Reply to Referee #2 interactive comment

The paper evaluates the data assimilation efficiency of SMAP brightness temperature data by updating the root-zone soil moisture with CLSM (the Catchment Land Surface Model) model and an RTM model (radiative transfer modeling). The result of soil moisture filed delta_R increments then identifies substantial factors that control this data assimilation efficiency, such as precipitation error and SSM-RZSM coupling strength error. I appreciate the motivation of this paper, and its conclusion and inference are probably attractive to the L-band TB data assimilation community. However, I cannot agree on the methodology part of this paper, and I don't think the findings would help with the further development of RZSM DA improvements.

We thank the reviewer for the constructive criticisms and helpful comments.

Note that the results using Pearson and Spearman's correlation are qualitatively consistent, as will be detailed in our response to Major comment #7.

We would also like to clarify that SMAP L4 is the only operational global DA system that assimilates L-band Tb and provides near-real-time root-zone soil moisture information. However, SMAP L4 has been mainly evaluated across sparse in-situ sites and SMAP cal/val sites (mainly in the US, Europe and Australia) and factors affecting its accuracy are still largely unknown, which is particularly true for RZSM.

Therefore, based on soil moisture observations from 2474 in-situ sites in China, this is the first study that comprehensively quantifies the SMAP L4 SSM and RZSM skill improvements and their major error sources, and further identifies the key priorities for future L4 development. For instance, SSM-RZSM coupling strength is identified as the most important factor in determining the L4 RZSM accuracy. Therefore, the appropriate representation of SSM-RZSM coupling strength should be considered as a priority for developing next generation of L4 system.

Additionally, we would like to point out that LSMs, RTMs and different variants of DA algorithms typically share similar structures. Therefore, our findings are not limited to the SMAP L4 system but are expected to be transferable for diagnosing and improving general soil moisture DA systems and enhancing their RZSM accuracies. We will add more emphasis on these aspects in the revised manuscript.

Major comments:

1. Line 23 & Line 75-79, Line 144 & Line 152: how do the authors select these eight control factors?

Re. the <u>comment about the choice of the control factors</u>: We provide the same response to Major comment #5 by Reviewer #1. For easy reference, please see below:

As mentioned in the abstract, the modeling portion of the SMAP L4 system consists of two components: land surface modelling (LSM) and radiative transfer modeling (RTM). Therefore, we select control factors from each of the two components.

For the LSM component, the errors can be attributed to: 1) model input forcing errors of a) precipitation and b) LAI; 2) model structure errors in a) characterizing SSM-RZSM coupling strength and b) the presence of vertical variability in soil properties; 3) model output error of LE.

For the RTM component, errors are characterized by: 1) DA innovation, i.e., SMAP Tb observations minus RTM Tb simulations; 2) the environmental factors that complicate the DA analysis when assimilating Tb observations, which include the magnitude of a) microwave soil roughness and b) LAI.

These 8 control factors from the above-mentioned 5 aspects determine the crucial aspects of both the LSM and RTM components in the L4 system, and are readily quantifiable using remote sensing products in the study. Therefore, they are selected to investigate the mechanism underlying the L4 improvement in this study. We will further clarify this in the corresponding paragraph of Section 2.3 in the revised manuscript.

2. Line 80, please show which part of the paper corresponds to each sentence. For instance, "Next, the in-situ measurements…" As I see, only figure 1 is about the in-situ measurements.

In the revised manuscript, we will specifically clarify which section of the paper corresponds to each sentence of the guidance paragraph in Line 80-85.

In fact, the content in Figure 1 is directly plotted using in-situ measurements. To be clearer, we will add an illustrative figure showing the spatial distribution of in-situ measurements.

3. Still, Line 80-81, the soil moisture profile measurements from CMA networks can reach 100 cm. please refer to: Han Shuai, Shi Chunxiang, Jiang Lipeng, Zhang Tao, Liang Xiao, Jiang Zhiwei, Xu Bin, Li Xianfeng, Zhu Zhi, Lin Hongjin. The Simulation and Evaluation of Soil Moisture Based on CLDAS[J]. Journal of Applied Meteorological Science, 2017, 28(3): 369-378. I suggest the authors separate the sites that contain measurements with 100 cm and the rest in the analysis.

We will include the suggested reference in the revised manuscript.

