Reply to Referee #1 interactive comment

The author assesses the performance of the surface and root-zone soil moisture (SSM and RZSM) estimates by SMAP Level-4 DA system using an open loop (OL) simulations and two years in situ profile soil moisture observations at 2474 sites over mainland China. The anomaly Spearman's rank rather than Pearson correlation coefficient is calculated for comparisons and evaluations. In the following, to evaluate the efficiency of SMAP L4 DA system, the author chooses eight factors and uses methods of random forest regression and box plot comparisons to do the attribution analysis. Results show the improvement of SSM and RZSM estimates through the increased anomaly with in situ measurements, compared to OL based results. Three factors namely the standard deviation of the observation-minus-forecast Tb residuals, errors in precipitation forcing data and microwave soil roughness parameter H are found dominantly affecting the efficiency for SSM and RZSM estimates by SMAP Level-4 DA system. Furthermore, the SSM-RZSM coupling strength characterizing the surface to subsurface physics in CLSM is evaluated based on in situ measurements and OL and DA estimates.

Although it is enough to understand what 'went on', the scientific and English language is imprecise in various places as well as some cited information. I have given some examples below and labeled some in the attachment, but the authors should go throughout the entire manuscript carefully, and check that the descriptions and citations are as exact as possible. On the other hand, the author often uses different tenses in a paragraph even in one sentence, making presentations a bit messy. Additionally, too many brackets are used to present information. Please do the appropriate revisions, as a reader, I tend to get accurate information rather than having a hesitation on whether I shall ignore/keep the information, and thereby guess how does each step be carried on and may lose interest. I am sorry. I would say, maybe some main contents are ignored by the reviewer because of the weak presentation.

We sincerely thank the reviewer for the constructive and thoughtful comments.

Many of the reviewer's comments are, unfortunately, based on misunderstandings. We apologize if the original text was not sufficiently clear. To address the comments, we will undertake major revisions of the text, including a careful revision of the imprecise expressions and the tenses throughout the manuscript.

Major comments:

1. In line 77, please specify key CLSM parameters and give the reason why you choose these parameters.

The CLSM simulated soil moisture is affected by a range of vegetation and soil parameters. For example, Dong et al. (2019) demonstrated that soil moisture DA within the CLSM system is strongly affected by LAI. Therefore, LAI is used here to represent the vegetation representation error impacts.

Root-zone soil moisture dynamics are controlled by their connection with the surface soil moisture, soil hydraulic properties and soil water transport. The bulk relationship between root-zone and surface soil moisture can be captured by vertical soil moisture

coupling strength (Kumar et al., 2009).

Therefore, this study uses LAI and surface-rootzone coupling strength to characterize the vegetation and soil parameter error impacts on CLSM. We will further clarify this in the revised manuscript.

[1] Dong, J., Crow, W.T., Reichle, R., Liu, Q., Lei, F., and Cosh, M.: A global assessment of added value in the SMAP Level 4 soil moisture product relative to its baseline land surface model, Geophys. Res. Lett., 46, 6604-6613, doi:10.1029/2019GL083398, 2019.

[2] Kumar, S.V., Reichle, R.H., Koster, R.D., Crow, W.T., and Peters-Lidard, C.D.: Role of subsurface physics in the assimilation of surface soil moisture observations, J. Hydrometeorol., 10, 1534-1547, doi:10.1175/2009JHM1134.1, 2009.

2. In line 105, please clarify whether the OL run is conducted in this study? In line 29, I am sorry I cannot understand, what does "error in Tb observation space" mean? Please also explicitly clarify Tb error. In line 108-111, please clarify/specify "microwave soil roughness parameters, a vegetation structure parameter, and the microwave scattering albedo". "Soil roughness parameters" are used throughout the paper but without explaining what they are. Does it refer to both h and N, or s and L, or others? Additionally, please keep the cited information correct (equation A1 instead of B1). Please carefully check throughout the manuscript.

Yes, the OL run is conducted in this study.

Tb error refers to the difference between Tb observations from SMAP sensor and the Tb simulations obtained via a radiative transfer model.

The "microwave soil roughness parameters" refers to parameter *h* that accounts for dielectric properties that vary at the subwavelength scale.

We will further explain these expressions in the revised manuscript. In addition, we will check throughout the manuscript and make corresponding corrections, including the one in the Appendix as pointed out by the reviewer.

