Response to Reviewer #3's comments on the manuscript egusphere-2023-1597

RC3: 'Comment on egusphere-2023-1597', Anonymous Referee #3, 23 Nov 2023

The manuscript presents an intercomparison between eight root zone soil moisture (RZSM) products and in situ measurements. The differences are discussed in the light of the uncertainty in precipitation, air temperature and soil properties.

Overall, I think the study could be a valuable scientific contribution. However, at the current status, I think the manuscript should be improved in several parts before reaching good quality for a possible publication. Specifically, I think the Authors should put major effort into improving the descriptions of the products, the methods should be extended, the discussion should be integrated accordingly. Below I provide additional details about my major concerns followed by specific comments.

The authors thank the Reviewer #3 for her/his constructive and insightful comments that help us improve the quality of the manuscript. The original comments from Reviewer #3 are in black font, and our responses are in blue font.

GENERAL COMMENTS

Irrigation

A major confusion in my opinion in the intercomparison is the role of irrigation. At L119 and L124 it stated that the study is conducted over a highly irrigated area. At L135 it is stated that soil moisture sensors are in areas without irrigation. It is not clear by reading if the RZSM products account for irrigation or not and if this can be a concern. Only later in the discussion (L452), it is stated that the overestimation of RZSM by ERA5 (Fig. 3) could be a signature of irrigation because the in situ RZSM observations do not capture irrigation. Does this mean that some RZSM products are based on model that take into account irrigation and others not? On the one hand, this information should better explained and discussed. On the other hand, I wonder what is the scientific meaning of comparing soil moisture in rainfed area to model that are accounting for irrigation.

Response: We completely agree with this comment from reviewer 3, which is also proposed by reviewer 2. For the sake of clarification, we will not emphasize the role of irrigation any more. L135 "Stations are located in areas without irrigation" and L452-453 "The overestimation of RZSM by ERA5 (Fig. 3) could be a signature of irrigation because the in situ RZSM observations do not capture irrigation" will be deleted in the revised manuscript.

And we try to address the confusion mentioned above by reviewer 3.

First, the Huai River Basin is highly irrigated.

Second, L135 "Stations are located in areas without irrigation", it means that the soil moisture probes are installed away from crops, which intends to avoid capturing the irrigation signal and obtain the natural soil moisture states.

Finally, the RZSM products don't take irrigation into account. And L452-453 are incorrect statements and will be removed. For example, ERA5 didn't model irrigation (Lavers et al., 2022).

Soil map and soil parameters (section 4.3.2 and section 5.2).

The assessment of the soil properties is valuable. It should be noted, however, that high discrepancies come with the use of different pedotransfer functions (PTF) to derive soil hydraulic parameters (retention curve and hydraulic conductivity). I guess these parameters are used in each RZSM product but estimated with different PTFs. This could also explain part of the uncertainty. This information is missing in the manuscript but should be integrated.

Response: Thank you for the valuable suggestion, we will add the statements about PTFs in section 5.2.

"The soil hydraulic parameters (SHPs), such as the hydraulic conductivity and matric potential, are crucial parameters for the transport of soil moisture between soil layers through the Richards' equation within the LSMs. Generally speaking, the SHPs are derived from a combination of soil properties (clay, sand, silt fractions and organic content, etc.) with pedotransfer functions (PTFs). However, different input variables and functional forms of the continuous PTFs are used in the LSMs. For example, The PTFs could be constructed by multivariate regression models, nonlinear regression models or artificial neural networks (Harrison et al., 2012). The SHPs can also be derived from a combination of soil properties with PTFs or remotely sensed soil moisture retrievals (Santanello et al., 2007). Therefore, soil moisture show great uncertainty and varies considerably across different LSMs. SMAP L4 soil moisture product adopts PTFs provided by Wösten et al. (2001) which takes the organic carbon affecting soil hydraulic and thermal parameters into consider. MERRA-2 adopts PTFs adapted from Cosby et al. (1984) without consideration of organic carbon (De Lannoy et al., 2014). Therefore, the PTFs could also explain part of the uncertainty.

Spatial aggregation and comparison at each site should be clarified and improved

As far as I have understood (L267), the comparison is conducted between the spatial average of the 58 in situ observations. It is not well reported how much is the spatial extent of the 58 stations but looking at figure 1 I guess the station covers an area of around 300 x 200 km2. I then deduce that this spatial average is compared to the spatial average of the gridded products (i.e., each product has different resolutions, but more than one cell of the gridded products covers the area of the 58 stations). So first of all it should be better explain how many cells have also been aggregated for each product. Only later in the results section (L317) I discovered that a comparison has also been performed without aggregating spatially. So, first of all, this information should be provided also before in the methods. Moreover, I would also considering moving some plots that are now in supplement to the main manuscript to strengthen the analysis. Anyway, I'm confused by the fact that the comparison is performed at each station. Does this mean that you have always one station against one cell of the gridded products? Please clarify.