Re. the comment about separating the sites with 100 cm measurements: According to the in-situ measurements data provider, constrained by soil profile condition, only a fraction (<50%) of 2474 sites have complete soil moisture measurements up to 100 cm. In addition, in the SMAP L4 validation procedure using SMAP core validation sites, although a few of the sparse-network sites have deeper-layer measurements (typically in the 80cm to 100cm range), Reichle et al. (2019) used measurements from sensors placed within 50 cm from the surface, as they found the spotty in-situ measurements time series at the deeper depths of relatively little use. Therefore, we routinely used in-situ data only down to 50 cm to evaluate the RZSM estimates of L4 and OL.

[1] 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 SMAP Level-4 soil moisture algorithm and data product, J. Adv. Model Earth Sy., 11(10), 3106-3130, doi:10.1029/2019MS001729, 2019.

4. Section 2.2, the repetitiveness of in-situ soil moisture measurements is questionable. Line 128, the atmospheric elements such as air temperature, humidity, etc. of these stations cover different land-use types. When it comes to soil moisture, due to high spatial difference, the in–situ soil moisture profile measurements may vary a lot to the station outside. In standard, all these CMA stations should only have grassland or bare soil land types. Other land covers are impossible, and this affects precipitation, evaporation, draining, etc., as authors said in Line 144-146. The scale mismatch between CLSM outputs and in-situ measurement would exceed the accuracy indicates in the evaluation.

We thank the reviewer for the insight into the CMA data.

For the random point-to-grid upscaling errors of sampling, we agree that using in-situ soil moisture data to validate the large-scale L4 product is challenged by spatial representativeness error. However, as demonstrated in Dong et al. (2020), the representativeness error is essentially random and can be averaged out when sampled across multiple sites. Therefore, it mainly affects the absolute soil moisture evaluation metrics, and has no impacts on the relative accuracy. Given that this study is mainly interested in the relative accuracy of OL and L4 data and approximately 2474 sites have been used, representativeness error is expected to have a minor impact on our conclusions.

For the systematic sampling errors, we analyze the site "representativeness" using the 500m MODIS Land Cover product (MCD12Q1 V6) in 2017, IGBP dataset. First, we take the land cover (LC) of the MODIS grid cell where a given in-situ site is located as the ground-based LC. Next, we search all the MODIS LC grid cells that fall within the SMAP 9km EASE grid cell in which this in-situ site is located. The latter area consists of about 20 x 20 = 400 MODIS grid cells. We calculate the fraction of these 400 MODIS grid cells that have the same LC type as the ground-based LC, and define this fraction as the site representativeness. Fig. 1 shows a histogram of the site representativeness of all 2474 sites used. Similar results were found for 2018 (not shown).

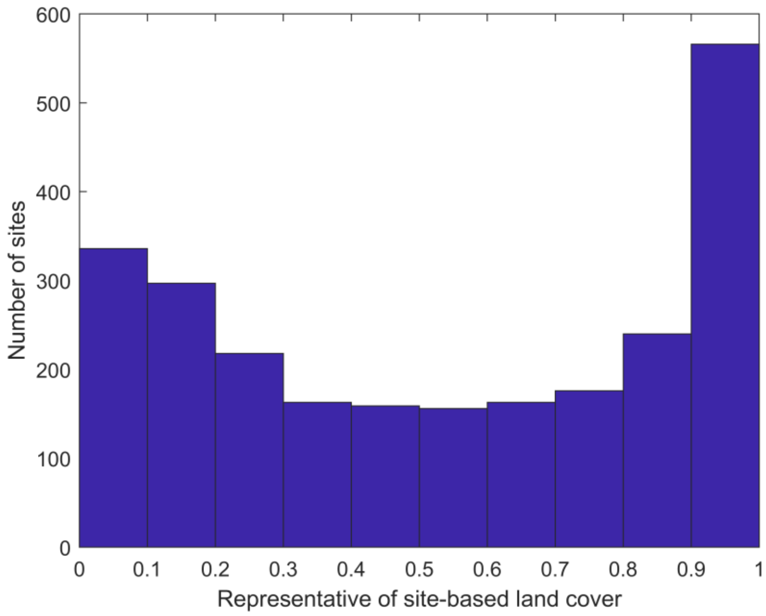

Fig. 1 Site representativeness of all 2474 sites. See text above for details.

It shows 59% of sites have site representativeness higher than 40%, and this fraction is 52% for site representativeness higher than 50%, and 46% for site representativeness higher than 60%. Furthermore, if we select sites with site representativeness higher than 50%, the RF attribute analysis show that the top 3 factors controlling skill improvement ($R_{L4} - R_{OL}$) R_{L4}, and R_{OL} are still the same, although the precipitation error becomes the top influencer for R_{L4} (Fig. 2).