3. In line 120, LH and SH are mentioned. LH error is seen, please if possible, give the reason why SH error is disregarded.

Note that LH and SH are typically (strongly) anti-correlated. Therefore, it is not suitable to include both of them in random forest analysis, which will yield biased high weights on LH and SH. We will further clarify this in the corresponding paragraph of Section 2.1 in the revised manuscript.

4. In line 127, if possible, please give the figure plotting the distribution of CASMOS

as new Fig. 1.

We will add a separate figure showing the distribution of CASMOS as new Fig. 1.

5. In line 133-134, I cannot be convinced by the described reason about the use of Spearman correlation rather than Pearson correlation. Could you explain more? Wikipedia says that the Spearman correlation concerns the rank and Pearson correlation the mean. Do you calculate Pearson correlation based results? Please give the definition of outliers excluded in this study. In line 144-147, please clarify why these five control factors are chosen, and why the difference in clay fraction across the vertical can be used to quantify vertical variability in soil properties.

Note that Pearson correlation assumes the linear consistency of underlying 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. 1-2 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.

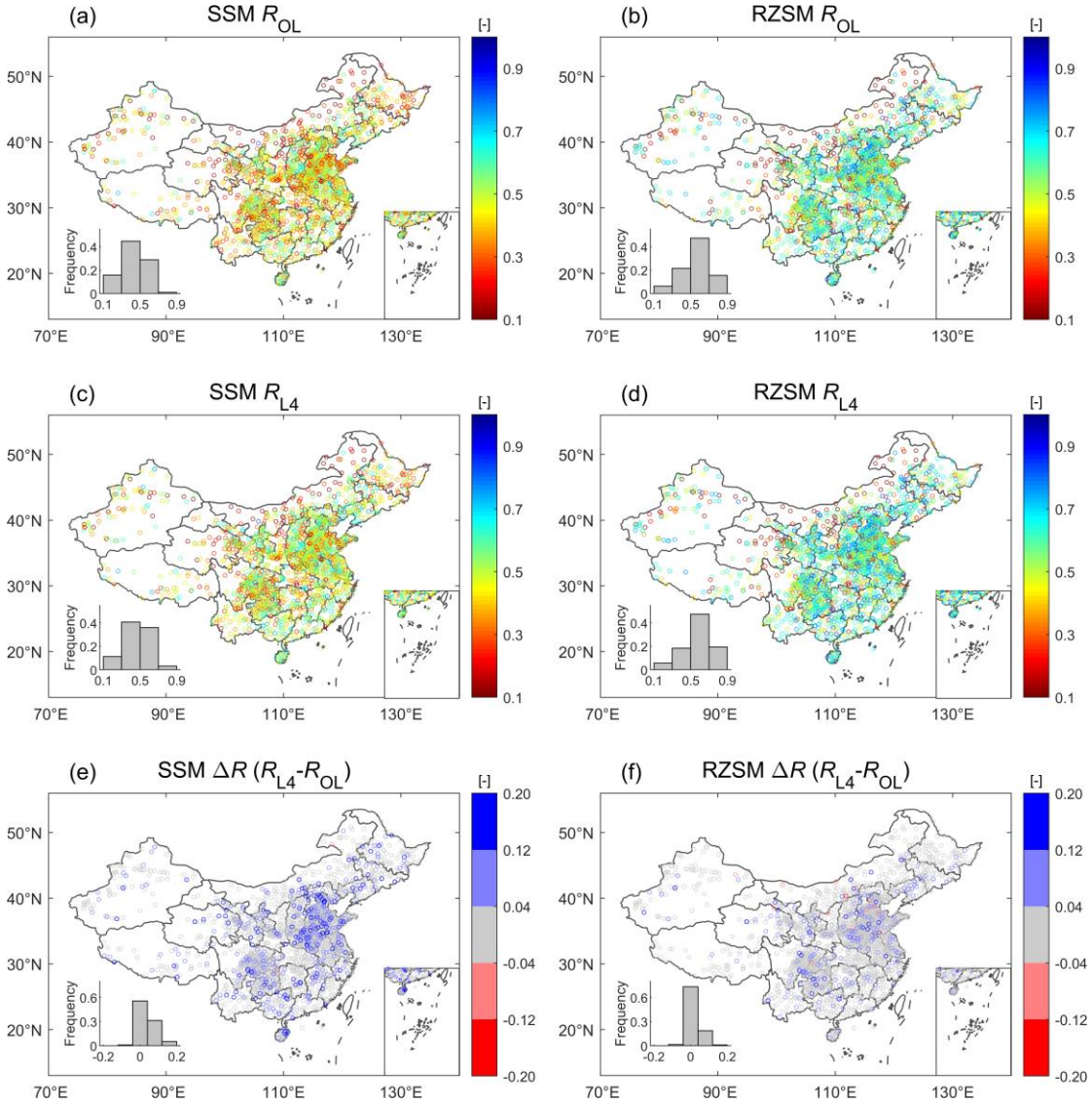

Fig. 1 Same content as in Fig. 1 of the manuscript, except that the correlation between in-situ soil moisture measurements and SMAP is measured using Pearson correlation.

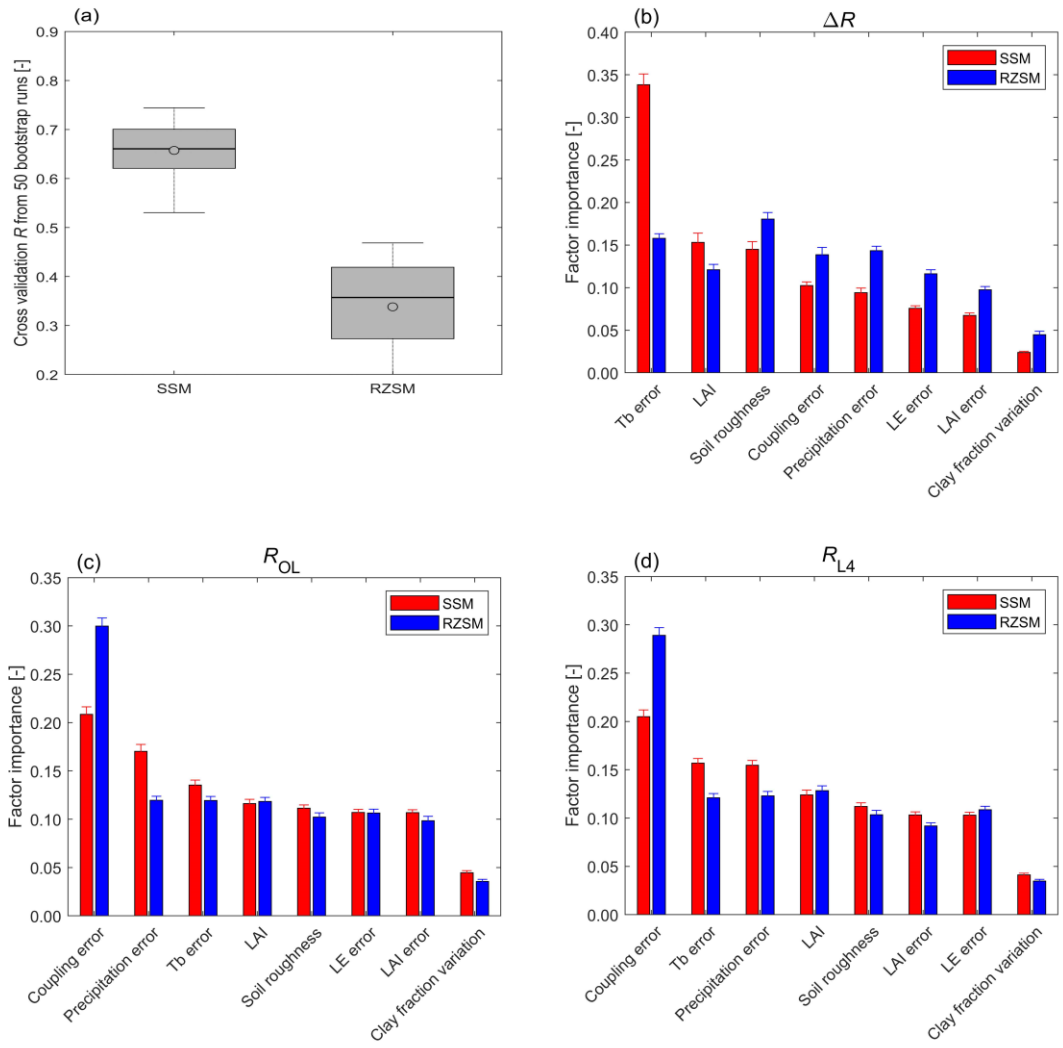

Fig. 2 Same content as in Fig. 3 of the manuscript, except that the correlation between in-situ soil moisture measurements and SMAP is measured using Pearson correlation.