Response: We feel sorry to make you confused. And we will try to resolve your confusion.

1. The spatial extent of the 58 stations is approximately 310×330 km^2.
2. For different RZSM products, different numbers of grids are aggregated. For example, CLDAS: 58, GLDAS_CLSM: 50, GLDAS_NOAH: 50, ERA5: 48, MERRA-2: 50, NCEP CFSv2: 55, SMAP L4: 58, SMOS L4: 51.
3. In this manuscript, Fig.2 and 3 represents that the comparison of the RZSM time series

averaged over all in situ stations with the RZSM time series averaged over all model grids where the stations are located. L317 represents the single point-grid validation, the measurements at each station are compared directly with the grid values where the station is located.

The following text (section 3.3 Validation strategies) will be added in chapter 3 Methods.

"In terms of the temporal resolution, except for the RZSM products (e.g., GLDAS_CLSM, SMOS L4) provided on daily time steps, the other sub-daily RZSM datasets (hourly/3-hourly/6-hourly time steps, shown in Table 1) are aggregated to daily average values. Therefore, the aggregated RZSM products could match the observations at daily time intervals. In terms of spatial resolution, we didn't change the spatial resolution of any RZSM products and used the original grid resolution. Two validation strategies were used in the study. The first is to compare the RZSM time series averaged over all in situ stations with the RZSM time series averaged over all model grids where the stations are located (Fig.2 and 3 shown in this study). The second one is the single point-grid validation, the measurements at each station are compared directly with the grid values where the station is located. If there is more than one station within a grid, the measurements of each station that located in the grid are compared to the grid values separately. The point-grid validation has been provided in the supplement (Fig. S2 and S3)."

Section 2.4: description of the eight RZSM products

The description of the eight products should be improved in several parts. The information provided for each product is not always consistent. Some products are better described and with more details than others. E.g., for MERRA-2, the description focuses on the precipitation. Instead, ERA5 does not have any information about. NCEP CFSv2 description is very short. Who provided that? What are the main properties? Some characteristics provided should also be put more in relation to the focus of the paper. E.g., for ERA5, the data assimilation system is described in detail. Is that relevant for the purpose of the paper. If yes, it should be clarified. Overall, the main differences between the products relevant for the present study (e.g., soil map, precipitation, land use, irrigation etc.) should be better highlighted. Table 1 should be extended accordingly.

Response: We will revise the whole section 2.4 thoroughly and Table 1, and provide more focused and consistent description of the eight RZSM products in the revised manuscript.

Please note that some relevant information are discussed only later but in my opinion they should be moved to the method section. This would help understanding and strengthening the discussion of the results. E.g., L410 The soil properties data used in the eight RZSM products were all derived from the FAO/UNESCO soil map of World except for CLDAS, which used the soil data developed by Shangguan et al. (2013), and SMAP L4, which used the HWSD soil properties over China. L426. Global precipitation and air temperature forcing data are used in the production of all RZSM products except for SMOS L4. L452. The overestimation of RZSM by ERA5 (Fig. 3) could be a signature of irrigation because the in situ RZSM observations do not capture irrigation.

Response: We will add the following contents to section "3.3 Validation strategies" to make it easier for the readers to follow and understand. L452 will be removed.

"The Global precipitation and air temperature forcing data are used in the production of all RZSM products except for SMOS L4, which are validated against the China Daily Gridded Ground Precipitation and Air Temperature dataset V2.0 described in section 2.2. The soil properties data used in the eight RZSM products were all derived from the FAO/UNESCO soil map of World except for CLDAS, which used the soil data developed by Shangguan et al. (2013), and SMAP L4, which used the HWSD V1.2 soil properties over China. The China soil dataset developed by Shangguan et al. (2013) is used as a reference to evaluate the accuracy of FAO/UNESCO and HWSD V1.2 soil properties (clay and sand content, organic carbon content and bulk density)."

Figure quality

Figures are not always readable and meaningful. I suggest putting some more effort into evaluating how to present the results. E.g.,

Fig 1. could also shows the pixel size of the products. This would help understanding spatial extend and intercomparisons.

Response: Fig. 1 has been revised and shows the grid size of 0.25 degree covering the in situ stations.

Fig. 4 is not readable at all. Plots and texts are too small.

