12.6 and 32.7, with a mean value of 20.6. At the 9 km scale, 8 of the 22 reference pixels are based on just 4 individual sensor profiles, while the rest of the 9 km reference pixels consist of about 10 sensor profiles each. The relative sampling density is therefore considerably lower for the 9 km reference pixels. For most reference pixels, individual sensor profiles tend to drop out temporarily. This leads to undesirable discontinuities in the reference pixel average soil moisture. To mitigate this effect, a minimum of 8 individual sensor profiles were required (after quality control) to compute the reference pixel average, provided at least 8 sensor profiles were in the ground. For the 8 reference pixels that are based on just 4 sensor profiles, data from all 4 sensors (after quality control) was required to compute the reference pixel average. Similarly, for the reference pixel (#16010913) with just 7 sensor profiles, data from all 7 were required to compute the average.

Table 2 lists the various skill metrics for all reference pixels, both for the L4_SM product and the Nature Run v4 (NRv4; section 5.1). The table is primarily provided for reference. A detailed discussion of the skill at select core validation sites is given in sections 6.2.2-6.2.5. The results for individual reference pixels reveal many features that support the quality of the L4_SM data product and indicate potential avenues for improvement. This is followed in section 6.2.6 by a discussion of the summary metrics obtained from averaging across the skill computed at all reference pixels, along with general conclusions from the core site validation.

Table 1. Core validation sites and reference pixels for L4_SM validation. 36 km reference pixels are shown with darker gray shading.

Site Name	Country	Climate Regime	Land Cover	Reference Pixel: ID	Latitude [degree]	Longitude [degree]	Horizontal Scale [km]	Depth of Deepest Sensor [m]	Number of Sensor Profiles: Min	Mean	Max
REMEDHUS	Spain	Temperate	Croplands	03013602	41.28	-5.41	36	0.05	8	14.7	17
				03010903	41.42	-5.37	9	0.05	4	4.0	4
				03010908	41.32	-5.27	9	0.05	4	4.0	4
Yanco	Australia (New South Wales)	Arid	Cropland / natural mosaic	07013601	-34.85	146.17	36	0.75	9	26.0	28
				07010902	-34.72	146.13	9	0.05	8	10.8	11
				07010916	-34.98	146.31	9	0.05	8	10.6	11
Carman	Canada (Manitoba)	Cold	Croplands	09013601	49.61	-97.94	36	0.05	8	19.3	20
				09010906	49.67	-97.98	9	0.05	8	10.7	11
Walnut Gulch	USA (Arizona)	Arid	Shrub open	16013603	31.68	-110.04	36	0.05	8	20.5	24
				16010906	31.72	-110.09	9	0.05	8	8.7	9
				16010907	31.72	-109.99	9	0.05	8	10.4	11
				16010913	31.83	-110.90	9	0.05	7	7.0	7
Little Washita	USA (Oklahoma)	Temperate	Grasslands	16023602	34.88	-98.09	36	0.45	8	16.1	18
				16020907	34.92	-98.04	9	0.45	4	4.0	4
Fort Cobb	USA (Oklahoma)	Temperate	Grasslands	16033602	35.42	-98.62	36	0.45	8	12.6	13
				16030911	35.38	-98.57	9	0.45	4	4.0	4
				16030916	35.29	-98.48	9	0.45	4	4.0	4
Little River	USA (Georgia)	Temperate	Cropland / natural mosaic	16043602	31.60	-83.59	36	0.30	12	19.6	21
				16040901	31.72	-83.73	9	0.05	8	8.0	8
St Josephs	USA (Indiana)	Temperate	Croplands	16060907	41.45	-84.97	9	0.05	8	8.4	9
South Fork	USA (Iowa)	Cold	Croplands	16073602	42.47	-93.39	36	0.50	8	18.9	20
				16070909	42.42	-93.53	9	0.50	4	4.0	4
				16070910	42.42	-93.44	9	0.50	4	4.0	4
				16070911	42.42	-93.35	9	0.50	4	4.0	4
Tonzi Ranch	USA (California)	Temperate	Savannas woody	25013601	36.47	-121.00	36	0.05	8	19.9	26
				25010911	38.43	-120.95	9	0.05	11	24.8	43
Kenaston	Canada (Saskatchewan)	Cold	Croplands	27013601	51.45	-106.46	36	0.50	8	26.2	28
				27010910	51.39	-106.51	9	0.50	8	8.0	8
				27010911	51.39	-106.42	9	0.50	8	13.6	14
Valencia	Spain	Cold	Savannas woody	41010906	39.57	-1.26	9	0.05	8	8.0	8
TxSON	USA (Texas)	Temperate	Grasslands	48013601	30.31	-98.78	36	0.50	18	32.7	36
				48010902	30.43	-98.82	9	0.50	8	10.8	14
				48010911	30.27	-98.73	9	0.50	8	14.2	15

Table 2. Metrics at individual reference pixels. 36 km reference pixels are shown with gray shading. L4_SM metrics are shown in bold. Confidence intervals are corrected for autocorrelation in the data (Reichle et al. 2015b).

	Reference Pixel		Surface Soil Moisture									Root Zone Soil Moisture								
			ubRMSE [m^3m^{-3}]			Bias [m^3m^{-3}]			R [-]			ubRMSE [m^3m^{-3}]			Bias [m^3m^{-3}]			R [-]		
Site Name	**ID**	**Horiz. Scale [km]**	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval
REMEDHUS	03013602	36	0.027	**0.025**	0.006	0.061	**0.064**	0.008	0.78	**0.82**	0.08	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	03010903	9	0.021	**0.019**	0.007	0.141	**0.156**	0.009	0.68	**0.69**	0.12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	03010908	9	0.035	**0.035**	0.008	0.005	**0.010**	0.010	0.77	**0.77**	0.09	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Yanco	07013601	36	0.054	**0.033**	0.022	0.019	**0.036**	0.027	0.82	**0.92**	0.10	0.018	**0.023**	0.021	-0.100	**-0.076**	0.022	0.89	**0.95**	0.52
	07010902	9	0.073	**0.050**	0.028	0.001	**0.020**	0.035	0.76	**0.88**	0.12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	07010916	9	0.050	**0.038**	0.021	0.048	**0.073**	0.027	0.78	**0.87**	0.12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Carman	09013601	36	0.024	**0.036**	0.004	-0.021	**-0.006**	0.006	0.61	**0.27**	0.14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	09010906	9	0.023	**0.032**	0.005	0.048	**0.072**	0.007	0.63	**0.35**	0.16	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Walnut Gulch	16013603	36	0.022	**0.023**	0.003	0.056	**0.060**	0.004	0.71	**0.78**	0.08	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	16010906	9	0.029	**0.033**	0.004	0.036	**0.044**	0.006	0.60	**0.60**	0.12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	16010907	9	0.026	**0.030**	0.004	0.045	**0.053**	0.005	0.62	**0.63**	0.11	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	16010913	9	0.033	**0.033**	0.006	0.077	**0.077**	0.009	0.54	**0.61**	0.16	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Little Washita	16023602	36	0.036	**0.030**	0.007	-0.005	**-0.015**	0.010	0.77	**0.85**	0.07	0.030	**0.023**	0.011	-0.043	**-0.047**	0.014	0.85	**0.90**	0.19
	16020907	9	0.036	**0.031**	0.008	-0.017	**-0.027**	0.011	0.75	**0.85**	0.09	0.025	**0.020**	0.010	-0.037	**-0.037**	0.013	0.91	**0.91**	0.14
Fort Cobb	16033602	36	0.037	**0.032**	0.007	0.029	**0.023**	0.009	0.77	**0.83**	0.07	0.022	**0.024**	0.009	0.025	**0.023**	0.012	0.88	**0.85**	0.20
	16030911	9	0.043	**0.034**	0.010	0.033	**0.029**	0.013	0.75	**0.84**	0.10	0.026	**0.027**	0.010	0.029	**0.031**	0.013	0.80	**0.86**	0.23
	16030916	9	0.030	**0.028**	0.005	0.003	**-0.006**	0.008	0.78	**0.82**	0.06	0.017	**0.019**	0.006	-0.023	**-0.024**	0.008	0.92	**0.87**	0.13
Little River	16043602	36	0.042	**0.032**	0.005	0.093	**0.085**	0.007	0.63	**0.73**	0.13	0.031	**0.022**	0.006	0.063	**0.054**	0.008	0.79	**0.85**	0.14
	16040901	9	0.043	**0.034**	0.006	0.119	**0.109**	0.009	0.66	**0.69**	0.15	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
St Josephs	16060907	9	0.042	**0.048**	0.014	0.135	**0.106**	0.018	0.55	**0.62**	0.29	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
South Fork	16073602	36	0.043	**0.039**	0.010	0.056	**0.044**	0.014	0.66	**0.71**	0.09	0.019	**0.024**	0.008	0.004	**-0.008**	0.010	0.72	**0.61**	0.31
	16070909	9	0.047	**0.040**	0.009	-0.005	**-0.013**	0.012	0.54	**0.71**	0.09	0.026	**0.025**	0.007	-0.071	**-0.078**	0.009	0.40	**0.59**	0.31
	16070910	9	0.043	**0.038**	0.009	0.046	**0.040**	0.013	0.56	**0.65**	0.10	0.020	**0.022**	0.008	0.031	**0.026**	0.011	0.55	**0.62**	0.35
	16070911	9	0.055	**0.052**	0.011	0.030	**0.021**	0.015	0.63	**0.63**	0.09	0.022	**0.026**	0.006	-0.006	**-0.015**	0.008	0.64	**0.54**	0.28
Tonzi Ranch	25013601	36	0.031	**0.028**	0.019	0.021	**0.035**	0.023	0.93	**0.92**	0.12	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	25010911	9	0.034	**0.031**	0.013	0.026	**0.034**	0.016	0.90	**0.91**	0.10	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Kenaston	27013601	36	0.032	**0.028**	0.005	0.008	**0.011**	0.007	0.59	**0.71**	0.10	0.016	**0.021**	0.009	-0.049	**-0.045**	0.011	0.74	**0.61**	0.43
	27010910	9	0.021	**0.031**	0.006	0.018	**0.020**	0.008	0.61	**0.45**	0.17	0.007	**0.021**	0.004	-0.069	**-0.067**	0.006	0.81	**0.70**	0.25
	27010911	9	0.035	**0.036**	0.010	-0.019	**-0.016**	0.013	0.55	**0.53**	0.15	0.014	**0.020**	0.006	-0.075	**-0.072**	0.008	0.70	**0.66**	0.35
Valencia	41010906	9	0.024	**0.024**	0.007	0.093	**0.086**	0.009	0.50	**0.52**	0.23	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
TxSON	48013601	36	0.042	**0.032**	0.012	0.088	**0.090**	0.016	0.84	**0.92**	0.12	0.036	**0.033**	0.044	0.048	**0.052**	0.043	0.95	**0.90**	0.63
	48010902	9	0.042	**0.038**	0.010	0.131	**0.145**	0.013	0.74	**0.83**	0.14	0.035	**0.030**	0.024	0.076	**0.091**	0.027	0.85	**0.83**	0.51
	48010911	9	0.054	**0.041**	0.013	0.126	**0.124**	0.017	0.79	**0.89**	0.14	0.029	**0.031**	0.022	0.086	**0.084**	0.025	0.92	**0.89**	0.44

6.2.2 Little Washita (Oklahoma)

The Little Washita watershed in Oklahoma has been utilized for many validation studies of microwave soil moisture retrievals. Several in situ measurement campaigns addressed both in situ sensor calibration and upscaling. Therefore, confidence in the quality of the in situ estimates is very high for this site, and performance at this site is considered to be an important factor in the L4_SM algorithm assessment.

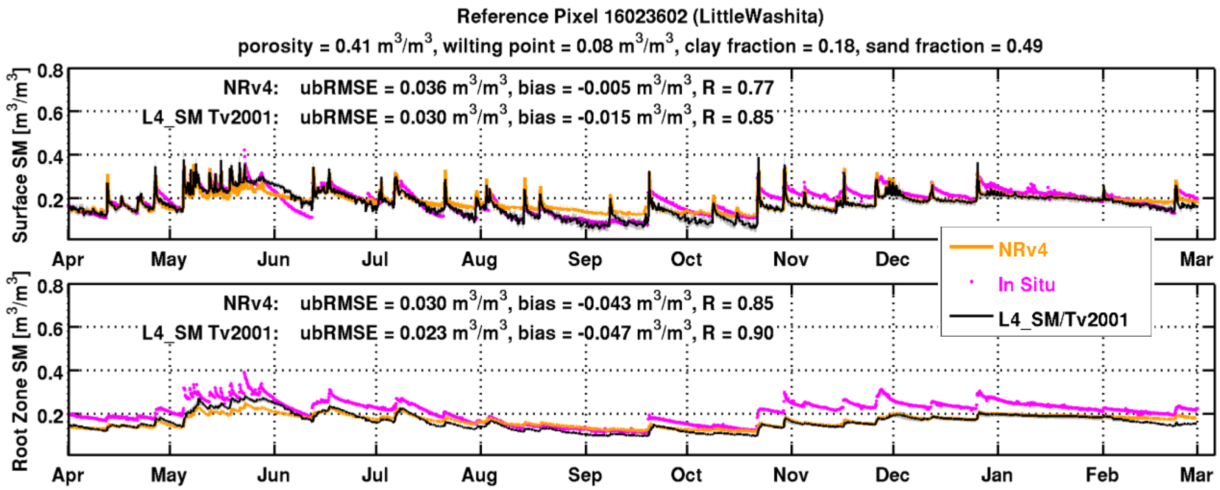

Figure 4. (Top) Soil moisture from (black solid line) L4_SM Tv2001, (orange solid line) NRv4, and (magenta dots) in situ measurements at the 36 km Little Washita reference pixel #16023602. The grey shading around the black solid line indicates the ensemble-based uncertainty estimates for surface soil moisture that are provided with the L4_SM product. (Bottom) Same as top panel but for root zone soil moisture.

There are long gaps in the in situ measurements for the 9 km reference pixel for Little Washita (#16020907), but the results at the 9 km reference pixel are similar to those at the 36 km reference pixel at that site (Table 2). Figure 4 therefore shows the L4_SM, NRv4, and in situ time series for the 36 km reference pixel (#16023602). Soil moisture varies considerably during the validation period, owing to the exceptionally wet conditions during May, which were preceded by relatively dry conditions in April and followed by a 3-month general drying trend. The L4_SM and NRv4 estimates clearly capture the overall variability, as well as the timing of the major rainstorms. However, neither the NRv4 nor the L4_SM estimates are fully capturing the wet conditions starting in late October and lasting through the winter. Nevertheless, the time series correlation coefficients are therefore very high, with R values of 0.85 for L4_SM surface soil moisture and 0.90 for L4_SM root zone soil moisture, which is an improvement over the already high values of 0.77 and 0.85 for NRv4 surface and root zone soil moisture, respectively.

The improvement is also reflected in the ubRMSE metric, which decreases from 0.036 m^3m^{-3} for NRv4 surface soil moisture to 0.030 m^3m^{-3} for L4_SM, and from 0.030 m^3m^{-3} for NRv4 root zone soil moisture to 0.023 m^3m^{-3} for L4_SM. The improvements are mostly due to the increased dynamic range and the generally faster dry-downs of the L4_SM estimates that result from the assimilation of the SMAP observations and better match the in situ measurements. Bias values are very low for surface soil moisture (around 0.01 m^3m^{-3} for L4_SM and NRv4). Root zone soil moisture, however, is generally more biased, with a higher value of -0.047 m^3m^{-3} for L4_SM than the -0.043 m^3m^{-3} bias for NRv4.