The above response also applies to the Major comment #7 from Reviewer #2.

Re. the comment about the choice of the control factors: 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.

The above response also applies to the Major comment #1 by Reviewer #2 and Major comment #1 by Reviewer #3.

Re. the comment about the difference in clay fraction across the vertical: As stated in the manuscript, CLSM assumes that soil texture and related properties are vertically homogeneous within the soil column. However, the more realistic condition reflected by Harmonized World Soil Database generally shows variation in soil properties along vertical profile. Therefore, the vertical heterogeneity of soil texture may affect the accuracy of CLSM soil moisture accuracy. We will further clarify this in the corresponding paragraph of Section 2.3 in the revised manuscript.

6. In Table 1, please clarify why different LAI products are used? What is the relationship between these two LAI datasets? Why does SMAP L4 LAI be used for LSM rather than RTM, which simulates Tb that is used for comparisons to SMAP Tb.

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: 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. NASA Goddard Space Flight Center, Greenbelt, Md. Available at https://ntrs.nasa.gov/search.jsp?R=20160002967, 2015.

[2] Reichle, R. H., and Coauthors: 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, 2017.

The above response also applies to Major comment #5 by Reviewer #2.

7. In line 153, why is there a joint error in SMAP Tb observations and RTM Tb simulations? Sorry if I misunderstood something, what does "joint" mean? How do you quantify this joint error and what is the rationality behind?

The expression of "joint" is meant to refer to the combined error in the SMAP Tb observations and RTM Tb simulations. The DA innovation is estimated by subtracting the SMAP Tb observations from RTM Tb simulations. In the revised manuscript, we will modify this expression as "i) estimates of the DA innovation, namely difference between SMAP Tb observations and RTM Tb simulations".

8. In line 154, "the magnitude of LAI (as a proxy for the vegetation optical depth at microwave frequencies, which modulates the sensitivity of the observed Tb to SSM conditions)". The description is inaccurate. LAI should be as a proxy for the estimation of vegetation optical depth. Please clarify how vegetation optical depth modulates the sensitivity of the observed Tb to SSM conditions, it is hard to make the audience understand who does not be familiar with the zero-order RTM.

We thank the reviewer for bringing this to our attention. In the revised manuscript, we will clarify this expression.

9. In line 156-160, please make expressions precise. You give "e.g.," may I ask what else do you use, please list every item as accurate as possible, as such, readers and the author are on the same page. In line 160, I fully doubt "because increased LAI is associated with decreased soil moisture information content in microwave observations", is it true? How do you explain, for example, when vegetation is mature, the soil experiences drying and wetting processes? Please make expressions accurate.

In line 156-160, the first category of factors addresses errors fed into the L4 system include: 1) error in CLSM rainfall forcing data; 2) error in SSM-RZSM coupling strength; 3) vertical variability of clay fraction; 4) SMAP L4 LAI error; 5) output LE error; 6) Tb error. The second category of factors is based on the magnitude of the variable itself and include microwave soil roughness and annual mean LAI. We will make it clearer in the revised manuscript.

Regarding to the comment for Line 160, please see our reply to Major comment #8.

10. In line 204-205, please clarify the reason.

We have clarified why using the difference in clay fraction across the vertical soil

profile, i.e., the clay fraction difference between topsoil and deep-layer soil to quantify vertical variability in soil properties. Please see our reply to Major comment #5.

11. In line 210, why the anomaly SSM and RZSM are not used for Eq. 1, because in previous it is mentioned that anomaly Spearman's rank correlation is calculated with in-situ observations.

Indeed, the anomaly SSM and RZSM are used in the Eq. 1. We will make it clearer in the revised manuscript.

12. In line 214, "Cases with negative CP do not exist." I have litter doubt whether the in situ measurements will show that α is greater than 2.0, then CP can be negative? Please confirm this.

Based on the in-situ measurements during our 2-year study period, we do not observe any negative CP.

13. In line 227, please explicitly clarify "error" in FLUXCOM LE. Does this error refer to the uncertainties mentioned in line 186?

