Response to Editor Comments on HESS submission

Evaluation of reanalysis soil moisture products using Cosmic Ray Neutron Sensor observations across the globe

Yanchen Zheng[1,2], Gemma Coxon[1], Ross Woods[2], Daniel Power[2], Miguel Angel Rico-Ramirez[2], David McJannet[3], Rafael Rosolem[2,4], Jianzhu Li[5] and Ping Feng[5]

Corresponding to: Yanchen Zheng (yanchen.zheng@bristol.ac.uk)

General Response:

We thank the editor and reviewers for taking the time to review the paper and provide the valuable comments, which have helped to improve this manuscript and the clarity of the research.

In response to the reviewers comments we have made two key changes to the paper.

Firstly, in response to Reviewer 1, we have added a new section '2.1.1 CRNS sensor calibration' to describe the calibration steps in greater detail. In the revised supporting information of this paper, the calibrated N0 values, the number of calibration days used to derive N0 and lattice water values for all CRNS sites used in this study are provided according to the reviewer comments. The detailed response to this comment can be found in the 'Response to Reviewer1.docx' file.

Secondly, in response to Reviewer 2, we have now included SMAP L4 in our analysis. We have calculated the metrics for SMAP L4, reanalysed the results, updated all the figures and revised the manuscript carefully for the paper and also the supporting information. Our results show that SMAP L4 show better performance in arid regions and in cropland. Also, SMAP L4 is the optimal option for Australia. However, nearly 70% of the CRNS sites were established before 2015, which means the evaluation period for SMAP L4 is shorter than for the other reanalysis products. Thus, a fair comparison between SMAP L4 and other products might be less robust. We have added this to the limitations of the paper. The detailed response addressing this can also be found in the 'Response to Reviewer2.docx' file.

Response to Reviewer #1 Comments on HESS submission

Evaluation of reanalysis soil moisture products using Cosmic Ray Neutron Sensor observations across the globe

Yanchen Zheng[1,2], Gemma Coxon[1], Ross Woods[2], Daniel Power[2], Miguel Angel Rico-Ramirez[2], David McJannet[3], Rafael Rosolem[2,4], Jianzhu Li[5] and Ping Feng[5]

Corresponding to: Yanchen Zheng (yanchen.zheng@bristol.ac.uk)

Detailed responses to all comments are provided as follows. Author responses are in **bold** and any modifications to the manuscript are in *italic* and highlighted in blue colour below each of the comments. A tracked changes version of the paper is also uploaded as the attachment.

Reviewer #1 comments:

The authors present a noteworthy study that compares CRNS stations across multiple continents with remote sensing/reanalysis soil moisture products. Utilizing the CRSPY tool, the authors reprocess the raw CRNS data, addressing gaps in information, such as the atmospheric moisture correction (absent in the COSMOS USA network), and harmonize the dataset. Undoubtedly, this tool holds promise for advancing the CRNS method and enhancing the global CRNS community. The article is well-crafted, with comprehensive data documentation in both the appendix and online. However, I have some suggestions and comments that should be addressed before publication.

Response: We are grateful for your valuable and positive feedback on our paper. Your insightful comments are highly appreciated, and we have carefully revised our paper according to the points you have mentioned.

Major Suggestion:

1) I find some ambiguity regarding the reprocessing of CRNS data with CRSPY, particularly in the computation of the N0 parameter. The main article and supplemental material lack any mention of N0. To calculate N0, each site typically requires at least one gravimetric calibration survey, composed of 12-18 soil profiles within the footprint, sampled every 5 cm down to 30 cm. The gravimetric survey data is then weighted (following Kohli 2015/2017, etc.), and the Desilets 2010 function is inverted to determine N0. Most CRNS sites have gravimetric calibration data for one or more sample periods. Please include a detailed description of the process for estimating N0 at each site, specifying the gravimetric calibration dataset used (single or average of multiple calibrations, etc.). The accuracy of the N0 estimate is crucial, as incomplete or non-representative data may introduce significant bias in the comparison with reanalysis products. Best practices for

N0 involve 2-3 calibration dates (Iwema 2015), though this is labor-intensive and not consistently implemented.

Thanks for this comment. In this revision, we have added a new section '2.1.1 CRNS sensor calibration' to describe the calibration steps in greater detail. The sensor calibration process and the computation of N0 parameter for each CRNS network are summarized in the papers (Hawdon et al., 2014; Bogena et al., 2022; Cooper et al., 2021; Power et al., 2021; Zreda et al., 2012) which have been added for reference. Calibration of the COSMOS-EUROPE and COSMOS-UK data was undertaken by the respective networks with both networks using the most current weighting procedure (Cooper et al., 2021, Bogena et al., 2022). The CosmOz and COSMOS USA networks were recalibrated through crspy to ensure the method matches that of the other networks, which is possible thanks to the open data availability of both networks. Note that the CosmOz network recently updated their procedures to utilise the revised Schrön et al., (2017) method, however as they are using an alternative method for incoming neutron intensity correction, we continue to use the reprocessed data from crspy (https://cosmoz.csiro.au/about, last accessed 12/01/2024). The calibrated N0 values are given in the supporting information, along with the number of calibration days used to derive N0 (see 'CRNSsiteDataNEWR1.xlsx' file).

