Manuscript ID: egusphere-2023-1548
Response to the Reviewers

Reviewer comments are presented in italics, while the authors' responses are in blue.

1 Referee Comments and Responses

Anonymous Referee 1:

The authors have investigated how cosmic-ray neutron soil moisture data can help improve the simulation of both soil moisture and evapotranspiration with a hydrological model. They used three relevant methods to incorporate neutron counts into large-scale hydrological modelling and compared their performance. Using cosmic-ray neutron soil moisture data to calibrate the hydrological model improved the simulation of both soil moisture and evapotranspiration.

I have found the manuscript interesting and mostly well written. The significance of the work is clear to me. I do have a few suggestions to make the contribution of the presented research stronger. These include both suggestions to improve readability and some suggestions to solidify the outcomes with some more elaborate explanations of the methodologies and a few small additional analyses.

We appreciate the reviewer comments and detailed suggestions. We greatly value your efforts to improve the paper structure. We will make the recommended revisions to the introduction and methodology section. We are especially grateful for the referee positive feedback. In this document, we present our comprehensive responses and outline our strategy for addressing the reviewer's comments in a future revision of this manuscript.

Major comments:

1. *Please, consider shortening the paragraphs of lines 32-48, 49-60, and 61-72 of the Introduction on pages 2 and 3, to help the reader understand the story line better. It is now a broader literature review that might help the reader to get the key message. Some references that are highly relevant can (and in many cases do already) enter the story in the Results and Discussion.*

   Thank you for your suggestion to improve the Introduction. We will shorten paragraphs as recommended and consider moving relevant references to the Results and Discussion section for improved clarity.

2. *Introduction, page 3, L83: "... neutron counts at scales of 1.2 x 1.2 km^2" and Conclusion and future outlook, page 26, lines 483-484 "... for simulating neutron counts at the 0.01562^o x 0.01562^o grid ...". Please, clarify in which way the neutron count simulations evaluated at this scale. Scale mismatch between model grid cell size, different model inputs, and different model calibration/validation data assimilation data should be an important aspect of this study. Please, include a discussion on the impact of scale mismatches in the manuscript. Clarification can be done in the Introduction and/or Materials and Methods and Discussion.*

   We are grateful for the reviewer's feedback. We will explain the mHM model setups, including spatial resolution, in more detail.
   L1 and L2: 0.01562° x 0.01562° is eq.$\sim$ 1.2 x 1.2 km^2. Level 1 (L1) describes the spatial resolution, as which dominant hydrological processes are modelled and Level 2 (L2) describes the resolution of the meteorological forcing data.
   L0: 0.001953125° × 0.001953125°. Level 0 (L0) describes the subgrid variability of relevant basin characteristics, which includes information on the soil as well as land use, topography and geology.

3. *If the model produces other output than soil moisture and evapotranspiration, meaning other water fluxes, can the authors discuss how the estimation of these fluxes changes under calibration with CRNS-data? If observations are available, please include these in comparison, or at least mention such analyses as recommendations for next research*

steps. It is important to verify that other model outputs do not deteriorate, or better, actually improve simultaneously with evapotranspiration simulation.

For this study, we had access to eddy-covariance measurement data only at the Hohes Holz site, which allowed us to perform the cross-validation of evapotranspiration simulations at this site. In the discussion section, we will add how the correlations between observed and simulated evapotranspiration vary in different seasons. The correlation coefficients (r) for each season are as follows: autumn [SON] (r = 0.79), spring [MAM] (r = 0.77), summer [JJA] (r = 0.42), and winter [DJF] (r = 0.87). It is worth noting that winter shows the highest correlation between observed and simulated ET, while summer exhibits the lowest correlation. The most significant deviation in terms of RMSE is evident during the summer, when evapotranspiration is highest, while the smallest difference is in winter when evapotranspiration has less impact. The model slightly overestimates evapotranspiration in summer and spring, possibly because of the absence of a dynamic vegetation growth module in the mHM, also discussed for evapotranspiration in Zink et al. (2017).