The FLUXCOM LE product is generated via merging energy flux measurements from FLUXNET eddy covariance towers with remote sensing and meteorological data. The error of FLUXCOM LE could stem from each data source and also from the merging process. This error also refers to the uncertainties mentioned in Line 186.

14. In line 235, please clarify "three independent sources (x, y and z)", does it refer to geographic location or one of the variables mentioned in your study? Please also explicitly explain two instrumental variables I and J. I did not see the time information mentioned in Eq. 2. Please is "(I and J, i.e., It = αxPt-1 + Bx + εxt-1, Jt = αyPt-1 + By + εyt-1)" important in the calculation, if so, please list it as an independent equation. Please clarify ε_(xt-1) or do you mean ε_(x,t-1)? Additionally, too much information is listed in brackets, shall readers ignore/keep this information? Please do revisions.

In Line 235, the expression of "three independent sources (x, y and z)" refer to any of three geophysical variables that are not linearly correlated in each of their time series.

The instrumental variable *I* refer to the lag-1 time series of variable *x*, and instrumental variable *J* refer to the lag-1 time series of variable *y*.

To be clearer, we will list the following equation originally listed in the bracket: $I_t = \alpha_x P_{t\text{-}1} + B_x + \varepsilon_{x,t\text{-}1}$, $J_t = \alpha_y P_{t\text{-}1} + B_y + \varepsilon_{y,t\text{-}1}$ as new Eqs. 3 and 4. In addition, $\varepsilon_{xt\text{-}1}$ and $\varepsilon_{yt\text{-}1}$ will be more precisely denoted as $\varepsilon_{x,t\text{-}1}$ and $\varepsilon_{y,t\text{-}1}$ respectively.

In correspondence with Eqs. 3 and 4, the Eq. 2 will be shown with time information,

which is $x_t = \alpha_x P_t + B_x + \varepsilon_{x,t}$

In addition, in the revised manuscript content within the brackets are rearranged to be clearer to readers in Section 2.5.

15. In line 255, "based on the output of RF", as a reviewer, I do not know more about RF, what are inputs for RF? I think the introduction of RF is too general and not informative. Please do revisions. Taking this paragraph as a case, past and present tenses are mixed used. Please do revisions.

As a machine learning based regression approach, RF is using the selected 8 control factors as regressors to estimate, or regress the DA efficiency (i.e., the difference of OL and DA soil moisture accuracy) for both SSM and RZSM estimates. Therefore, the input for RF is the 8 control factors (see Table 1) that covers two perspective of L4, and the output of RF is the DA efficiency in L4 SSM and RZSM sampled at 2714 sites. Note that, by training the 8 control factors to capture observed DA efficiency, RF can also summarize the relative importance of each control factor in controlling the L4 DA efficiency.

In addition, we will revise tenses throughout the manuscript.

16. In Fig. 1a-d, what is the maximum value for R? Can it reach 0.9? If not, please adjust the scalar. Please rewrote the caption of Fig.1.

The maximum of R in Fig. 1a-d can reach to approximately 0.9, so their common maximum will not be adjusted.

17. In line 261, "an increasing trend of SSM estimation skill moving from northwest to southeast China", if possible, please write a short sentence to explain the reason.

The reason for the observed "increasing trend of SSM estimation skill moving from northwest to southeast China" is most likely due to the similar spatial pattern of gauge density. We will briefly explain this in the revised manuscript.

18. In line 280-281," Errors in the CLSM precipitation forcing are relatively higher in northern and northwestern areas of China (Fig. 2a), where the gauge density is generally more sparse than southern China." I agree with this point. But I am sorry if I misunderstood. The magnitude of precipitation on the northwestern part may be smaller than on the southern part, as such, there is a possibility that errors may present a reverse trend, is this a case? Please confirm.

First of all, there is no direct evidence that the error of precipitation is linearly related to its magnitude. Secondly, by comparing the gauge density of northern and southern

China, we can observe a clear difference that the former is sparser than the latter, which inevitably result in higher interpolating error of precipitation forcing in northern China.

*19. Figure 2g, please revise the title as "soil roughness parameter *". In Fig. 2h, the maximum value of LAI is 2.0 m2/m2, please confirm. Fig. 2f, please revise the title as "the standard deviation of O-F Tb residuals". I think the meaning of "O-F Tb residuals" is different from Tb error itself.*

We will revise the titles of the subplot as suggested.