It's important to note that calibration values (N0) are influenced by the specific calibration methodologies employed by different networks, which can affect their direct comparability. Notably, COSMOS-UK calibrations utilize the water vapor content present on the calibration day as a reference, in contrast to other networks that use a dry atmosphere reference for atmospheric water vapor correction. Additionally, the approach to biomass correction varies. Crspy and COSMOS-EUROPE each treat biomass as a static value given the lack of dynamic data availability. In the case of crspy a biomass value is obtained from ESA CCI biomass data (given in kg/m^2), whereas COSMOS-EUROPE presumes a biomass value of 0. In each case the biomass correction is applied statically and so does not impact calculated soil moisture values. It is important to note however that whilst these described methodological differences do not impact the calculated soil moisture values, caution should be given when directly comparing corrected neutron counts and N0 values between the different networks. We have added this as a short note in the 'CRNSsiteDataNEWR1.xlsx' file where we provide the N0 data.

Modifications to the paper are as follows:

*"**2.1.1 CRNS sensor calibration***

The inverse relationship between fast neutrons and hydrogen atoms means that as neutron counts rise (fall), we infer that the moisture content of the soil is decreasing (increasing). However, in order to convert this signal into volumetric soil moisture values, calibration of each sensor is required. This involves obtaining multiple samples of soil moisture profiles within the sensor footprint that are together combined to provide an average moisture content (Zreda et al., 2012). Each sample is subjected to oven-drying, providing us with gravimetric soil moisture values, which can be converted to volumetric soil moisture when multiplied by the dry soil bulk density of the soil sample. As our *understanding of the sensor signal has grown, improvements to this calibration step have been developed which have been shown to provide more accurate results. In particular revised weighting schemes have been derived that consider the increased sensitivity of the signal to soil moisture nearer the sensor (Köhli et al., 2015; Schrön et al., 2017), as well as research showing the benefit in conducting multiple calibration campaigns across different seasons (Iwema et al., 2015). Ultimately this calibration step will provide us with the so called N0 number (i.e., the theoretical neutron count found in absolutely dry conditions), which is calculated by comparing the averaged field scale soil moisture value derived through the sampling campaign with the count rate at the time of sampling. This N0 number is used to derive the ratio between the actual counting rate (N) and the theorised maximum counting rate (N0) in the Desilets et al (2010) equation for converting neutrons to soil moisture values. It's important to note therefore, that changes in this number, or differences in how this number is derived, can lead to biases in soil moisture values.*

The COSMOS-UK and COSMOS-EUROPE datasets each use the aforementioned revised weighting schemes to calibrate the sensors (Bogena et al., 2022; Cooper et al., 2021). To ensure comparability the COSMOS-USA and CosmOz sites were updated to utilise the revised scheme using crspy (Power et al., 2021), which was possible thanks to the openly available calibration data provided by each of the networks (Zreda et al., 2012, Hawdon et al., 2014). When multiple calibration days where available calibration would be taken on more than one day, with the N0 number being the value that reduced the error across all calibration days. It should be noted that recently the CosmOz network updated their data to utilise the revised weighting scheme, however there is still a difference in incoming neutron intensity correction, necessitating harmonization through crspy (https://cosmoz.csiro.au/about, last accessed 12/01/2024). More detailed of CRNS data reprocessing can be found in section 2.1.2. The calibrated N0 values, along with information on how many calibration days where used are given in the supporting information (see 'CRNSsiteDataNEWR1.xlsx' file)."

2) I am a bit confused by the description of soil organic carbon. Does this include both the organic carbon and mineral lattice water values? The variation in lattice water was found to be important for many of the original CRNS networks (Zreda 2012 COSMOS, Hawdon 2014 COSMOZ, etc.). The paper and appendix does not include the description of lattice water and should be clarified or added to the metadata.

Thanks for pointing this out. The soil organic carbon in our paper does not include the lattice water values. In this revision, the lattice water values are also provided in the appendix. We added clarifications of the concept for 'soil organic carbon' and 'lattice water' in the paper. The lattice water values for each CRNS site are provided in the supporting information ('CRNSsiteDataNEWR1.xlsx' file).

"Soil properties data, i.e., bulk density, soil organic carbon content and lattice water, are provided in metadata from Power et al. (2021) and Bogena et al. (2022). Soil organic carbon represents the total organic carbon in the soil at the site, while lattice water represents the hydrogen contained in the mineral structures of the soil. In studies from Power et al. (2021) and Bogena et al. (2022), local measurements of soil properties data

are collected for the majority of CRNS sites (bulk density: 98% sites; soil organic carbon content: 94% sites, lattice water: 98% sites), while the global raster-based SoilGrids soil dataset (Hengl et al., 2017) was used to provide data for the sites with missing measurements."

