Response to Reviewer #1's comments on the manuscript egusphere-2023-1597

RC1: 'Comment on egusphere-2023-1597', Anonymous Referee #1, 07 Nov 2023

The authors attempted to evaluate the predictive performance of eight satellite data-derived root-zone soil moisture data. The authors provided a quantitative statistical analysis of each RZSM product based on average values of in-situ, and remote sensing estimates for the Huai River Basin. The authors concluded that the GLDAS CLSM RZMS products outperform other RZSM products. While the information contained in the manuscript is quantitative, I ended up questioning the potential contribution of this paper to the readers. I have tried to address why I think

The authors thank the Reviewer #1 for her/his constructive and insightful comments that help us improve the quality of the manuscript. The original comments from Reviewer #1 are in black font, and our responses are in blue font.

- Major comments:

1. Although the study provides a significant amount of comparative analysis between remote sensing-derived RZSM products as well as against observational data, the mechanistic understanding and explanation of each result are widely missing across the manuscript. Therefore, the author's rationale about the causes of differences was often too obvious or uncertain.

   Response: Thank you for your valuable comment. In this study, we presented an intercomparison between eight root zone soil moisture (RZSM) products and in situ observations. Although the different RZSM products are evaluated quantitatively, it should be noted that the focus of this study is to investigate the sources of error of the RZSM products. The SMOS Level 4 (L4) RZSM is produced by combining a modified exponential filter and SMOS Level 3 surface soil moisture (SSM). The other seven RZSM products (except SMOS L4) are produced by land surface models driven by surface atmospheric forcing data from atmospheric general circulation models; the ability of the land surface model to simulate states (e.g., RZSM) and fluxes is limited by uncertainties in meteorological forcing and parameters, as well as inadequate model physics. It is difficult to quantify the model physics because different land surface models are used to produce different RZSM products. Therefore, we analyzed the atmospheric forcing data (especially for precipitation, which dominates the terrestrial water cycle), which is considered as the most important factor in determining the accuracy of the modeled RZSM, and the model static parameters (soil properties), which strongly influence the movement of water in the vadose zone, and local underlying surface conditions. The objective of this study is to provide some insights on how to improve the ability of land surface models to simulate land surface states and fluxes by finding the common characteristics of bias existing in different datasets used in land surface models.

2. Also, the authors argued that GLDAS CLSM-derived RZSM outperforms other RZSM estimates. However, rather than trying to explain the different predictive performances of each RZSM product (against in-situ observations) in relation to soil properties, land cover/use, and vegetation in the study catchment, the authors just used the average of 58 in-situ data as well as satellite products, resulting in 'all-lumped' single time series for each dataset. Thus, it is not convincible to say a certain remote sensing-derived estimates outperform others since the performance differences can be revealed differently depending on soil properties, vegetation, land cover/use, etc.

Response: Most of the in situ stations are located in the Huaibei Plain, which is a major grain production area. According to the land cover map of the Huai River basin (Figure 1c), most of them (56 of 58) is located in the cropland regions, we will add it in the supplement (see Table S1 below). In terms of soil properties, the lime concretion black soil is the main soil type in the Huaibei plain, which is shown in line 482. Therefore, we believe that soil properties, vegetation, and land cover/use are homogeneous across different *in situ* stations. In addition, two validation strategies were used in the study. The first is to compare the mean RZSM averaged over all in situ stations with the mean RZSM averaged over all grids. The second one is the point-grid validation, the station measurements are compared directly with the grid value where the station is located, if more than one station, the measurements of these stations are averaged. The point-grid validation has been provided in the supplement and draws the same conclusion as the station-averaged validation that GLDAS_CLSM outperforms other RZSM products. The section “3.3 Validation strategies” will be added in chapter 3 Methods.

Table S1 Overview of in situ stations in Huai River Basin

Station Name	Longitude (E)	Latitude (N)	Elevation (m)	Land cover
Taolaoba	117.16	32.18	48	Irrigated Crop
Chahua	116.02	33.03	39	Rainfed Crop
Hanting	116.32	33.02	28	Rainfed Crop
Songji	115.27	32.82	39	Rainfed Crop
Funan	115.57	32.64	33	Rainfed Crop
Santa	115.70	32.81	33	Rainfed Crop
Yaoli	116.17	31.82	58	Irrigated Crop
Guanting	116.85	31.80	51	Irrigated Crop
Zhuangmu	117.11	32.36	27	Irrigated Crop
Guiji	116.62	32.78	23	Irrigated Crop
Xiaji	116.54	32.65	25	Rainfed Crop
Shuangfu	115.57	33.34	37	Rainfed Crop
Fentai	115.73	33.45	35	Rainfed Crop
Santang	115.83	33.31	32	Rainfed Crop
Lixin	116.21	33.14	28	Rainfed Crop