4. *A comparison with in-situ soil moisture observations is now briefly discussed in the Discussion, page 23, line 411. I suggest that the authors move this forward and make it more prominent by showing a comparison in a figure and expand the discussion. If in-situ soil moisture data were available at the other sites, these should be discussed too. If such data are not available, please mention this explicitly. Given the grid cell size of >1 km, satellite remote sensed soil moisture data is relevant too. Please discuss the relevance of CRNS data compared to satellite data at this modelling scale. To my opinion, this issue should be discussed. Implementing actual calibration and/or validation/ data assimilation with point scale soil moisture data and satellite remote sensing soil moisture data, I think should be a recommendation in the final chapter of this manuscript and should be considered by the researchers as interesting future work.*

Thank you for your feedback. In the COSMOS-Europe data paper (Bogena et al., 2022; Boeing et al., 2022), which is a key reference in soil moisture studies across Europe, soil moisture data from 66 CRNS stations deployed across Europe (referred to as COSMOS-Europe) is presented. This paper also includes the study sites that we focused on. In our paper, our primary focus was to establish a framework to invert soil hydraulic parameterization in mHM/COSMIC by directly comparing modelled neutron counts with measured ones. The on-site intensity of epithermal neutrons is directly linked to the soil moisture within the vertical and horizontal CRNS footprint.

Neutron count measurements capture soil moisture variability, as they are closely inter-linked Zreda et al. (2008); Desilets et al. (2010); Shuttleworth et al. (2013). Comparing modeled soil moisture (SM) with observations presents challenges due to scale mismatches, both in spatial extent and vertical depth. Neutron count measurements effectively overcome these challenges, making them a more suitable choice for comparison. The great advantage of CRNS over satellite data is that CRNS not only covers the few top centimeters of the soil as satellite measurements ('surface soil moisture'), but provides information on a vertical integral of soil moisture for about 15-50 cm. Furthermore, CRNS time series have a much higher temporal resolution than current satellite data. We will add this suggestion to the concluding section of the manuscript, emphasizing the significance of such future work for researchers in this field. Due to the reasons mentioned above including scale mismatch between model simulations and observations of soil moisture, we will put less emphasis on soil moisture simulation comparisons in the revised manuscript. This will also help concentrating the main focus of the study towards the neutron count simulations.

5. *Results section, page 12, lines 269-275, to the reader it is now not crystal clear which parameters were calibrated? Just the neutron related parameters or also other mHM parameters? Please clarify this textually and include a manuscript main text table (or other mechanism) to make this instantly clear.*

Thank you for your question. We will provide a comprehensive explanation of the parameter sets used in the methods section to ensure clarity and transparency. In our study, we used a total of 29 parameters for the Desilets method and 31 parameters for the COSMIC method, which include snow, soil moisture, and neutrons modules. For clarity, we have included box plots showing the calibrated range of all parameters in Figure S3 and Table S1 in the supplementary materials.

6. *Please, consider creating either one section Results and Discussion, or move bits of preliminary discussion (p 15, lines 312-315, 318-324, 325-330, 370-376) from the Results section to the Discussion section.*

   Thank you for your feedback. We will reorganize the manuscript by moving the mentioned preliminary discussion segments (p 15, lines 312-315, 318-324, 325-330, 370-376) from the Results section to the Discussion section as requested.

Other, specific comments:

7. *Title: "Improved representation of soil moisture simulations..." I doubt if the word "representation" in relation to "soil moisture simulations" is well chosen. 'representation of soil moisture processes' or 'representation of soil moisture measurements' sounds logic, but here it seems as if the representation of soil moisture simulations is improved. Please think if this is really what you mean and if so, please consider if will be understood by the wider audience.*

   Thank you for your feedback, regarding the title of our manuscript we would like to move to a revised title as: "Improved representation of soil moisture processes ...".

8. *Abstract: P1,L12-14: "A Monte Carlo simulation with Latin hypercube sampling approach . . . " Please, consider removing this sentence or writing it in more understandable wording for the audience. It is now hard to see the exact relevance of the technical details given, like 'N = 100 000'. What does such a number tell the audience?*

   Thank you. We will revise the texts to make it more clear: "We use a Monte Carlo simulation method, specifically the Latin hypercube sampling approach with a large sample size (N = 100 000)". Furthermore, to avoid confusion with 'N' representing neutron counts, we will switch to the notation 'S' for the sample size to improve the clarity of the text.

9. *Abstract, P1, L15-17: "We find that the non-uniform weighting scheme in the Desilets method provides the most reliable performance, whereas the more commonly used approach with uniformly weighted average soil moisture overestimates the observed CRNS neutron counts". How did COSMIC perform compared to the two Desilets methods?*

   In Table 4, we present a comprehensive evaluation of model performance based on three different approaches. Although we did not elaborate on COSMIC in the abstract, we will address the performance of COSMIC compared to the two Desilets methods in the abstract as suggested.