Response: Fig. 4 has been revised with improved plots and texts.

Fig. 5 can be improved by having only 8 histograms of the RZSM products and overlapping each histogram with the observation's histogram. This could help to visualize the differences.

Response: Fig. 5 has been revised.

Fig.6. The plots are hardly comparable. It could be evaluated to present one plot with all the cumulative precipitation, or histogram of the precipitation etc.

Response: We have presented one new plot with all the cumulative precipitation. Since most of the daily precipitation ranges from 1 to 10 mm day^{-1}, the histogram or probability density function of different precipitation datasets can't be well distinguished from each other. Therefore, it is not included in the plot.

Fig.7. It is not clear to me what is actually presented. This is in line to the general comment above about aggregation. Are you comparing spatial averaged precipitation? What does standard deviation refer to? If you compare each rain gauge to pixel wise, how have you aggregated?

Response: For Fig. 7, we compare each rain gauge to pixel wise where the rain gauge is located. The 6 statistical metrics in Fig.7 are calculated at each station, so there are 58 data points for each statistical metric. The histogram represents the median of 58 data points for each statistical metrics between modelled precipitation and observations. The stand deviation represents the variability in the statistical metrics. We will add more detailed descriptions to the legend of Fig. 7.

Take-home-message

I'm expecting to read the overall take-home-message. After performing this intercomparison, what can you conclude? Could we trust this products? Where and which conditions? How would you suggest further improving, studies etc.? This is missing throughout the manuscript but should be understandable from the abstract and more extended at the conclusions.

Response: We will provide concise and focused statement about the intercomparison in the abstract, and the main limitations and inspirations of the intercomparison will be illustrated in the conclusions. Overall, the GLDAS_CLSM outperforms the other RZSM products and could be used for drought monitoring and flood forecast in the Huaibei Plain. The intercomparison (revised Fig. 4) shows that SMOS L4 shows the very low correlation (around 0.4) and high dry bias with any one of the other RZSM products. However, except for SMOS L4, the model-based RZSM products show high consistency (R above 0.7) with each other. This phenomenon could be caused by the fact that the precipitation is not used in the production of SMOS L4. So, the SMOS L4 RZSM has a weaker response to precipitation than other RZSM products. Moreover, SMOS L4 RZSM is consistently lower than any one of the other RZSM products. The correlation R between SMAP L4 RZSM and MERRA-2 RZSM is the highest among any two of the seven LSM-based RZSM products, which could be related to both RZSM products use the same CLSM and meteorological forcing derived from GEOS-5 model system. The revised Fig. 5 shows the frequency distribution of normalized soil moisture for eight RZSM products and in situ observations. It is clear that the histograms MERRA-2, GLDAS_CLSM and SMAP L4 shows the better consistence with observations, although they also slightly overestimate the frequency of wet soil moisture. However, all of them didn't capture the peak and underestimate the frequency of normalized soil moisture ranging from 0.2 to 0.4. The other RZSM products show an obvious offset towards wet soil moisture. The data provider should produce less wet soil moisture and more dry soil moisture.

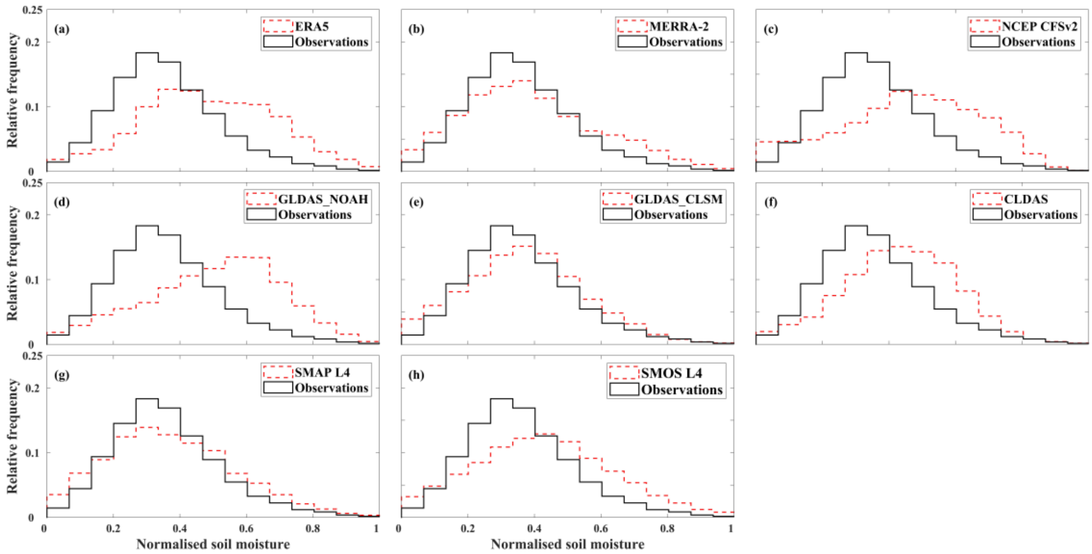

Fig. 5 The histograms of normalised RZSM products (dashed and red lines) and in situ observations (black and solid lines).