3) Comment: The influence of rapidly growing crops on the CRNS observations remains a challenge. There have been several attempts to provide correction factors (either on N0 itself or on the moderated counts) but nothing definitive has been adopted by the community. The authors mention this in the limitations. I hope the community comes to a consensus soon about how best to deal with CRNS data in croplands. The influence of forest biomass seems to be an even more challenging problem but equally important.

   Thanks for agreeing to this point. We also share the hope that the research community will collaboratively work towards developing more effective solutions to address and mitigate the impact of vegetation on CRNS observations in future studies.

Minor comment:

4) Figure 2. Label the 4 subplots a-d and list what geographical region they are from. This was not clear from the description.

   Modified. We have added the region names and also (a)/(b)/(c)/(d) to this figure, as shown below.

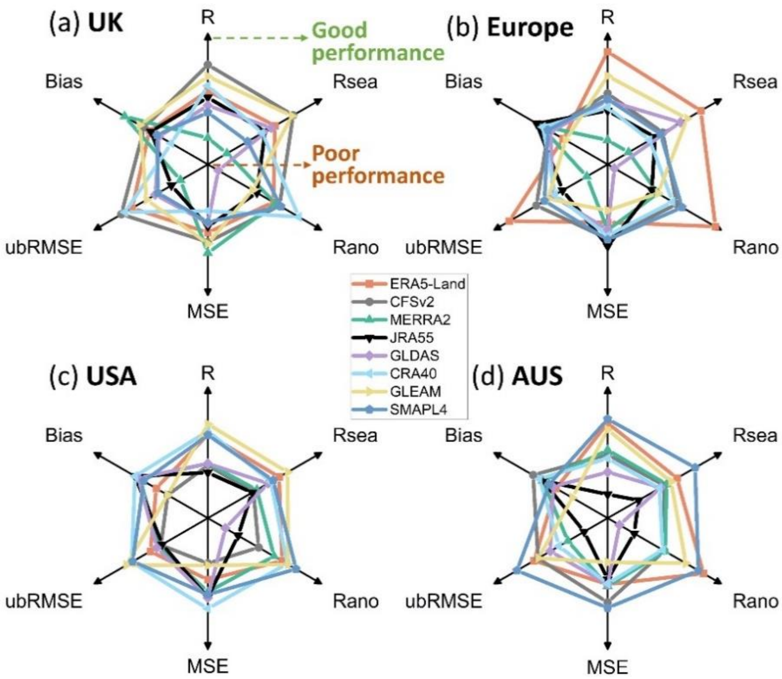

Figure 2: Brunke ranking results for a total of 8 products performance in terms of 6 statistical metrics across different regions, i.e., (a) UK, (b) mainland Europe, (c) USA and (d) Australia (AUS).

References:

Bogena, H. R., Schrön, M., Jakobi, J., Ney, P., Zacharias, S., Andreasen, M., Baatz, R., Boorman, D., Duygu, M. B., and Eguibar-Galán, M. A.: COSMOS-Europe: a European network of cosmic-ray neutron soil moisture sensors, Earth Syst. Sci. Data, 14, 1125-1151, https://doi.org/10.5194/essd-14-1125-2022, 2022.

Cooper, H. M., Bennett, E., Blake, J., Blyth, E., Boorman, D., Cooper, E., Evans, J., Fry, M., Jenkins, A., and Morrison, R.: COSMOS-UK: national soil moisture and hydrometeorology data for environmental science research, Earth Syst. Sci. Data, 13, 1737-1757, https://doi.org/10.5194/essd-13-1737-2021, 2021.

Hawdon, A., McJannet, D., and Wallace, J.: Calibration and correction procedures for cosmic-ray neutron soil moisture probes located across Australia, Water Resour. Res., 50, 5029-5043, https://doi.org/10.1002/2013WR015138, 2014.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., and Bauer-Marschallinger, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS one, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Köhli, M., Schrön, M., Zreda, M., Schmidt, U., Dietrich, P., and Zacharias, S.: Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons, Water Resour. Res., 51, 5772-5790, https://doi.org/10.1002/2015WR017169, 2015.

Power, D., Rico-Ramirez, M. A., Desilets, S., Desilets, D., and Rosolem, R.: Cosmic-Ray neutron Sensor PYthon tool (crspy 1.2.1): an open-source tool for the processing of cosmic-ray neutron and soil moisture data, Geosci. Model Dev., 14, 7287-7307, https://doi.org/10.5194/gmd-14-7287-2021, 2021.

Schrön, M., Köhli, M., Scheiffele, L., Iwema, J., Bogena, H. R., Lv, L., Martini, E., Baroni, G., Rosolem, R., Weimar, J., Mai, J., Cuntz, M., Rebmann, C., Oswald, S. E., Dietrich, P., Schmidt, U., and Zacharias, S.: Improving calibration and validation of cosmic-ray neutron sensors in the light of spatial sensitivity, Hydrol. Earth Syst. Sci., 21, 5009-5030, https://doi.org/10.5194/hess-21-5009-2017, 2017.