6.2.3 TxSON (Texas)

While Little Washita is one of the oldest sites, TxSON (Texas) is one of the newest. The site was designed specifically to satisfy the validation of SMAP soil moisture products at the 3 km, 9 km, and 36 km spatial scales. Figure 5 shows the results for one of the 9 km TxSON reference pixels. The results are generally similar at the two other (9 km and 36 km) TxSON reference pixels. The figure shows that the precipitation pattern at TxSON was similar to that in Little Washita (section 6.2.2), with a dry April followed by an exceptionally wet May, an extended drydown period through the summer, and exceptionally wet conditions again starting in mid-October and lasting through 2015.

As with the Little Washita reference pixels, the time series correlation coefficients for the TxSON reference pixels are very high, with R values of 0.83 for L4_SM surface and root zone soil moisture. The assimilation again results in faster drydowns and a greater dynamic range compared with NRv4, and therefore a generally better agreement with the in situ measurements. For the reference pixel shown in Figure 5, the improvements are reflected in the ubRMSE values for surface and root zone soil moisture as well as in the R values for surface soil moisture. The improvements manifest themselves somewhat differently at the other TxSON reference pixels (Table 2), but overall the assimilation of SMAP observations is clearly beneficial. Unlike for Little Washita, the TxSON bias values are very high, ranging from 0.08 m^3m^{-3} to 0.14 m^3m^{-3} depending on the specific variable, data set, and reference pixel (Table 2). The bias is related to the model soil parameters for this site, including the relatively high clay fraction of 0.53 and wilting point of 0.24 m^3m^{-3}, which reflect average values across the entire 0-100 cm root zone because vertical gradients in soil texture are not represented in the soil hydrological component of the model.

One notable feature is the rapid increase in the L4_SM uncertainty estimates for surface soil moisture once the surface and root zone soil moisture values drop below the model's wilting point of 0.24 m^3m^{-3} around 15 July 2015. At that point, the model's transpiration shuts down, and modeled root zone soil moisture remains at the wilting point until a significant rain event results in sufficient infiltration to raise root zone soil moisture again. While root zone soil moisture remains stagnant at the wilting point, bare soil evaporation still taps into the surface layer soil moisture, which is no longer replenished from below and becomes highly sensitive to the perturbations in the surface meteorological forcings and the soil moisture prognostic variables, which in turn results in a dramatically increased ensemble spread in surface soil moisture.

Figure 5 also illustrates residual issues with the processing of the in situ measurements. Surface soil moisture increases slightly on July 8 and August 7, which would suggest a minor rain event. However, the increase is not due to rainfall. An analysis of the measurements from the individual sensors reveals that around the dates in question, data from several of the sensors reach extremely dry conditions and are then flagged by the quality control, which results in an average that is based on only the sensors with the somewhat wetter measurements.

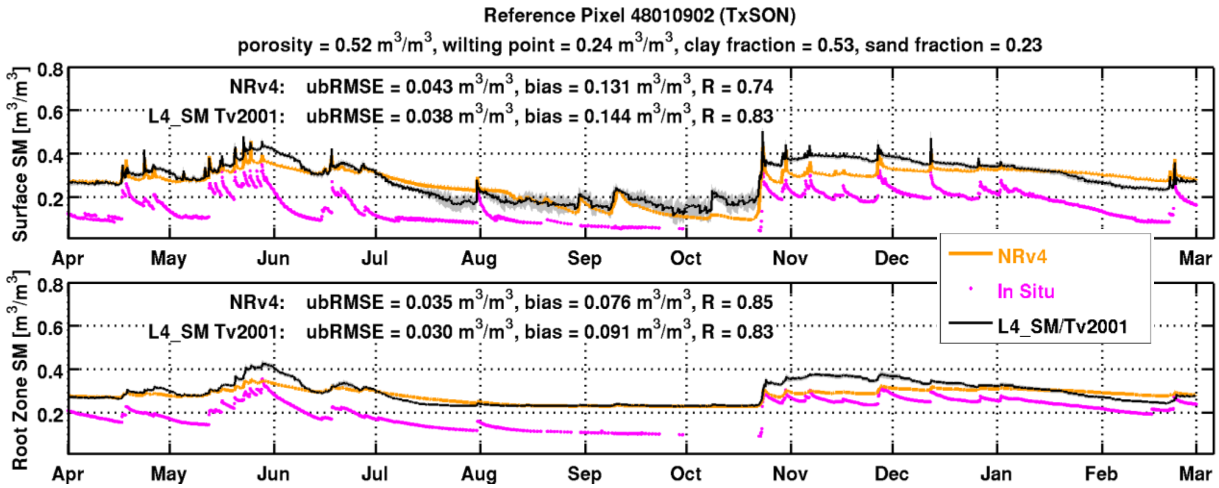

Figure 5. Same as Figure 4 but for the 9 km TxSON reference pixel #48010902.

6.2.4 Little River (Georgia)

Little River, Georgia, has been providing in situ soil moisture measurements since the launch of AMSR-E on Aqua in 2002 (Jackson et al. 2010). The site is unique in that it represents a humid agricultural environment. It also includes a substantial amount of tree cover, has very sandy soils, and is subject to irrigated agriculture. There are no in situ measurements of root zone soil moisture at the 9 km reference pixel for Little River, and the surface soil moisture results are similar for the 9 km and 36 km reference pixels. Figure 6 therefore shows the time series for the 36 km reference pixel (#16043602).

All time series reflect a drop from somewhat moister conditions in April during a long drydown in May, followed by somewhat drier conditions with frequent yet typically modest rain events during the rest of the summer and fall, followed by typical, wetter conditions during winter. The frequent wetting and drying events shown in the in situ measurements are reasonably captured by the L4_SM and NRv4 estimates, but the exact timing and magnitude of the storms and drydowns is less certain, with lower R values than for the Little Washita and TxSON reference pixels. Surface soil moisture has an R value of only 0.63 for NRv4, which improves to 0.73 for L4_SM with the assimilation of the SMAP observations. The correlation for root zone soil moisture is higher, with R values of 0.79 for NRv4 and 0.85 for L4_SM. The assimilation also improves the ubRMSE values of surface soil moisture estimates from 0.041 m^3m^{-3} for NRv4 to 0.032 m^3m^{-3} for L4_SM and of root zone soil moisture estimates from 0.031 m^3m^{-3} for NRv4 to 0.022 m^3m^{-3} for L4_SM. Bias values are relatively high at 0.09 m^3m^{-3} for surface soil moisture and 0.06 m^3m^{-3} for root zone soil moisture.

Figure 6 also reveals issues with the in situ measurements. Between May 17 and June 5, the reference pixel average root zone soil moisture shows somewhat erratic behavior. In this particular case, bad data from one sensor passed the automated quality control, and sensors also drop out repeatedly during the period in question.

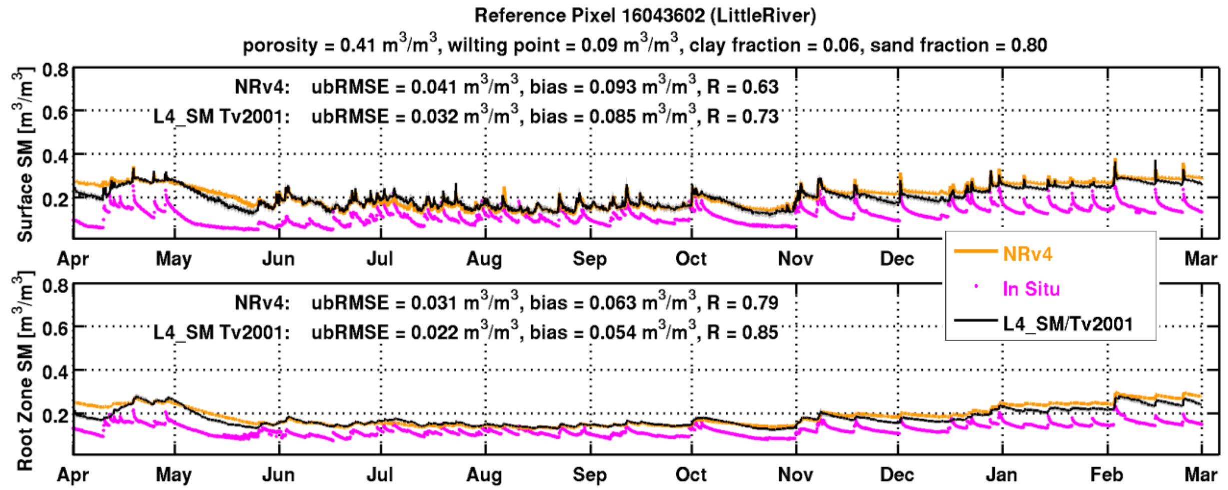

Figure 6. Same as Figure 4 but for the 36 km Little River reference pixel #16043602.

6.2.5 South Fork (Iowa)

South Fork, Iowa, is an agricultural region dominated by summer crops of corn and soybeans. Conditions in April 2015 were mostly bare soil or stubble, followed by intensive tillage that created large surface roughness. Such variations in surface roughness are difficult to capture in the microwave radiative transfer model parameters of the L4_SM algorithm. The roughness decreases again with subsequent soil treatments and rainfall, and becomes less of an issue as the growing season proceeds and crops cover the surface. By early July, corn would typically have high vegetation water content (around 3 kg m^{-2}) while that of soybeans would typically be much smaller (around 0.3 kg m^{-2}) (Jackson et al. 2004). It should also be noted that the agricultural fields are equipped with tiles to improve drainage, a process that is not captured in the global-scale Catchment land surface model of the L4_SM algorithm.

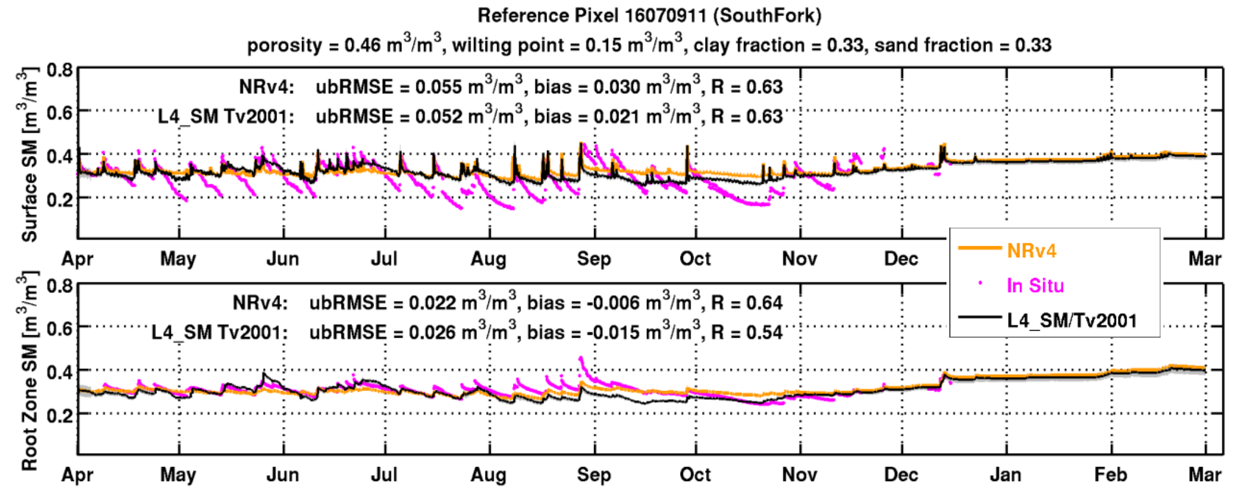

Figure 7. Same as Figure 4 but for the 9 km South Fork reference pixel #16070911.

Figure 7 shows soil moisture time series for the 9 km reference pixel #16070911, one of three 9 km reference pixels at South Fork, for which there is also a 36 km reference pixel (Table 2). Soil moisture conditions during the warm season are dominated by approximately weekly rain events with subsequent drydowns, except for a generally wetter period from mid-June into early July. The L4_SM surface soil moisture estimates capture this pattern reasonably well and present a small improvement over NRv4 in terms of ubRMSE values, which decrease from 0.055 m^3m^{-3} for NRv4 to 0.052 m^3m^{-3} for L4_SM. The surface soil moisture R value remains unchanged at 0.63 for NRv4 and L4_SM. Root zone metrics, however, are worse for L4_SM than for NRv4, both in terms of ubRMSE values, which increase from 0.022 m^3m^{-3} for NRv4 to 0.026 m^3m^{-3} for L4_SM, and in term of R values, which decrease from 0.64 for NRv4 to 0.54 for L4_SM.

Generally, the NRv4 and L4_SM estimates do not capture the larger dynamic range of the in situ observations, which may be a reflection of the tile drainage. Bias values are generally low and are around 0.02 m^3m^{-3} for L4_SM and NRv4 surface and root zone soil moisture. This is encouraging because an extensive study involving sensor calibration and additional point sampling was conducted that clearly demonstrated that the network represents the average soil moisture of the 0-5 cm soil layer of the SMAP grid cell (M. Cosh 2015, USDA Hydrology and Remote Sensing Laboratory, personal communication).

6.2.6 Summary Metrics

Table 3 lists the summary metrics for surface and root zone soil moisture. The summary metrics are provided separately for the 9 km and 36 km reference pixels and are obtained by averaging across the metrics from all individual reference pixels at the given scale (Table 2). The key findings for the summary metrics (Table 3) generally match those obtained for the sample reference pixels discussed above (sections 6.2.2-6.2.5). Perhaps the most important result is that the ubRMSE values for surface and root zone soil moisture for L4_SM as well as NRv4 and at both the 9 km and the 36 km scales all meet the accuracy requirement of 0.04 m^3m^{-3}.

Table 3. Metrics averaged across core validation site reference pixels. The product (NRv4 vs L4_SM) with the better skill score is indicated by green shading. L4_SM metrics are shown in bold. Confidence intervals are corrected for autocorrelation in the data (Reichle et al. 2015b).

	Horiz. Scale	Number of Reference Pixels	ubRMSE [m^3m^{-3}]			Bias [m^3m^{-3}]			R [-]		
			NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval
Surface Soil Moisture	9 km	22	0.038	**0.035**	0.010	0.051	**0.053**	0.013	0.67	**0.70**	0.13
	36 km	11	0.035	**0.031**	0.009	0.037	**0.039**	0.012	0.74	**0.77**	0.10
Root Zone Soil Moisture	9 km	10	0.022	**0.024**	0.010	-0.006	**-0.006**	0.013	0.75	**0.75**	0.30
	36 km	7	0.025	**0.024**	0.015	-0.007	**-0.007**	0.017	0.83	**0.81**	0.35

For a more in-depth analysis, we first compare the skill of the L4_SM and NRv4 estimates. The color-coding of the summary metrics in Table 3 indicates whether the L4_SM or NRv4 skill is higher. For the ubRMSE and R metrics and at the 9 km and the 36 km scales, the surface soil moisture skill of L4_SM exceeds that of NRv4, albeit not by a statistically significant margin. For example, at the 9 km scale the ubRMSE for L4_SM is 0.035 m^3m^{-3}, compared to 0.038 m^3m^{-3} for NRv4. The corresponding R values are 0.70 for L4_SM and 0.67 for NRv4. The bias metric is slightly worse for L4_SM than NRv4.

The summary metrics for root zone soil moisture show a more mixed picture. At the 9 km scale, the L4_SM ubRMSE (0.024 m^3m^{-3}) is slightly higher than that of NRv4 (0.022 m^3m^{-3}), and the R value for L4_SM is the same as that of NRv4 (0.75). The numbers are reversed for the 36 km scale, where the L4_SM ubRMSE is better but the L4_SM R value is worse than the corresponding NRv4 metrics.