10. *Introduction, page 2, L49-41: Please, improve textually by building a logical bridge between the paragraph of lines 32-48 and of lines 49-69. As is, HYDRUS-1D is introduced suddenly and in a way that makes is seem as a very key model, without being clear why so.*

    Thank you for the suggestion. We will revise the introduction to improve the flow of text between lines 32-48 and lines 49-69, as well explaining the relevance and significance of HYDRUS-1D in the context of our study.

11. *Introduction, page 3, L73-74: The word "Eventually" and the wording with which the mHM model is introduced, at the start of this sentence and paragraph, make it seems as if the mHM model is a key hydrological model, that is the logical end-point of a discovery process and that is the standard that every reader should instantly know. It might be a well-known model, but it is one of many. Please, to help the reader understand the position of the mHM model, rewrite this to a more neutral wording.*

    Thank you. We will make the suggested change in the revised manuscript.

12. *Introduction, page 3, L83: "The COSMIC method is complex . . . ". What is meant by complex here? Please, clarify for the reader.*

    We will revise the sentence in the introduction to avoid using the term 'complex' and provide a more descriptive explanation instead i.e., "The COSMIC method enables a comprehensive representation of the neutron generation process, which is computationally more demanding than using the analytical Desilets equation ".

13. *Materials and Methods, page 4, L105: "four sites". Why does this number differ from that on line 90 of the introduction (page 4), which says "three"?*

    Here the three different sites mentioned are according to the respective landcover states: i.e, agriculture, deciduous forests, and grasslands. Four sites are mentioned in terms of CRNS locations from where we utilized the measured neutron counts data.

14. *Materials and Methods, page 4, line 111: "... producing methane fluxes". How is this relevant to the research presented? If it is relevant, it should be mentioned here and maybe discussed later on.*

    The idea to mention the methane flux was related to provide few site-specific environmental conditions of the Grosses Bruch site. But we agree with the reviewer, that mentioning them at this point probably rather confuses and does not illuminate much to the study. We will take these parts out in the revision of the manuscript.

15. *Materials and Methods, page 5, Table 1: Please say in the caption that precipitation and temperature are yearly averages and in the table itself, say '[mm/year]' for Precipitation.*

    We will make the suggested change in the revised manuscript.

16. *Materials and Methods, page 6-7: The first reference to figure 2 is now on line 151. I think that by referring the reader earlier (from line 138 onwards), it will be easier for them to understand the methodology, with this key figure in hand.*

    Thank you for your suggestion. We will make the recommended change by referring to Figure 2 (Flowchart) earlier in section 2.3 on Model setup.

17. *Materials and Methods, page 7, figure 2: In this figure, a 'Neutrons' module now appears in the upper part (modules of mesoscale hydrologic model mHM) and below, where the different neutron models are mentioned. Is this how the modelling actually works? Is there one neutron module in the mHM and then, the outcomes (neutron counts) of these are fed to the neutron models? Please adjust the figure and/or make very clear in the manuscript text how the different bits are actually connected. In addition, please clarify if the short arrows between the left and right bits connect 'Spatial Data' to 'Model Setup' and 'Model Setup' to 'Performance Matrix' or if the connections are actually 'Spatial Data' to 'mesoscale hydrologic model' and 'CRNS-methods output' to 'Performance Matrix'*

    In our study, we considered all mHM calibration parameters related to snow, soil moisture, and neutrons modules, leading to a total of 29 parameters employed for the Desilets method and 31 parameters for the COSMIC method. The simulation of soil water content considered these three mHM modules to estimate neutron counts. To comprehensively cover the parameter ranges, we sampled 100 000 (prior) parameter sets. Finally, we focused on the top 10 best performing (posterior) parameter sets based on the objective function, $KGE_{\alpha\beta}$, for further analysis and evaluation. We intend to enhance the clarity of the model setup section in the revised manuscript by providing additional descriptions, ensuring a clear understanding for readers.

18. *Materials and Methods, page 8, lines 158-159: "We compared these simulated values with the measured soil water content obtained through CRNS" This suggests soil moisture values were compared. Is this true or were actually neutron counts computed from mHM soil moisture simulation compared to neutron counts?*

    In the revised manuscript, we will clearly state that we are comparing neutron counts, not soil moisture. Our main objective is to optimize the parameterization of soil hydraulic properties in mHM/Cosmic based on the comparison between measurement and modelled neutron counts. We will adjust the text in the revised manuscript to ensure that this objective is clearly reflected, thereby eliminating any potential confusion.

