Hydrology and
Earth System
Sciences

Open Access

EGU

Discussions

1   **The added value of brightness temperature assimilation for the**
2   **SMAP Level-4 surface and root-zone soil moisture analysis over**
3   **mainland China**

4   Jianxiu Qiu[1,2], Jianzhi Dong[3], Wade T. Crow[3], Xiaohu Zhang[4,5], Rolf H. Reichle[6], Gabrielle J.
5   M. De Lannoy[7]

6   [1]Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun
7   Yat-sen University, Guangzhou, 510275, China

8   [2]Southern Laboratory of Ocean Science and Engineering (Guangdong, Zhuhai), Zhuhai, 519000, China

9   [3]USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705, USA

10   [4]National Engineering and Technology Center for Information Agriculture, Nanjing Agricultural University, Nanjing,
11   China

12   [5]Jiangsu Key Laboratory for Information Agriculture, Nanjing Agricultural University, Nanjing, China

13   [6]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

14   [7]Department of Earth and Environmental Sciences, KU Leuven, Heverlee, Belgium

15

16   *Correspondence to*: Jianxiu Qiu (qiujianxiu@mail.sysu.edu.cn)

Abstract. The Soil Moisture Active Passive (SMAP) Level-4 Surface Soil Moisture and Root-Zone Soil Moisture (L4) product provides global estimates of surface soil moisture (SSM) and root-zone soil moisture (RZSM) via the assimilation of SMAP brightness temperature (Tb) observations into the Catchment Land Surface Model (CLSM). Here, using in-situ measurements from 2474 sites in mainland China, we evaluate the performance of soil moisture estimates from L4 and from a baseline "open-loop" (OL) simulation of CLSM without Tb assimilation. Using random forest regression, the efficiency of the L4 ~~data assimilation~~ (DA) system (i.e., the performance improvement in L4 relative to OL) is attributed to 8 control factors related to the land surface modelling (LSM) and radiative transfer modeling (RTM) components of the L4 system. Results show that 77% of the 2287 9-km EASE grid cells in mainland China that contain at least one ground station exhibit an increase in the Spearman rank correlation skill (R) with in-situ measurements for L4 SSM compared to that of OL, with an average R increase of approximately 14% (ΔR = 0.056). RZSM skill is improved for about the same percentage of 9-km EASE grid cells, but the average R increase for RZSM is only 7% (ΔR = 0.034). Results further show that the SSM DA efficiency is most strongly related to the error in Tb observation space, followed by the error in precipitation forcing and microwave soil roughness. For RZSM DA efficiency, the three dominant control factors remain the same, although the importance of soil roughness exceeds that of the Tb error. For the skill of the L4 and OL estimates themselves, the top control factors are the precipitation error and the SSM-RZSM coupling strength error (in descending order of factor importance for R_{OL}), both of which are related to the LSM component of the L4 system. Finally, we find that the L4 system can effectively filter out errors in precipitation. Therefore, future development of the L4 system should focus on improving the characterization of the SSM-RZSM coupling strength.

Keywords. SMAP Level 4, soil moisture, data assimilation, attribute analysis, random forest regression

1 Introduction

Soil moisture modulates water and energy feedbacks between the land surface and the lower atmosphere by determining the partitioning of incoming net radiation into latent and sensible heat (Seneviratne et al., 2010, 2013). High-quality, global-scale soil moisture products have become increasingly available in recent years (Gruber et al., 2020). In particular, the L-band NASA Soil Moisture Active Passive (SMAP) satellite mission (Entekhabi et al., 2010; Piepmeier et al., 2017) has significantly improved the skill of available, global-scale soil moisture products. However, the SMAP observations contain temporal data gaps and are only representative of conditions within the top 5 cm of the vertical soil moisture column. To address these limitations, the SMAP Level-4 Surface and Root-Zone Soil Moisture (L4) algorithm assimilates SMAP brightness temperature (Tb) observations into the NASA Catchment Land Surface Model (CLSM) to derive an analysis of surface (0–5 cm) and root-zone (0–100 cm) soil moisture estimates with global, 3-hourly coverage (Reichle et al., 2017a; Reichle et al., 2017b; Reichle et al., 2019).

49 However, the performance of a land data assimilation (DA) system is sensitive to its parameterization and requires

50 careful assessment. For instance, Reichle et al. (2008) demonstrate that DA based on incorrect assumptions of modeling

51 and observation errors can degrade soil moisture estimates, compared with the case of not performing any DA.

52 Theoretically, the optimality of DA can be evaluated using so-called innovations, or observations-minus-forecast

53 residuals; however, an investigation of the innovations alone is often insufficient to determine if the soil moisture

54 analysis is optimal (Crow and Van Loon, 2006).

55 Recently, Dong et al. (2019a) proposed a novel statistical framework for evaluating the performance of a soil moisture

56 DA system. Specifically, they demonstrated that the relative skill of surface soil moisture (SSM) estimates acquired

57 with and without DA can be estimated using the ratio of their correlations with just one noisy but independent ancillary

58 remote sensing product. This approach was applied to the SMAP L4 system using ASCAT soil moisture retrievals.

59 Their results show that the added value of SMAP DA is closely related to both rain gauge and vegetation density.

60 However, due to the limited availability of independent root-zone soil moisture (RZSM) products for performing

61 statistical error estimation, this method is only applicable for SSM estimates.

62 Relative to SSM, the efficiency of assimilating land surface observations to improve RZSM is complicated by model

63 structural error that affects the ability of the DA to update unobserved model states. For instance, Kumar et al. (2009)

64 identified the surface–root zone coupling strength, which is the result of a model-dependent representation of processes

65 related to the partitioning of rainfall into infiltration, runoff, and evaporation components, as an important factor for

66 determining RZSM improvement associated with the assimilation of SSM retrievals. Their synthetic experiments

67 suggest that – faced with unknown true subsurface physics – overestimating the surface–root zone coupling in the land

68 model is a more robust strategy for obtaining skill improvements in the root zone than under-estimating the coupling.

69 Likewise, Chen et al. (2011) suggested that their Soil and Water Assessment Tool significantly under-predicts the

70 magnitude of vertical soil water coupling in the Cobb Creek Watershed in southwestern Oklahoma, USA, and this lack

71 of coupling impedes the ability of DA to effectively update deep-layer soil moisture, groundwater flow and surface

72 runoff. In the context of the present paper, the evaluation of L4 RZSM estimates has been limited to relatively few

73 SMAP core validation and sparse network sites (Reichle et al., 2017a; Reichle et al., 2017b; Reichle et al., 2019). With

74 such limited sample sizes, the RZSM skill of the L4 product at the global scale remains uncertain.

75 The primary objective of this study is to determine the DA efficiency, i.e., performance improvement in DA results

76 relative to the open-loop (OL) baseline of the L4 product, as a function of a variety of system aspects, including errors

77 in CLSM forcing (e.g., precipitation), errors in key CLSM parameters (e.g., relating to vegetation), mean errors in

78 CLSM structure (e.g., surface and root-zone coupling), and errors in the radiative transfer modeling (RTM) that links

79 the modeled soil moisture and temperature estimates to the observed Tb.

80 To this end, we first evaluate the performance of L4 SSM and RZSM estimates using a very large number ($n = 2474$)

81 of soil moisture profile measurement sites (generally acquired at sub-surface depths between 10 and 50 cm) within

82 mainland China. Next, the in-situ measurements are used to assess the DA efficiency of the L4 system, which is defined

83 as the skill difference between the L4 estimates and model-only estimates derived without SMAP Tb assimilation.

84 Additionally, we apply a machine-learning technique to quantify by how much various control factors drive the spatial

85 variations in the efficiency of the L4 system. In this way, we seek to prioritize future enhancements to the L4 system.

86 **2 Data and Methods**

87 This section briefly describes the SMAP L4 soil moisture product (Section 2.1), the extensive network of in-situ soil

88 moisture observations over mainland China (Section 2.2) and the ancillary data sources and metrics used in the skill

89 assessment (Sections 2.3 and 2.4). Next, we introduce the double instrumental variable (IVd) method employed to

90 determine the errors in control factors that cannot be determined using ground observations (Section 2.5). Finally, we

91 describe the random forest (RF) regression method used to identify the main factor(s) (out of the 8 control factors from

92 both CLSM and RTM aspects) that affect the spatial variations in SMAP L4 DA efficiency and L4 performance

93 (Section 2.6).

94 **2.1 SMAP L4 soil moisture product**

95 The SMAP L4 soil moisture product (version 4; Reichle et al., 2019) is generated by assimilating the SMAP L1C

96 Radiometer half-orbit 36 km EASE-Grid brightness temperature (Tb) observations (Version 4 SPL1CTB; Chan et al.,

97 2016) into the CLSM. The SMAP Tb observations are assimilated at 3-h intervals using a spatially distributed, 24-

98 member ensemble Kalman filter (Reichle et al. 2017b). The surface meteorological forcing data are from the global

99 Goddard Earth Observing System (GEOS) Forward Processing atmospheric analysis (Lucchesi, 2013), with

100 precipitation corrected using the daily, 0.5-degree, gauge-based Climate Prediction Center Unified (CPCU) product

101 (Xie et al. 2007). The L4 product provides global, 9-km, 3-hourly surface (0–5 cm) and root-zone (0–100 cm) soil

102 moisture estimates along with related land surface fields and analysis diagnostics. For the present study, we aggregated

103 all soil moisture estimates to daily-average (00:00 to 23:59 UTC) data. A baseline, model-only, ensemble CLSM

104 simulation without the assimilation of SMAP Tb observations (but using the same perturbations as in the L4 system)

105 is referred to as the "open-loop" (OL) run.