A closer look at the ubRMSE metric for the individual reference pixels (Table 2) reveals that the root zone soil moisture skill meets the 0.04 m^3m^{-3} threshold at all reference pixels for both L4_SM and NRv4. Surface soil moisture estimates from NRv4 fail to meet the 0.04 m^3m^{-3} threshold at 14 of the 33 reference pixels, including those at Yanco (3 out of 3), Fort Cobb (1 out of 3), Little River (2 out of 2), St. Josephs (1 out of 1), South Fork (4 out of 4) and TxSON (3 out of 3). By contrast, L4_SM surface soil moisture estimates fail to meet the threshold at only 4 of 33 reference pixels, including 9 km pixels at Yanco, St. Josephs, South Fork, and TxSON. This result further illustrates the key role played by the assimilation of SMAP observations in improving the skill of the surface soil moisture estimates beyond the levels obtained strictly from the land surface model's integration of meteorological forcing.

Next, we compare the skill values at 9 km to those at 36 km. Here, the picture is also clearer for surface soil moisture. The L4_SM and NRv4 skill at 36 km is better for all three metrics than that at 9 km, which is consistent with the fact that the model forcing data and the assimilated SMAP brightness temperature observations are all at resolutions of about 30 km or greater. The information used to downscale the assimilated information stems from the land model parameters, which are at the finer, 9 km resolution. It is therefore not a surprise that the estimates at 36 km are more skillful than those at 9 km.

For root zone soil moisture, the results are again mixed. The R values for NRv4 and L4_SM at 9 km are lower than those at 36 km. However, the ubRMSE metric for NRv4 is better at 9 km than at 36 km. It should be noted, though, that the differences are not statistically significant, and that there are far fewer pixels available for evaluating root zone soil moisture than for surface soil moisture. As the time series become longer and core validation sites that are not automated (e.g., Tibet) are added, the results may become clearer.

Finally, we compare the skill of the surface estimates to that of the root zone estimates. Across all scales and metrics and for the L4_SM and NRv4 estimates, the skill of the root zone soil moisture estimates is always better than that of the surface estimates. This result makes sense because there is much more variability in surface soil moisture.

In summary, the results discussed here demonstrate that the validated-release L4_SM product is of sufficient maturity and quality for dissemination to the public.

6.3 Sparse Networks

6.3.1 Method and Overview

The locally dense networks of the core validation sites are complemented by regional to continental-scale sparse networks. The defining feature of the sparse networks is that there is usually just one sensor (or profile of sensors) located within a given 9 km (or 36 km) EASE v2 grid cell. Although sparse networks are not ideal for soil moisture validation for a variety of reasons, they offer in situ measurements in a larger variety of environments and provide data operationally with very short latency. See Reichle et al. (2015b) for further discussion of the advantages and limitations of using sparse networks in the L4_SM validation process.

This Assessment Report focuses on metrics obtained from a direct comparison of the L4_SM product to in situ measurements, that is, metrics derived without using Triple Co-location approaches that attempt to correct for errors in the in situ measurements (Chen et al. 2016). The values of the time series correlation metrics provided here are thus lower than those that would be obtained with the aid of Triple Co-location, and they are therefore conservative estimates of the true skill. Note also that the *relative* performance of the products under investigation does not depend on the use of Triple Co-location approaches.

The L4_SM assessment against sparse network data was conducted in two ways, using two different spatio-temporal masks. First, the skill of the L4_SM estimates was computed using all available in situ measurements (after quality control) at 3-hourly time steps, and this skill was then compared to that of the NRv4 estimates. Note that quality control generally excludes in situ measurements when the ground is frozen (see Reichle et al. 2015b, Appendix C). Instantaneous L4_SM data from the “aup” Collection and NRv4 data were taken directly from the standard 9 km EASEv2 grid cell that includes the sensor location (that is, the data product estimates are not interpolated bilinearly or otherwise to the precise location of the in situ sensor locations). Metrics were computed for surface and root zone soil moisture against in situ measurements from the SCAN, USCRN, OK Mesonet, and OZNet-Murrumbidgee networks (Table 4).

Second, the skill of the L4_SM data in reproducing available in situ measurements was compared to that of the surface soil moisture retrievals from the SMAP L2 Passive Soil Moisture (L2P) data product (CRID: T12400; “SCA-V” baseline) after cross-masking the L4_SM data to the times and locations for which L2P retrievals with a “good quality” summary flag were available. For this comparison, 3-hourly time-average L4_SM data from the “gph” Collection were first aggregated to the standard 36 km EASEv2 grid cells of the L2P product. Next, time-matched L4_SM and L2P data were taken directly from the standard 36 km EASEv2 grid cell that includes the sensor location. Metrics were computed for the SCAN, USCRN, OK Mesonet, GPS, COSMOS, SMOSMania, Pampas, and MAHASRI networks (Table 4).

Table 4. Overview of sparse networks used in L4_SM validation, with indication of the number of sites (N), data periods, and sensor depths used here.

Network	Area	Sensor or Sensing Depths [cm]	NRv4 Skill Comparison			L2P Skill Comparison	
			N		Period	N	Period
			Sur-face	Root zone		Sur-face	
SCAN	USA	5, 10, 20	123	108	1 Apr 2015 - 29 Feb 2016	94	1 Apr 2015 - 29 Feb 2016
USCRN	USA	5, 10, 20	111	85	1 Apr 2015 - 29 Feb 2016	56	1 Apr 2015 - 29 Feb 2016
OK Mesonet	Oklahoma	5, 25, 60	116	76	1 Apr 2015 - 29 Feb 2016	93	1 Apr 2015 - 29 Feb 2016
OZNet	Australia	4, 45	42	18	1 Apr 2015 - 29 Feb 2016	–	–
GPS	Western USA	5	–	–	–	50	1 Apr 2015 - 29 Feb 2016
COSMOS	Mostly USA	<50	–	–	–	33	1 Apr 2015 - 29 Feb 2016
SMOSMania	France	5	–	–	–	8	1 Apr 2015 - 3 Nov 2015
Pampas	Argentina	5	–	–	–	14	1 Apr 2015 - 29 Feb 2016
MAHASRI	Mongolia	3	–	–	–	13	1 Apr 2015 - 30 Aug 2015
All Networks			***392***	***287***		***361***	

Measurements used for L4_SM validation cover most of the contiguous United States (SCAN, USCRN, OK Mesonet, GPS, COSMOS), parts of the Murrumbidgee basin in Australia (OZNet), and regional areas in southern France (SMOSMania), Argentina (Pampas), and Mongolia (MAHASRI). The in situ measurements from the sparse network sites used in the NRv4 skill comparison were subjected to extensive automated and manual quality control procedures by the L4_SM team following (Liu et al. 2011), which removed spikes, temporal inhomogeneities, oscillations, and other artifacts that are commonly seen in these automated measurements. Table 4 also lists the number of sites with sufficient data after quality control. The in situ measurements used in the L2P skill comparison were subjected to automated quality control only, which is typically less effective.

For the NRv4 skill comparison, a total of 392 sites provided surface soil moisture measurements, and 287 provided root zone soil moisture measurements. Most of the sites are in the continental United States, including about 100 each in the USCRN and SCAN networks, and another 100 in just Oklahoma from the Mesonet. The OZNet network contributes 42 sites with surface soil moisture measurements, of which 18 sites also provide root zone measurements. For the L2P skill comparison, a total of 361 sites were used. There are fewer sites from SCAN, USCRN, and the OK Mesonet in the L2P than in the NRv4 skill comparison because the availability of L2P retrievals (unlike that of the L4_SM and NRv4 products) is limited by satellite overpass times and the requirement for surface conditions that support the soil moisture retrieval.

Table 4 also lists the sensor depths that were used to compute the in situ root zone soil moisture in the NRv4 skill comparison. As for the core validation sites, vertical averages for SCAN, USCRN, and OK Mesonet are weighted by the spacing of the sensor depths within the 0-100 cm layer corresponding to the L4_SM estimates, and the average is only computed if all sensors within a given profile provide measurements after quality control. For SCAN and USCRN sites, measurements at 50 cm (and occasionally 100 cm) depth are available, but these deeper layer measurements are not of the quality and quantity required for L4_SM validation and are therefore not used here. In future assessments, longer validation periods may facilitate the use of the measurements at 50 cm depth. For OZNet, in situ root zone soil moisture is given by the measurements at the 45 cm depth, that is, no vertical average is computed.

Because of the larger number of sparse network locations compared to the core validation site data, it is possible to examine the results stratified by general characteristics, including land cover and topographic complexity. One key distinction used in the NRv4 skill comparison is whether a site is within the mask for which the formal accuracy requirement applies (section 4). A site falls outside the mask if it is in an area with mountainous topography or dense vegetation, or if it is in an urban area. The delineation used here is based on the maximum climatological LAI, the land cover class, and the variance of the elevation within the 36 km EASEv2 grid cell (a measure of topographic complexity) that contains the site. These parameters are readily available in the L4_SM modeling system. Specifically, a site is within the mask if the maximum climatological LAI is less than 5, if the land cover is not forest, wetlands, or urban (that is, if the site has IGBP class 6-10, 12, or 14), and if the elevation variance around the site is less than 5,000 m^2 (that is, if the standard deviation is less than 71 m).

The L2P skill comparison breaks down the metrics based on vegetation water content (VWC) conditions at the time of the retrieval, where the VWC information is from the MODIS-based estimates provided in the L2P product. For this binning, the time series at each site is first split into portions for which VWC is within a certain range, and only this part of the time series is used to compute the skill metric for the site (provided the partial time series contains at least 4 or more data points). The average skill for a VWC bin is then computed by averaging the metrics computed from each site's partial time series. In the "All VWC" category, the metrics for each site are computed from the complete time series and then averaged across all sites (that is, the "All VWC" metrics are *not* the average of the metrics for the VWC bins).

6.3.2 NRv4 Skill Comparison

Figure 8 illustrates the ubRMSE values for the L4_SM estimates at the sparse network sites. The background shading in the figure also indicates whether a site is within the mask of the formal accuracy requirement (section 6.3.1). The resulting delineation (Figure 8) suggests, for example, that sites in the topographically complex western United States mountain areas and in the more densely vegetated portions of the eastern United States fall outside the mask, which is commensurate with expectation.

Overall, ubRMSE values range from 0.02 m^3m^{-3} to 0.07 m^3m^{-3}, with generally lower values for root zone soil moisture than for surface soil moisture (Figure 8). Errors are generally lowest in the dry and mountainous areas of the western United States, where the soil moisture variability is typically low, thus naturally limiting the ubRMSE values. The R values for the sparse network sites, shown in Figure 9, range from 0.3 to 0.9, with generally similar values for surface and root zone soil moisture. There is no obvious spatial pattern.

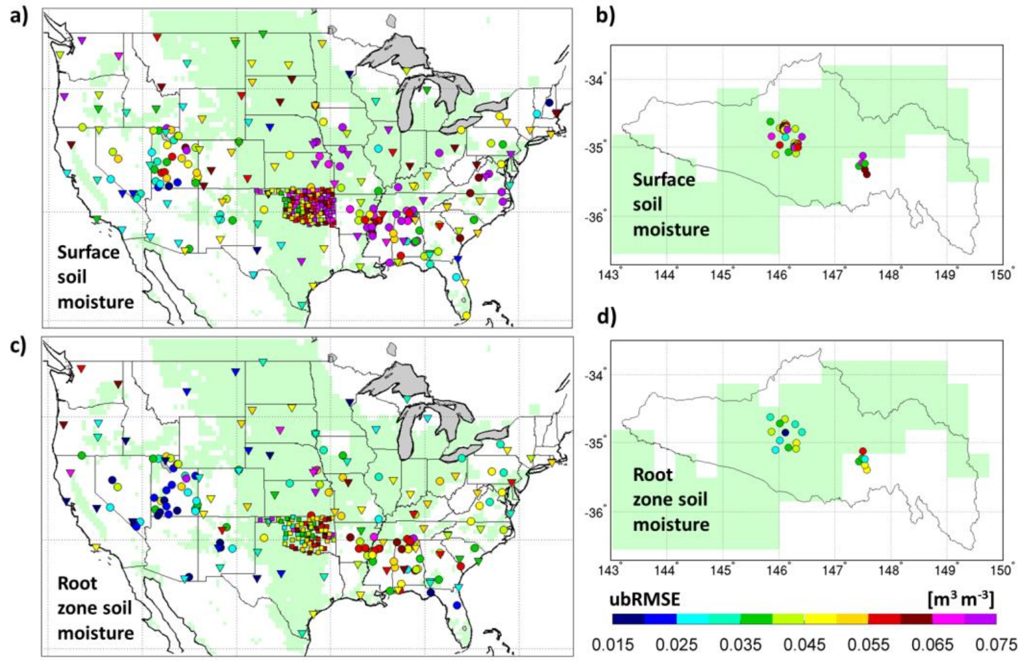

Figure 8. ubRMSE for L4_SM (a,b) surface and (c,d) root zone soil moisture. (a,c) The United States sites include (dots) SCAN, (inverted triangles) USCRN, and (squares) OK Mesonet. (b,d) The Australian sites are from OZNet. Green shading indicates areas with low or modest vegetation cover and topographic complexity that are within the mask of the SMAP accuracy requirement. See text for details.

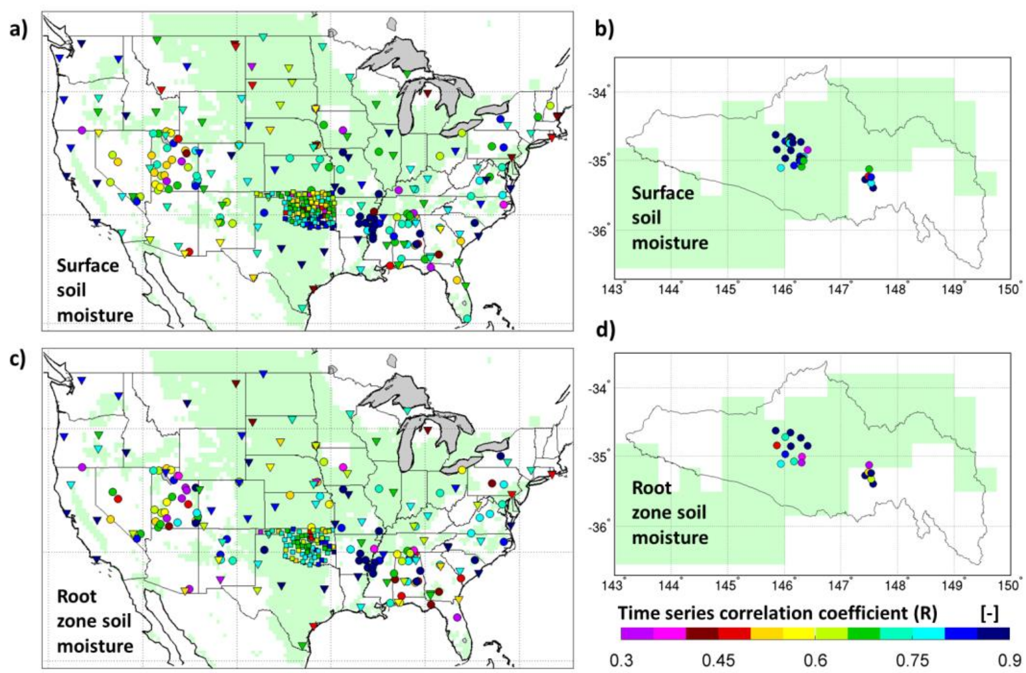

Figure 9. Same as Figure 8 but for the time series correlation coefficient (R).

Figure 10 shows the average L4_SM metrics across the sites from all networks broken down by the exclusion mask of the accuracy requirement (as indicated by the shading in Figures 8 and 9). The average metrics are computed based on a clustering algorithm that assigns the weights given to each location based on the density of sites in the surrounding region (De Lannoy and Reichle 2016). As suggested by the map plots above, Figure 10 illustrates that the L4_SM ubRMSE values are lower at the sites outside the mask, with values of 0.046 m^3m^{-3} for surface soil moisture and 0.037 m^3m^{-3} for root zone soil moisture, compared to 0.052 m^3m^{-3} and 0.042 m^3m^{-3} for surface and root zone soil moisture, respectively, at sites within the mask. Again, this result is related to the much lower variability of soil moisture in the arid regions of the western United States, which also happen to lie largely in mountainous terrain. The result is reversed for the network-average absolute bias, where values are much lower within the mask (0.05-0.07 m^3m^{-3}) than outside the mask (0.07-0.09 m^3m^{-3}). The values for the time series correlation coefficients are more similar inside and outside the mask and generally range between 0.65 and 0.70.