19. *Materials and Methods, page 8, 2.4.1: Please, restructure this paragraph, such that parameter names are mentioned after this equation, to improve the readability.*

    Thank you for the feedback, we will change the Desilets method section as suggested in the revised manuscript.

20. *Materials and Methods, page 9, lines 185-186: "Organic water equivalent ... ". Please rephrase this sentence.*

    We will rephrase the sentence in the Materials and Methods section as: "regarding the variables of soil organic carbon (SOC) and biomass, it is important to note that these variables are often not readily available, especially when it comes to biomass data."

21. *Materials and Methods, page 10, line 204: "... does not get too small and SWC is not too high". Please, quantify.*

    We will revise the text in the manuscript to specify that the lower limit for bulk density (BD) was defined as 1.0 g/cm^3, addressing the reviewer's concern.

22. *Materials and Methods, page 11, line 234: COSMIC parameter alpha is mentioned here, but was also mentioned on line 221. This seems confusing. Please, check and improve/clarify.*

    Thank you for pointing this out. It was a typo, and we will remove the mention of α_{COSMIC} from line 234.

23. *Materials and Methods, page 11, lines 237-238: The parameters within the formula on line 237 seem not to match the parameters on line 238.*

    We agree, there was a missing equation after the one presented in line 237. We will make the necessary revisions in the manuscript by including the missing equation(s) after line 237. Thanks for spotting this.

24. *Materials and Methods, page 11, lines 248-249: "However, we modify KGE (Eq. 15) by removing the correlation coefficient rho, as it is just a measure of temporal signature and is largely dominated by seasonality alone". Why should seasonality not be included? Why is the correlation coefficient not relevant? Please, clarify this better for the reader.*

    Thank you for your comment. Gupta et al. (2009) proposed the KGE as a weighted combination of the three components (bias, variability, and correlation terms), given that our simulation already exhibited satisfactory correlation due to strong seasonality, we opted not to consider it in our assessment (objective function), as it accounted for 33% of the total weighting in the overall KGE score.
    Seasonality is an inherent characteristic in the northern hemisphere where precipitation minus evaporation is mostly driven by evapotranspiration. Even if a random parameter is selected correlation will always be higher because the meteorological forcing is the precipitation - evaporation is seasonal. The study by Cinkus et al. (2022) examined the limitations of commonly used hydrological performance criteria, particularly the Kling-Gupta Efficiency (KGE) and its variants, in model calibration and evaluation. In the revised version we will better explain why it is necessary to exclude the correlation from the KGE.

25. *Results, p12, line 273: "... in all 10 000 simulations": why was this number chosen? How do we know it is sufficient, insufficient, or too large? If only the N-parameters from the neutron models were calibrated, this seems like a large number.*

    Thank you for your comment. The choice of 100 000 simulations was determined to ensure reasonably good coverage of the parameter sets within their prescribed range, given the relatively high number of parameters involved in our study. We sampled 29 parameters for the Desilets method and 31 parameters for the COSMIC method that are not only related to neutron count module (N_0) but also to other snow, soil, and vegetation processes that affects the soil water dynamics in mHM.

26. *Results, p12, line 270: "... parameter distributions that almost cover the entire prior ... ". Why the word almost here, what is meant with it? Why is it significant to mention 'almost'?*

    In response to this question, we will extend the relevant section in the manuscript with additional explanations. We used the wording 'almost' to recognize that we couldn't be completely sure we sampled every possible parameter set, we meant that this large sample size was chosen to comprehensively explore the parameter sets and capture a wide range of possible parameter combinations.

27. *Results, p12, lines 277-278: "... Most of the high-sensitive parameters show more peaked densities in a narrower range of parameter values, reflecting the significance of variations in model parameter values ... ". Please, explain exactly why the statement is true.*

Thank you for mentioning this aspect. We will include the explanation in the manuscript as mentioned. Our study aimed to determine optimal N_0 values by refining the parameter range for N_0 using the three approach ($N_{\mathrm{Des,U}}$, $N_{\mathrm{Des,W}}$, and N_{COSMIC}), the parameter set range set for N_0 ranges between (600–1500) for Desilets method and (100–400) for COSMIC method. Through a calibration process, we adjusted these parameters to align more closely with observed data. From the iteration of 100 000 parameter sets, we selected the top 10 sets that yielded a narrower range of N_0 values, providing the best fit to the observed data. By 'more peaked densities,' we mean that following calibration from the posterior distribution, the figure (Fig. 4) displays the x-axis in gray, representing the original parameter range (600–1 500) prior distribution for Desilets method and (100–400) for COSMIC method. Meanwhile, the colored sections in orange, green, and purple indicate the parameter values of the top-performing sets for each study site.