Zreda, M., Shuttleworth, W., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R.: COSMOS: The cosmic-ray soil moisture observing system, Hydrol. Earth Syst. Sci., 16, 4079-4099, https://doi.org/10.5194/hess-16-4079-2012, 2012.

Response to Reviewer #2 Comments on HESS submission

Evaluation of reanalysis soil moisture products using Cosmic Ray Neutron Sensor observations across the globe

Yanchen Zheng[1,2], Gemma Coxon[1], Ross Woods[2], Daniel Power[2], Miguel Angel Rico-Ramirez[2], David McJannet[3], Rafael Rosolem[2,4], Jianzhu Li[5] and Ping Feng[5]

Corresponding to: Yanchen Zheng (yanchen.zheng@bristol.ac.uk)

Detailed responses to all comments are provided as follows. Author's general responses are in red and point-to-point responses are in **bold**. Any modifications to the manuscript are in *italic* and highlighted in yellow colour below each of the comments. A tracked changes version of the paper is also uploaded as the attachment.

Reviewer #2 comments:

The paper analyses the performance of global reanalysis soil moisture products against soil moisture obtained from cosmic ray neutron sensors. The results are analysed using different performance criteria, region specific and climate specific. The paper addresses highly relevant scientific questions, is exceptionally well structured, provides novel insights thereby presenting novel concepts and setting the global data sets into new context. The conclusions are reached, to the point and substantial.

General response:

We express our sincere appreciation for the valuable and constructive feedback offered by Reviewer 2. We acknowledge the significance of addressing these comments, as we think that incorporating these suggestions will enhance the overall quality of our work.

However, there are two main improvements to be made before acceptance.

First: I suggest the authors consider to include SMAP Level 4 data as reanalysis product for evaluation. SMAP Level 4 comes as data assimilation product and is considered state-of-the-art soil moisture for the recent years. Without SMAP Level 4, the study seems incomplete.

Thanks for this suggestion. We agree that the SMAP L4 is considered as state-of-the-art soil moisture reanalysis product for the recent years. Thus, in this revision, we have included SMAP L4 in our analysis. We have calculated the metrics for SMAP L4, reanalysed the results, updated all the figures and revised the manuscript carefully for the paper and also the supporting information.

Our results show that SMAP L4 performs better in arid regions, especially for the metrics R_{ano} and MSE. SMAP L4 is the optimal option for Australia, since it exhibits good temporal correlation with both original, seasonal measured soil moisture timeseries and also performs well in terms of *MSE*, *ubRMSE* . In addition, SMAP L4 also shows good performance in cropland land cover type, particularly in terms of R_{ano}. We have highlighted these findings in the abstract, results, discussion and conclusions of the paper. The revised abstract is presented as follows:

"Overall, ERA5-Land, CRA40, CFSv2, SMAP L4 and GLEAM exhibit superior performance compared to MERRA2, GLDAS-Noah and JRA55. We recommend ERA5-Land and CFSv2 could be used in humid climates, whereas SMAP L4 and CRA40 perform better in arid regions. SMAP L4 has good performance for cropland, while GLEAM is more effective in shrubland regions. Our findings also provide insights on directions for improvement of soil moisture products for product developers."

However, one key point of difference is that the SMAP L4 data begins in 2015 and therefore contains a shorter time period compared to the rest of the seven reanalysis products analysed in this paper. There is a total of 93 CRNS sites that were established before 2015, accounting for nearly 70% of the total sites. Since SMAP L4 starts to provide data from March 2015, the evaluation period for SMAP L4 is shorter than that for other reanalysis products, with 10 CRNS sites exhibiting temporal overlaps of less than 2 years. This indicates that the fair comparison between SMAP L4 and other products might be less robust. We have added this to the limitations of the paper.

"Regarding temporal coverage, our assessment of SMAP L4 product is limited by the period of record which begins in 2015. Since 70% of our CRNS sites were established before 2015, the evaluation period for SMAP L4 is shorter than that for other reanalysis products, with 10 CRNS sites exhibiting temporal overlaps of less than 2 years. SMAP L4 is included in this analysis because it is a state-of-the-art data assimilation soil moisture reanalysis product in recent years, yet it should be noted that it is evaluated over a shorter time period compared to the other reanalysis products."

Second: the terminology in Figure 10 as "highly recommended", "recommended" and "not recommended" must be adapted and if needed must be adapted throughout the text. All reanalysis products show high bias, and hence, the ERA5-Land cannot be highly recommended for humid sites as the authors have done in Figure 10. It should rather be recommended with care. I provide a suggestion below.

We agree that the current terminology in Figure 10 is not suitable. In our revision, we adapted these into "Higher performance", "Ok" and "Lower performance". We have applied these changes consistently throughout Figure 10, entire paper and the supporting information file.