SPECIFIC COMMENTS IN ORDER OF APPEARANCE (L = LINE NUMBER)

L13. I would not use the term “direct validation” but “assessment”.

Response: “A direct validation” will be replaced by “The assessment”.

L29. Th abstract focused on describing the actual results. This is fine but I would also expect at the end to read the take-home-message. E.g., what do we learn by this study? Can we trust, use, apply RZSM products? Where? In which conditions? What in our view and based on this study would further suggest to improve the performances?

Response: The abstract will be replaced by the following text in the revised manuscript.

“Root zone soil moisture (RZSM) is critical for water resource management, drought monitoring and sub-seasonal flood climate prediction. While RZSM is not directly observable from space, several RZSM products are available and widely used at global and continental scales. This study conducts a comprehensive quantitative evaluation of eight RZSM products over the Huai River Basin (HRB) in China. The assessment is performed using observations from 58 in situ soil moisture stations from 1 April 2015 to 31 March 2020. Attention is drawn to the potential factors that contribute to the uncertainties of model-based RZSM, including errors in atmospheric forcing (precipitation, air temperature), vegetation parameterizations, soil properties, and spatial scale mismatch, etc. The results show that the Global Land Data Assimilation System Catchment Land Surface Model (GLDAS_CLSM) outperforms other RZSM products with the highest correlation coefficient (R=0.69) and the lowest unbiased root mean square error (ubRMSE=0.018 m^3 m^{-3}), respectively. All RZSM products tend to overestimate in situ soil moisture values, except for the Soil Moisture and Ocean Salinity Level 4 (SMOS L4) product, which underestimates RZSM. The underestimation of Surface Soil Moisture (SSM) in SMOS Level 3 (L3), caused by underestimated physical surface temperature and overestimated ERA interim soil moisture, may contribute to the underestimation of RZSM in SMOS L4. The other model-based RZSM products show an overestimation of in situ observations, which could be associated with the overestimation of the precipitation amounts and precipitation events (drizzle effects) and the underestimation of air temperature. In addition, the biased soil texture (organic carbon, clay and sand fractions) and flawed vegetation parameterizations (e.g., canopy and root tissues) affect the hydrothermal transport processes represented in different LSMs, leading to inaccurate soil moisture. The intercomparison of the eight RZSM products shows that MERRA-2 and SMAP L4 RZSM have the highest correlation, which could be attributed to the fact that both products use the catchment land surface model and the atmospheric forcing provided by the Goddard Earth Observing System Model, version 5 (GEOS-5), although the versions differ slightly. This in situ validation shows that GLDAS_CLSM could be used for drought monitoring and flood forecast in Huaibei Plain. Moreover, the RSZM intercomparison indicates that the model should focus on increasing the frequency of dry soil moisture, decreasing the frequency of wet soil moisture and the ability to capture the frequency peak of soil moisture. The uncertainty analysis implies that the model-based RZSM can be improved by correcting precipitation, using more accurate soil properties and more prefect vegetation parameterization schemes, etc.”

L108-109. Are these two lines really needed here? I would integrate this information later when you speak about comparison. E.g., L133 for the definition of RZSM.

Response: L108-109 will be deleted. L133 will be replaced by “Since the study aims to evaluate the accuracy of eight RZSM products (0-100 cm) which are summarised in Table 1, the in situ soil moisture measurements at the four depths are depth-weighted averaged to obtain the 0-100 cm soil moisture data."

L128. Table S1 shows the results of the assessment, and it does not provide additional information about the in situ stations. I would remove this cross reference here and rather add Figure 1 where the locations of the in situ stations are shown.

Response: "Table S1" will be replaced by "Fig. 1".

L263. I would extend a bit on the meaning and interpretation for PD, FAR and CSI.

Response: The following text will be added in section "3.1 Statistical metrics".

"POD is the proportion of real precipitation events simulated by AGCM relative to the actual precipitation events, reflecting the ability of AGCM to detect precipitation. FAR is the fraction of unreal precipitation events out of the total precipitation events simulated by AGCM. CSI is a more balanced score that combines the characteristics of false alarms and missed events, representing the probability of successful simulation of AGCM precipitation.