Figure 10 also shows the skill of the NRv4 estimates. Across the board, the L4_SM skill in terms of ubRMSE and R is slightly higher than that of NRv4, reflecting the additional information contributed by the assimilation of the SMAP brightness temperature observations in the L4_SM system. The skill differences are small, though, and not statistically significant because of the relatively short data record. As for the core validation sites, the typically small differences between L4_SM and NRv4 estimates reflect the fact that the sparse network measurements are located in areas where the surface meteorological forcing takes advantage of high-quality, gauge-based precipitation measurements. Larger improvements from the assimilation of SMAP observations can be expected in areas where the precipitation forcing inputs are not as informed by gauge measurements.

Table 5 provides average skill metrics broken down by land cover as well as by the individual networks. The breakdown by network is provided for completeness, but it is difficult to interpret because of the large differences in the number and location of individual sites within each network.

The breakdown by land cover follows the IGBP classes. In the NRv4 skill comparsion, there are no sparse network sites in the closed shrublands, savannas, permanent wetlands, and snow/ice classes (IGBP classes 6, 9, 11, 15). Urban/built-up and barren/sparse classes include only a few sites. We lumped the five IGBP classes for forests, including evergreen/deciduous/needleleaf/broadleaf and mixed forest (IGBP classes 1-5), into a single “forest” class. Besides this lumped forest class, there are five additional IGBP classes for which between 27 and 169 sites are available (Table 5).

As stated above, the sparse network metrics are best interpreted in terms of time series correlation coefficients, which discounts some of the errors that arise because a given in situ site is not representative of the grid cell average soil moisture. The R values for the L4_SM product range between 0.67 and 0.76 for surface soil moisture and between 0.62 and 0.70 for root zone soil moisture (Table 5). The R values of the L4_SM surface soil moisture estimates exceed those of NRv4 for all IGBP classes, by an average of 0.04. For root zone soil moisture the R values of L4_SM are better than or close to those of NRv4 except for grasslands, where NRv4 is noticeably better. Averaging across all sites, the R value for L4_SM root zone soil moisture matches that of NRv4. The average ubRSME metrics across all networks similarly indicate a very small improvement in surface soil moisture (by 0.002 m^3m^{-3}) and a still smaller improvement in root zone soil moisture. These results mirror the key finding of the core validation site analysis: the assimilation of SMAP brightness temperatures primarily improves surface soil moisture estimates, and, on average, does not (yet) improve the skill of the modeled root zone soil moisture.

The bias values listed in Table 5 suggest that across the four networks, the mean soil moisture from the L4_SM and NRv4 estimates is biased high (that is, wet) by about 0.06 m^3m^{-3} for the surface and by about 0.02 m^3m^{-3} for the root zone. The root zone bias in particular is remarkably small and provides some confidence in the skill of the model-based estimates. Note, however, that the typical bias at an individual site (as measured by the mean of the absolute bias) is around 0.08 m^3m^{-3} for the surface and 0.06 m^3m^{-3} for the root zone (not shown). This is at least partly a reflection of the fact that sparse network sites are not necessarily representative of the conditions in the 9 km grid cell for which the L4_SM and NRv4 estimates are valid. This is particularly true for the forest class, because measurement sites are typically on grassy areas, regardless of the surrounding land cover. For the forest class the L4_SM and NRv4 estimates have the highest bias values, around 0.10 m^3m^{-3} for surface soil moisture and around 0.07 m^3m^{-3} for root zone soil moisture (not considering the much higher bias at the single site in the urban class).

Overall, the skill values for the sparse network sites yield results that are very similar to those obtained from the core validation sites. The beneficial impact of assimilating SMAP brightness temperature observations is greatest for surface soil moisture. Furthermore, root zone soil moisture estimates are not getting worse when SMAP brightness temperatures are assimilated. Finally, it is important to keep in mind that the skill metrics presented here underestimate the true skill because these metrics are based on a direct comparison against in situ measurements (which are subject to error). Therefore, the sparse network ubRMSE values suggest that the L4_SM estimates would meet the formal accuracy requirement across a very wide variety of surface conditions, beyond those that are covered by the few core validation sites that have been available to date for formal verification of the accuracy requirement. The sparse network results thus provide additional confidence in the conclusions drawn from the core validation site comparisons.

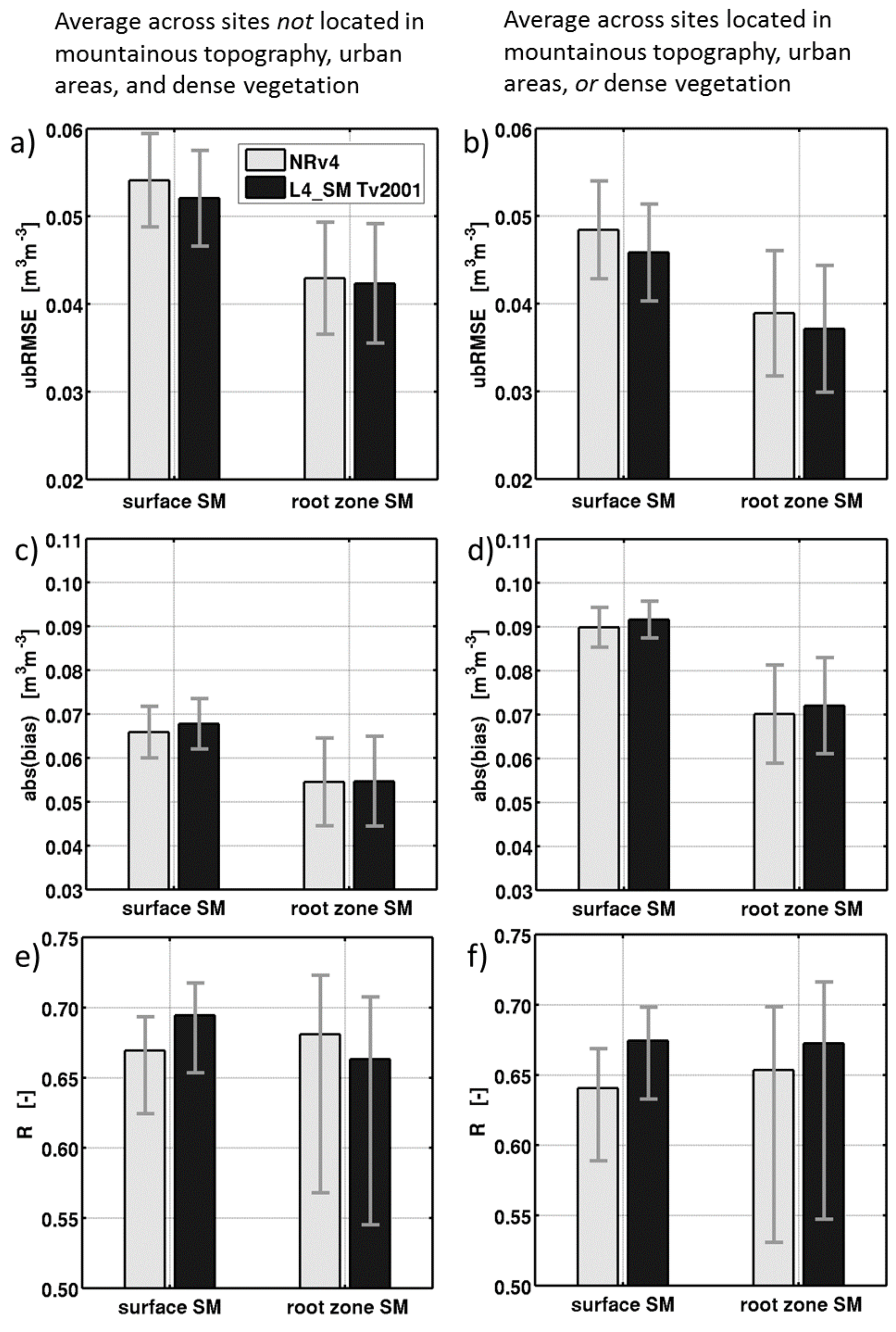

Figure 10. Skill metrics for L4_SM and NRv4 averaged over the sparse network sites listed in Table 4. (a,b) ubRMSE, (c,d) absolute bias, and (e,f) time series correlation coefficient (R). Summary metrics are averaged over sites (a,c,e) within and (b,d,f) outside the mask of the formal accuracy requirement. See text for details.

Table 5. Sparse network metrics for L4_SM and NRv4 by land cover (IGBP class), by network, and by the mask used for the (core validation site) accuracy requirement. Confidence intervals are corrected for autocorrelation in the data (Reichle et al. 2015b).

Sparse Network Subset	Surface Soil Moisture										Root Zone Soil Moisture									
		ubRMSE [m^3m^{-3}]			Bias [m^3m^{-3}]			R [-]				ubRMSE [m^3m^{-3}]			Bias [m^3m^{-3}]			R [-]		
	Number of sites	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval	Number of sites	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval	NRv4	**L4_SM Tv2001**	95% Conf. Interval
Forests (IGBP 1-5)	40	0.054	**0.051**	0.008	0.094	**0.091**	0.007	0.67	**0.69**	0.05	31	0.048	**0.045**	0.009	0.051	**0.051**	0.012	0.66	**0.68**	0.11
Open shrublands (IGBP 7)	27	0.037	**0.034**	0.005	0.017	**0.031**	0.005	0.61	**0.67**	0.05	20	0.027	**0.025**	0.013	0.003	**0.014**	0.013	0.62	**0.62**	0.16
Woody savannas (IGBP 8)	28	0.060	**0.056**	0.011	0.047	**0.036**	0.009	0.67	**0.70**	0.11	22	0.053	**0.048**	0.016	0.029	**0.010**	0.014	0.69	**0.70**	0.21
Grasslands (IGBP 10)	169	0.049	**0.049**	0.007	0.042	**0.051**	0.007	0.66	**0.68**	0.05	124	0.035	**0.038**	0.014	-0.001	**0.009**	0.021	0.69	**0.65**	0.13
Croplands (IGBP 12)	78	0.058	**0.054**	0.007	0.037	**0.036**	0.009	0.64	**0.67**	0.05	52	0.042	**0.043**	0.008	0.012	**0.014**	0.012	0.69	**0.68**	0.12
Urban/built-up (IGBP 13)	4	0.070	**0.061**	0.018	0.065	**0.044**	0.007	0.52	**0.67**	0.15	3	0.047	**0.048**	0.021	0.087	**0.069**	0.008	0.64	**0.69**	0.32
Crop/natural (IGBP 14)	39	0.054	**0.053**	0.006	0.039	**0.034**	0.006	0.69	**0.70**	0.05	31	0.044	**0.042**	0.006	0.014	**0.008**	0.007	0.68	**0.70**	0.11
Barren/sparse (IGBP 16)	2	0.034	**0.027**	0.008	0.003	**0.007**	0.005	0.66	**0.76**	0.09	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
SCAN	123	0.051	**0.049**	0.004	0.029	**0.034**	0.005	0.63	**0.64**	0.04	108	0.040	**0.038**	0.006	-0.002	**0.004**	0.006	0.64	**0.62**	0.09
USCRN	111	0.050	**0.048**	0.006	0.066	**0.068**	0.005	0.68	**0.71**	0.03	85	0.043	**0.041**	0.007	0.031	**0.032**	0.014	0.70	**0.71**	0.07
OZNet-Murrumbidgee	42	0.065	**0.055**	0.025	0.020	**0.037**	0.028	0.70	**0.79**	0.15	18	0.028	**0.038**	0.070	-0.088	**-0.067**	0.067	0.64	**0.69**	0.68
OK Mesonet	116	0.063	**0.059**	0.002	0.011	**0.004**	0.002	0.55	**0.67**	0.14	76	0.046	**0.049**	0.004	-0.012	**-0.013**	0.004	0.69	**0.69**	0.47
Inside mask	268	0.054	**0.052**	0.005	0.042	**0.042**	0.006	0.67	**0.69**	0.03	188	0.043	**0.042**	0.007	0.009	**0.010**	0.010	0.68	**0.66**	0.08
Outside mask	124	0.048	**0.046**	0.006	0.074	**0.079**	0.004	0.64	**0.67**	0.04	99	0.039	**0.037**	0.007	0.040	**0.044**	0.011	0.65	**0.67**	0.08
Average (all sites)	**392**	**0.050**	**0.049**	**0.005**	**0.055**	**0.059**	**0.005**	**0.65**	**0.68**	**0.03**	**287**	**0.041**	**0.040**	**0.007**	**0.020**	**0.023**	**0.012**	**0.67**	**0.68**	**0.08**

6.3.3 L2P Skill Comparison

Figure 11 illustrates the skill values of the L4_SM surface soil moisture estimates at the sparse network sites when cross-masked to the L2P retrievals and broken down by vegetation water content (VWC). The median ubRMSE values for L4_SM and L2P surface soil moisture are generally comparable and range from 0.04 to 0.05 m^3m^{-3}, with generally similar skill for L4_SM and L2P across all VWC classes and across all sites. Contrary to expectation, the ubRMSE values do not generally increase with VWC for the L2P (nor the L4_SM) estimates. The median bias values range between -0.03 m^3m^{-3} and 0.03 m^3m^{-3}, with slightly less absolute bias for L4_SM than L2P. The median R values range between 0.6 and 0.8, with the skill of L2P exceeding that of L4_SM by about 0.04 for all sites. For all metrics, the 25th and 75th percentile skill values generally indicate the same relative performance of the L2P and L4_SM estimates as that reflected in the median skill values.

Table 6 lists the average skill values for L4_SM and L2P broken down by vegetation class. The L2P estimates are slightly more skillful than the L4_SM estimates in terms of ubRMSE across most vegetation classes, but the performance is comparable when averaged over all sites (0.049 m^3m^{-3} for L2P and 0.050 m^3m^{-3} for L4_SM). In terms of R values, L2P is more skillful than L4_SM across most vegetation classes and when averaged across all sites (0.67 for L2P and 0.63 for L4_SM), but L4_SM has typically less absolute bias than L2P. It should be noted, however, that the L2P skill comparison includes only conditions with VWC less than 5 kg m^{-2}, that is, conditions for which the L2P retrievals are expected to do well. While 95% confidence intervals were not computed for the L2P skill comparison, the confidence intervals are likely to be larger than those of the NRv4 skill comparison, because the latter included more data (section 6.3.2). The differences between the L4_SM and L2P skill values are therefore not statistically significant.

Regardless of the skill differences, the L4_SM estimates offer the advantage of more complete spatial and temporal coverage than the L2P retrievals. The L4_SM product also provides a variety of related land surface variables, most importantly root zone soil moisture, that are consistent with the SMAP observations.

Table 6. Sparse network metrics for L4_SM and L2P surface soil moisture by land cover (IGBP class).