106 The SMAP L4 assimilation system includes a zero-order "tau-omega" forward RTM (De Lannoy et al., 2013) that

107 converts SSM and surface soil temperature into L-band brightness temperature estimates. Selected parameters of the

108 L4 RTM, including microwave soil roughness parameters, a vegetation structure parameter, and the microwave

109 scattering albedo, were calibrated using multi-angular L-band brightness temperature observations from the Soil

110 Moisture Ocean Salinity (SMOS) mission (De Lannoy et al., 2014). The L4 RTM parameterizes microwave soil

111 roughness as a function of SSM (De Lannoy et al., 2013, their equation B1). Here, we used this parameterization to

112 compute the 2017-2018 time-averaged microwave soil roughness estimates as one potential indicator of DA efficiency

113 (Section 2.3). The necessary parameters were obtained from L4 "Land-Model-Constants" output Collection (last

114 access: 8 July 2020; DOI: https://doi.org/10.5067/KGLC3UH4TMAQ; Reichle et al., 2018a). The L4 "Analysis-

115 Update-Data" output Collection includes RTM predictions of Tb and the assimilated SMAP Tb observations (last

116 access: 8 July 2020; DOI: https://doi.org/10.5067/60HB8VIP2T8W; Reichle et al., 2018b).

117   To avoid the impact of seasonality, we performed our analysis using anomaly time series, derived by subtracting a

118   seasonally-varying (daily) climatology from each raw time series. The climatology of a given time series was obtained

119   by sampling the mean value of all soil moisture estimates that fall within a 31-day moving window centered on a

120   particular day-of-year. Moreover, L4 estimates of land latent heat flux (LE), land sensible heat flux (SH) and the

121   climatological LAI inputs to CLSM and the RTM, were obtained from the L4 "Geophysical-Data" output Collection

122   (last access: 6 April 2020; DOI: https://doi.org/10.5067/KPJNN2GI1DQR; Reichle et al., 2018c). These datasets were

123   also used to compute control factors to explain spatial variations in the DA efficiency of the L4 system (Section 2.3).

124   **2.2 Soil moisture validation data**

125   In-situ soil moisture measurements during 2017 and 2018 were collected from a national network of Chinese

126   Automatic Soil Moisture Observation Stations (CASMOS) maintained by the Chinese Meteorological Administration

127   (CMA). In total, soil moisture measurements from 2474 separate stations arrayed across mainland China, and covering

128   different land use types, were collected. At each CASMOS site, frequency domain reflectometry-based instruments

129   were used to record hourly volumetric soil moisture content within the following vertical depth ranges: 0–10, 10–20,

130   20–30, 30–40, and 40–50 cm below the surface. These hourly estimates (at multiple depths) were then aggregated into

131   daily values and linearly averaged (vertically) to produce 0-10 cm (SSM) and 0-50 cm (RZSM) in situ soil moisture

132   measurements – which were subsequently used to validate the L4 and OL SSM (0-5 cm) and RZSM (0-100 cm)

133   estimates. Note that Spearman correlation rather than Pearson correlation is used for L4 and OL validation, in order to

134   avoid impact of outliers in the time series and prior assumptions about soil moisture distributions.

135   Ground observations falling within the same 9-km EASE grid were averaged for comparisons against the collocated

136   9-km L4 and OL soil moisture estimates. A total of 2287 individual 9-km EASE grid cells within mainland China are

137   included in the analysis. Among them, 92.35% of grid cells contain one in-situ site, 7.26% contain two sites, 7 grid

138   cells contain three sites, and the remaining two grid cells contain four and five sites respectively.

139   **2.3 Explanatory data products**

140   As discussed above, our hypothesis is that the efficiency of the SMAP L4 system will be sensitive to the ability of the

141   ensemble-based L4 analysis in filtering errors that exist in the OL (that is, CLSM), in the model forecast Tb (that is,

142   the RTM), and in the SMAP Tb observations. We therefore considered two separate categories of factors that

143   potentially control spatial variations in DA efficiency. The factors are summarized in Table 1.

144   The first category represents a range of factors known to affect the skill of soil moisture estimates derived from LSM

145   (in this case, CLSM). The five control factors in this category are: i) the error in precipitation forcing, ii) the error in

146   (input) LAI, iii) the error in (output) LE, iv) the magnitude of mean error in CLSM SSM-RZSM coupling strength,

147   and v) the presence of vertical variability in soil properties (defined as the difference in clay fraction across the vertical

148   soil profile). Note that such variability represents a potential source of error because CLSM assumes that soil texture

149   and the associated soil parameters are vertically homogeneous within the soil column, with the exception of some

150   surface-layer moisture transport parameters. The soil texture information is from Harmonized World Soil Database

151   (HWSD) v1.2.

152   The second category contains three factors that affect radiative transfer modeling (RTM) and therefore DA updates.

153   These include: i) estimates of the joint error in SMAP Tb observations and RTM Tb simulations, ii) the magnitude of

154   microwave soil roughness, and iii) the magnitude of LAI (as a proxy for the vegetation optical depth at microwave

155   frequencies, which modulates the sensitivity of the observed Tb to SSM conditions).

156   The control factors take a variety of forms. Some factors are based on estimates of the errors fed into the L4 system as

157   (e.g., the error in CLSM rainfall forcing data). Other factors consist of the magnitude of the variable itself (e.g., the

158   vertical variability of clay fraction). Note that LAI is used in both ways: LAI error is used to predict OL performance

159   (because LAI is an important input into CLSM) while mean LAI is used to explain DA performance (because increased

160   LAI is associated with decreased soil moisture information content in microwave observations).

161   Note that the LAI used in the L4 system is a climatology derived from satellite observations of the Normalized

162   Difference Vegetation Index. Therefore, to indicate the magnitude by which each grid cell's LAI typically deviates

163   from its long-term climatology, we use the temporal standard deviation of anomaly time series of the benchmark LAI

164   (from SPOT VGT product) as a measure of the error in the LAI used in L4. Owing to the lack of reference Tb

165   observations at similar satellite overpass times and locations, Tb errors are gauged using the time series standard

166   deviation of the observation-minus-forecast (O-F) Tb residuals, which indicate the typical misfit between the model

167   forecast Tb and the (rescaled) SMAP Tb observations. This metric measures the total error in Tb observation space.

168   The exact data sets and the metrics utilized for evaluating these 8 control factors are summarized in Table 1.

169

Table 1 Benchmark data sets and metrics used for evaluating control factors of SMAP L4

Factor category	Control factor	Dataset/Benchmark	Temporal resolution	Spatial resolution	Data range	Metrics
	Precipitation error	Rain gauge (CGDPA)	daily	0.25 °	2017-2018	Spearman's rank correlation R
	SSM-RZSM coupling strength error	CASMOS	daily	NA	2017-2018	ΔCP (see Section 2.4)
LSM	Vertical variability of clay fraction	HWSD	NA	9 km	NA	Difference in clay fraction between topsoil (0-30 cm) and root-zone (0-100 cm) layers
	SMAP L4 LAI error	SPOT LAI	10 d	1 km	2017-2018	Temporal standard deviation of SPOT VGT LAI anomaly
	LE error	FLUXCOM	daily	(1/120) °	2017-2018	IVd-based R
	Tb error	SMAP L4	daily	9 km	2017-2018	Temporal standard deviation of O-F Tb residuals
RTM	Microwave soil roughness	SMAP L4	daily	9 km	2017-2018	Temporal average based on De Lannoy et al. (2013)
	Annual mean LAI	MODIS/Geoland-based product	daily	9 km	2017-2018	Climatological mean

170

2.3.1 Gauge-based precipitation gridded product

Errors in the GEOS precipitation data used to force the CLSM within the SMAP L4 system were estimated via Spearman's rank correlation with available rain-gauge observations. These network observations are based on an analysis of ∼2400 rain gauge stations distributed unevenly over mainland China. Recently, the China Gauge-based Daily Precipitation Analysis (CGDPA) with a spatial resolution of 0.25 °×0.25 °based on this network was constructed and has been made operational over mainland China. CGDPA uses a modified interpolation method of climatology-based optimal interpolation (OI) with topographic correction proposed by Xie et al. (2007). In this process, daily precipitation climatology over mainland China is optimized and is rebuilt using the 30-year average precipitation observations from ∼2400 gauges of the period 1971–2000 (Shen et al., 2010). CGDPA is shown to have smaller bias and root mean square error than the CPCU product used in L4, which is based on fewer than 400 gauge sites over mainland China (Shen et al., 2015).

2.3.2 FLUXCOM LE estimates

The FLUXCOM ensemble of global land-atmosphere energy fluxes was used to evaluate the error of the L4 LE estimates. This ensemble merges energy flux measurements from FLUXNET eddy covariance towers with remote sensing and meteorological data based on a machine learning method to estimate global gridded net radiation, latent and sensible heat and their related uncertainties (Jung et al., 2019). The resulting FLUXCOM database has a 0.0833 ° spatial resolution when applied using MODIS remote sensing data. The monthly energy flux data of all ensemble members, as well as the ensemble estimates from the FLUXCOM initiative, are freely available (CC4.0 BY license) from the Data Portal (http://fluxcom.org/), while the daily- and 8-day FLUXCOM products are available upon request from dataset provider Martin Jung. To calculate the LE error, we've collected the daily, high spatial resolution FLUXCOM product and extracted the estimates where in-situ soil moisture sites located.

2.3.3 SPOT VGT LAI

The data set used as a benchmark for assessing leaf area index (LAI) errors present in the SMAP L4 analysis was derived from SPOT/VEGETATION and PROBA-V LAI products (version 2) that are generated every 10 days at spatial resolution of 1 km. The SPOT LAI version 2 product capitalizes on the development and validation of already existing products: CYCLOPES version 3.1 and MODIS collection 5 and the use of neural networks (Baret et al., 2013; Verger et al., 2008). The version 2 products are derived from top of canopy daily (S1-TOC) reflectances instead of normalized top of canopy 30-day composited reflectances as in the version 1. Compared to version 1, the compositing step is performed at the biophysical variable level instead of reflectance level. This ensures reduced sensitivity to missing observations and avoids the need for a BRDF model.

