Dear Editor and Reviewers,

We would like to thank the Editor and three Reviewers for their efforts in handling the manuscript and for their valuable comments to improve the manuscript. We have revised our manuscript thoroughly according to the comments and provide a point-by-point response to the comments from three Reviewers below. The original comments from three Reviewers are in black font, and our responses are in blue font.

On behalf of all co-authors,
En Liu

Response to Reviewer #2's comments on the manuscript egusphere-2023-1597

RC2: 'Comment on egusphere-2023-1597', Anonymous Referee #2, 12 Mar 2024

I believe that the manuscript has been improved based on the reviewer comments. I still have difficulty in grasping the key messages and contributions from the paper. I think this can be overcome by critically removing text that does not directly contribute to the main conclusion. This would make the paper more readable. Besides, I think that strengthening (and shortening) the abstract and conclusion can definitely help to better convey the main message.

Response: The authors thank the Reviewer #2 for her/his constructive and insightful comments that help us improve the quality of the manuscript. The original comments from Reviewer #2 are in black font, and our responses are in blue font.

See below some specific comments:

- The abstract is relatively long and not entirely understandable by itself. For example L28: 'The seven model based products' raises questions by the reader; too many details on SMOS for the abstract; a lot of emphasis on the limitations. I believe that addressing these issues in the abstract will make it more clear and impactful.

Response: We have strengthened and shortened the abstract and conclusion to make it more concise and readable.

The abstract is replaced by

"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 and quantitative evaluation of eight RZSM products using observations from 58 *in situ* soil moisture stations over the Huai River Basin (HRB) in China. Attention is drawn to the potential factors that contribute to the uncertainties of model-based RZSM, including the errors in atmospheric forcing, vegetation parameterizations, soil properties, and spatial scale mismatch. The results show that the Global Land Data Assimilation System Catchment Land Surface Model (GLDAS_CLSM) outperforms the other RZSM products with the highest correlation coefficient (R = 0.69) and the lowest unbiased root mean square error (ubRMSE = 0.018 $m^3 m^{-3}$). While SMOS Level 4 (L4) RZSM shows the worst performance among eight RZSM products. The RZSM products based on land surface models generally perform better in the wet season than in the dry season due to the enhanced ability to capture of the temporal dynamics of *in situ* observations in the wet season and the inertia of remaining high soil moisture values even in the dry season. While SMOS L4 RZSM product, derived from SMOS L3 surface soil moisture (SSM) combined with exponential filter method, performs better in the dry season due to the attenuated ground microwave radiation signal caused by the increased water vapor absorption and scattering in the wet season. The underestimated SMOS L3 SSM triggers the underestimation of RZSM in SMOS L4. The overestimated RZSM products based on land surface models could be associated with the overestimated precipitation amounts and frequency, the underestimated air temperature and ratio of transpiration to the total terrestrial evapotranspiration. In addition, the biased soil properties and flawed vegetation parameterizations affect the hydrothermal transport processes represented in different LSMs and lead to inaccurate soil moisture simulation. The scale mismatch between point and footprint also introduces representative errors. The comparison of frequency of normalized soil moisture between RZSM products and *in situ* observations indicates that the LSMs should focus on reducing the frequency of wet soil moisture, increasing the frequency of dry soil moisture and the ability to capture the frequency peak of soil moisture. The study provides some insights into how to improve the ability of land surface models to simulate the land surface states and fluxes by taking into account the issues mentioned above. Finally, these results can be extrapolated to other regions located in the similar climate zone, as they share the similar precipitation patterns that dominate the terrestrial water cycle."

The conclusion is replaced by

"In this study, eight RZSM products were quantitatively evaluated against observations from 58 *in situ* soil moisture stations over the HRB in China. The impact of several potential confounding factors on the uncertainty of RZSM products was investigated, including meteorological forcing variables, soil properties, soil stratification, vegetation parameterization and spatial scale mismatch. Nevertheless, there are still some shortcomings to be overcome in this study. The land cover type affects the dynamics of soil moisture, future study should focus on the effect of different land cover types on soil moisture simulation. The main conclusions of this study are as follows:

(1) GLDAS_CLSM outperformed the other RZSM products over the HRB, followed by MERRA-2, CLDAS, SMAP, ERA5, NCEP CFSv2, and GLDAS_NOAH. The SMOS L4 product presented the worst performance due to the fact that SMOS L4 does not contain precipitation information and has a weaker response to precipitation. Seven RZSM products based on land surface models overestimated the *in situ* observations with median bias values ranging from 0.033 $m^3 m^{-3}$ (SMAP L4) to 0.116 $m^3 m^{-3}$ (CLDAS). While SMOS L4 underestimated the RZSM with a median bias value of -0.050 $m^3 m^{-3}$.