Jieshou	115.36	33.27	42	Rainfed Crop
Yangqiao	115.39	33.02	28	Rainfed Crop
Guangwu	115.33	33.37	42	Rainfed Crop
Huangling	115.13	33.04	37	Rainfed Crop
Quanyang	115.44	33.11	35	Rainfed Crop
Kanheliu	115.85	33.10	33	Rainfed Crop
Kouziji	116.09	32.84	26	Rainfed Crop
Sanshilipu	116.11	32.70	27	Rainfed Crop
Xiaqiao	116.38	32.64	26	Rainfed Crop
Hengpaitou	116.36	31.59	72	Woodland
Xianghongdianxia	116.18	31.58	116	Woodland
Wangchenggang	116.53	31.74	76	Irrigated Crop
Lumiao	115.80	34.00	39	Rainfed Crop
Dasi	115.87	33.80	42	Rainfed Crop
Youhe	115.79	33.63	38	Rainfed Crop
Huagou	116.06	33.51	33	Rainfed Crop
Dahu	116.35	33.52	31	Rainfed Crop
Chenqiao	116.56	33.09	25	Rainfed Crop
Heliu	116.97	33.03	25	Rainfed Crop
Linhuanzha	116.57	33.67	29	Rainfed Crop
Guzhenzha	117.33	33.30	18	Rainfed Crop
Wudaogou	117.34	33.16	21	Rainfed Crop
Hexiangzha	117.18	33.00	18	Rainfed Crop
Tancheng	116.56	33.44	29	Rainfed Crop
Xibakou	117.87	33.15	11	Rainfed Crop
Xulouzha	116.75	33.92	30	Rainfed Crop
Suxianzha	117.08	33.67	28	Rainfed Crop
Gukouzha	116.45	34.27	39	Rainfed Crop
Kuaitanggou	117.55	33.75	20	Rainfed Crop
Yanglou	116.78	34.32	39	Rainfed Crop
Langanji	117.23	33.93	25	Rainfed Crop
Dulou	116.85	34.20	37	Rainfed Crop
Xiangyang	117.58	33.47	24	Rainfed Crop
Shuangdui	116.90	33.42	25	Rainfed Crop
Shuoli	116.90	34.03	32	Rainfed Crop
Huangmiao	117.65	33.08	19	Rainfed Crop
Baoji	117.11	33.16	22	Rainfed Crop
Dinghouying	117.34	33.46	24	Rainfed Crop
Xuanmiao	116.27	34.52	54	Rainfed Crop
Longhai	116.35	34.40	45	Rainfed Crop
Zhangzhuangzhai	116.60	34.12	37	Rainfed Crop
Sixian	117.92	33.43	16	Rainfed Crop
Dazhuang	117.87	33.67	20	Rainfed Crop

3. It is also not indicated how each satellite-based soil moisture (at multiple depths) and RZSM 'with different spatiotemporal resolutions were aggregated (again, spatially and temporally) to come up with the sets of time series that require consistent temporal scales between them. The method used for spatial aggregation of the gridded-RZSM also needs to be manifested (i.e., methods).

Response: 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)."

4. As this is site-specific, it sounds even less convincing that CLSM-derived soil moisture products outperform, and thus it gets more confusing what the authors want to argue from the RZSM products comparison.

Response: On the one hand, we want to evaluate the performance of eight RZSM products in the agricultural crop area, which could provide a more accurate RZSM dataset for agricultural drought monitoring. The results show that GLDAS_CLSM outperforms the other RZSM products. However, it doesn't mean CLSM-derived soil moisture outperforms, because SMAP L4 and MERRA-2 also use CLSM. More importantly, the focus of this study is to investigate the sources of error of the different RZSM products, which could provide some insights about how to improve the ability of land surface models to simulate the land surface states and fluxes.

5. There is significant inconsistency (due to the randomness in estimating RZSM from the remote sensing data) between RZSM estimation methods. For example, the authors tried to estimate RZSM using a depth-weighted method, but equation 1 used for in-situ RZSM is different from equation 2, which was used for RZSM estimation from satellite-derived modeled soil moisture.