2.3.4 HWSD soil texture

The HWSD attribute database (v1.2) is a 30 arc-second raster database with 15773 different soil-mapping units. It provides information on the standardized soil parameters for topsoil (0–30cm) and subsoil (30-100 cm) separately. In

204   this study, we use the difference of clay fractions between topsoil (0-30cm) and the aggregated 0-100cm layer to

205   measure the vertical clay fraction variation at each 9-km grid cell.

2.4 Vertical coupling metric

207   The RZSM time series generally show decreased temporal dynamics relative to SSM. As a result, overestimated SSM-

208   RZSM coupling tends to spuriously increase the (correlation-based) similarity of SSM and RZSM time series, and

209   thereby, overestimate RZSM temporal variability. Therefore, analogous to Kling-Gupta efficiency (Gupta et al., 2009),

210   we defined the SSM-RZSM coupling strength (CP) as:

$$CP = 1 - \sqrt{(R\text{-}1)^2 + (\alpha\text{-}1)^2} \qquad (1)$$

211   where R is the Spearman's rank correlation between SSM and RZSM, and α is the ratio of temporal standard deviation

212   of SSM to that of RZSM. A CP value of one represents the extreme case where RZSM is identical to SSM, i.e., a

213   strongly coupled case. Likewise, a CP of zero represents the opposing case of completely uncoupled time series. Cases

214   with negative CP do not exist.

215   Observed CP (CP_{obs}) was based on comparisons between 0-10 cm "surface" estimates and 0-50 cm "root-zone" in situ

216   observations and used as a benchmark. In contrast, SMAP L4 CP estimates (CP_{OL}) was based on the comparison of 0-

217   5 cm "surface" estimates and 0-100 cm "root-zone" estimates. Therefore, the surface versus root-zone storage contrast

218   in the observation time series is less than that of the L4 estimates. This will likely cause the observed correlation

219   between surface and root-zone time series to be systematically higher than the analogous vertical correlation

220   calculation for L4 estimates. However, this bias is partially corrected for by the second term in Eq. (1) – since the

221   observed α ratio will, by the same token, tend to be smaller (i.e. closer to one) than α sampled from the L4 analysis.

222   Such ability to compensate for vertical depth differences is a key reason we apply CP, rather than *simple correlation*,

223   as a vertical coupling strength metric. Nevertheless, it should be noted that our main interest here lies in describing

224   spatial variations in (CP_{OL} - CP_{obs}) and care should be taken when interpreting raw (CP_{OL} - CP_{obs}) differences as an

225   *absolute* measure of L4 vertical coupling bias.

2.5 Double instrumental variable (IVd) method

227   The benchmark data set of FLUXCOM LE described above contains error that is (likely) of a similar order of

228   magnitude as the L4 LE dataset it is applied to evaluate. Therefore, in an attempt to correct for the impact of this error,

229   the LE error used here as a control factor is obtained via a double instrumental variable (IVd; Dong et al., 2019b)

230   analysis approach that minimizes the spurious impact of random errors in benchmark data sets. As shown in Dong et

231   al. (2019b), for the evaluation of two time series with auto-correlation in both of them, IVd is more robust than single

232   instrumental variable based algorithm, therefore we apply IVd to evaluate the LE error.

Hydrology and
Earth System
Sciences
Discussions

Open Access

EGU

233   IVd is a modified version of triple collocation (TC) analysis. In TC analysis (McColl et al., 2014), geophysical

234   variables obtained from three independent sources (x, y and z) are assumed to be linearly related to the true signal P

235   as:

$$x = \alpha_x P + B_x + \varepsilon_x \tag{2}$$

236   where the α_x is a scaling factor; B_x is a temporal constant bias and ε_x is zero-mean random error.

237   As opposed to the TC method, IVd uses only two independent products (x, y) to characterize geophysical data product

238   errors. This method introduces two instrumental variables (I and J, i.e., $I_t = \alpha_x P_{t-1} + B_x + \varepsilon_{xt-1}$, $J_t = \alpha_y P_{t-1} + B_y + \varepsilon_{yt-1}$),

239   which are based on the lag-1 (day) time series (at day t) of x and y, respectively. Therefore, assuming that the errors of

240   two independent products are serially white, the covariance between instrumental variables and products can be written

241   as follows:

$$C_{Ix} = \alpha_x^2 L_{PP} \tag{3}$$

$$C_{Jy} = \alpha_y^2 L_{PP} \tag{4}$$

242   where C represents the covariance of the subscript products. For instance, C_{Ix} represents the covariance of x and its

243   instrumental variable I. Variable L_{PP} is the lag-1 auto-covariance of the true signal. Combining Eqs. (3) and (4), the

244   scaling ratio s_{ivd} of the two products x and y can be written as:

$$s_{ivd} = \sqrt{\frac{C_{Ix}}{C_{Jy}}} \tag{5}$$

245   Based on Eq. (5), their correlation with truth can be estimated as:

$$R_{Px}^2 = \frac{C_{xy} s_{ivd}}{C_{xx}} \tag{6}$$

$$R_{Py}^2 = \frac{C_{xy}}{C_{yy} s_{ivd}} \tag{7}$$

246   In this way, the error in the L4 LE (measured by IVd-based correlation with truth) can be estimated robustly using the

247   FLUXCOM LE product described in Section 2.3.2.

248   **2.6 Random forest regression**

249   A random forest (RF) regression approach was used to rank and quantify the importance of the 8 control factors

250   introduced above (Table 1) for describing spatial patterns in DA efficiency for both SSM and RZSM estimates. The

251   RF method is a supervised learning algorithm based on an averaged ensemble of decision trees (Breiman, 2001). Unlike

252   linear regression approaches, RF can capture non-linear interactions between the features and the target. In addition,

253   the normalization (or scaling) of data is not necessary in RF application. Another advantage of the RF algorithm is that

254   it can readily measure the relative importance of each feature on the estimates, which makes it highly suitable for an

255   attribution analysis. Therefore, based on the output of RF, key control factors determining the efficiency of SMAP

256   DA were evaluated and ranked. The RF estimates are based on a 10-fold cross-validation approach.

257   **3 Results**

258   **3.1 Validation of SMAP L4 and OL estimates of SSM and RZSM anomalies**

259   Figure 1 maps validation results (i.e., anomaly Spearman's rank correlation with in-situ observations, R) for SMAP L4

260   and associated OL soil moisture estimates. The skill patterns for OL and L4 are, in general, quite spatially consistent.

261   Both are characterized by an increasing trend of SSM estimation skill moving from northwest to southeast China (Fig.

262   1a and 1b). In relative terms, the L4 product surpasses the baseline OL's SSM skill within 77% of the 2287 9-km

263   EASE grid cells containing ground observations – with a mean R increase of $\Delta R = 0.056$ [-] and mean relative

264   improvement versus R_{OL} of 14%.

265   Similar spatial patterns are observed for RZSM skill. As with SSM, generally higher consistency with in-situ RZSM

266   measurements is found in southeast China relative to northern China. However, relative to SSM, the added value of

267   SMAP data assimilation (i.e. L4) is reduced for RZSM and the mean relative R improvement falls to 7% ($\Delta R = 0.034$

268   [-]) (compare Fig. 1e and 1f). This is not surprising since assimilated SMAP Tbs are primarily sensitive to soil moisture

269   conditions in the surface (0-5 cm) layer.
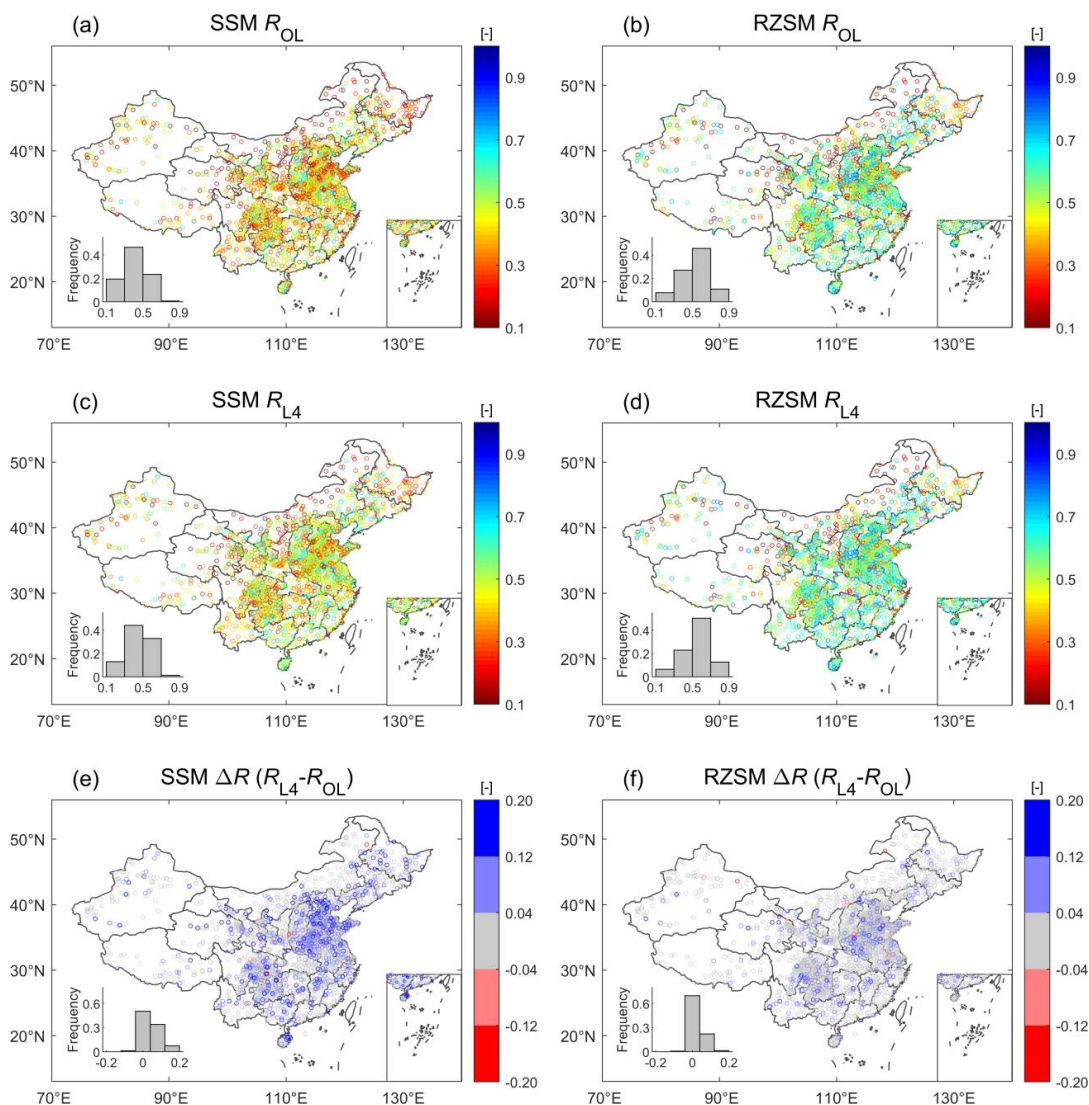
Figure 1: OL (a, b) and L4 (c, d) skills (R values) for SSM (left column) and RZSM (right column). DA efficiency ($\Delta R = R_{L4}$ - R_{OL}) for (e) SSM and (f) RZSM. Blue (red) colors in (e) and (f) indicate grid cells where L4 estimates are better (worse) than OL. Non-significant differences (based on a 1000-member bootstrapping analysis) are colored grey. The lower left inset in each subplot indicates the frequency of binned R-values across all 9-km EASE grid cells containing ground observations.