28. *Results, p13, lines 291-292: "..., indicating that the model has the potential to generate accurate cosmic-ray soil moisture estimates even under dry conditions." Please, explain why 'even under dry conditions'? Is high performance under these conditions a surprise? If so, why?*

We mention that "even under dry conditions" emphasizes the model (mHM) performs well under dry conditions, we highlight the model ability to simulate a wide range of moisture conditions. In contrast, some hydrological models, such as HBV and PREVAH (PREecipitation Runof EVApotranspiration Hydrological response unit model; Viviroli et al., 2009), have shown weaker performance in simulating soil moisture, particularly during dry conditions, as demonstrated by Orth et al. (2015), with slightly better agreement with observations during wet conditions.

29. *Results, p13, line 293: "..., a loss of the physical meaning of the parameters in question would be very critical". Why would this be critical?*

Thank you for pointing that out, we agree that the sentence does not contribute to the clarity of the text and we will remove it.

30. *Results, p13, line 295: "One of the important additions of this work ... ". Was incorporating lattice water count added by this study for the first time?*

Yes, the addition of lattice water to the neutron counts module of mHM is a novelty of our study.

31. *Results, page 14, figure 4: Please, try to make this figure easily readable in greyscale, this would help readers who print to read the paper carefully.*

Thank you for this suggestion. We will accordingly update the figure in the revised manuscript.

32. *Results, page 14, line 308: "Furthermore, the behavioral simulation ensembles captured more variations in the COSMIC method compared to the Desilets method after the application of the objective function (i.e. KGEalpha,beta)". Do you know why? Please discuss here if Results and Discussion are combined.*

Thank you for your question. The broader confidence interval, indicating a greater range of variations, implies a higher degree of uncertainty in the COSMIC method (N_{COSMIC}). The COSMIC approach explicitly accounts for water content snow, vegetation interception, and root-zone soil processes that may likely lead to a better representation of observed neutron count variation compared to Desilets that empirically represent such processes.

33. *Results, page 16, lines 318-322. How are you sure that surface ponding and shallow groundwater and other mentioned factors are a major cause of uncertainty? Was an uncertainty analysis performed? Please, if so, discuss these briefly. If not, on which observations is this discussion based?*

No, a formal uncertainty analysis was not performed. Our discussion regarding these factors, particularly for the Grosses Bruch site, is based on prior observations of field data explained in Schrön et al. (2017). Ponding in the wet season is a common phenomenon on this site and these effects are explicitly not considered in the mHM model; and therefore we identify and mention them for future model development.

34. *Results, page 16, lines 325-330: Please, provide references to support the discussion in this paragraph.*

    Thank you for pointing this out, we will add following references to support the statement (Massoud et al., 2019 and Zink et al., 2017).

35. *Results, page 17, figure 5: The figure could be interpreted more easily and quicker if the choice of colours stated in the caption (red and black) for the different Desilets daily neutron counts, are put in the figure legend.*

    Thank you for the hint and suggestion to improve the clarity of figure 5. We will make the necessary adjustments to the figure legend, with the name $N_{(Des,W)}$, and $N_{sim(Des,Uni)}$.

36. *Results, page 18, line 340: An LHS sample 100 000 seems a lot for just the N-parameters from the three neutron models. Why was this sample size chosen?*

    The choice of 100 000 simulations was determined to ensure thorough coverage of the parameter sets, given the relatively large number of parameters involved in our study. 29 parameters for the Desilets method and 31 parameters for the COSMIC method, we aimed to comprehensively explore the possible combinations of these parameter sets values. These 100 000 simulations enable us to fully capture the distribution of parameter values.

37. *Results, page 19, table 4: A figure could help the reader to get a clear overview of these results quickly. Please, consider a parallel coordinates plot or something alike.*

    Thank you for your suggestion, we have already included a boxplot in Figure 7 to illustrate model performance for KGE and objective function $KGE_{\alpha\beta}$. In Table 4, we have highlighted the best-performing values to aid readers in quickly identifying the best values for each method. Additionally, we have improved the caption of Table 4 for enhanced clarity.

38. *Results, page 20, figure 7: Please, add horizontal axis title.*

    Thank you for the suggestion, we will add a horizontal axis title to Figure 7 in the manuscript.