Response: The in situ soil moisture measurements are available at four depths (10, 20, 40 and 100 cm). However, in addition to the GLDAS_CLSM, MERRA-2, SMAP L4 and SMOS L4, which directly provide the 0-100 cm RZSM, the other model-based soil moisture datasets are provided in different soil layers, i.e., NCEP CFSv2, CLDAS and GLDAS_NOAH ($\theta_{0-10\,cm}$, $\theta_{10-40\,cm}$, $\theta_{40-100\,cm}$), ERA5 ($\theta_{0-7\,cm}$, $\theta_{7-28\,cm}$, $\theta_{28-1\quad cm}$). The in situ measurements are for each soil depth, but the model-based RZSM products are for each soil layer. They are not consistent. Therefore, the study uses two different equations to calculate the RZSM. The in situ RZSM is calculated using a depth-weighted mean of the measurements at four soil depths (10, 20, 40 and 100 cm). This method (equation 1) has been used in the study by Gao et al., (2017) and Xing et al., (2021). The model-based RZSM is calculated with a weighted average of the 0-100 cm RZSM. This method (equation 2) has been used in the study by González-Zamora et al., (2016) and Xing et al., (2021) and calculation of SMOS L4 RZSM (Al bitar et al., 2021).

-Specific comments:

line 39-40: is this sentence needed?

Response: We will delete this sentence in the revised manuscript.

line 42: duplicate definition of RZSM?

Response: We will delete this sentence in the revised manuscript.

line 99-100: by this sentence, do you intend not to include any process-based explanation for the soil moisture products? What about attempting to explain the performance differences found among the RZSM products (as this is essentially modeled data) in relation to model structure? Why does CLSM outperform other land models in terms of RZSM products?

Response: This study attempts to investigate the error sources of RZSM products without considering the model structure. While only evaluating the atmospheric forcing, soil texture, and local conditions, we analyze the effects of these error sources on RZSM estimation from the perspective of physical processes. For example, overestimated precipitation tends to lead to overestimated water-related states (soil moisture) or fluxes (runoff). The clay exhibits stronger water retention capacity compared to sand at the same matric potential, and high soil organic carbon leads to high soil porosity. Therefore, the overestimated clay fraction and soil organic carbon lead to higher water stored in the soil. In addition, different land surface models are used to produce different RZSM products. For example, ERA5 (HTESSEL), MERRA-2 (CLSM), NCEP CFSv2 (Noah), GLDAS_NOAH (Noah), GLDAS_CLSM (CLSM), CLDAS (CLM, CoLM, Noah-MP), SMAP L4 (CLSM), SMOS L4 (exponential filter, not land surface model). Even the same CLSM land surface model is used for both MERRA-2 and SMAP L4, but the model version is a bit different. Therefore, it is difficult to directly quantify the effect of model structure on RZSM. In this study, the GLDAS_CLSM RZSM product outperforms other model-based RZSM products, but this doesn't mean that CLSM outperforms other land surface models. The accuracy of RZSM depends on the meteorological forcing, the structure of the land surface model, and the parameterization scheme.

Chapter 2.4: the information on the spatial and temporal resolution of each data needs to be revisited and clearly indicated.

Response: The information on the spatial and temporal resolution of eight RZSM products is shown in Table 1.

Chapter 3.2: why did you estimate satellite-derived RZSM different from in-situ RZSM? Why equation 1 and 2 are different? How convincing are the RZSM comparisons based on equation 1 and 2?

Response: The in situ soil moisture measurements are available at four depths (10, 20, 40 and 100 cm). However, in addition to the GLDAS_CLSM, MERRA-2, SMAP L4 and SMOS L4, which directly provide the 0-100 cm RZSM, the other model-based soil moisture datasets are provided in different soil layers, i.e., NCEP CFSv2, CLDAS and GLDAS_NOAH ($\theta_{0-10\ cm}$, $\theta_{10-40\ cm}$, $\theta_{40-10\ \ cm}$), ERA5 ($\theta_{0-7\ cm}$, $\theta_{7-28\ cm}$, $\theta_{28-100\ cm}$). The in situ measurements are for each soil depth, but the model-based RZSM products are for each soil layer. They are not consistent. Therefore, the study uses two different equations to calculate the RZSM. The in situ RZSM is calculated using a depth-weighted mean of the measurements at four soil depths (10, 20, 40 and 100 cm). This method (equation 1) has been used in the study by Gao et al., (2017) and Xing et al., (2021). The model-based RZSM is calculated with a weighted average of the 0-100 cm RZSM. This method (equation 2) has been used in the study by González-Zamora et al., (2016) and Xing et al., (2021) and calculation of SMOS L4 RZSM (Al bitar et al., 2021).

line 293: instead of averaging all in-situ stations, can you think of disaggregating the study basin (and stations) using any available information such as surface soil properties, orography (e.g., slope, and elevation), land cover, and/or vegetation? That will help the readers get more generalizable information and references.