3.2 Spatial distribution of potential factors controlling SMAP L4 DA performance

As described in Section 2.3, we selected 8 control factors that potentially influence the skill of SMAP L4 soil moisture estimates. Using the attribution analysis described in Section 2.6, these factors will be used to explain the spatial

279   variations in skill and DA efficiency seen in Fig. 1. As a first step, this section examines the spatial patterns inherent
280   in the 8 control factors. Errors in the CLSM precipitation forcing are relatively higher in northern and northwestern
281   areas of China (Fig. 2a), where the gauge density is generally more sparse than southern China. Among the factors
282   representing CLSM structural errors, a pre-dominantly negative bias is observed in SSM-RZSM coupling strength
283   generally across China (i.e., lower CP_{OL} compared to CP_{obs}), while a very small number of grid cells show a positive
284   coupling strength bias in eastern China (dark green dots in Fig. 2b). This is expected since at the coarse resolution, the
285   model's vertical coupling strength should be much less than at any single point. In addition, this may be partly
286   attributed to the layer depths differences, since CLSM represents surface and root-zone depths of 0-5 cm and 0-100
287   cm, whereas the corresponding in-situ observations represent the 0-10 cm and 0-50 cm layers, and it can be expected
288   that CP_{OL} should thus be smaller than CP_{obs}. In addition, the vertical variability of the clay fraction seems to show little
289   spatial variation across mainland China (Fig. 2c). With respect to CLSM LAI error, regions in southern China that
290   have generally higher LAI show larger standard deviation in SPOT LAI time series (Fig. 2d and 2h). The IVd-based
291   estimates of SMAP L4 LE error, which represent a potential control factor for water-balance errors in CLSM, generally
292   show low-level of error across mainland China (Fig. 2e).

293   For O-F Tb residuals describing RTM-related error, a higher standard deviation of O-F Tb residuals is observed in the
294   North China Plain (Fig. 2f), which is very consistent in spatial distribution with areas displaying the highest and most
295   significant SSM prediction improvement (Fig. 1c). This is expected, as mentioned above, because O-F Tb residuals
296   are the basis for the soil moisture corrections (or increments) that are applied in the DA system as part of the L4
297   analysis. The 2017-2018 mean of soil roughness and the 2017-2018 mean LAI show higher values in southwest and
298   southeast China (Fig. 2g-h).

Hydrology and
Earth System
Sciences
Discussions

Open Access
EGU
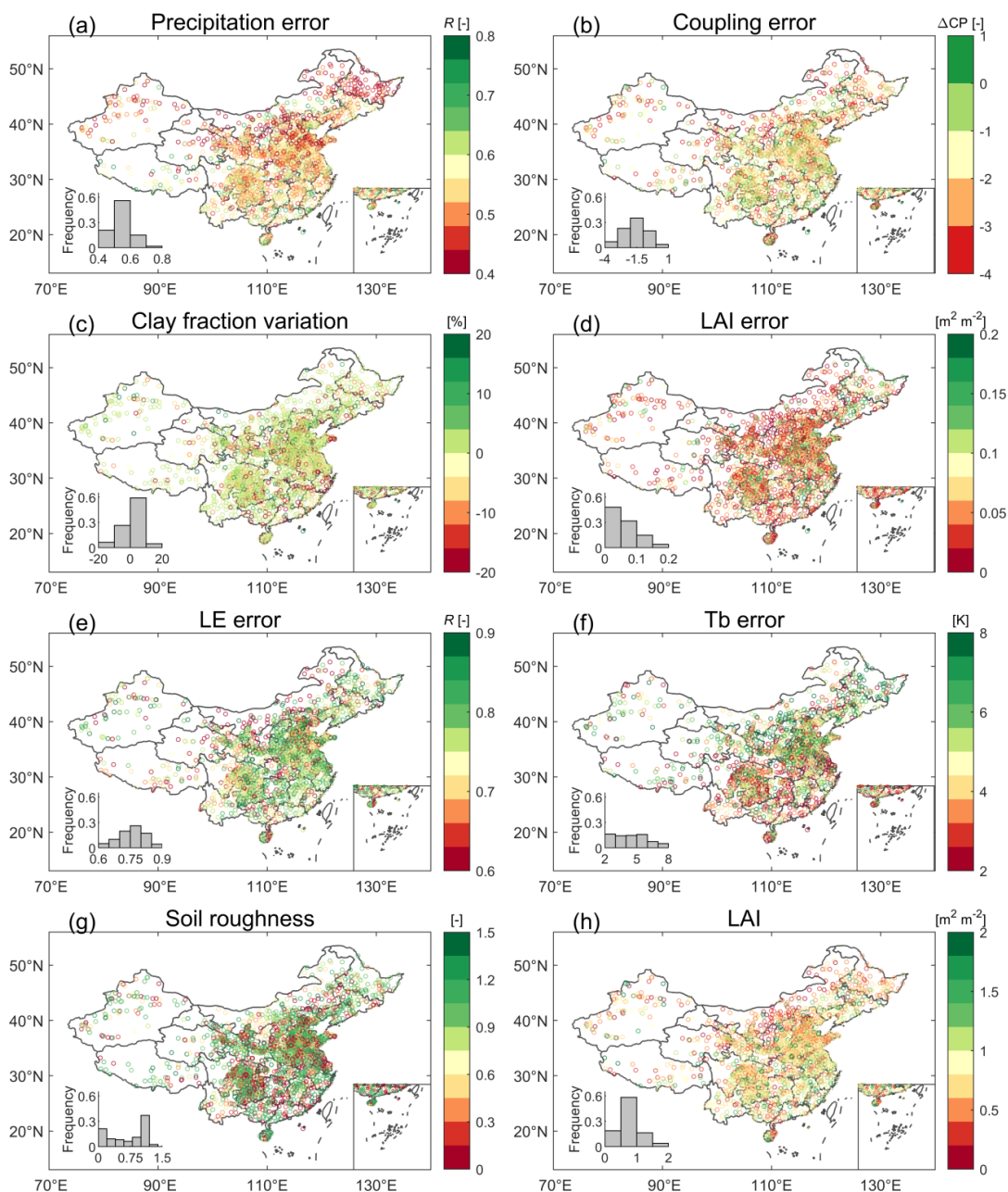
Figure 2: Factors potentially influencing SMAP L4 performance over mainland China: (a) CLSM precipitation error measured by the Spearman's rank correlation between CLSM precipitation and ground observations; (b) SSM-RZSM coupling strength error (CP_{OL} minus CP_{obs}); (c) clay fraction variation (difference) across the soil profile; (d) error in LAI input to L4; (e) IVd-based error of LE from L4; (f) Tb error; (g) L4 microwave soil roughness; (h) climatology mean of LAI input to L4.

305

3.3 Attribution of SMAP L4 versus OL performance to control factors

3.3.1 Attribution using random forest regression

As mentioned above, RF regression was used to identify the relative importance of our 8 control factors for determining the efficiency of SMAP L4 DA (i.e., $\Delta R = R_{L4} - R_{OL}$), and also L4 ($R_{L4}$) and OL performances ($R_{OL}$). To start, we first investigate the robustness of RF for predicting ΔR. To estimate the magnitude of randomness in the RF algorithm, we use 50 bootstrap runs. As shown in Fig. 3a, the 10-fold cross-validation test (228 validation samples) shows that the predicted and in-situ-based ΔR have a mean correlation of 0.72 and 0.46 for SSM and RZSM, respectively.