IGBP Class	**Num-ber of sites**	**ubRMSE** [m³/m³]		**Bias** [m³/m³]		**R** [-]	
		L2P	**L4_SM**	**L2P**	**L4_SM**	**L2P**	**L4_SM**
Open shrublands (IGBP 7)	45	0.046	**0.049**	-0.036	**0.028**	0.68	**0.62**
Woody savannas (IGBP 8)	9	0.044	**0.046**	-0.017	**0.049**	0.72	**0.73**
Grasslands (IGBP 10)	149	0.049	**0.049**	-0.035	**0.019**	0.68	**0.63**
Croplands (IGBP 12)	115	0.051	**0.053**	-0.031	**0.024**	0.63	**0.63**
Crop/natural (IGBP 14)	39	0.045	**0.047**	-0.017	**0.019**	0.66	**0.63**
Barren/sparse (IGBP 16)	4	0.041	**0.050**	-0.017	**0.014**	0.76	**0.47**
All	***361***	*0.049*	***0.050***	*-0.031*	***0.022***	*0.67*	***0.63***

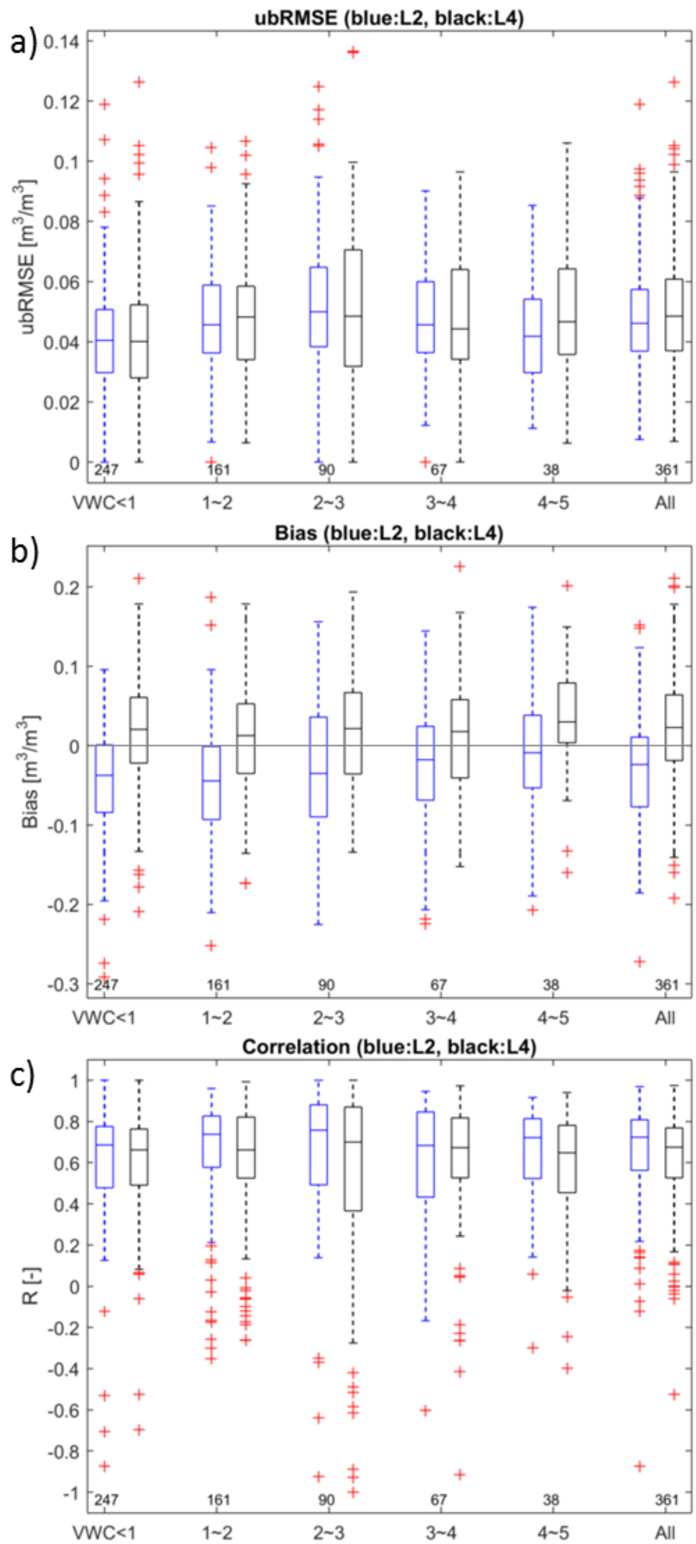

Figure 11. Skill metrics for L4_SM and L2P surface soil moisture averaged over the sparse network sites listed in Table 4. (a) ubRMSE, (b) bias, and (c) time series correlation coefficient (R). Binning is by vegetation water content (VWC; kg m^{-2}). See text for details.

6.4 Data Assimilation Diagnostics

This section provides an evaluation of the L4_SM data assimilation diagnostics, including the statistics of the observation-minus-forecast (O-F) residuals, the observation-minus-analysis (O-A) residuals, and the analysis increments. Because the L4_SM algorithm assimilates brightness temperature observations, the O-F and O-A diagnostics are in terms of brightness temperatures (that is, in "observation space"). Strictly speaking, the analysis increments are in the space of the Catchment model prognostic variables that make up the "state vector", including the "catchment deficit", "root zone excess", "surface excess", and "top-layer ground heat content" (Reichle et al. 2014b). For the discussion below, the increments have been converted into equivalent soil moisture and soil temperature terms.

A key element of the analysis update is the downscaling and inversion of the observational information from the 36 km grid of the assimilated brightness temperatures into the modeled geophysical variables on the 9 km grid, based on the modeled error characteristics, which vary dynamically and spatially. An example and illustration of a single analysis update can be found in (Reichle et al. 2015b, their section 6.4.1).

6.4.1 Observation-Minus-Forecast Residuals

Figure 12 shows the global coverage of the SMAP L1C_TB observations that were used in the L4_SM analysis for 7 June 2015, 0z. The analysis window includes brightness temperature observations between 22:30z on 6 June 2015 and 01:30z on 7 June 2015. Within this window, 8,976 observations were used in total, including 2,679 H-pol and 2,725 V-pol observations from two descending half-orbits over eastern Russia and Indonesia and 1,775 H-pol and 1,797 V-pol observations from two ascending half-orbits over the Americas.

SMAP L1C_TB observations were not used over China where L-band radio-frequency interference (RFI) is common. SMAP is equipped with a variety of hardware and software tools to deal with RFI and generally provides near-global coverage. However, the L4_SM algorithm requires knowledge of the long-term L-band brightness temperature climatology to address observation-minus-forecast bias in the system (Reichle et al. 2014b). The necessary climatological information is derived from observations provided by the Soil Moisture Ocean Salinity (SMOS) mission, which cannot provide good quality observations in the RFI-affected areas.

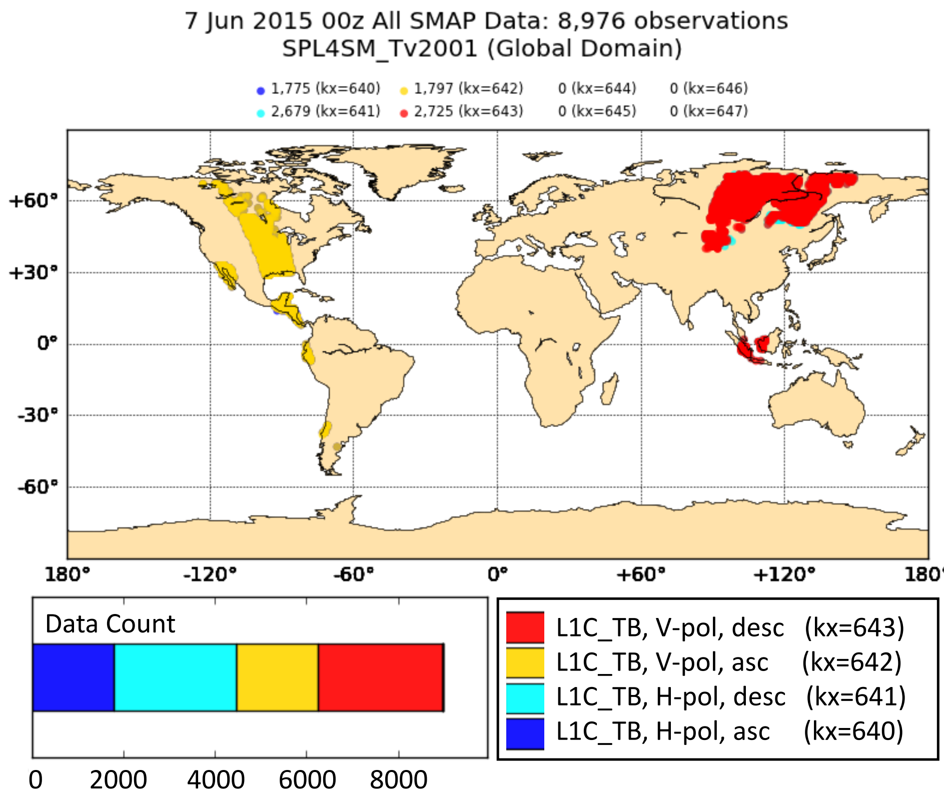

Figure 12. (Top) Coverage of the assimilated SMAP L1C_TB brightness temperature observations that were used in the L4_SM analysis for 7 June 2015, 0z (Tv2001). In the map, the coverage of V-pol observations obscures the areas where H-polarized observations are used. (Bottom) Number of brightness temperature observations assimilated in this analysis.

The gaps in spatial coverage are further illustrated in Figure 13, which shows the total number of L1C_TB observations that were assimilated during the assessment period (1 Apr 2015, 0z, to 1 Mar 2016, 0z). This count includes H- and V-pol observations from ascending and descending orbits. The average data count across the globe is approximately 361 for the 335-day period, but no data are assimilated across large areas in eastern Europe and the southern half of continental Asia due to the lack of a SMOS brightness temperature climatology because of RFI. Moreover, few or no SMAP brightness temperatures are assimilated in mountainous areas, including the Rocky Mountains and the Andes, in the vicinity of lakes, such as in northern Canada, and next to major rivers, including the Amazon and the Congo. Despite the much shorter warm (unfrozen) season at high-latitudes, northern areas exhibit relatively high counts of assimilated brightness temperature observations because of SMAP's polar orbit, which results in more frequent revisit times there.

The spatial gap in the assimilation of SMAP observations will be addressed in future versions (section 7.1). Note, however, that the L4_SM product provides soil moisture estimates everywhere, even if in some regions the L4_SM estimates are not based on the assimilation of SMAP observations and rely only on the information in the model and forcing data.

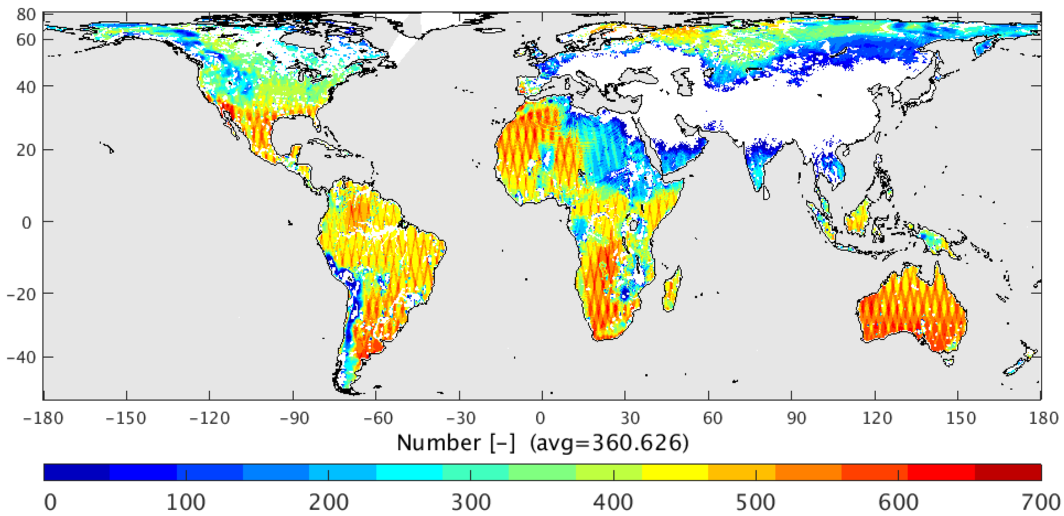

Figure 13. Number of L1C_TB observations used in the L4_SM algorithm (Tv2001) during the validation period (1 Apr 2015, 0z, to 1 Mar 2016, 0z). Data counts include H-pol and V-pol data from ascending and descending half-orbits.

Next, Figure 14 shows a time series of the global observations counts for 1 April 2015 to 1 March 2016, again including H-pol and V-pol observations from ascending and descending half-orbits. The data counts vary with season and time of day. There are 8 analysis times per day (at 0z, 3z, …, 18z, and 21z), with data counts varying depending on the time of day, primarily based on the amount of land surface area where the local time is close to 6am or 6pm local time, when SMAP crosses the Equator. Each 3-hourly L4_SM analysis typically ingests between 5,000 and 15,000 observations, with a mean close to 8,000 observations.

The bottom panel of Figure 14 shows the time series of the O-F and O-A statistics. Global mean O-F values (cyan bars) typically range from -3 K to 3 K, with a long-term average value of just 0.5 K. Mean O-A values (orange bars) are slightly smaller than mean O-F values and have long-term average value of 0.4 K. Overall, the relatively small mean O-F and O-A values suggest that the assimilation system is reasonably bias-free in a global average sense.

Typical magnitudes of the O-F residuals, indicated by the RMS values (blue bars), range between 3 K and 10 K, with a long-term average value of 6.2 K. The RMS values of the O-A residuals (red bars) are generally lower and have a long-term average value of 4.2 K, thereby reflecting the reduction in uncertainty obtained from the analysis.

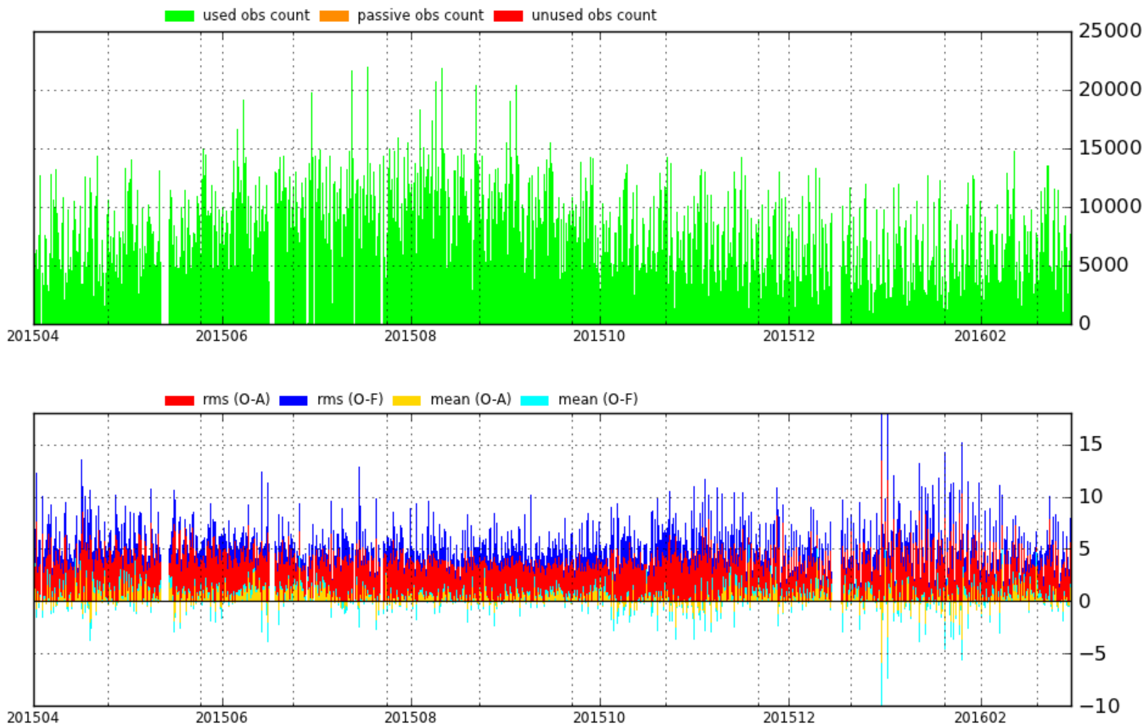

Figure 14. (Top) Counts of L1C_TB observations assimilated into L4_SM (Tv2001) between 1 Apr 2015 and 1 Mar 2016, including H-pol and V-pol data from ascending and descending orbits. (Bottom) Statistics of the corresponding O-F and O-A residuals, where the mean and RMS values are computed separately for each 3-hourly analysis by averaging across the global domain.