These rankings are calculated based on the Brunke ranking method and also updated with adding the new SMAP L4 product. In this revision, "Ok" represents the top two optimal options among all evaluated reanalysis products in this study based on all 6 statistical metrics. The soil moisture reanalysis products ranked last are labelled with "Lower performance". In particular, the soil moisture products of Brunke ranking scores less than 3.2 (i.e., the product with this threshold score indicates its ranking is within or around the top 3 in terms of all 6 metrics across most of the sites in this category) are labelled with "Higher performance".

Although all reanalysis products show relatively high bias, for humid sites, ERA5-Land is the one exhibiting the optimal performance among all evaluated reanalysis products according to all 6 statistical metrics not only just considering bias. Also because we changed the terminology, so we believe it is appropriate to give ERA5-Land "Higher performance" for humid sites in the revised Figure 10.

The revised Figure 10 is shown as follows.

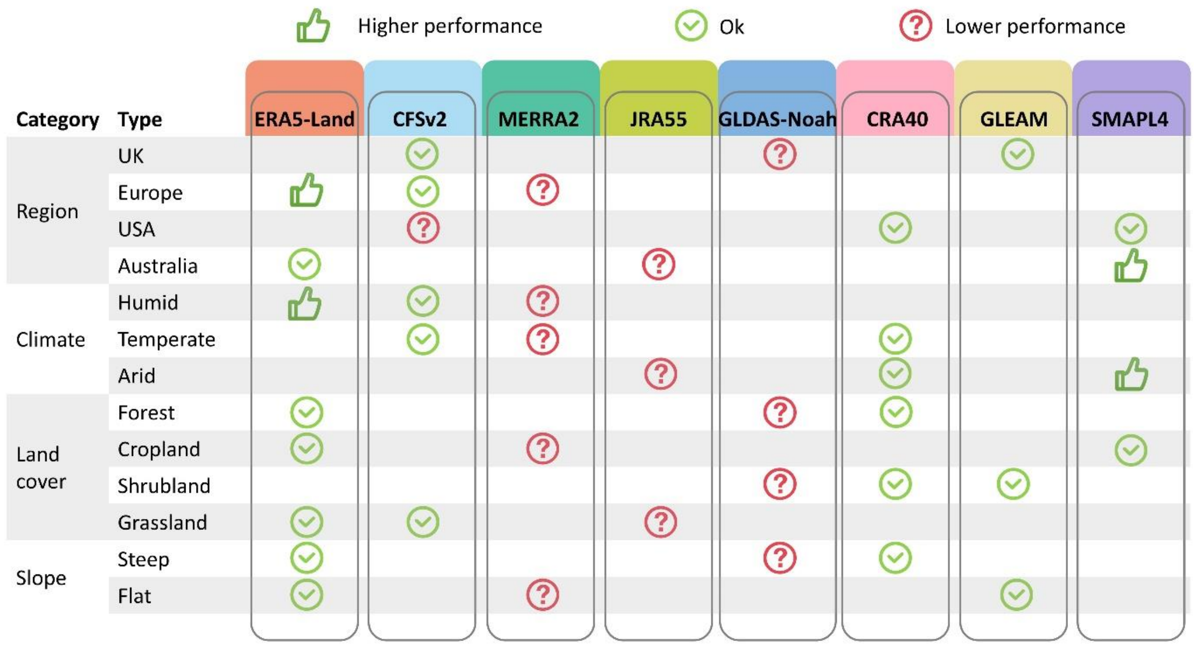

Figure 10: Recommendations for choosing 8 reanalysis soil moisture products under various regions, climate, land cover and topographic slope conditions based on the average Brunke ranking scores.

Minor comments:

1) Line 11: It is recommended to shorten the abstract. Lines 11-19 can be reduced to 3 lines. This will focus the reader on the methods, novelty, and key findings. These are mentioned in the abstract.

   Revised. We have shorten the lines 11-19 of abstract into 3 lines:

   "Reanalysis soil moisture products are valuable for diverse applications but their quality assessment is limited due to scale discrepancies when compared to traditional in-situ point-scale measurements. The emergence of Cosmic Ray Neutron Sensors (CRNS) with field-scale soil moisture estimates (~250m radius, up to 0.7m deep) is more suitable for the product evaluation owing to its larger footprint."

2) Line 21: Although UK has its own network, it reads awkwards to read "sites from UK, Europe, USA and Australia" as a list because the UK is included in Europe and should not be listed beside Europe as separate geogrphic entity. Please not that this list refers to CRNS networks COSMOS-UK, COSMOS-Europe...

Modified. We have revised the sentence to:

"In this study, we perform a comprehensive evaluation of eight widely-used reanalysis soil moisture products (ERA5-Land, CFSv2, MERRA2, JRA55, GLDAS-Noah, CRA40, GLEAM and SMAP L4 datasets) against 135 CRNS sites from the COSMOS-UK, COSMOS-Europe, COSMOS USA and CosmOz Australia networks."

3) Line 42: Please also list this newer review on reanalysis, were a definition of reanalysis products is given https://doi.org/10.1029/2020RG000715

Modified. We have added this new reference for the definition of reanalysis products.