Given the sampling errors of ΔR, which is based on a two-year validation period, and the relatively low spatial variability in RZSM skill (Figs. 1f), the performance of RF is acceptable. In addition, ground-measurement upscaling error is likely a significant contributor to unexplainable spatial variability for ΔR in Fig. 1. In fact, Chen et al. (2016) found large spatial variability in the ability of point-scale SSM ground observations to describe grid cell-scale SSM dynamics. In-situ observations sites associated with larger upscaling errors will introduce a spurious low bias into sampled estimates of ΔR values (see Appendix B in Dong et al., 2020). Therefore, some of the ΔR spatial variability observed in Fig. 1 is unrelated to any aspect of the L4 system and is therefore unexplainable via the 8 control factors we have selected.

321
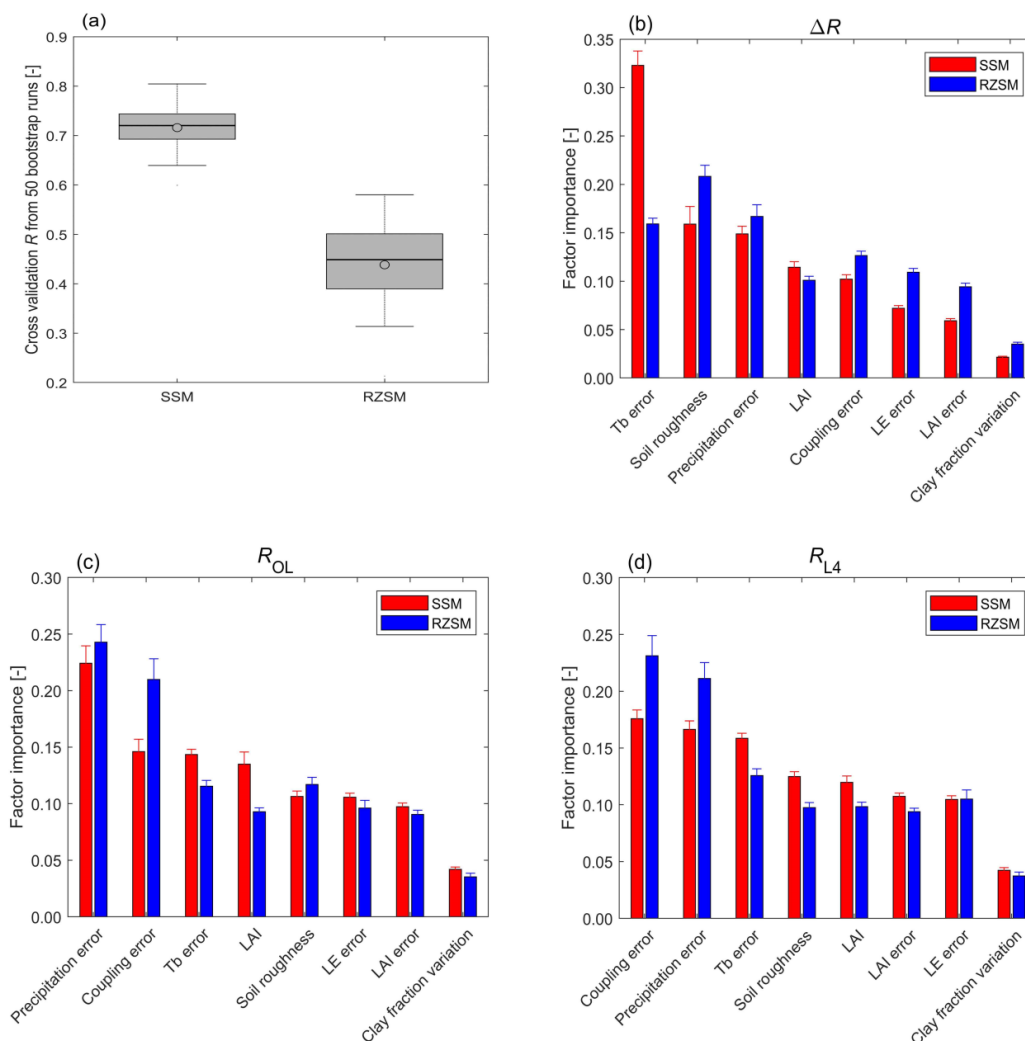
Figure 3: Attribution analysis of SMAP L4 DA efficiency: (a) Cross-validation of RF regression method in predicting DA efficiency ΔR= R_{L4} - R_{OL} based on our 8 control factors (Table 1). Relative importance of 8 control factors determining spatial patterns in (b) DA efficiency (ΔR), (c) OL performance (R_{OL}), and (d) L4 performance (R_{L4}). Red (blue) bars represent predictor importance for SSM (RZSM). Error bars reflect the standard deviation from 50-member bootstrapping of the RF importance estimates.

Based on the RF results, the Tb error is quantified as the most prominent factor in determining DA efficiency (i.e., ΔR = R_{L4} - R_{OL}) – followed by precipitation error and microwave soil roughness (Fig. 3b). The RF-derived ranking of control-factor importance for RZSM is similar to that of SSM in that the same three factors are still the most

332   explanatory. However, in contrast to SSM, the importance of Tb error for RZSM decreased dramatically from >30%

333   to ~15%. Other modeling error sources (e.g., the vertical variability of soil properties) have only very limited impact

334   on SMAP DA improvement.

335   As seen in Fig. 3c, for the OL performance (R_{OL}), the most important factors identified by RF include precipitation

336   error, SSM-RZSM coupling error, and Tb error (microwave soil roughness) for SSM (RZSM). Note that although the

337   Tb error is identified as third important factor for R_{OL} in SSM skill, this is an instance where there is correlation (poorer

338   skill happens to coincide with higher Tb error), but this does not imply a causal relationship. Specifically, it is normal

339   that Tb (O-F) errors are higher where the OL performs worse, but a high Tb error is not the cause of a low OL

340   performance. The same applies to the relationship between microwave soil roughness and OL skill for RZSM

341   estimation. The SMAP L4 system is able to reduce the predominant impact of precipitation errors on both SSM and

342   RZSM estimation skill, rendering SSM-RZSM coupling error the most important factor for R_{L4} (Fig. 3d). In addition,

343   in the L4 system, the high vegetation density effect on SSM and RZSM estimation is clearly reduced, as the fourth

344   most important factor of LAI is replaced by Tb error.

345   The qualitative rankings provided by the RF analysis in Fig. 3 are relatively robust to our particular choice of

346   benchmark data set to define the 'error' of various control variables. For instance, we replaced the CGDPA

347   precipitation benchmark with the CMORPH-merge product (Version 1, last access: 6 April 2020; DOI:

348   https://doi.org/10.25921/w9va-q159; Xie et al., 2019), which is the 0.1 degree merging product of CMORPH and

349   observations from more than 30,000 automatic weather stations in mainland China. For this case, the predictive power

350   of the regression model established by the RF is not affected (similar to Fig. 3a), and the qualitative rankings of the

351   precipitation error in R_{OL} and R_{L4} are not impacted (similar to Fig. 3c-d).

352

3.3.2 Attribution using box plot comparisons

353

354   As stated in Section 2.5, the RF method is adept at summarizing the impact of multiple (co-varying) control factors

355   simultaneously in the established regression model and thus provides more comprehensive insights than the

356   examination of how the target variable (DA improvement) fluctuates with each individual control factor. However, it

357   does not allow the investigation of the sign of the relationship between DA improvement and each control factor –

358   which is important for understanding exactly how each factor influences the DA system. In addition, since the net

359   impact of various factors can enhance DA efficiency by either degrading the OL or enhancing the ability of DA to add

360   more value, it is important to decompose the source of variations in ΔR. Therefore, in addition to examining how

361   SMAP DA efficiency, i.e., $\Delta R = R_{L4} - R_{OL}$, varies as a function of the most prominent control factors identified in the

362   above Section 3.3.1 (i.e., Tb error, precipitation forcing, and microwave soil roughness), we also examine how

363   precipitation error as a control factor affects the OL performance, i.e., R_{OL}.

364 To minimize the uncertainty caused by large errors in each of the control factors, we exclude samples with errors
365 (separately for each control factor) ranking above the 80th percentile in the following analysis. The relationship
366 between Tb errors and L4 DA efficiency is straightforward: higher Tb errors are associated with higher ΔR, with ΔR
367 generally larger for SSM than for RZSM (Fig. 4a-b).
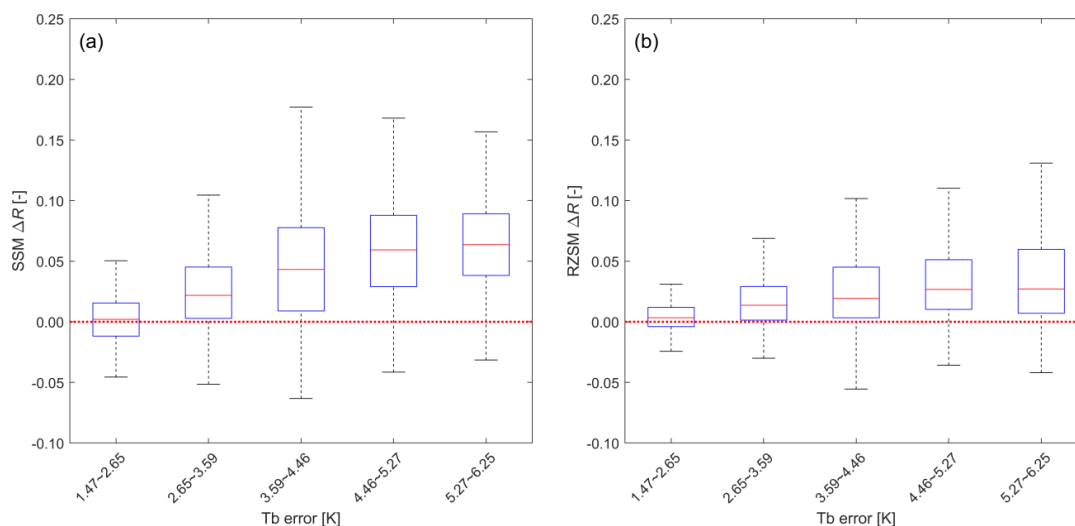
368

369 **Figure 4: SMAP L4 DA efficiency ($\Delta R = R_{L4} - R_{OL}$) as a function of Tb error for (a) SSM and (b) RZSM. Samples with Tb**
370 **error ranking above the 80th percentile are excluded from the analysis.**

371

372 For precipitation, this decomposition is illustrated in Fig. 5. Note that, as expected, low-quality precipitation tends to
373 degrade the skill (i.e., correlation versus ground observations) of OL SSM and RZSM estimates (see Fig. 5a-b). This
374 degradation provides an enhanced opportunity for SMAP L4 DA to provide added value. As a result, ΔR tends to be a
375 proportional function of precipitation skill (i.e., higher precipitation skill leads to lower ΔR, see Fig. 5c-d). This inverse
376 relationship is a well-known tendency for land data assimilation systems (Liu et al., 2011; Bolten and Crow, 2012;
377 Dong et al., 2019a). Precipitation quality has a diminished impact on RZSM estimation skill compared to SSM
378 estimation skill. This is expected since RZSM is (essentially) the result of applying a low-pass time series filter to
379 precipitation. As such, it is less sensitive to high-frequency errors in precipitation products than SSM is.