The O-F RMS values show occasional spikes exceeding 10 K. The two largest such spikes can be seen for the 30 December 2015, 21z analysis and for the 1 January 2016, 21z analysis, extending to about 30 K and 22 K, respectively. (The spikes are cut off in the figure.) These spikes are accompanied by large negative O-F mean values, with individual O-F values reaching -90 K (not shown). The spikes correspond to major rain events in central Australia that were missed in the (CPCU gauge-corrected) L4_SM precipitation forcing data but are somewhat reflected in higher-quality local gauge observations from the Australian Bureau of Meteorology (http://www.bom.gov.au/jsp/awap/rain/). For the 30 December 2015, 21z analysis, in particular, the coverage of the L1C_TB observations was limited to a single, descending half-orbit across central Australia and Indonesia, so that the global statistics shown in Figure 14 for this analysis are dominated by the missed rain event in a portion of Australia. These events highlight the potential for SMAP to provide valuable information about soil moisture and rainfall in areas where rain gauge observations are sparse.

Figure 15 shows the global distributions of the time series mean and standard deviation of the O-F residuals. The time mean values of the O-F residuals are typically small and mostly range from -2 K to 2 K. Overall, there is a positive bias of 0.5 K, with fewer areas exhibiting negative mean O-F values. Larger values are found in the Sahel and in central and southern Africa. Over Africa, the L4_SM precipitation forcing is not corrected to the gauge-based product (Reichle and Liu 2014). The L4_SM algorithm therefore relies heavily on the consistency of the present forcing data from the ¼ degree GEOS-5 operational forward processing (FP) system (GEOS-5.13) and the historic forcing data from the ½ degree reprocessing (RP-IT/FP-IT) system (GEOS-5.9) that was used to derive the brightness temperature rescaling factors in the calibration of the L4_SM algorithm. High values are also seen in the center of the

United States, Argentina, Australia, and portions of Siberia, which indicates that the L4_SM system would benefit from further calibration.

The time series standard deviation of the O-F residuals ranges from a few Kelvin to around 15 K, with a global (spatial) average of about 5.9 K (Figure 15b). The highest values are found in central North America, the Sahel, central Asia, and in Australia. These regions have sparse or modest vegetation cover and typically exhibit strong variability in soil moisture conditions. The O-F residuals are generally smallest in more densely vegetated regions, including the eastern United States, the Amazon basin, and tropical Africa. Small values are also found in high-latitudes, including Alaska and Siberia, and in the Sahara desert. The spatially averaged time series standard deviation of the O-A residuals is 4.0 K (not shown), which again reflects the impact of the SMAP observations on the L4_SM system.

Note that the numbers for the average RMS magnitudes of the O-F residuals derived from Figures 14 and 15 are slightly different because they are derived by temporally averaging spatial RMS values and by spatially averaging temporal RMS values, respectively. Note also that the 2.6 K value for the average RMS magnitude of the beta-release (Vb1004) O-A residuals quoted on page 5 of (Reichle et al. 2015b) is incorrect due to a processing error. The correct numbers is close to 4 K.

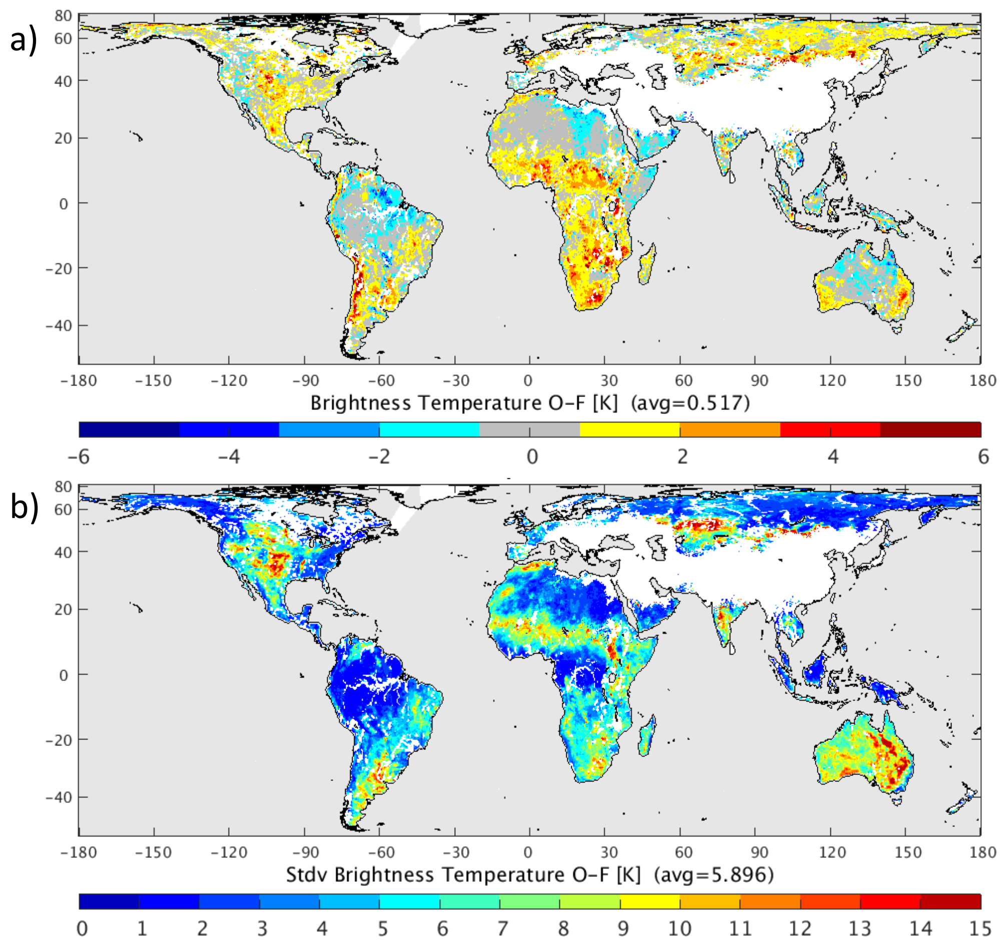

Figure 15. (a) Mean and (b) standard deviation of the O-F residuals from the L4_SM algorithm (Tv2001) for 1 Apr 2015, 0z, to 1 Mar 2016, 0z.

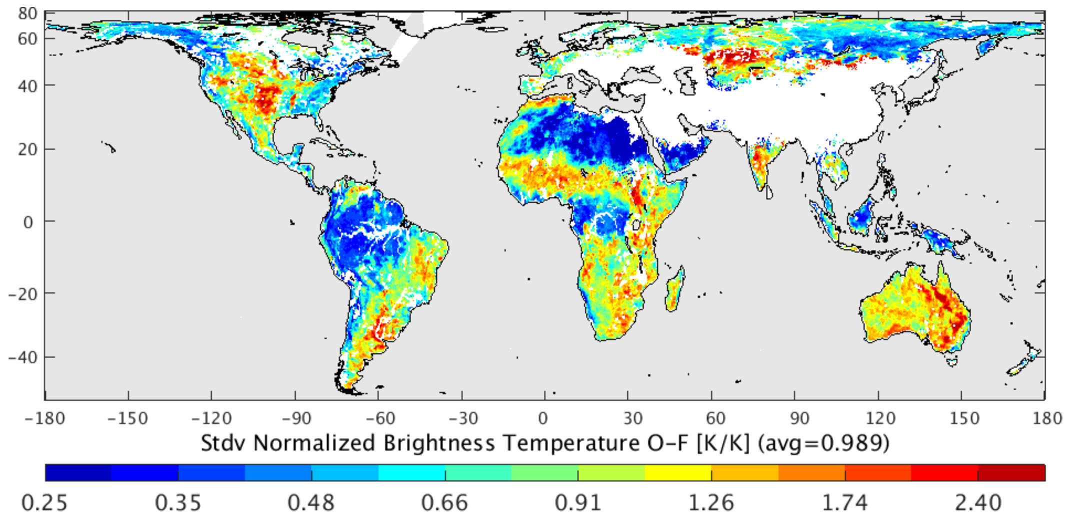

Figure 16. Standard deviation of the *normalized* O-F residuals from the L4_SM algorithm (Tv2001) for 1 Apr 2015, 0z, to 1 Mar 2016, 0z.

Finally, Figure 16 shows the standard deviation of the *normalized* O-F residuals, which measures the consistency between the expected (modeled) errors and the actual errors. Specifically, the normalization of the O-F residuals is with the standard deviation of their expected total error, which is composed of the error in the observations (including instrument errors and errors of representativeness) and errors in the brightness temperature model forecasts (Reichle et al. 2015b, their Appendix B). The parameters that determine the expected error standard deviations are key inputs to the ensemble-based L4_SM assimilation algorithm. If they are chosen such that the modeled errors are fully consistent with the actual errors, the metric shown in the figure should be unity. If the metric is less than one, the actual errors are overestimated by the assimilation system, and if the metric is greater than one, the actual errors are underestimated.

The global average of the metric is 0.99 (Figure 16), which would suggest that, on average, the modeled errors are consistent with the actual errors. The metric, however, varies greatly across the globe. Typical values are either too low or too high. In the Amazon basin, the eastern US, tropical and southern Africa, and the high northern latitudes, values range from 0.25 to 0.5, and thus errors there are considerably overestimated. Conversely, in central North America, the Sahel, India, and southeastern Australia, values range from 1.5 to 4, meaning that errors in these regions are considerably underestimated. Future work will focus on improving the calibration of the input parameters that determine the model and observation errors in the L4_SM system (section 7.1).

6.4.2 Increments

Figure 17 shows the average number of increments that the L4_SM algorithm generated per day during the validation period. The global mean is 0.74, which means that for a given location, there are approximately three increments applied every four days on average, either from an ascending or a descending overpass. The overall pattern of the increments count follows that of the count of the assimilated observations shown in Figure 13. The coverage of the increments, however, is somewhat greater than that of the observations due to the spatial interpolation and extrapolation of the observational information in the distributed analysis update of the L4_SM algorithm. The figure also reveals the diamond patterns resulting from SMAP's regular 8-day repeat orbit.

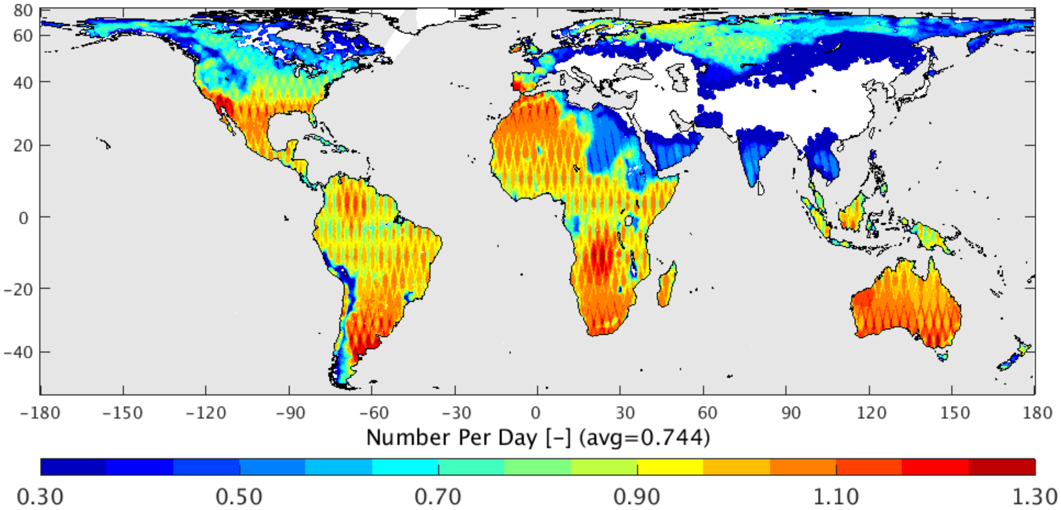

Figure 17. Average number of increments per day generated by the L4_SM algorithm (Tv2001) during the validation period (1 Apr 2015, 0z, to 1 Mar 2016, 0z). The result applies equally to all elements of the control vector, including the model prognostic variables related to surface soil moisture, root zone soil moisture, surface temperature, and top-layer soil temperature.

Figure 18 shows the time mean values of the analysis increments for surface and root zone soil moisture as well as for the surface layer soil temperature. In the long-term average, the increments for root zone soil moisture and surface soil temperature vanish nearly everywhere. Only the increments in surface soil moisture exhibit a slight bias in some regions, including the US Great Plains, the Sahel, southern Africa, and Australia, with extreme values of around -0.01 m^3m^{-3}. These drying increments are a reflection of the slight warm bias in the O-F residuals (Figure 15a). Overall, Figure 18 suggests that the analysis system is very nearly unbiased.

Figure 19 shows the time series standard deviation of the increments in surface and root zone soil moisture as well as surface soil temperature. This metric measures the typical magnitude of instantaneous increments. Typical increments in surface soil moisture are on the order of 0.02 m^3m^{-3} in the western US, central Mexico, the Sahel, southern Africa, southern India, and most of Australia. They are around 0.005 m^3m^{-3} in the eastern US, Argentina, and the high northern latitudes. Over the tropical forests, surface soil moisture increments are generally negligible, reflecting the fact that in those areas the measured SMAP brightness temperatures are mostly sensitive to the dense vegetation and are only marginally sensitive to soil moisture and soil temperature.

Typical increments in root zone soil moisture (Figure 19b) show a global pattern that is very similar to that of the surface soil moisture increments, albeit with smaller magnitudes that again reflect the lesser sensitivity of the L-band brightness temperatures to the deeper layer soil moisture. The magnitude of the average root zone soil moisture increments rarely exceeds 0.01 m^3m^{-3}, with a global average value of about 0.003 m^3m^{-3}. Finally, increments for the top layer soil temperature (Figure 19c) and surface (skin) temperature (not shown) also exhibit a pattern similar to that of the surface soil moisture increments.