"Reanalysis products provide soil moisture data over long time periods (Li et al., 2005; Baatz et al., 2021) and typically merge soil moisture observations and land surface model output by adopting data assimilation techniques, which often results in better soil moisture estimation than satellite products (Naz et al., 2020; Beck et al., 2021; Mahto and Mishra, 2019)."

4) Line 50: This is well formulated. However, SMAP Level 4 is missing in this list https://doi.org/10.1029/2019MS001729

Revised. We have added SMAP Level 4 in this list of reanalysis products mentioned in the introduction.

"Currently, many reanalysis products exist including ERA5-Land (Muñoz-Sabater 2019, Muñoz-Sabater et al. 2021), CFSv2 (Saha et al. 2011, Saha et al. 2014), MERRA2 (GMAO 2015, Gelaro et al. 2017), JRA55 (JMA 2013, Kobayashi et al. 2015), GLDAS-Noah (Rodell et al. 2004, Beaudoing 2020), CRA40 (Liu et al. 2017, Li et al. 2021), GLEAM (Miralles et al. 2011, Martens et al. 2017) datasets and SMAP Level 4 datasets Reichle et al. (2019) etc (one should note that technically speaking GLDAS-Noah and GLEAM datasets are global land model-based products, we termed them as 'reanalysis products' in this paper for consistency)."

5) Line 115: rephrase to "harmonized processing of CRNS datasets".
Corrected. We have rephrased this sentence to:
"The notable deviation in two networks indeed highlights the importance of harmonized processing of CRNS datasets."

6) Line 132: Where is SMAP Level 4? Please give strong reason for not using the most recent Soil Moisture Product https://doi.org/10.1029/2019MS001729 SMAP Level 4 is a reanalysis product as it includes meteorological variables, and satellite observations that are used to update soil moisture in a data assimilation framework.

   See the detailed reply in the general response. In this revision, we have added the SMAP L4 product into our analysis.

7) In fact, I strongly suggest to add SMAP Level 4. As such, the study is out-dated. Adding SMAP Level 4 should be a short exercise - multiplying the scientific impact of this study by a factor. Hopefully SMAP Level 4 performs best amongst all Reanalysis products. Although it is not 20 years length, it is the SM reanalysis product that will be used for recent years rather than any of the other products analyzed. The technical definition of 20+ years (line 42) by a reference of 2005 is not sufficient to not use the most up-to-date SM reanalysis product in this highly relevant global study.

   Modified. We have removed "20+ years" in the definition of reanalysis products in the introduction and added the newer references. See the detailed reply in the general response. We have added the SMAP L4 product into our analysis in this revision.

8) Line 164: This depends on the source of uncertainty. Daily averaging causes loss of signal and proper filtering maintains signal while reducing uncertainty/noise. High measurement uncertainty can be compensated for by applying temporal filtering methods or simple daily averaging e.g. https://doi.org/10.3390/s22239143 .

   Revised. We have rephrased this sentence and also added this reference:

   "Due to the nature of the CRNS technology, the hourly measurements might contain higher uncertainty compared to daily measurements (Zreda et al., 2008; Desilets et al., 2010; Iwema et al., 2021), in general, high measurement uncertainty can be compensated by applying simple daily averaging (Davies et al., 2022)."

9) Line 174: Rephrase towards more active voice: "Spatial scale matching"

   Modified. We have rephrased this subtitle to:

   "3.1.2 Spatial scale matching"

10) Line 184: see comment before.

   Modified. We have rephrased this subtitle and also corrected the corresponding words in the supplementary information.

   "3.1.3 Vertical footprint matching"

11) Figure 2: Add region names and a/b/c/d to the figure (or names of the networks).

Revised. We have added the region names and also (a)/(b)/(c)/(d) to this figure, as shown below.