380

381
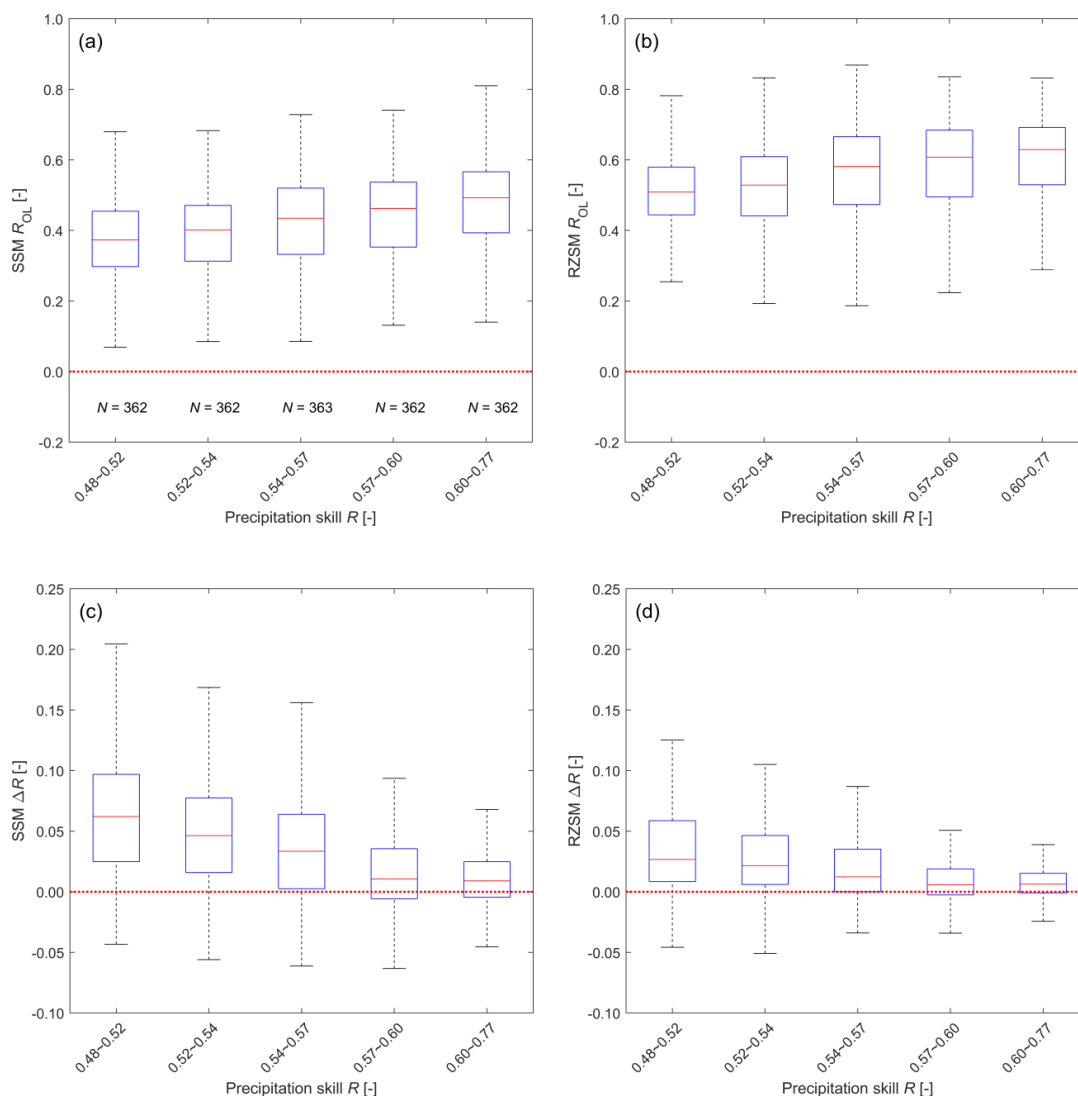
Figure 5: OL performance (R_{OL}) as a function of precipitation forcing skill R for (a) SSM and (b) RZSM. SMAP L4 DA efficiency ($\Delta R = R_{L4} - R_{OL}$) as a function of precipitation skill for (c) SSM and (d) RZSM. Samples with precipitation skill ranking below the 20th percentile are excluded from the analysis.

382
383
384

385

386 Figure 6 is analogous to Fig. 4 but shows skill differences ΔR as a function of microwave soil roughness. Similar to

387 Tb errors, it is as expected that this control factor of microwave soil roughness has little impact on the OL performance,

388 except that R_{OL} shows slight decreasing tendency with increasing soil roughness (not shown). Given the fact that the

19

389　OL does get worse with increasing roughness, there is more room for improvement as the roughness increases, which

390　makes it plausible that ΔR increases with increasing soil roughness (see Fig. 6a-b).


391
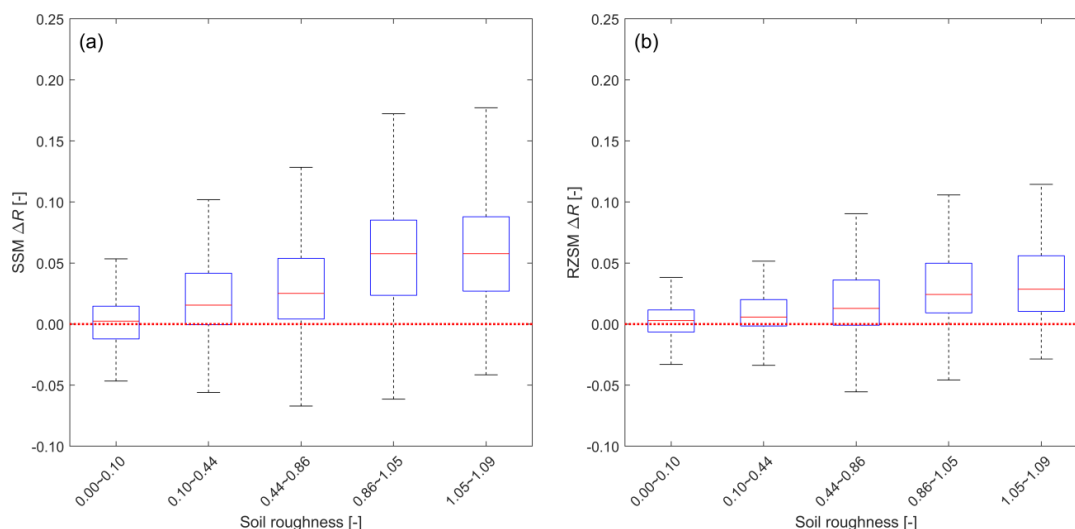
Figure 6: As in Fig. 4 but for ΔR as a function of microwave soil roughness.

393

394　Besides the above three control factors that dominate the DA efficiency, we also examine the top factor that affects

395　SMAP L4 performance, i.e., vertical-coupling errors (Fig. 7). As expected, larger (absolute) bias in SSM-RZSM

396　coupling in CLSM tends to be associated with degraded OL estimates of both SSM and RZSM (see Figs. 7a-b),

397　although the analysis does not prove such a causal relationship. Similar to precipitation errors above, decreased OL

398　skill (seen on the left-hand-side of the figures) provides an opportunity for increased DA efficiency – which is clearly

399　seen in Fig. 7. However, such increases are much larger for SSM than for RZSM.