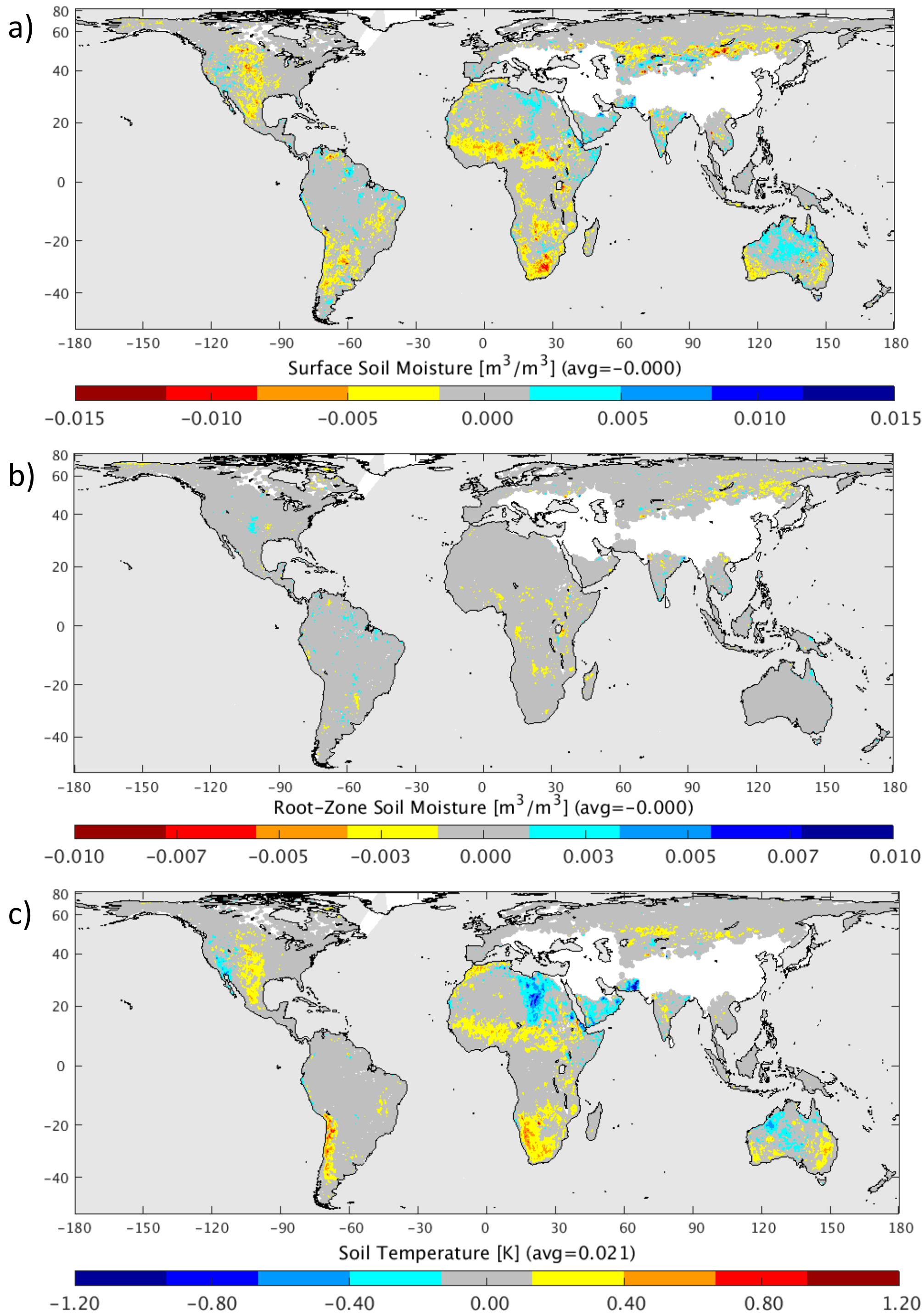

Figure 18. Time series mean of the increments for (a) surface soil moisture, (b) root zone soil moisture, and (c) top layer soil temperature from the L4_SM algorithm (Tv2001) for 1 Apr 2015, 0z, to 1 Mar 2016, 0z. Note the difference in colorscale for (a) and (b).

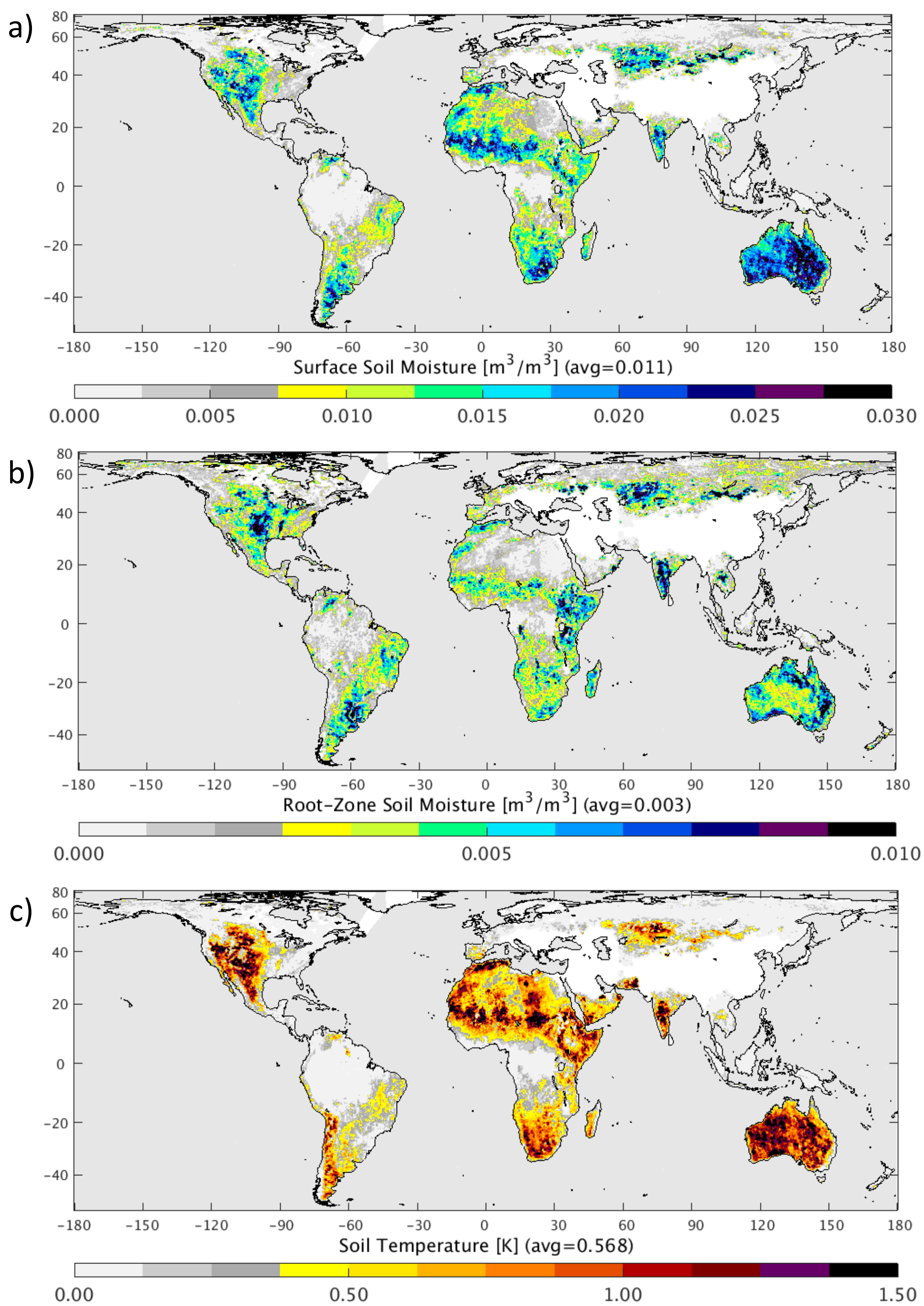

Figure 19. Same as Figure 18 but for time series standard deviation of the increments. Note the difference in colorscale for (a) and (b).

6.4.3 Uncertainty Estimates

The L4_SM data product also includes error estimates for key output variables, including surface and root zone soil moisture as well as surface soil temperature. These uncertainty estimates vary dynamically and geographically because they are computed as the standard deviation of a given output variable across the ensemble of land surface states at a given time and location. (The ensemble is an integral part of the ensemble Kalman filter employed in the L4_SM algorithm, and the ensemble mean provides the estimate of the variable under consideration.) By construction, the uncertainty estimates represent only the random component of the uncertainty. Bias and other structural errors such as errors in the dynamic range are not included.

Figure 20 shows the time mean of the uncertainty estimates for the validation period. Across the globe, surface soil moisture uncertainty typically ranges from 0.01 m^3m^{-3} to 0.03 m^3m^{-3}, with the larger uncertainties concentrated in the driest regions such as the Sahara desert. Uncertainty is also large where few or no SMAP brightness temperatures are assimilated in the L4_SM system so that the ensemble spread is never reduced through analysis updates. These regions include most of eastern Europe and the southern half of continental Asia, where the lack of climatological information from SMOS prevents the assimilation of SMAP observations this early in the mission (section 6.4.1). This limitation will be removed later in the mission when a sufficient number of SMAP observations will be available to derive rescaling parameters based solely on SMAP information (section 7.1). Uncertainty is also high in the portions of the northern high-latitudes where few observations are assimilated, presumably because of frozen or snow-covered conditions during a considerable portion of the validation period.

Uncertainty in root zone soil moisture (Figure 20b) is generally smaller than for surface soil moisture, with typical values ranging from 0.005 m^3m^{-3} to 0.025 m^3m^{-3}. The global pattern of uncertainty in root zone soil moisture is quite different from that of surface soil moisture. The driest areas are associated with low values of uncertainty, because in arid regions the deeper layer soil moisture is mostly constant, with random errors in surface forcings (precipitation and radiation) having only a small impact. Uncertainty is highest in southern China, where root zone soil moisture is variable but SMAP observations cannot be assimilated. Finally, uncertainty in surface soil temperature (Figure 20c) ranges from 0.5 K to 2 K and is again largest in dry (and hot) regions and in regions where SMAP brightness temperatures are not assimilated.

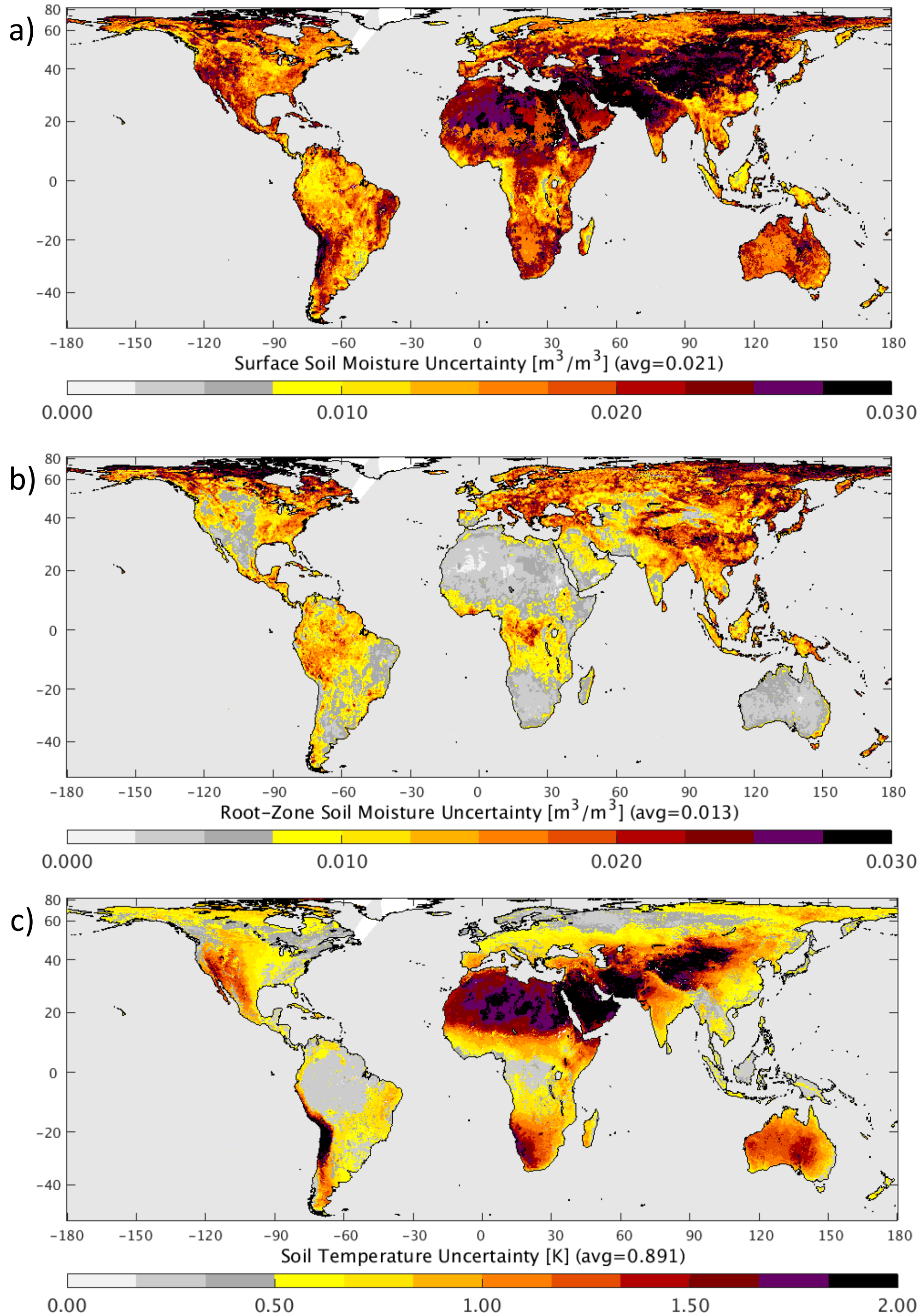

Figure 20. L4_SM (Tv2001) uncertainty estimates for (a) surface soil moisture, (b) root zone soil moisture, and (c) top-layer soil temperature averaged across the validation period (1 April 2015, 0z to 1 Mar 2016, 0z). Uncertainty estimates are computed as the standard deviation across the model ensemble that is used in the L4_SM assimilation algorithm.

6.5 Summary

The SMAP L4_SM Tv2001 data product is a test stream for the "Version 2" validated release scheduled for 29 April 2016 that covers the period from 1 April 2015, 0z to 1 March 2016, 0z. The test product was validated using in situ soil moisture measurements from core validation sites and sparse networks. The product was further evaluated through an assessment of the data assimilation diagnostics generated by the L4_SM algorithm, such as the observation-minus-forecast residuals and the increments.

Based on the comparisons with the core validation site measurements, the L4_SM Tv2001 estimates of surface and root zone soil moisture meet the accuracy requirement (ubRMSE < 0.04 m^3m^{-3}). For surface soil moisture the ubRMSE is 0.035 m^3m^{-3} at the 9 km scale and 0.031 m^3m^{-3} at the 36 km scale. For root zone soil moisture, the ubRMSE is 0.024 m^3m^{-3} at the 9 km and 36 km scales.

The assimilation of SMAP brightness temperatures in the L4_SM algorithm is beneficial primarily for surface soil moisture estimates, where the improvements over the model-only SMAP Nature Run (NRv4) are consistent across the 9 km and 36 km scales and the ubRMSE and R metrics. Note, however, that the improvements are not statistically significant at the 5% level. For root zone soil moisture, L4_SM and NRv4 estimates have essentially the same skill.

The comparison with in situ measurements from a global set of sparse networks corroborate the results obtained for the core validation sites, thus supporting compliance with the CEOS Stage 2 Validation requirements (section 2).

The data assimilation diagnostics further broaden the validation to the global domain and indicate that the L4_SM system is reasonably unbiased in the global average sense. The time mean, globally averaged analysis increments in surface and root zone soil moisture and in the skin and surface soil temperatures are negligibly small. However, the observation-minus-forecast residuals of brightness temperature reveal a modest warm bias of a few Kelvin in the L4_SM product in select regions, which is reflected in a small (absolute) bias in the surface soil moisture increments of up to 0.01 m^3m^{-3} in those regions. The assimilation diagnostics further reveal that, on a regional basis, the actual errors in brightness temperature are typically over- or underestimated considerably by the L4_SM system, even though the global average of that diagnostic would suggest unbiased error estimates.

Uncertainty estimates for the analyzed surface soil moisture, root zone soil moisture, surface temperature, and top layer soil temperature are also provided with the product. These uncertainty estimates are designed to reflect the random error in key geophysical product fields. While the uncertainty estimates appear reasonable, it is not yet clear how well they reflect actual uncertainties.

It is important to keep in mind that the comparisons against the in situ measurements are impacted by the fact that the in situ measurements themselves are prone to errors. The metrics presented here therefore underestimate the true skill of the product. In fact, the ubRMSE results presented in the report should be interpreted as the unbiased RMS *difference* between the model estimates and the in situ measurements, rather than errors with respect to actual (or true) soil moisture conditions.

Based on the results presented in this report, the public release of the Version 2 validated L4_SM data product is recommended. The results also uncovered limitations in the Version 2 product and possible avenues for future development. These limitations and developments are addressed in the next section.