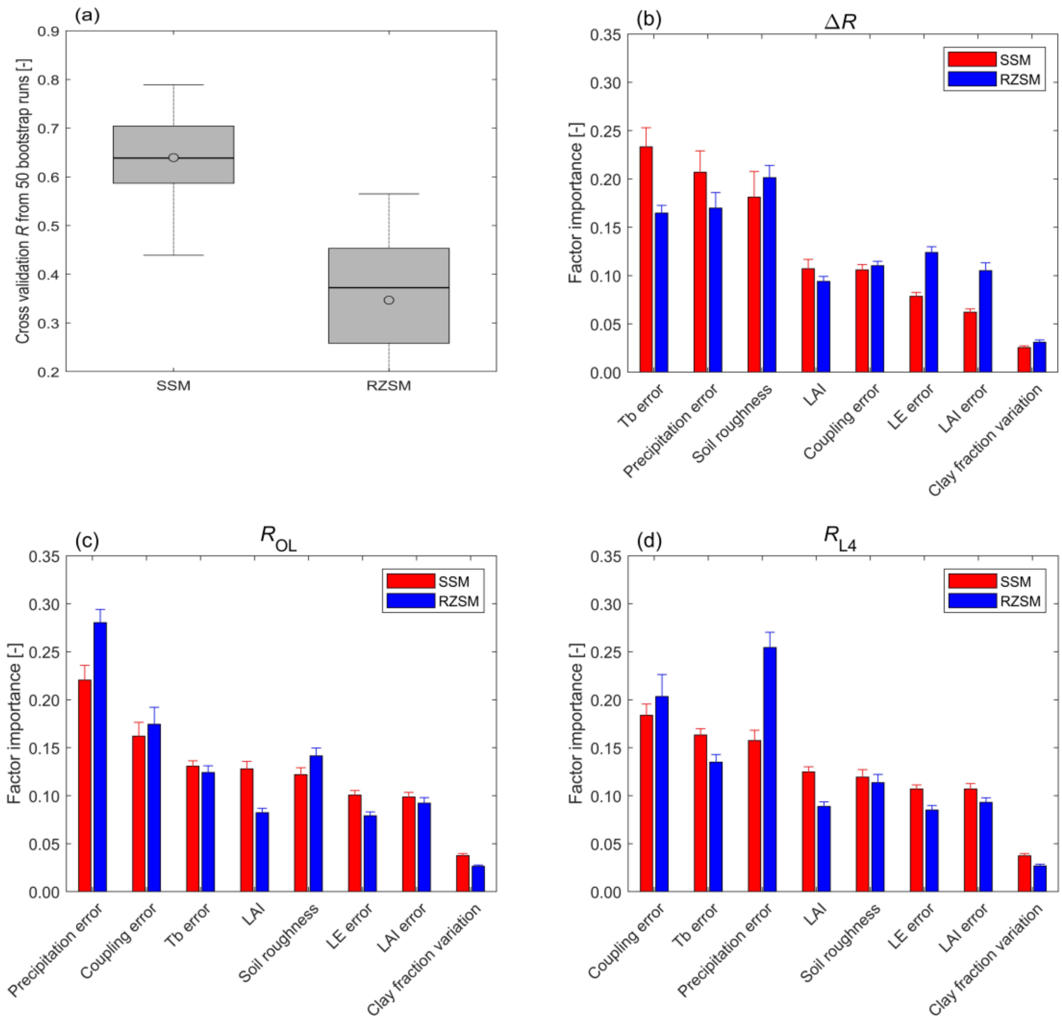

Fig. 2 Same content as in Fig. 3 of the manuscript, except that we use sites with site representativeness higher than 50%.

We will further clarify this in Section 2.2 in the revised manuscript.

[1] Dong, J., Crow, W.T., Tobin, J. K., Cosh, H. M., Bosch, D. D., Starks, J. P., Seyfried, M., and Collins, H. C. Comparison of microwave remote sensing and land surface modeling in surface soil moisture climatology estimation, Remote Sens. Environ., 242, 111756, doi :10.1016/j.rse.2020.111756, 2020.

5. Line 161, Table 1 & Line 192, could we use the same LAI data? As well as the rainfall data.

Re. the <u>comment about LAI dataset</u>: We provide the same response to Major comment #6 by Reviewer #1, for easy reference, please see below:

The inherent LAI in SMAP L4 system is merged from a MODIS/Geoland-based data product (Mahanama et al., 2015; Reichle et al., 2017).