The maximum of annual mean LAI is higher than 2.0 m^2/m^2. In the original manuscript, we set the colorbar maximum to be 2.0 m^2/m^2, so that the spatial difference in LAI magnitude can be easily observed in Fig. 2h. We will recover the colorbar maximum of 4.0 m^2/m^2 and change the colormap to be nonlinear to reconcile the two issues. In addition, it should be noted that the SMAP LAI time series during growing season frequently exceed 4.0 m^2/m^2 as expected, whereas our Fig. 2h shows the annual mean LAI covering both growing and non-growing seasons and hence showing lower maximum value of LAI.

20. In line 297-298, "The 2017-2018 mean of soil roughness and the 2017-2018 mean LAI show higher values in southwest and southeast China (Fig. 2g-h)." The sentence is not informative. Please revise.

We will revise the original expression as: "The 2017-2018 mean of soil roughness shows a relatively scattered spatial pattern (Fig. 2g), while the 2017-2018 mean LAI shows higher values in southwest and southeast China (Fig. 2h)".

21. In line 335-336, OL run does not implement DA, why "Tb error (microwave soil roughness)" are involved. Please clarify. I am sorry if I misunderstood something.

Indeed, the OL run does not involve the implementation of DA, and its errors are therefore not related to Tb error or microwave soil roughness. We had tried to explain this point in the subsequent text, which stated that the observed high correlation between OL run skill and these two factors does not imply causality. However, showing the feature importance of Tb error and microwave soil roughness to the OL skill, or R_{OL} in Fig. 3c seems to be misleading anyway. Therefore, to avoid confusion, we will remove the result of RTM-related feature importance to R_{OL} in Fig. 3c.

22. In line 389, "OL does get worse with increasing roughness, there is more room for improvement as the roughness increases", please clarify whether the increase of roughness is physically reasonable.

In this context, we are not implying to increase soil roughness, which would be

physically unreasonable, as it is a function of soil moisture. The logic should be that in areas with higher soil roughness, the possibility of improving OL skill is higher. We will further clarify this point in the revised manuscript.

23. In line 441, "it is unclear whether or not the observed SSM-RZSM coupling strength biases are real in an absolute sense – or simply reflect inconsistencies in the depth of modelled versus observed SSM and RZSM time series". I am sorry, I am confused whether the coupling strength based on in situ measurements can represent the real?

Ideally, when comparing the SSM-RZSM coupling strength of in-situ measurement and that of CLSM, SSM and RZSM data for identical depths from both in-situ measurements and CLSM should be used. However, the depth of first-layer SSM measurement is 0~10cm, which is thicker than CLSM SSM of 0~5cm. This discrepancy could inherently result in higher fluctuation of CLSM SSM simulation than that of SSM measurement, and consequently lower SSM-RZSM coupling of CLSM simulation (CP_{OL}) than that of measurement (CP_{obs}). Therefore, we cannot conclude that the observed lower CP_{OL} compared to CP_{obs} is due to the negative bias of SSM-RZSM coupling strength, or the depth inconsistencies of CLSM modelled versus observed SSM and RZSM time series. We will further clarify this point in the revised manuscript.

24. In Conclusions, the second and fourth paragraphs have duplicate content. Please do revisions.

We will remove the duplication in the revised manuscript.

25. In line 451-452, "the partitioning of the available energy into latent and sensible heat (LE error) and the microwave radiative transfer modeling (Tb error)." is not informative.

We will revise the original expression as follows: "…additional focus should thus be placed on improving the model's characterization of the microwave radiative transfer modeling (Tb error), together with the partitioning of the available energy into latent and sensible heat (LE error)."

Minor comments:

1. Please give the full name for abbreviations when they appear for the first time. The examples are SPOT VGT and EASE. Please carefully check throughout the manuscript.

We will carefully check the first-time abbreviations and make corresponding revisions throughout the manuscript.

2. In line 216, SMAP L4 CP estimates (CP_{OL}), please confirm. You mentioned SMAP L4 is the assimilation experiment.

We will revise the original expression as "CP estimates of OL (CP_{OL})…"

3. Please confirm the use of RTM-related, R-values, and so on throughout the whole paper, as well as the use of "their" and "our".

We will carefully check those occurrences and make corresponding revisions throughout the manuscript.

Re. the comments in the annotated manuscript pdf file:

We will make corresponding revisions in the manuscript.