7 OUTLOOK AND PLAN FOR FUTURE IMPROVEMENTS

The assessment of the Version 2 L4_SM product presented in the previous section revealed a number of current limitations as well as avenues for future development.

7.1 Bias and L4_SM Algorithm Calibration

Figure 15a showed that there are regions with a modest residual bias between the predicted brightness temperatures from the L4_SM modeling system and the SMAP observations. The magnitude of this bias is reduced from that of the (beta release) Version 1 product (Reichle et al. 2015b) due to the improvements in the L1 radiometer calibration. Nevertheless, there are additional concerns related to bias that can be addressed in the L4_SM system. First, the surface meteorological forcing data that are used in the operational L4_SM system are based on data from the current GEOS-5 Forward Processing ("FP") system (GEOS-5.13; ¼ degree resolution). The corresponding retrospective SMAP Nature Run version 4 (NRv4) uses data from the GEOS-5 Reprocessing and Forward Processing for Instrument Teams ("RP/FP-IT") system (GEOS-5.9; ½ degree resolution). The vintage of the two systems is quite different, and the climatological parameters derived from NRv4 data do not necessarily reflect the true climatology of the L4_SM modeling component. This climatological inconsistency negatively impacts the quality of the soil moisture output in *percentile* units from the "gph" Collection, which at this time should be used only with caution.

The second reason for the residual biases could be differences between the SMAP brightness temperatures used in the L4_SM Tv2001 product and the SMOS v504 brightness temperatures that were used in the calibration of the microwave radiative transfer parameters of the L4_SM system.

To address the biases in the L4_SM algorithm, the calibration of the system will be revisited. The L4_SM team has already constructed a new version of the SMAP Nature Run (NRv5) that shows slightly better soil moisture skill than NRv4. The new version (NRv5) is forced with the newly available MERRA-2 reanalysis data (GEOS-5.12; ½ degree resolution), which is closer to the current L4_SM forcing from the FP system (Bosilovich et al. 2015). NRv5 also includes a revised approach to precipitation corrections, along with updates to the GEOS-5 land model and its ancillary data. Preliminary results with the recalibrated system show that the soil moisture output in percentile units from the "gph" Collection would be greatly improved relative to the current system.

Based on NRv5 and the latest version of the SMOS brightness temperatures (v620), the parameters of the L4_SM microwave radiative transfer model will be recalibrated. The objective of these revisions is to provide an improved L4_SM modeling component along with retrospective data that is as consistent as possible with the present-day data in terms of its climatology.

Because SMOS observations are impacted by RFI and thus do not provide a climatology of L-band brightness temperatures in large parts of Asia and Europe, the L4_SM algorithm cannot yet assimilate SMAP data in those regions (Figure 13). (Note that the L4_SM product provides soil moisture estimates everywhere, even if in some regions the L4_SM estimates are not based on the assimilation of SMAP observations and rely only on the information in the model and forcing data.) This limitation manifests itself in two steps in the calibration of the L4_SM system. First, the L4_SM microwave radiative transfer model parameters cannot be calibrated locally. Second, the brightness temperature scaling parameters cannot be computed for the affected regions. The first limitation can be addressed by setting the microwave radiative transfer model parameters in the regions in question according to land cover (vegetation) class (De Lannoy et al. 2013, 2014). The second limitation is more difficult to overcome, as the residual seasonal model biases in brightness temperature cannot be addressed through the pentad rescaling factors used in the L4_SM algorithm. Ultimately, the solution is to calibrate the L4_SM system using a history of SMAP observations, and a robust calibration based solely on SMAP data can only be accomplished towards the end of the SMAP mission. In the meantime, we will investigate whether the L4_SM system can be provisionally calibrated in the RFI-affected regions by using land cover-based microwave radiative transfer model parameters along with a short history from SMAP to scale out the residual seasonal biases.

Another aspect of algorithm calibration involves the tuning of the L4_SM observation and model error parameters. Initial research into tuning the observation error parameters using a poor man's adaptive filter has been inconclusive. Such tuning resulted, by construction, in improved assimilation diagnostics, but it has so far failed to consistently improve the skill metrics obtained from the comparisons against independent in situ observations.

7.2 Impact of SMAP Observations and Ensemble Perturbations

The assessment of the L4_SM product presented herein uses NRv4 estimates as the model-based reference. Both estimates use the same gauge-corrected precipitation forcing. Not surprisingly, the assessment of the L4_SM (and NRv4) estimates versus in situ soil moisture observations is focused on regions for which the model forcing data can take advantage of typically dense and reliable precipitation gauge observations. The generally good skill of the NRv4 estimates therefore leads to an underestimation of the impact of the SMAP brightness temperature observations in the L4_SM assimilation system. In regions with poor precipitation data, the impact of the SMAP observations should be larger, but because of the lack of observations of any kind in those regions, the precise impact remains unknown.

Secondly, the NRv4 estimates and the L4_SM estimates differ partly because the NRv4 estimates are from a single-member model run without perturbations, whereas the L4_SM estimates are based on an ensemble of model realizations that experiences perturbations to its model forcing and prognostic variables. An undesirable, yet at this time unavoidable, side effect of the perturbations regime is that it leads to biases between the ensemble mean estimates and the estimates from the unperturbed NRv4 model integration. This is particularly acute in very arid regions such as the Sahara desert, where the perturbations in soil moisture are, by construction, biased wet because the unperturbed, single-member model run typically remains at the lowest possible soil moisture value, thereby making negative (that is, drying) perturbations unphysical. Some of the differences between the NRv4 and L4_SM estimates will therefore partly reflect the impact of the perturbations regime rather than the use of SMAP observations.

To address these issues, the assessment will be refined by expanding the comparison of the L4_SM skill to that of additional model-only data sets. To better assess the impact of the SMAP observations, a model-only run will be conducted without the gauge-based precipitation corrections. Furthermore, an ensemble integration will be conducted in which the perturbations are applied but no SMAP data are assimilated. By comparing the L4_SM estimates to the different model-only runs, it will be possible to more clearly identify the impact of the SMAP observations. These refined assessment are planned to coincide with the release of the recalibrated, NRv5-based L4_SM system (section 7.1).

7.3 Expanded Site Locations, Record Length, and Data Sets

The assessment of the "Version 2" validated release data is limited by the period of record. Only 11 months of data have been available for this assessment report, which does not yet cover a full annual cycle. As the SMAP observatory continues to provide measurements, the length of the data record that can be assessed, and therefore the reliability of the assessment, will continue to increase. With time, we also expect to see a further increase in the volume of validation data, with additional measurements becoming available from core sites and sparse networks that are still under development.

Time and resources permitting, the L4_SM data product will also be evaluated by comparison with model-based estimates from other data providers, including the ERA-Land reanalysis data from the European Centre for Medium-Range Weather Forecasts and data products from Environment Canada.

Resources permitting, a skill comparison of L4_SM and L2P estimates using core validation site measurements will also be included in future assessments.

7.4 L4_SM Algorithm Refinements

Despite its overall complexity, the L4_SM algorithm includes many simplifications. For example, SMAP brightness temperatures are not assimilated when the water fraction of the observed field-of-view exceeds a threshold of 5%. Rather than discarding such water-contaminated observations, the L4_SM algorithm could be refined to include the brightness temperatures of open water in its forward operator. This would require a dynamic model of the surface temperature of lakes and large rivers in the L4_SM modeling system, as well as a corresponding radiative transfer model, based, for example, on the model by Klein and Swift (1977).

Another simplification in the current L4_SM algorithm is the use of flags in the L1 Radiometer product. Currently, the L4_SM system only checks the summary flag for the assimilated L1C_TB brightness temperatures. More specific binary flags, including dynamic surface flags, are available in the input data product and could perhaps be used in a refined L4_SM algorithm. Finally, the L4_SM algorithm uses the L1C_TB data product because of its convenient data format and posting on the 36 km EASE v2 grid. Since the L4_SM algorithm operates on the finer, 9 km EASE v2 grid, and since the SMAP radiometer greatly oversamples the brightness temperature relative to the size of its instantaneous field-of-view, the L4_SM algorithm might benefit from the direct assimilation of the time-ordered brightness temperatures from the L1B_TB data product.

As mentioned in section 5.2, the SMAP radar anomaly that occurred on 7 July 2015 means that radar-based data products from SMAP are only available for a short period at the beginning of the SMAP mission. The L4_SM algorithm originally included plans for a freeze-thaw analysis using the SMAP radar-based freeze-thaw retrievals. The SMAP Project is considering the development and generation of a freeze-thaw data product using the passive microwave observations. If an operational freeze-thaw product becomes available, the L4_SM algorithm could be expanded to assimilate it, provided adequate resources are available.

ACKNOWLEDGEMENTS

This report was made possible by the contributions of many individuals from the SMAP Project, the SMAP Science Team, the SMAP Cal/Val Partner Program, and the NASA Global Modeling and Assimilation Office. The NASA Soil Moisture Active Passive mission and the NASA Modeling, Analysis, and Prediction program supported the research. Computational resources were provided by the NASA High-End Computing Program through the NASA Center for Climate Simulation at the Goddard Space Flight Center.

REFERENCES

Bosilovich, M. G., S. Akella, L. Coy, R. Cullather, C. Draper, R. Gelaro, R. Kovach, Q. Liu, A. Molod, P. Norris, K. Wargan, W. Chao, R. Reichle, L. Takacs, Y. Vikhliaev, S. Bloom, A. Collow, S. Firth, G. Labow, G. Partyka, S. Pawson, O. Reale, S. D. Schubert, and M. Suarez (2015), MERRA-2: Initial Evaluation of the Climate, *NASA Technical Report Series on Global Modeling and Data Assimilation, NASA/TM-2015-104606, Vol. 43*, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 139pp.

Colliander, A., S. Chan, N. Das, S. Kim, S. Dunbar, T. Jackson, C. Derksen, K. McDonald, J. Kimball, E. Njoku, R. Reichle, and B. Weiss (2014), SMAP L2-L4 Data Products Calibration and Validation Plan, Soil Moisture Active Passive (SMAP) Mission Science Document. JPL D-79463, Jet Propulsion Laboratory, Pasadena, CA.

CEOS (2015), Committee on Earth Observation Satellites (CEOS) Working Group on Calibration and Validation (WGCV): http://calvalportal.ceos.org, CEOS WGCV Land Products Sub-Group: http://lpvs.gsfc.nasa.gov. Accessed 7 October 2015.

Chen, F., W. T. Crow, A. Colliander, M. Cosh, T. J. Jackson, R. Bindlish, R. H. Reichle, S. K. Chan, D. D. Bosch, P. J. Starks, D. C. Goodrich, M. Seyfried (2016), Application of Triple Collocation in Ground-based Validation of Soil Moisture Active/Passive (SMAP) Level 2 Data Products, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, submitted.

De Lannoy, G. J. M, R. H. Reichle, and V. R. N. Pauwels (2013), Global Calibration of the GEOS-5 L-band Microwave Radiative Transfer Model over Nonfrozen Land Using SMOS Observations, *Journal of Hydrometeorology*, *14*, 765-785, doi:10.1175/JHM-D-12-092.1.

De Lannoy, G. J. M., R. H. Reichle, and J. A. Vrugt (2014), Uncertainty Quantification of GEOS-5 L-Band Radiative Transfer Model Parameters using Bayesian Inference and SMOS Observations, *Remote Sensing of Environment, 148*, 146-157, doi:10.1016/j.rse.2014.03.030.

De Lannoy, G. J. M., and R. H. Reichle (2016), Global Assimilation of Multiangle and Multipolarization SMOS Brightness Temperature Observations into the GEOS-5 Catchment Land Surface Model for Soil Moisture Estimation, *Journal of Hydrometeorology, 17*, 669-691, doi:10.1175/JHM-D-15-0037.1.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow (2010), Performance Metrics for Soil Moisture Retrievals and Application Requirements, *Journal of Hydrometeorology, 11*, 832-840, doi:10.1175/2010JHM1223.1.

Entekhabi, D., and Coauthors (2014), SMAP Handbook, *JPL Publication, JPL 400-1567*, NASA Jet Propulsion Laboratory, Pasadena, California, USA, 182 pp.

Gruber, A., C.-H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner (2016), Recent advances in (soil moisture) triple collocation analysis, *International Journal of Applied Earth Observation and Geoinformation*, *45B*, 200-211, doi: 10.1016/j.jag.2015.09.002.

Jackson, T., D. Chen, M. Cosh, F. Li, M. Anderson, C. Walthall, P. Doraiswamy, and E. R. Hunt (2004), Vegetation water content mapping using Landsat data derived normalized difference water index (NDWI) for corn and soybean, *Remote Sensing of Environment*, *92*, 475-482.

Jackson, T. J., M. Cosh, R. Bindlish, P. Starks, D. Bosch, M. Seyfried, D. Goodrich, S. Moran, and D. Du (2010), Validation of Advanced Microwave Scanning Radiometer soil moisture products. *IEEE Trans. Geosci. Rem. Sens.*, *48*, 4256 – 4272.

Jackson, T. J., A. Colliander, J. Kimball, R. Reichle, C. Derksen, W. Crow, D. Entekhabi, P. O'Neill, and E. Njoku (2014), SMAP Science Data Calibration and Validation Plan (Revision A), Soil Moisture Active Passive (SMAP) Mission Science Document. JPL D-52544, Jet Propulsion Laboratory, Pasadena, CA.

Klein, L., and C. T. Swift (1977), An improved model for the dielectric constant of sea water at microwave frequencies, *IEEE Transactions on Antennas and Propagation*, 25, 104-111, doi:10.1109/TAP.1977.1141539.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res.*, *105*, 24809–24822, doi:10.1029/2000JD900327.

Liu, Q., R. H. Reichle, R. Bindlish, M. H. Cosh, W. T. Crow, R. de Jeu, G. J. M. De Lannoy, G. J. Huffman, and T. J. Jackson (2011), The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system, *Journal of Hydrometeorology*, *12*, 750-765, doi:10.1175/JHM-D-10-05000.1.

Reichle, R. H., and Q. Liu (2014), Observation-Corrected Precipitation Estimates in GEOS-5, *NASA Technical Report Series on Global Modeling and Data Assimilation, NASA/TM-2014-104606, Vol. 35*, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 18pp.

Reichle, R. H., G. J. M. De Lannoy, B. A. Forman, C. S. Draper, and Q. Liu (2014a), Connecting Satellite Observations with Water Cycle Variables through Land Data Assimilation: Examples Using the NASA GEOS-5 LDAS, *Surveys in Geophysics*, *35*, 577-606, doi:10.1007/s10712-013-9220-8.

Reichle, R. H., R. Koster, G. De Lannoy, W. Crow, and J. Kimball (2014b), SMAP Level 4 Surface and Root Zone Soil Moisture Data Product: L4_SM Algorithm Theoretical Basis Document (Revision A), Soil Moisture Active Passive (SMAP) Mission Science Document. JPL D-66483, Jet Propulsion Laboratory, Pasadena, CA.

Reichle, R. H., R. A. Lucchesi, J. V. Ardizzone, G.-K. Kim, E. B. Smith, and B. H. Weiss (2015b), Soil Moisture Active Passive (SMAP) Mission Level 4 Surface and Root Zone Soil Moisture (L4_SM) Product Specification Document. GMAO Office Note No. 11 (Version 1.4), 82 pp., NASA Goddard Space Flight Center, Greenbelt, MD, USA.

Reichle, R. H., G. J. M. De Lannoy, Q. Liu, A. Colliander, A. Conaty, T. Jackson, J. Kimball, and R. D. Koster (2015b), Soil Moisture Active Passive (SMAP) Project Assessment Report for the Beta-Release L4_SM Data Product, *NASA Technical Report Series on Global Modeling and Data Assimilation*, *NASA/TM-2015-104606*, *Vol. 40*, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA, 63pp.