(2) The intercomparison of RZSM products shows that the correlation coefficient R between any two of the seven model-based RZSM products varied from 0.68 (ERA5 vs. CLDAS) to 0.95 (SMAP L4 vs. MERRA-2). In contrast, SMOS L4 presented a lower correlation with the other seven RZSM products with R ranging from 0.30 (MERRA-2) to 0.41 (GLDAS_NOAH). The comparison of the frequency distribution between eight RZSM products and *in situ* observations indicates that all RZSM products overestimate the frequency of wet soil moisture and underestimate the frequency of dry soil moisture.

Besides, the frequency peaks of eight RZSM products are underestimated and show an obvious offset towards wet soil moisture compared to the *in situ* observations. Therefore, the Richard's equation in LSMs should focus on producing less wet soil moisture and more dry soil moisture.

(3) Except for CLDAS, the overestimated RZSM products based on land surface models could be associated with the overestimated precipitation amounts and frequency, underestimated air temperature and ratio of transpiration to the total terrestrial evapotranspiration existing in most earth system models, which consumes less water in the root zone for transpiration. The underestimation of the SMOS L4 RZSM is related to the underestimation of the SMOS L3 SSM.

(4) The model-based RZSM products generally perform better in the wet season than in the dry season due to the enhanced ability to capture of the temporal dynamics of *in situ* observations in the wet season and the inertia of remaining high soil moisture values even in the dry season. While SMOS L4 performs better in the dry season than in the wet season, as the ground microwave radiation signal is more attenuated in the wet season due to a substantial increase in water vapor absorption and scattering, which is propagated to SMOS L4 RZSM.

(5) Spatial-average validation could reduce the spatial noise of in situ soil moisture measured at different locations and improve the representativeness of soil moisture observations to model-based grid values.

(6) The study could provide some insights into how to improve the ability of land surface models to perform the land surface analysis by addressing the above issues. Furthermore, these results can be extended to other regions to improve the numerical simulation capability of land surface models at global scale."

- L28: we have eight products, but only seven model-based products? I assume SMOS is not considered as model based? Please clarify this throughout the text and in Table 1

Response: Yes, SMOS L4 RZSM product is not considered as model based. SMOS L4 RZSM product is obtained from SMOS L3 surface soil moisture combined with exponential filter method, which have been mentioned in abstract, Table 1 and section 2.4.8.

- L266 doesn't -> does not

Response: Correction done.

- Table 2: which correlation coefficient? Pearson?

Response: Correction done.

- Fig 4: Why are the outliers red, and not the same color as the corresponding box?

Response: The outliers are abnormal values, which is used to distinguish from the normal values.

- 5.3 title: capital W in 'What'

Response: Correction done.

-L740-743: this would fit better at the end of the conclusion

Response: Correction done.

- The authors have put a lot of effort in evaluating potential causes for mismatches between the different datasets, could the authors also put this into perspective in the conclusion? I mean: what is needed to overcome the major issues?

Response: Yes, the following text is added into the conclusion.

“Nevertheless, there are still some shortcomings to be overcome in this study. The land cover type affects the dynamics of soil moisture, future study should focus on the effect of different land cover types on soil moisture simulation.”

- Can you put the potential causes for the mismatches into perspective while considering the spatial resolution of the models assessed here? For example, it is not likely that the models will ever cover all the details on the soil textures as presented in Sect. 5.2. The answer to the section title: 'are the soil properties correctly represented' is obviously no. Can you refer to literature that is trying to overcome this issue in the model world? - L104: 2.84m should be 2.89m

Response: Due to the influence of precipitation and underlying surface conditions, soil moisture shows great spatial heterogeneity. Therefore, the validation of point-scale site observations and grid-scale soil moisture simulation values results in spatial scale mismatch. How to address the issue has been illustrated in section 4.3.3 (Line 535-538).

“ In addition, upscaling the sparse ground-based observations to the footprint-scale satellite soil moisture retrieval or model grid scale through the temporal stability concept, block kriging, field campaign data, or LSM, reduces the uncertainty of spatial resampling and further improves the reliability of soil moisture validation (Crow et al., 2012).”

2.84m have been replaced by 2.89m.

Reference

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D. and Walker, J. P.: Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, Rev. Geophys., 50, RG2002, https://doi.org/10.1029/2011rg000372, 2012.