Response: It is a very good and useful suggestion. However, this study pays more attention to soil moisture measured at 58 stations rather than a specific station. The underlying surface conditions (e.g. surface soil properties, orography, land cover and vegetation) is considered as homogeneous. 56 of 58 in situ stations are located in crop lands of the Huaibei Plain, and the elevation is quite similar (Table S1). The lime concretion black soil is the main soil types in the Huaibei plain. In future study, we will attempt to investigate the effect of different underlying surface conditions (vegetation types, etc.) on soil moisture estimations for specific station.

line 306-309: This needs to be rephrased. It is hard to understand what is meant.

Response: The text (line 306-309) will be rephrased from “Figure 3 shows time series of in situ RZSM observations averaged over all in situ stations with its spatial variability, and of 3 RZSM products, ERA5, SMOS L4, and GLDAS_CLSM, presenting a marked overestimation, a marked underestimation, and the best overall agreement with in situ observations, respectively. Other products can be seen in Fig. S1.”

To “Figure 3 shows the time series of observed and model-based RZSM averaged over all in situ stations and the grids where the *in situ* stations are located. ERA5, SMOS L4, and GLDAS_CLSM show overestimation, underestimation, and the best overall agreement with in situ observations, respectively. Other products are shown in Fig. S1”.

line 311: can you explain why SMOS L4 showed less rapid changes and smoother trends?

Response: It is well known that the SSM shows a faster response to atmospheric variations than RZSM, especially for precipitation. Therefore, RZSM shows less rapid changes and smoother trends than SSM, which shows a strong variability. On the one hand, SMOS L4 RZSM is estimated from SMOS L3 SSM together with a modified exponential filter with different parameter T (characteristic time length) proposed by Wagner et al., (1999). The exponential filter can smooth the trend of SSM, the higher the T value, the smoother the RZSM trend. On the other hand, precipitation with high spatial and temporal variability is the main forcing input of other model-based RZSM, which show a strong response to precipitation. Precipitation is not used in producing SMOS L4 RZSM. Therefore, SMOS L4 shows less rapid changes and smoother trends.

line 321: can you explain why they did a better job in the wet season compared to the dry season?

Response: In the Huai river basin, more than 60 % of the annual precipitation falls between June and September (wet season), which significantly increases the RZSM. According to Figure 1 and S1, it is obvious that RZSM shows a strong response to precipitation events. In general, the model-based RZSM datasets increase with the increasing in situ observations after a precipitation event. The model-based RZSM datasets show strong variability and a good agreement with observations in wet season. However, the in situ RZSM shows stronger variability than the model-based RZSM datasets in dry season, the model-based RZSM datasets show little variability and don’t capture the temporal trend of in situ observations. It indicates that the land surface models are sensitive to precipitation events than no precipitation events and show better skill in simulating RZSM when there is a precipitation event. This could explain the better performance of model-based RZSM datasets in the wet season than that in the dry season.

line 360: can you explain why individual satellite-based RZSM products showed different probabilistic distributions? Some are log-normal and the others are normal. Can you add more explanation on this matter?

Response: The peak of the relative frequency for model-based RZSM products ranges from 0.3 to 0.6. RZSM products with log-normal distribution show that low values dominate the RZSM time series, which could be caused by low precipitation. The precipitation field derived from the atmospheric general circulation model (AGCM) generally has too many drizzle events (<1 mm day^{-1}). The modeled RZSM is affected by meteorological forcing, model structure and parameterization, etc. The RZSM estimates are subject to random error and systematic bias, and it is difficult to directly quantify which factor affects the probability distributions of different RZSM products. In addition, the probability distribution of RZSM may depend on the research periods, the probability distribution of RZSM in wet season may be different from that in dry season.

line 375: how does this ground-based observation of precipitation (840 mm/year) represent the average precipitation of the basin area? You also compared gridded-precipitation with this in-situ precipitation observation (line 430). Can you clarify how solid the comparison of this in-situ precipitation with gridded precipitation is?