39. *Results, page 20, lines 351-353: Were eddy-covariance measurements available at the other sites? If so, the same analysis should be done and presented for those sites, for complete insights from this research.*

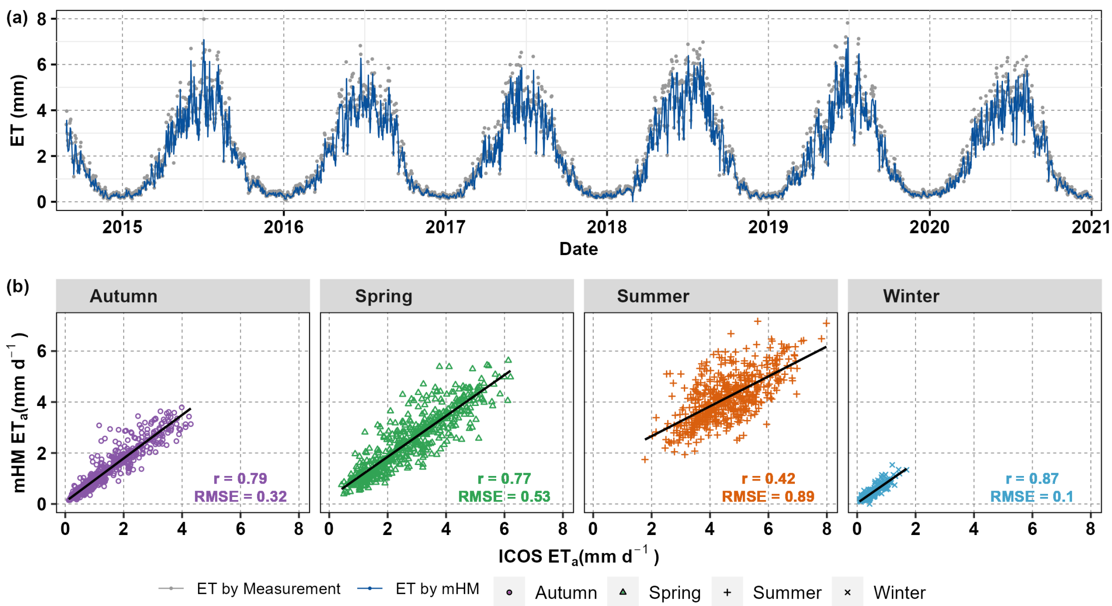
    No, for this study eddy-covariance measurements were only available at the Hohes Holz site. Therefore, we were able to evaluate the evapotranspiration simulations at this specific site only.

40. *Results, page 21, lines 354-355: 'Panel C displays the scatter plot that reveals no systematic over or underestimation of the observed actual evapotranspiration": The dashed line in the figure does not show the 1:1-line. How then does the scatter plot reveal no over or underestimation?*

    Thank you for pointing this out. The reviewer is right – the dashed lines in the figure do not represent the 1:1 line (identity line). Instead, they correspond to the best-fit regression lines corresponding to the data for the growing and non-growing seasons. These two regression lines provide insights into how well our models capture ET variations during these distinct seasons. We can also specifically estimate some summary statistics reflecting the over/underestimation of simulated values of ET. We will change the line in the revised manuscript.

41. *Results, page 21, lines 357-360. Given pieces of discussion that occur in the current Results chapter, please discuss the differences in correlations between observed and simulated evapotranspiration between different seasons.*

    In the discussion section, we will add how the correlations between observed and simulated evapotranspiration vary in different seasons. This plot provides insights into the seasonal variations in the relationship between observed and simulated ET. It suggests that the model performs best during winter, while its performance during summer is comparatively weaker. The correlation coefficients (r values) for each season are as follows: autumn [SON] ($r = 0.79$), spring [MAM] ($r = 0.77$), summer [JJA] ($r = 0.42$), and winter [DJF] ($r = 0.87$). It is worth noting that winter shows the highest correlation between observed and simulated ET, while summer exhibits the lowest correlation. The most significant deviation in terms of RMSE is evident during the summer, when evapotranspiration is highest, while the smallest difference is in winter when evapotranspiration has less impact. The model slightly overestimates evapotranspiration in summer and spring, previously addressed in the response to question 41.



42. *Results, page 21, lines 370-376; This paragraph seems to relate to the paragraph and results I mentioned in my previous comment. If this is correct, please restructure the text so this becomes clearer.*

    Thank you for your feedback. We will update the figure in the revised manuscript accordingly.