To correctly characterize error in LAI of SMAP L4, we use LAI product from an entirely independent source, i.e. from the SPOT satellite. The prominent difference between SMAP L4 LAI and SPOT LAI is that the former uses an LAI climatology from the period 1999-2011, while the latter is the actual LAI time series with inter-annual variation.

Note that besides the LAI from SMAP L4 system, we only use one external LAI product of SPOT VGT. We have correctly listed both LAI datasets in Table 1, and will further clarify in Section 2.3: “The LAI used in the SMAP L4 system is a merged climatology from MODIS and Geoland data, based on satellite observations of the Normalized Difference Vegetation Index (Mahanama et al., 2015; Reichle et al., 2017)”.

[1] Mahanama, S. P., and Coauthors, 2015: Land boundary conditions for the Goddard Earth Observing System model version 5 (GEOS-5) climate modeling system—Recent updates and data file descriptions. NASA/TM-2015-104606, Vol. 39, 55 pp., https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160002967.pdf.

[2] Reichle, R. H., and Coauthors, 2017. Assessment of the SMAP Level-4 surface and root-zone soil moisture product using in situ measurements. J. Hydrometeorol. 18(10), 2621–2645, 10.1175/JHM-D-17-0063.1.

Re. the <u>comment about rainfall dataset</u>: The result of precipitation error presented in the manuscript only involves rain gauge (CGDPA) data, other rainfall data are used to prove the robustness of analysis results from CGDPA.

6. Line 248, what is the DA efficiency? Line 309, it says “the efficiency of SMAP L4 DA (i.e., R = RL4 - ROL)”. The data efficiency is not that simple, indeed. Please refer to (Nearing et al. 2018) for the definition of data assimilation efficiency, or provide where this “the efficiency of SMAP L4 DA (i.e., R = RL4 - ROL)” comes from in citations. Nearing, G., Yatheendradas, S., Crow, W., Zhan, X., Liu, J., & Chen, F. (2018). The Efficiency of Data Assimilation. Water Resources Research, 54, 6374-6392.

We will change “DA efficiency” to “DA skill improvement”, which is a more precise terminology for quantifying R_{L4} and R_{OL} differences.

7. Figure 1 & Section 3.1, one of the conclusions in this paper that RZSM is improved by assimilating brightness temperature. Figure 1 & Section 3.1 are the only evidence to support this view, which is vital to the following paragraphs. By any two datasets, the increased correlation coefficient is hard to address improvement. unRMSE, bias, and other characters shall also be accounted as SMAP evaluate its soil moisture products. As in Line 52-54, "observations-minus-forecast residuals" may not be sufficient, but it doesn’t mean it is unnecessary. Besides, Spearman’s rank correlation coefficient is very loose in statistics. Pearson correlation can assess linear relationships hypothesized in ordinary DA filters. Line 133-134 is not solid for support the advantage of Spearman’s and it should clarify what the outliers are.

We thank the reviewer for the constructive comments. We would like to clarify that the goal of this manuscript is not to prove increased RZSM accuracy, but to evaluate L4 RZSM and understand the mechanism that controls L4 RZSM accuracy.

Nonetheless, it is indeed our finding that L4 DA system improves OL RZSM accuracy, which is supported by evaluation sampled from 2474 sites at $p = 0.05$ significance level (based on a 1000-member bootstrapping analysis). However, as stated in the abstract, that we do agree the improvements in RZSM is slight ($\Delta R = 0.034$, or 7% in relative term) over 74% of soil moisture in-situ sites. Moreover, this statistically significant improvement in RZSM – albeit small – is also true for metrics of ubRMSE (2.3% reduction in relative term over 65% of in-situ sites). We will further clarify this in Section 3.1 in the revised manuscript.

Re. the <u>comment on using Spearman's rank correlation vs. Pearson correlation</u>: We provide the same response to Major comment #5 by Reviewer #1, which is reproduced here for easy reference:

Note that Pearson correlation assumes the linear consistency of underlining variables. However, this assumption may be adversely affected by outliers. To avoid ad-hoc thresholds, we did not exclude any soil moisture outliers and employed Spearman's rank correlation, which is less sensitive to such outliers. Nonetheless, we repeated the analysis based on Pearson correlation (see Figs. 3-4 below). The Pearson-based results are quantitatively consistent with the results using Spearman's correlation. We will further clarify this in the corresponding paragraph of Section 2.2 in the revised manuscript.