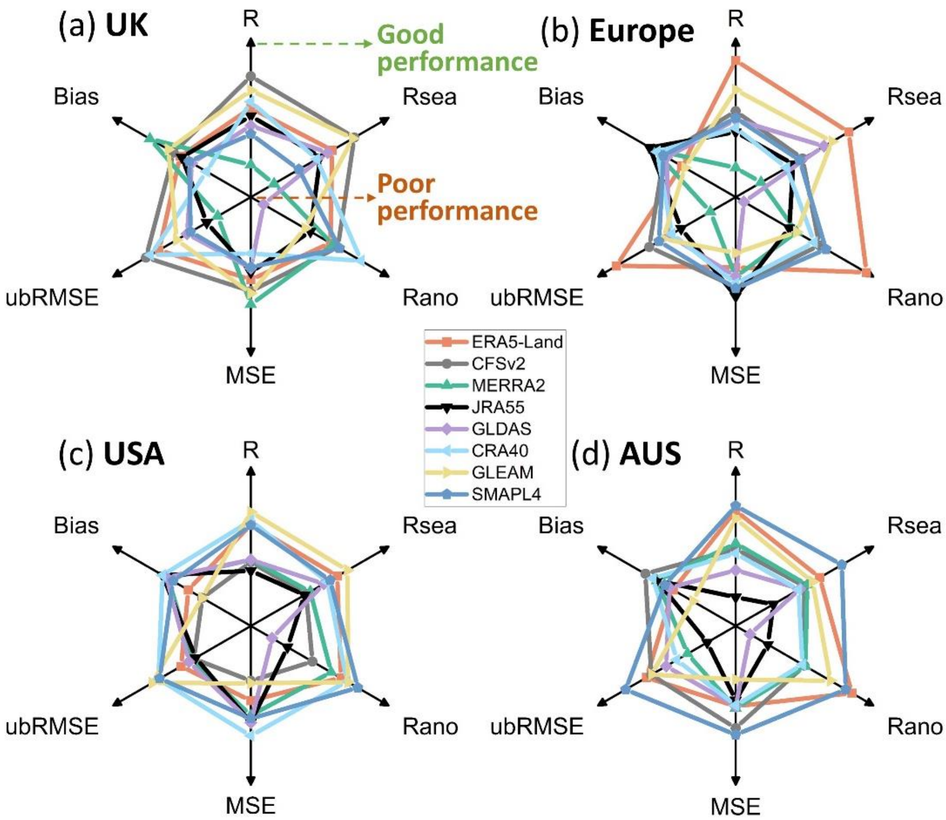

Figure 2: Brunke ranking results for a total of 8 products performance in terms of 6 statistical metrics across different regions, i.e., (a) UK, (b) mainland Europe, (c) USA and (d) Australia (AUS).

12) Line 290: Clarify meaning of "Figure 4cd". And typically it is called "US sites" rather than "USA sites".

Clarified. We have corrected "Figure 4cd" to "Figure 4(c) and (d)", which represents the subplot (c) and (d) of the Figure 4. The "USA sites" in this sentence has been corrected to "US sites"

"Figure 4(c) and (d) show the timeseries comparison of two US sites with low average R values."

13) Line 306: Clarify the use of "main paper" which implies, there is a secondary paper.
Clarified. We have corrected this sentence to:

"The distribution of some possible factors (i.e., seasonality, snow), which shows insignificant influence, is not presented in this paper."

14) Line 318: "low temporal correlation" directly contradicts the abstract line 22: "all reanalysis products exhibit good temporal correlation with the measurements". In the abstract, I suggest to add "products generally exhibit" to weaken this statement in the abstract.
Modified. We have added "generally" to this sentence in the abstract.
"Results show that all reanalysis products generally exhibit good temporal correlation with the measurements, with the median of temporal correlation coefficient (R) values spanning from 0.69 to 0.79, though large deviations are found at sites with seasonally varying vegetation cover."

15) Line 325: Please rephrase. Sites cannot be a reason for low performance. Reasons for low performance can only be process related or of technical nature. A site itself cannot be the reason for low performance.

Modified. We have rephrased this sentence in this revision:

"All reanalysis products tend to have lower performance in terms of R, R_{sea} and ubRMSE metrics at shrubland and several sites at cropland, indicating that the reanalysis products exhibit poor performance in regions characterized by high mean annual temperature, low mean annual precipitation, and high altitude (Figure 6)."

16) Line 329: CFSv2 or CFSRv2?

Corrected. CFSv2 is the second version of the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFS), this products are archived as an extension of CFSR (Saha et al., 2014; Saha et al., 2011). We have corrected all the names to "CFSv2" particularly for Figure 7, Figure 8, Figure 9 and also keep consistent throughout the paper.

17) Line 330: I contradict, a model cannot perform best in all statistical metrics except for bias. High bias produces high MSE. With a poor bias, the MSE should be poor as well. Please clarify.
Corrected. Given we have added the SMAP L4 in the analysis, the ranking of metrics has changed. We have carefully checked the statistical metrics values for this CFSv2 product, CFSv2 performs best in grassland in terms of R and ubRMSE according to the updated results.
"The Bias in grassland from a total of 6 reanalysis products is primarily negative, which means that the reanalysis products tend to underestimate the soil moisture observations in grassland (Figure 7f). CFSv2 performs best in grassland in terms of R and ubRMSE."

18) Line 345: better compared to what? Please specify.

Clarified. We have included additional details for specificity.

"In general, compared to the humid and temperate climate, all reanalysis products perform noticeably better in terms of R_{ano} *in arid climates but overestimate the CRNS measurements in Bias (Figure 5c and 5f)."*

19) Line 383: Please add that vegetation correction were proposed e.g. https://doi.org/10.1002/2014WR016443

Revised. We have added this reference.

"Besides, the lower correlation of reanalysis or satellite soil moisture products over densely vegetation regions are reported in studies from (Hagan et al., 2019; Beck et al., 2021; Kim et al., 2015; Kim et al., 2020), indicating the need for improving the vegetation parameters in land surface model or soil moisture retrieving algorithms (Baatz et al., 2015)."