43. *Results, page 22, figure 8: If the two RMSE boxplots are combined into a single one with a single vertical axis domain, could this help the comparison?*

    We believe that merging the two RMSE boxplots into a single plot with a single vertical axis domain is not suitable in this case. The reason is that the Y-axis values for ET during the growing and non-growing seasons significantly differ (due to differences in ET values between these two seasons). Combining them into one plot would result in the non-growing season boxplot being too small to visualize and making it difficult to distinguish the mean values within the boxplots. Therefore, we prefer to keep them separated for clarity.

44. *Discussion, page 23, lines 389-390: "Therefore, we extended this uniform-averaging scheme by a vertical weighting scheme to mimic the sensitivity of the neutrons to the upper layer" Was this a contribution done through the work in this research or should previous work be referenced here?*

    In our study, we incorporated the vertical weighting scheme for soil moisture in Desilets method, into the mHM model, and we applied and tested it across various landcover sites. In past studies, the techniques of both weighted and non-weighted soil moisture approaches in the context of CRNS have been discussed (rivera et al. 2014, baroni et al. 2015, schreiner et al. 2016, zreda et al. 2016, Schrön et al. 2017, vather et al. 2019, barbosa et al. 2021)

45. *Discussion, page 23, lines 407-408: "..., indicating that the dynamic vegetation effect is just a minor observational issue (...)". The abundant vegetation does affect the CRNS measurement precision. How does that affect the calibration process and further analysis of this study?*

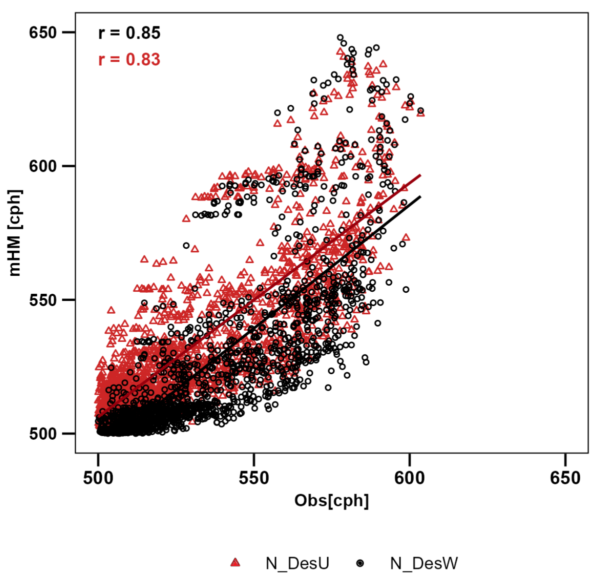
Thank you for your comment. We will remove these lines because they are related to CRNS data calibration at the field site, which is not directly relevant to our study on hydrological modeling.

46. *Discussion, page 24, lines 420-421: "… the results confirm the findings from Zink et al.(2017)." Please expand a small bit on this reference. Which type of soil moisture data did they use?*

Zink et al. (2017) utilized soil moisture observations, obtained from eddy covariance stations, to evaluate modeled soil moisture. These soil moisture measurements were collected using Time-Domain Reflectometer (TDR) or Frequency-Domain Reflectometer (FDR) sensors, which have a control volume of ten to hundreds of cubic centimeters only. Because of variations in spatial representativeness and sampling depth, they did not directly compare observed and simulated soil moisture. Instead, their objective was to analyze the temporal dynamics of soil moisture by normalizing the respective soil moisture time series (as described in Koster et al., 2009). We will expand the texts in the revise mansucript to include these aspects.

47. *Discussion, page 24, line 435-436: "… while the weighted approach N(Des,U) shows a slightly better performance that the other two methods … " How significant was the difference, i.e. What is meant with 'slightly'?*

The Grosses Bruch site with the uniformly weighted approach $N_{(\mathrm{Des,U})}$ shows a "slightly better" performance than the $N_{(\mathrm{Des,W})}$, means that in terms of correlation and another performance indices (i.e., KGE, NSE, PBIAS), as shown in Table 4. $N_{(\mathrm{Des,U})}$ (0.85, 0.69,0.7%) and $N_{(\mathrm{Des,W})}$ (0.81, 0.60, -1.3%).