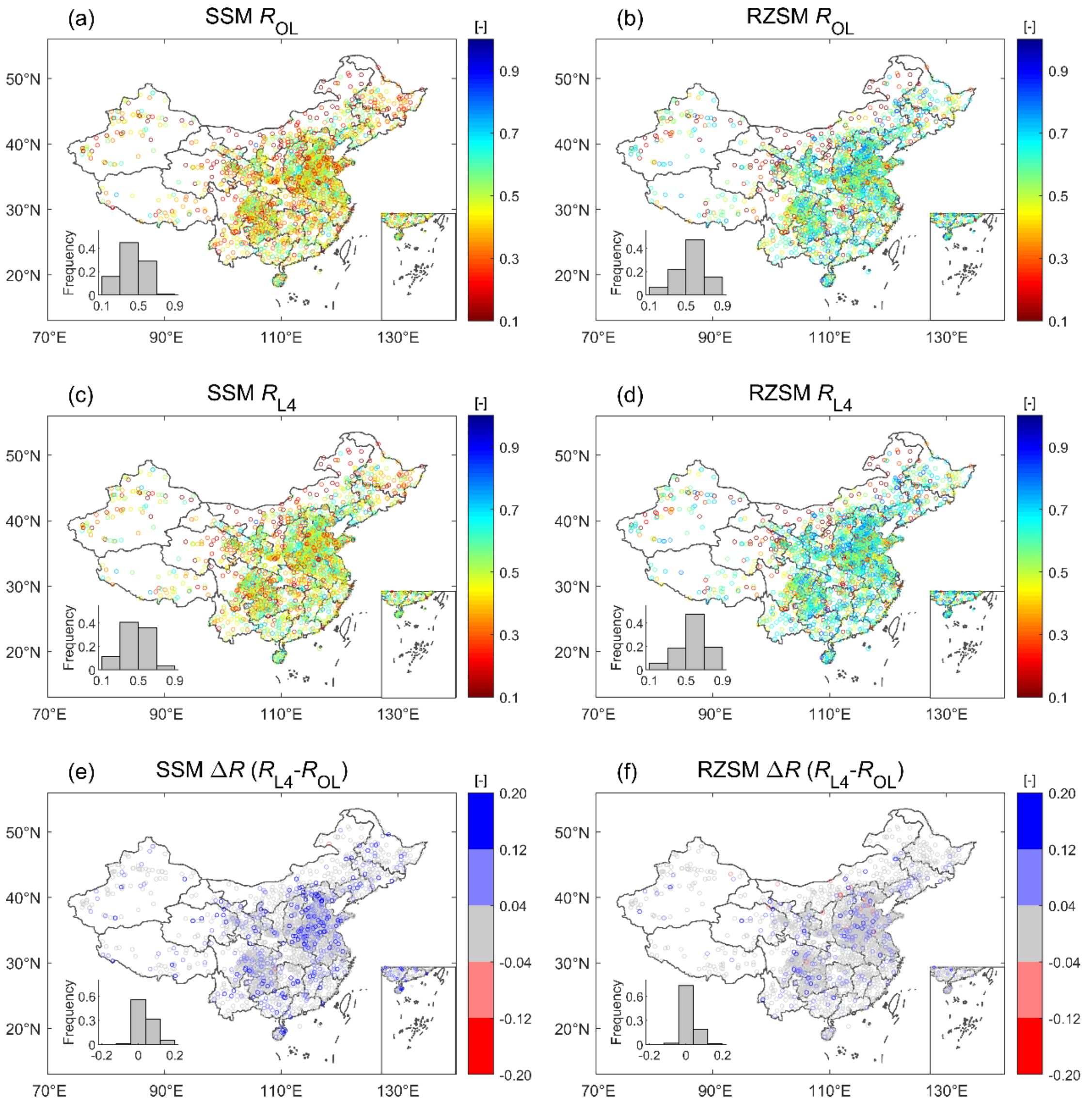

Fig. 3 Same content as in Fig. 1 of the manuscript, except that the correlation between in-situ soil moisture measurements and SMAP is measured using Pearson correlation.