L280. I was expecting to read more about the description of the equation after L280. Any text missing?

Response: We will add the following description after L280.

"where θ_{RZSM} denotes 0-100 cm RZSM (m^3 m^{-3}), θ_{0-10cm}, $\theta_{10-40cm}$ and $\theta_{40-100cm}$ denote the 0-10 cm, 10-40 cm and 40-100 cm soil moisture, respectively."

L293. I think here is a good place to cite table 3 as well.

Response: L293-294 will be replaced by "Figure 2 shows scatterplots of RZSM products against the in situ measurements averaged across all in situ stations over the HRB, from 1 April 2015 to 31 March 2020 (see Table 3)."

L295. It is stated that SMOS-L4 underestimates and the other overestimated the observations. By looking at figure 2, I see the opposite. Please double check what you are plotting.

Response: Thank you for bringing up the mistake. We reversed the labels for the x and y axes and we have revised the Figure 2.

L306. Figure 3 shows spatial average of in situ soil moisture and its spatial variability. As far as I have understood (see general comment above on the spatial aggregation), the spatial average of the RZSM products is shown but, if possible, it could be interesting to show here also the spatial variability of the RZSM products.

Response: L306 will be replaced by "Figure 3 shows time series of in situ RZSM observations averaged over all in situ stations". We will add one figure in the supplement regarding the spatial distribution of eight RZSM products averaged from 1 April 2015 to 31 March 2020.

L317. Description of the comparison at each site should be reported as well in the method (section 3.1) See also general comment above on the spatial aggregation.
Response: See the response in **Spatial aggregation and comparison at each site should be**

clarified and improved (Page 2).

L319. The method to calculate the anomalies is reported only in the supplement. I think should be moved to the method (section 3.1)

Response: The anomalies metrics are not used in the manuscript.

L585. After summarizing the mani conclusions, I suggest summarizing the outlook of the study. See also general comments above.

Response: The following text will be added in L585.

"(5) Eight RZSM products are evaluated over the Huaibei Plain under the homogeneous underlying surface conditions. However, different vegetation cover has a large impact on the soil moisture simulations. The role of vegetation cover types in the uncertainty of model-based soil moisture can be further explored, and more detailed vegetation parameterizations (canopy and root tissue) can be discussed at the point scale.

(6) Based on the main conclusions of the study, many factors contribute to the uncertainty of model-based RZSM simulations. Precipitation plays a crucial role. The AGCM-derived precipitation can be corrected in different ways to drive LSMs and hydrological models, and to analyze how the simulated land surface states and fluxes to respond the corrected precipitation."

Supplement. I think the supplement should have a title with the name of authors as well.

Response: "Supplement of

Evaluation of root-zone soil moisture products over the Huai river basin"

Liu En et al.

Correspondence to: Yonghua Zhu (zhuyonghua@hhu.edu.cn)" will be added in the supplement.

Reference

Cosby, B., G. Hornberger, R. Clapp and Ginn, T.: A statisticsl exploration of the relationships of soil moisture characteristics to the physical propertie of soil, Water Resour Research, 20, 682-690,1984.

De Lannoy, G. J. M., Koster, R. D., Reichle, R. H., Mahanama, S. P. P. and Liu, Q.: An updated treatment of soil texture and associated hydraulic properties in a global land modeling system, J. Adv. Model Earth Syst., 6, 957-979, https://doi.org/10.1002/2014ms000330, 2014.

Harrison, K. W., Kumar, S. V., Peters-Lidard, C. D. and Santanello, J. A.: Quantifying the change in soil moisture modeling uncertainty from remote sensing observations using Bayesian inference techniques, Water Resources Research, 48, 10.1029/2012wr012337, 2012.

Lavers, D. A., Simmons, A., Vamborg, F. and Rodwell, M. J.: An evaluation of ERA5 precipitation for climate monitoring, Quarterly Journal of the Royal Meteorological Society, 148, 3152-3165, https://doi.org/10.1002/qj.4351, 2022.

Santanello, J. A., Peters-Lidard, C. D., Garcia, M. E., Mocko, D. M., Tischler, M. A., Moran,

M. S. and Thoma, D. P.: Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed, Remote Sens. Environ., 110, 79–97, 10.1016/j.rse.2007.02.007, 2007.
Wösten, J. H. M., Pachepsky, Y. A. and Rawls, W. J.: Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics, J. Hydrol., 251, 123-150, https://doi.org/10.1016/S0022-1694(01)00464-4, 2001.