48. *Discussion, page 25, line 451: "We also included offset hydrogen pools in the form of lattice water to the N_0 calibration function, … " Was soil organic matter included? If not, why not? Another factor, vegetation (including intercepted water), was this corrected for in this study? If not why not? If so, what did the results indicate? How substantial was the effect of vegetation at the different sites?*

Unfortunately, soil organic matter was not explicitly parameterized in the version of mHM used for this study. The intercepted water on leaves and in the litter layer can be particularly challenging to quantify, especially in forested stations such as Hohes Holz (Bogena et al., 2013; Schrön et al., 2017). The assessment of mHM with evapotranspiration data from eddy covariance stations at Hohes Holz site showed deficiencies in mHM. Especially in summer and spring, deviations of the modeled and observed ET indicate room for improving the representation of vegetation dynamics within mHM. However, for other sites (Grosses Bruch, Hordorf, and Cunnerdorf), we did not have eddy covariance stations to check the evapotranspiration of the measured vs. simulated ones. In lines 405-452, we discussed in detail the simulation of neutron counts and the factors influencing these simulations in comparison to observations at our study sites. We explored this using the three approaches: Desiles (Uniform, Non-Uniform), and COSMIC.

49. *Conclusion and future outlook, page 27, lines 500-503: Different sources of uncertainty regarding the neutron modelling are mentioned here. I wondered, given modelling tools are available to give an estimation of the size of the contributions from the different factors on neutron intensity, were such estimations made within this study? If so, what did they tell?*

    Thank you for your question, our analysis primarily addresses the uncertainty of the model parameters, and we will clarify this in the revised manuscript. To assess parameter uncertainty in mHM with respect to neutron counts, we employed Latin hypercube sampling involving 100 000 parameter sets. we took the top 10 best parameter sets as a behavioral solution. In Supplement Figures S1-S5, we present the Probability Density Function (PDF) plots of all parameter sets for our study sites, both prior and posterior to the simulation. Our analysis result shows that $N_{0,\mathrm{Des}}$, $N_{0,\mathrm{COSMIC}}$, rootFractionCoefficient_pervious, and rootFractionCoefficient_forest e, are the most sensitive model parameters.

50. *Conclusion and future outlook, page 27, lines 509-511: "… provides a more realistic representation of soil moisture dynamics as well as evapotranspiration, particularly at the forest site". If I have understood the manuscript correctly, evapotranspiration was evaluated at one site only. The sentence here in the conclusions chapter seems to suggest a broad result for evapotranspiration. Please, rewrite to make this explicit.*

    Yes, evapotranspiration was evaluated at one site because we have the eddy covariance flux observation only at the Hohes Holz. Thank you for your feedback, we will update the texts in the revised manuscript accordingly.

51. *Please check for textual imperfections throughout the manuscript. Three examples from the abstract, introduction, and results:*

    - *P1, L3: "… due to their hectare scale footprint and … " -> "… due to its hectare scale footprint and … "*
    - *P2, L21: "the mass" -> water mass? Carbon mass? Both or more?*
    - *P13, L285: The words 'uniform prior distribution range for' should be repeated before "N_0,cosmic", or rephrase in another way*

    Thank you for your thorough review and your valuable feedback. We appreciate your attention to detail, and we will address these issues accordingly.

References

Boeing, F., Rakovec, O., Kumar, R., Samaniego, L., Schrön, M., Hildebrandt, A., Rebmann, C., Thober, S., Müller, S., Zacharias, S., Bogena, H., Schneider, K., Kiese, R., Attinger, S., and Marx, A.: High-resolution drought simulations and comparison to soil moisture observations in Germany, Hydrology and Earth System Sciences, 26, 5137–5161, https://doi.org/10.5194/hess-26-5137-2022, 2022.

Bogena, H. R., Schrön, M., Jakobi, J., Ney, P., Zacharias, S., Andreasen, M., Baatz, R., Boorman, D., Duygu, M. B., Eguibar-Galán, M. A., et al.: COSMOS-Europe: a European network of cosmic-ray neutron soil moisture sensors, Earth System Science Data, 14, 1125–1151, 2022.

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., and Goldscheider, N.: When best is the enemy of good–critical evaluation of performance criteria in hydrological models, Hydrology and Earth System Sciences Discussions, 2022, 1–25, 2022.

Desilets, D., Zreda, M., and Ferré, T. P.: Nature's neutron probe: Land surface hydrology at an elusive scale with cosmic rays, Water Resources Research, 46, 2010.

Shuttleworth, J., Rosolem, R., Zreda, M., and Franz, T.: The COsmic-ray Soil Moisture Interaction Code (COSMIC) for use in data assimilation, Hydrology and Earth System Sciences, 17, 3205–3217, 2013.

Zink, M., Kumar, R., Cuntz, M., and Samaniego, L.: A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty, Hydrology and Earth System Sciences, 21, 1769–1790, 2017.

Zreda, M., Desilets, D., Ferré, T., and Scott, R. L.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, Geophysical research letters, 35, 2008.