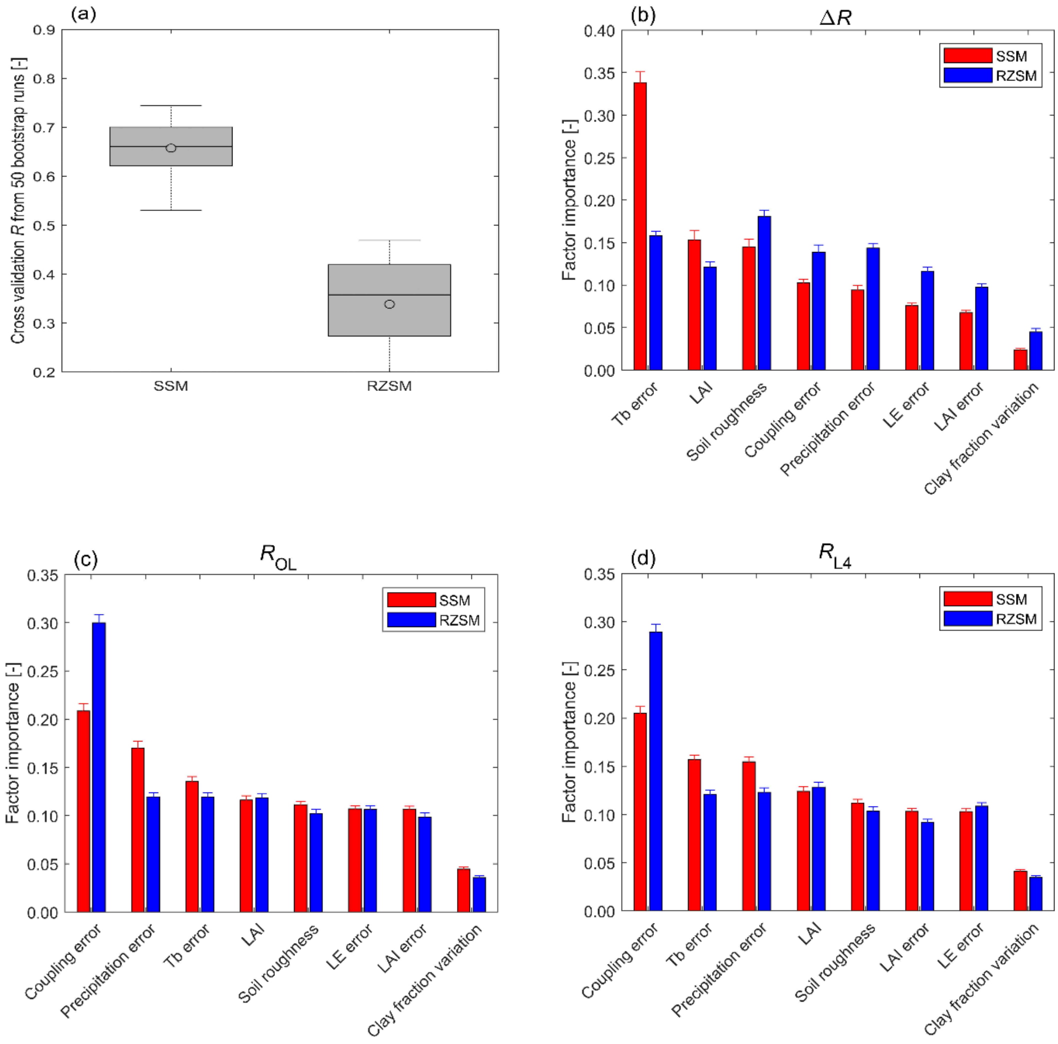

Fig. 4 Same content as in Fig. 3 of the manuscript, except that the correlation between in-situ soil moisture measurements and SMAP is measured using Pearson correlation.

Minor comments:

1. Line 1, please clarify what is the added value, is it a correlation coefficient or DA efficiency? The term "added value of … soil moisture…" is misleading because the study is based on delta_R, not soil moisture increment.

We will change the term “added value” to “benefit” in the revised manuscript.

2. Line 25, it should be "Spearman’s rank correlation coefficient" instead of "Spearman rank correlation skill". Skill is more sophisticated.

We will revise this expression into “Spearman’s rank correlation” in the revised manuscript.

3. Line 27, "the same percentage" is not clear.

We will revise this expression into specific number of “74%” in the revised manuscript.

4. Line 172-174, a citation is needed.

We will add the following reference in Line 174 of the manuscript.

[1] Shen, Y. and Xiong, A.: Validation and comparison of a new gauge-based precipitation analysis over mainland China, Int. J. Climatol., 36(1), 252-265, doi:10.1002/JOC.4341, 2015.

5. Line 258, clarify what kind of anomalies it is.

We defined anomaly in Line 117-118: “…we performed our analysis using anomaly time series, derived by subtracting a seasonally-varying (daily) climatology from each raw time series.” To avoid redundancy, we did not repeat this again in Line 258 or the following Section.