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.

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 than 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.

2. Introduction, page 3, L83: "… neutron counts at scales of 1.2 x 1.2 km2" and Conclusion and future outlook, page 26, lines 483-484 "… for simulating neutron counts at the 0.01562o x 0.01562o grid …". Please, clarify in which way the neutron count simulation is 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.

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

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.

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.

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.

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.

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 number tell the audience?

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?

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.

11. Introduction, page 3, L73-74: The word "Eventually" and the wording with which the mHM model is introduces, 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 the mHM-model, rewrite this to a more neutral wording.

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

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"?

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.

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.

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.

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'

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?

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.

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

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

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.

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.

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.

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.

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'?

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.

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?

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?

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

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.

32. Results, page 14, line 308: "Furthermore, the behavrioral simulation ensembles captured more variations in de 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.

33. Results, page 16, lines 318-322. How are you sure the 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?

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

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.

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?

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.

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

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.

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

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

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.

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?

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?

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?

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?

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'?

48. Discussion, page 25, line 451: "We also included offset hydrogen pools in the form of lattice water to the N0 calibration function, …" Was soil organic matter included? If now, 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?

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?

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.

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 "N0,cosmic", or rephrase in another way