400　For RZSM, SSM-RZSM coupling bias represents a double-edged sword. While such bias leads to an enhanced

401　opportunity to improve upon a degraded OL, it should also hamper the ability of DA to transfer SSM increments into

402　the root-zone – particularly when, like here, the bias reflects the lack of vertical coupling in the model (Kumar et al.,

403　2009). This means that some of the opportunity presented by the larger OL RZSM errors is squandered by sub-optimal

404　DA. As a result, the increase in RZSM DA efficiency associated with biased SSM-RZSM coupling (Fig. 7d) is smaller

405　than the analogous increase in SSM DA efficiency (Fig. 7c).
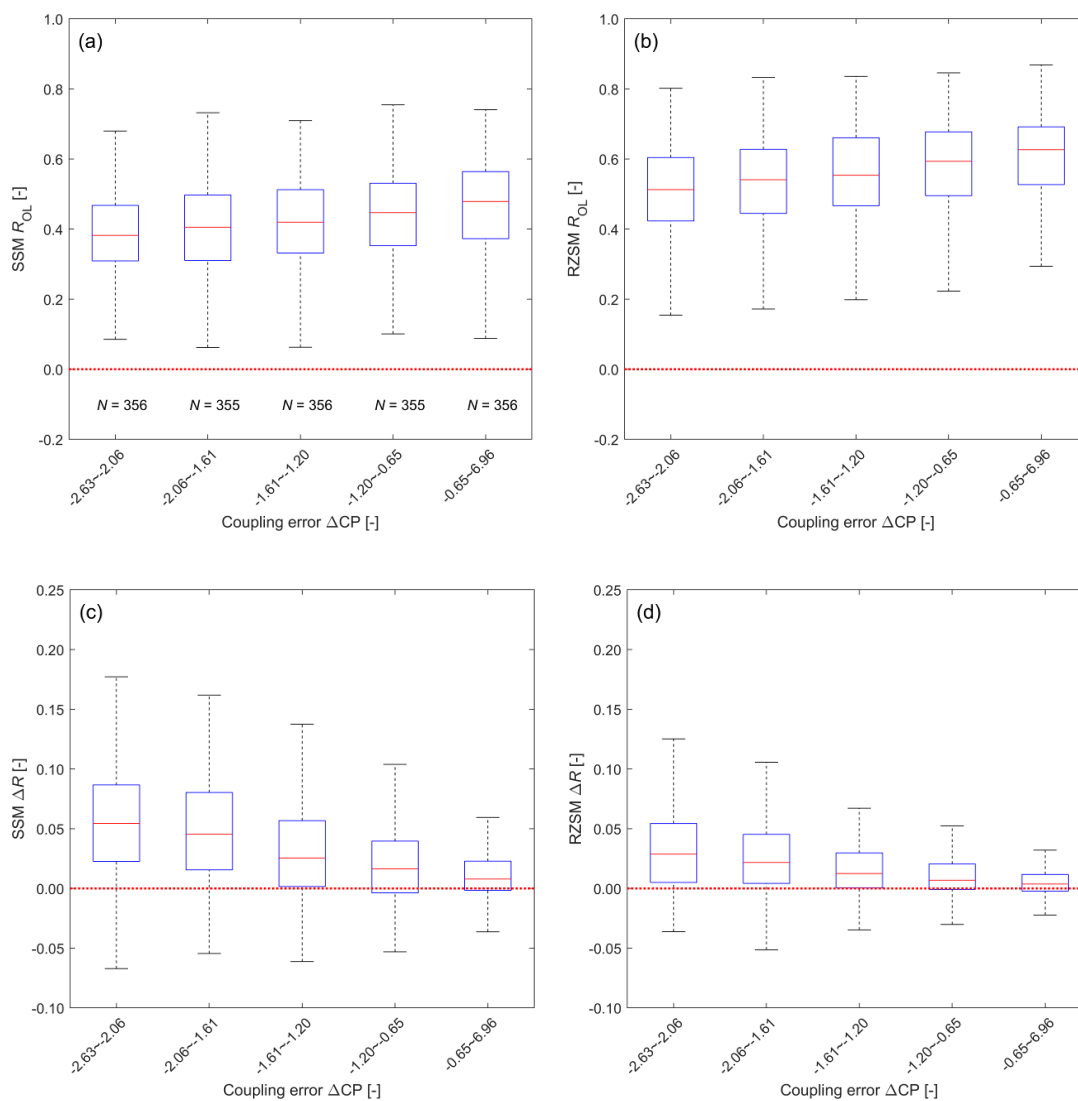
Figure 7: As in Fig. 5 but for R_{OL} and ΔR as a function of SSM-RZSM coupling error indicated by the CP difference ($\Delta CP = CP_{OL} - CP_{obs}$).

For the three strongest control factors that determine DA efficiency ΔR, i.e., Tb error, precipitation error and microwave soil roughness, we further conducted paired one-way analysis of variance (not shown). Results indicates that for each of the five binned groups separated by each of the above-mentioned three control factors, the inter-group difference in ΔR caused by each control factor is significant ($p<0.01$) for both SSM and RZSM. In addition, except for

415   the groups with lowest mean ΔR in Fig. 4a and Fig. 6a, the averages of ΔR from all groups are significantly higher

416   than 0 ($p<0.01$).

417   As expected, precipitation error is the dominant factor for explaining the skill of the OL estimates. In contrast, the

418   SSM-RZSM coupling error is the dominant factor for explaining the skill of the L4 results, which shows DA is able to

419   correct for precipitation errors.

420   **4 Conclusions**

421   The SMAP L4 algorithm assimilates L-band Tb observations into the Catchment Land Surface Model, to provide

422   surface and root-zone soil moisture estimates (i.e., SSM, RZSM) with global, 3-hourly coverage at 9-km resolution.

423   The performance of the L4 soil moisture estimates compared to a baseline model-only simulation (OL) is influenced

424   by multiple control factors associated with the land surface modelling (LSM) and radiative transfer modeling (RTM)

425   components of the L4 system. In this study, we assess the performance of SMAP L4 DA system using the 2 years of

426   in-situ soil moisture profile observations at 2474 sites across mainland China. We apply a random forest (RF)

427   regression to identify the dominant factors that control the spatial distribution of the DA efficiency (defined as the skill

428   difference between the L4 and OL estimates of SSM and RZSM as measured by their Spearman rank correlation with

429   in-situ measurements). Results show that L4 improves SSM prediction skill by 14% on average, with over 77% of the

430   2287 9-km EASE grid cells showing an increase in Spearman's rank correlation with in-situ observations. Similarly

431   widespread but smaller improvements are also observed in RZSM, with averaged R improvement of 7%.

432   Based on the RF regression analysis, the added value of SMAP L4 DA for SSM is primarily determined by Tb error

433   (measured by standard deviation of O-F Tb residuals), followed by microwave soil roughness and daily precipitation

434   error. These three factors are also the most prominent factors controlling SMAP DA improvement for RZSM, albeit

435   with the Tb error being the least important of these three factors for RZSM DA efficiency.

436   Generally, the OL performance clearly decreases with increasing precipitation error, whereas for L4 performance

437   precipitation error is not identified as the most dominant control factor. This indicates that the L4 system is able to

438   correct for errors in precipitation forcing. In addition, our results demonstrate that SMAP DA contributes the most

439   added value for cases where CLSM underestimates SSM-RZSM vertical coupling strength. However, due to the

440   difference in top-layer soil depth between the in-situ observations (10 cm) and the L4 analysis (5 cm), it is unclear

441   whether or not the observed SSM-RZSM coupling strength biases are real in an absolute sense – or simply reflect

442   inconsistencies in the depth of modelled versus observed SSM and RZSM time series. Nevertheless, it is worth

443   stressing that, despite the ambiguity with regards to their absolute magnitude/sign, relative variations in apparent SSM-

444   RZSM coupling biases explain a significant amount of the observed spatial variation in L4 performance. Therefore,

445   this finding clearly underpins the importance of properly specifying SSM-RZSM coupling strength in CLSM as a way

446   to improve the SMAP L4 product.

447   For SMAP L4 SSM skill, the next-most important factors (after SSM-RZSM coupling) are the precipitation error, the

448   Tb error and microwave soil roughness (Fig. 3d). For L4 RZSM skill, the next-most important factors (after SSM-

449   RZSM coupling) are the precipitation error, the Tb error and the LE error, with the latter two factors of comparable

450   importance (Fig. 3d). To enhance the L4 performance, additional focus should thus be placed on improving the model's

451   characterization of the partitioning of the available energy into latent and sensible heat (LE error) and the microwave

452   radiative transfer modeling (Tb error).

453   **Data availability**

454   The SMAP L4 datasets are available from https://nsidc.org/data/SPL4SMAU/versions/4. Gauge-based precipitation

455   dataset CGDPA is from http://data.cma.cn/data/cdcdetail/dataCode/SEVP_CLI_CHN_PRE_DAY_GRID_0.25.html.

456   The availabilities of other datasets are stated in their corresponding subsections.

457   **Author contributions**

458   Jianxiu Qiu and Jianzhi Dong conceptualized the study. Jianxiu Qiu carried out the analysis and wrote the first draft

459   manuscript, Wade Crow refined the work, Jianzhi Dong, Rolf Reichle, and Gabrielle De Lannoy helped with the analysis.

460   All authors contributed to the analysis, interpretation of the results and writing.

461   **Competing interests**

462   The authors declare that they have no conflict of interest.

463   **Acknowledgments**

468   **References**

469   Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., and Smets, B.: GEOV1: LAI, FAPAR

470   Essential Climate Variables and FCOVER global time series capitalizing over existing products. Part1: Principles of

471   development and production, Remote Sens. Environ., 137, 299-309, doi:10.1016/j.rse.2013.02.030, 2013.

472

473   Bolten, J.D. and Crow, W.T.,: Improved prediction of quasi-global vegetation conditions using remotely-sensed

474   surface soil moisture, Geophys. Res. Lett., 39(19), doi:10.1029/2012GL053470, 2012.

475

476   Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.

477

478   Chan, S., Njoku, E. G. and Colliander A.: SMAP L1C radiometer half-orbit 36 km EASE-Grid brightness temperatures,

479   version 3. NASA National Snow and Ice Data Center Distributed Active Archive Center, 10.5067/E51BSP6V3KP7,

480   2016.

481

482   Chen, F., Crow, W.T., Starks, P.J. and Moriasi, D.N.: Improving hydrologic predictions of a catchment model via

483   assimilation of surface soil moisture, Adv. Water Resources., 34(4), 526-536, doi:10.1016/j.advwatres.2011.01.011,

484   2011.