Response: In this study, the ground-based precipitation observation doesn't represent the average precipitation of the watershed area. We only compare the ground-based precipitation observation with the modeled precipitation of the grid where the station is located from the grid perspective.

line 432: do you think MERRA-2 and GLDAS-CLSM would outperform other satellite-derived RZSM in other basins (or area) as well? What if you perform a continental-scale study, will you still think there will be a certain winner? If not, how can you limit the scale of this sort of comparison study to be meaningful and convincing?

Response: In this study, MERRA-2 and GLDAS-CLSM outperform other model-based RZSM products in the Huaibei Plain, where cropland dominates. It is uncertain whether MERRA-2 and GLDAS-CLSM would still outperform other products if this study were conducted in other basins (areas) or on a continental scale. Because the precipitation data derived from the atmospheric general circulation model perform differently in different regions. For example, these large-scale atmospheric processes over the extra-tropics are better resolved in the AGCM than convective processes over the tropics. It is a study for specific underlying surface conditions (agricultural crop region). On the one hand, the evaluation of eight RZSM products in the agricultural crop region could provide a more accurate RZSM dataset for agricultural drought monitoring. More importantly, the focus of this study is to investigate the sources of error of the different RZSM products, which could provide some insights to improve the ability of land surface models to simulate the land surface states and fluxes.

line 436-438: the sentences need to be re-structured to clarify the argument.

Response: The text (line 436-438) will be rephrased from "The MERRA-2 model background precipitation corrected with NOAA CPCU gauge-based precipitation observations was implemented in the coupled land-atmosphere reanalysis system, which may also contribute to the high consistency with the ground-based precipitation"

To "Before driving the land surface water budget, the MERRA-2 model background precipitation was corrected with NOAA CPCU gauge-based precipitation in the coupled land-atmosphere reanalysis system, resulting in more accurate precipitation fields for MERRA-2".

line 453: in-situ RZSM observation does not capture irrigation effect? Can you explain how the irrigation water supply does not impact the soil moisture content?

Response: We didn't express it clearly. The original meaning of this sentence is that the in situ station does not capture the irrigation signal. Because the in situ stations are usually installed away from the cropland to avoid the effect of anthropogenic irrigation on the original soil water content supplied by precipitation. In addition, reviewer2 and reviewer3 also raise question about question. The overall comments from three reviewers indicate that the irrigation is not an issue in this paper, and should not be emphasized. We will delete relevant statements about irrigation.

line 485-489: can you add more information on how the soil properties could end up in certain ranges of soil moisture values?

Response: We have illustrated the effect of soil properties on ranges of soil moisture values (line 457-462). "In general, soil texture is closely related to the ability of the soil to retain water, as water molecules adhere more tightly to fine-textured clay particles than coarse-textured sand particles. Consequently, clay exhibits stronger water retention capacity and higher water content stored in the soil compared to sand at the same matric potential. In addition, the overestimated FAO/UNESCO soil organic carbon content (Fig. 9) leads to higher soil porosity and lower bulk density. As a result, water can infiltrate more quickly and more water can flow through the soil and can be retained in the soil".

Reference:

Gao, X., Zhao, X., Brocca, L., Huo, G., Lv, T., Wu, P., 2017. Depth scaling of soil moisture content from surface to profile: multistation testing of observation operators. Hydrol. Earth Syst. Sci. 1–25, https://doi.org/10.5194/hess-2017-292.

Xing, Z., Fan, L., Zhao, L., et al. A first assessment of satellite and reanalysis estimates of surface and root-zone soil moisture over the permafrost region of Qinghai-Tibet Plateau, Remote Sens. Environ., 265, 112666, https://doi.org/10.1016/j.rse.2021.112666, 2021.

González-Zamora, Á., Sánchez, N., Martínez-Fernández, J., Wagner, W., 2016. Root-zone plant available water estimation using the SMOS-derived soil water index. Adv. Water Resour. 96, 339–353. https://doi.org/10.1016/j.advwatres.2016.08.001.

Al Bitar, A., Mahmoodi, A., Kerr, Y., Rodriguez-Fernandez, N., Parrens, M. and Tarot, S.: Global Assessment of Droughts in the Last Decade from SMOS Root Zone Soil Moisture, 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 8628-8631, https://doi.org/10.1109/igarss47720.2021.9554773, 2021.

Wagner, W.; Lemoine, G.; Rott, H. A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data. Remote Sens. Environ. 1999, 70, 191–207, https://doi.org/10.1016/s0034-4257(99)00036-x.