20) Line 385: Please discuss in a sentence your study results with these results: https://doi.org/10.5194/hess-21-6201-2017

Added. Thanks for highlighting this reference. We found similarities between our studies and this paper from Beck et al. (2017). Their study also found that the reanalysis products tend to reproduce the season pattern of the variables well but hard to capture the anomalies.

"Moreover, Beck et al. (2017) evaluated 22 reanalysis and satellite precipitation datasets. Their study revealed that these precipitation products also tend to capture the monthly variation well but have lower performance in shorter timescales (i.e., Pearson correlation coefficient calculated for 3-day means, R_{3day}*). This aligns with our findings that the reanalysis products tend to reproduce the seasonal pattern of the variables well but that it is hard to capture the anomalies."*

21) Line 407: lower performance is preferred to "worse" performance which comes with a very negative connotation

Corrected. This sentence has been modified to:

"Our results reveal that all reanalysis products show lower performance in terms of all statistical metrics at the sites with low bulk density and high soil organic carbon (Figure S9), which are particularly from the humid regions in the UK."

22) Line 423: Pleas be specific for which use these reanalysis products are recommended. The reader leaves the study with the question - for what can these SM reanalysis be recommended?

Clarified. We have re-phrased and added the clarification here. We provide recommendations to users based on the Brunke ranking scores regarding which product demonstrates better overall performance under various regions, climate, land cover and topographic slopes. Since the Brunke ranking scores are calculated based on all 6 statistic metrics, a lower Brunke ranking score indicates the superior overall product performance. Our recommendations can help users to prioritize the product selection for their analysis. Modifications to the paper are as follows:

"To provide recommendations for the users, we classified the reanalysis products into three categories according to the Brunke ranking scores (Table S3, Figure 10), which are calculated based on all 6 statistic metrics. We provide recommendations based on these scores regarding which product demonstrates better overall performance under various regions, climate, land cover and topographic slopes. A lower Brunke ranking score indicates the superior overall product performance, suggesting users to prioritize its selection for their analysis."

23) Figure 10: Consider using more positive scoring such as +++ (optimal), ++ (ok) and + (least). No one wants to see "not recommended". Also, ERA5-Land in Humid seems to be rather not recommended according to your results (high bias) but must be considered optimal given there is no better product.

Modified. We have changed the wording to "Higher performance", "Ok" and "Lower performance". See the detailed reply and the revised figure in the general response.

24) Line 487: See above and add "generally" good agreement.

Corrected. We have added "generally" to this sentence.

"All reanalysis products generally exhibit good agreement in terms of temporal correlation with the median of R values over 0.7, whereas the lower performance with R_{ano} values are detected, indicating the weaker ability of capturing the soil moisture anomalies."

25) Line 498: There is no "balanced climate". What climate are you referring to here? I guess: temperate. Please correct.

Corrected. We have changed the "balanced climate" into "temperate" and revised carefully throughout the paper, figures and also the supporting information file.

References:

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrol. Earth Syst. Sci., 21, 6201-6217, 10.5194/hess-21-6201-2017, 2017.

Davies, P., Baatz, R., Bogena, H. R., Quansah, E., and Amekudzi, L. K.: Optimal Temporal Filtering of the Cosmic-Ray Neutron Signal to Reduce Soil Moisture Uncertainty, Sensors, 22, 9143, 2022.

Desilets, D., Zreda, M., and Ferré, T. P.: Nature's neutron probe: Land surface hydrology at an elusive scale with cosmic rays, Water Resour. Res., 46, https://doi.org/10.1029/2009WR008726, 2010.

Iwema, J., Schrön, M., Koltermann Da Silva, J., Schweiser De Paiva Lopes, R., and Rosolem, R.: Accuracy and precision of the cosmic-ray neutron sensor for soil moisture estimation at humid environments, Hydrol. Process., 35, e14419, https://doi.org/10.1002/hyp.14419, 2021.

Reichle, R. H., Liu, Q., Koster, R. D., Crow, W. T., De Lannoy, G. J. M., Kimball, J. S., Ardizzone, J. V., Bosch, D., Colliander, A., Cosh, M., Kolassa, J., Mahanama, S. P., Prueger, J., Starks, P., and Walker, J. P.: Version 4 of the SMAP Level-4 Soil Moisture Algorithm and Data Product, J. Adv. Model. Earth Sy., 11, 3106-3130, https://doi.org/10.1029/2019MS001729, 2019.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., and Iredell, M.: The NCEP climate forecast system version 2, J. Climate, 27, 2185-2208, https://doi.org/10.1175/JCLI-D-12-00823.1, 2014.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., H., v. d. D., Q., Z., W., W., M., C., and Becker, E.: NCEP Climate Forecast System Version 2 (CFSv2) Selected Hourly Time-Series Products. Research Data Archive at the National Center for Atmospheric Research [dataset], 2011.

Zreda, M., Desilets, D., Ferré, T., and Scott, R. L.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, Geophys. Res. Lett., 35, https://doi.org/10.1029/2008GL035655, 2008.