485

486   Chen, F., Crow, W.T., Colliander, A., Cosh, M.H., Jackson, T.J., Bindlish, R., Reichle, R.H., Chan, S.K., Bosch, D.D.,

487   Starks, P.J., and Goodrich, D.C.: Application of triple collocation in ground-based validation of Soil Moisture

488   Active/Passive (SMAP) level 2 data products, IEEE JSTARS., 99, 1-14, doi:10.1109/JSTARS.2016.2569998, 2016.

489

490   Crow, W.T. and Van Loon, E.: The impact of incorrect model error assumptions on the sequential assimilation of

491   remotely sensed surface soil moisture, J. Hydrometeorol., 8(3), 421-431, doi:10.1175/jhm499.1, 2006.

492

493   De Lannoy, G. J. M., Reichle, R. H., and Pauwels, V. R. N.: Global calibration of the GEOS-5 L-band microwave

494   radiative transfer model over nonfrozen land using SMOS observations, J. Hydrometeorol., 14(3), 765–785,

495   doi:10.1175/JHM-D-12-092.1, 2013.

496

497   De Lannoy, G. J. M., Reichle, R. H., and Vrugt, J. A.: Uncertainty quantification of GEOS-5 L-band radiative transfer

498   model parameters using Bayesian inference and SMOS observations, Remote Sens. Environ., 148, 146–157,

499   doi :10.1016/j.rse.2014.03.030, 2014.

500

501   Dong, J., Crow, W.T., Reichle, R., Liu, Q., Lei, F., and Cosh, M.: A global assessment of added value in the SMAP

502   Level 4 soil moisture product relative to its baseline land surface model, Geophys. Res. Lett., 46, 6604-6613,

503   doi:10.1029/2019GL083398, 2019a.

504

505   Dong, J., Crow, W.T., Duan, Z., Wei, L., and Lu, Y.: A double instrumental variable method for geophysical product

506   error estimation, Remote Sens. Environ., 225, 217-228, doi:10.1016/j.rse.2019.03.003, 2019b.

507

508   Dong, J., Crow, W.T., Tobin, J. K., Cosh, H. M., Bosch, D. D., Starks, J. P., Seyfried, M., and Collins, H. C.:

509   Comparison of microwave remote sensing and land surface modeling in surface soil moisture climatology estimation,

510   Remote Sens. Environ., 242, 111756, doi :10.1016/j.rse.2020.111756, 2020.

511

512   Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., and Edelstein, W. N.: The soil moisture active

513   passive (SMAP) mission, P. IEEE., 98(5), 704–716, doi:10.1109/jproc.2010.2043918, 2010.

514

515   Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J. C., and Draper, C.: Validation practices

516   for satellite soil moisture retrievals: What are (the) errors?, Remote Sens. Environ., 244, 111806,

517   doi:10.1016/j.rse.2020.111806, 2020.

518

519   Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE

520   performance criteria: Implications for improving hydrological modelling, J. Hydrometeorol., 377(1-2), 80-91,

521   doi:10.1016/j.jhydrol.2009.08.003, 2009.

522

523   Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., and Reichstein, M.: The FLUXCOM ensemble

524   of global land-atmosphere energy fluxes, Sci. Data., 6(1), 1-14, doi:10.1038/s41597-019-0076-8, 2019.

525

526   Kumar, S.V., Reichle, R.H., Koster, R.D., Crow, W.T., and Peters-Lidard, C.D.: Role of subsurface physics in the

527   assimilation of surface soil moisture observations, J. Hydrometeorol., 10, 1534-1547, doi:10.1175/2009JHM1134.1,

528   2009.

529

530   Lucchesi, R.: File specification for GEOS-5 FP, NASA GMAO Office Note 4 (version 1.0), 63 pp. Available at

531   https://ntrs.nasa.gov, 2013

532

533   McColl, K., Vogelzang, J., Konings, A.G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation:

534   Estimating errors and correlation coefficients with respect to an unknown target, Geophys. Res. Lett., 41(17), 6229-

535   6236, doi:10.1002/2014gl061322, 2014.

536

537   Piepmeier, J. R., Focardi, P., Horgan, K. A., Knuble, J., Ehsan, N., Lucey, J., Brambora, C., Brown, P. R., Hoffman,

538   P. J., French, R. T., Mikhaylov, R. L., Kwack, E. Y., Slimko, E. M., Dawson, D. E., Hudson, D., Peng, J., Mohammed,

539   P. N., de Amici, G., Freedman, A. P., Medeiros, J., Sacks, F., Estep, R., Spencer, M. W., Chen, C. W., Wheeler, K. B.,

540   Edelstein, W. N., O'Neill, P. E., and Njoku, E. G.: SMAP L-band microwave radiometer: Instrument design and first

541   year on orbit, IEEE T. Geosci. Remote., 55(4), 1954–1966, doi:10.1109/TGRS.2016.2631978, 2017.

542

543   Liu, Q., Reichle, R., Bindlish, R., Cosh, M.H., Crow, W.T., de Jeu, R., de Lannoy, G., Huffman, G.J. and Jackson,

544   T.J.: The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land

545   data assimilation system, J. Hydrometeorol., 12(5), 750-765, doi:10.1175/JHM-D-10-05000.1, 2011.

546

547   Reichle, R.H., Crow, W.T., Koster, R. D., Sharif, H. and Mahanama, S.: Contribution of soil moisture retrievals to

548   land data assimilation products, Geophys. Res. Lett., 35(1), doi:10.1029/2007GL031986, 2008.

549

550   Reichle, R. H., de Lannoy, G. J. M., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., Crow, W., Jackson, T. J.,

551   Jones, L. A., Kimball, J. S., Koster, R. D., Mahanama, S. P., Smith, E. B., Berg, A., Bircher, S., Bosch, D., Caldwell,

552   T. G., Cosh, M., González-Zamora, Á., Holifield Collins, C. D., Jensen, K. H., Livingston, S., Lopez-Baeza, E.,

553   Martínez-Fernández, J., McNairn, H., Moghaddam, M., Pacheco, A., Pellarin, T., Prueger, J., Rowlandson, T., Seyfried,

554   M., Starks, P., Su, Z., Thibeault, M., van der Velde, R., Walker, J., Wu, X., and Zeng, Y.: Assessment of the SMAP

555   Level-4 surface and root-zone soil moisture product using in situ measurements, J. Hydrometeorol., 18(10), 2621–

556   2645, doi:10.1175/JHM-D-17-0063.1, 2017a.

557

558   Reichle, R. H., de Lannoy, G. J. M., Liu, Q., Koster, R. D., Kimball, J. S., Crow, W. T., Ardizzone, J. V., Chakraborty,

559   P., Collins, D. W., Conaty, A. L., Girotto, M., Jones, L. A., Kolassa, J., Lievens, H., Lucchesi, R. A., and Smith, E. B.:

560   Global assessment of the SMAP Level-4 surface and root-zone soil moisture product using assimilation diagnostics, J.

561   Hydrometeorol., 18(12), 3217–3237, doi:10.1175/jhm-d-17-0130.1, 2017b.

562

563   Reichle, R. H., de Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., and Liu, Q.: SMAP L4 Global 9 km EASE-

564   grid surface and root zone soil moisture land model constants, Version 4, NASA National Snow and Ice Data Center

565   DAAC, https://doi.org/10.5067/KGLC3UH4TMAQ, 2018a

566

567   Reichle, R. H., de Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., & Liu, Q.: SMAP L4 global 3-hourly 9 km

568   EASE-grid surface and root zone soil moisture analysis update data, version 4, NASA National Snow and Ice Data

569   Center DAAC, https://doi.org/10.5067/60HB8VIP2T8W, 2018b

570

571   Reichle, R. H., de Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., & Liu, Q.: SMAP L4 global 3-hourly 9 km

572   EASE-grid surface and root zone soil moisture geophysical data, version 4, NASA National Snow and Ice Data Center

573   DAAC, https://doi.org/10.5067/KPJNN2GI1DQR, 2018c

574

575   Reichle, R. H., Liu, Q., Koster, R. D., Crow, W. T., De Lannoy, G. J., Kimball, J. S., and Kolassa, J.: Version 4 of the

576   SMAP Level-4 soil moisture algorithm and data product, J. Adv. Model Earth Sy., 11(10), 3106-3130,

577   doi:10.1029/2019MS001729, 2019.

578

579   Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., and Lehner, I.: Investigating soil moisture–climate

580   interactions in a changing climate: A review, Earth-Sci. Rev., 99, 125–161, doi:10.1016/j.earscirev.2010.02.004, 2010.

581

582   Seneviratne, S. I., Wilhelm, M., Stanelle, T., Hurk, B., Hagemann, S., and Berg, A.: Impact of soil moisture-climate

583   feedbacks on CMIP5 projections: First results from the GLACECMIP5 experiment, Geophys. Res. Lett., 40(19), 5212-

584   5217, doi:10.1002/grl.50956, 2013.

585

586   Shen, Y., Xiong, A., Wang, Y., and Xie, P.: Performance of high-resolution satellite precipitation products over China,

587   J. Geophys. Res-Atmos., 115(D2), doi:10.1029/2009JD012097, 2010.

588

589   Shen, Y. and Xiong, A.: Validation and comparison of a new gauge-based precipitation analysis over mainland China,

590   Int. J. Climatol., 36(1), 252-265, doi:10.1002/JOC.4341, 2015.

591

592   Verger, A., Baret, F., and Weiss, M.: Performances of neural networks for deriving LAI estimates from existing

593   CYCLOPES and MODIS products, Remote Sens. Environ., 112, 2789-2803, doi:10.1016/j.rse.2008.01.006, 2008.

594

595   Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., and Yang, S.: A gauge-based analysis of daily

596   precipitation over East Asia, J. Hydrometeorol., 8, 607-626, doi:10.1175/JHM583.1, 2007.

597

598   Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F., Lin, R..: NOAA CDR Program: NOAA Climate Data

599   Record (CDR) of CPC Morphing Technique (CMORPH) High Resolution Global Precipitation Estimates, Version 1.

600   NOAA National Centers for Environmental Information, 2019.