Answer to interactive comment by Anonymous Referee #1

We would like to thank the Referee for carefully reviewing the manuscript. An initial reply to the reviewer's comments is provided below following the original review comments shown in Italic font.

OVERVIEW

The presented manuscript investigates the potential impact of the measurement scale for calibration of a land surface model. For this purpose, observed and simulated land surface data at 12 sites on the continental US from several sources including Ameriflux, COSMOS and NLDAS was used. Point scale soil moisture data was compared to cosmic ray soil moisture retrievals. Furthermore, land surface simulations at the nine sites were done on an individual basis using JULES. At each sites, JULES was calibrated with cosmic ray data, point scale soil moisture data and eddy flux measurements. Model results were evaluated with eddy flux and soil moisture measurements. The case study demonstrates the added value of cosmic ray measurements at the model scale compared to local scale soil moisture measurements.

However, the study needs a major revision that addresses readability in the following: Reduce/clarify abbreviations, restructure part of the manuscript, improve English / sentence structure, remove speculations, be more specific /quantitative at a number of instances. There seems to be an issue with the data presented in Figure 7 concerning site MO.

The figures require further formatting. I suggest reducing the number of Figures. This allows the reader to focus on the essential messages of the study. I disagree with the outcome that coupling of soil moisture and latent heat flux is weak in JULES (e.g. see comment to Figure 9). Further suggestions in the Specific comments.

ANSWER:
We thank the referee for this evaluation of the manuscript. As also pointed out by other reviewers, we will address the issues mentioned above by the referee to increase the manuscript readability. The issue with Figure 7 is explained below in the 'specific comments'. We agree that our use of the word coupling was not appropriate as also pointed out by the other reviewers. Our interpretation is, in fact, in line with Reviewer #2 comments which should emphasize that the results obtained in our study are a consequence of calibrating soil parameters, rather than the soil moisture – evapotranspiration coupling in JULES.

GENERAL COMMENTS:
COMMENT: The paper exhibits a clear novelty by quantifying the impact of using cosmic ray soil moisture data for calibration as compared to local point soil moisture measurements. The study fits the scope of the HESS journal and deserves to be published in HESS after major revision.

The conclusions reached in the manuscript are not clear enough. I also found different conclusions from the data and results presented. For details: See Specific comments to Chapter 4. The scientific methods and assumptions were well chosen and represent state of the art.

ANSWER: We thank the reviewer for his/her positive comments regarding our manuscript and its relevance to the HESS community. We will address the issues regarding the clarity and inconsistency of our conclusions in the revised version as pointed out in detail by the reviewer's comments.

COMMENTS:

Description of experiments and calculations need to be revised. I suggest following new structure: Chapter 2.1 can remain there. Then, Chapter 2.3 should be changed to Chapter 2.2 as soil moisture data should be compared before calibration or modeling. Then explain JULES, then JULES forcing and initial conditions, the following Chapters can remain in place.

The results chapter needs a new structure. The results are presented in the right order but intermittent by discussions that are out of place because there IS a Chapter "3.8 Discussion". A more clear structure would be either consistently "3. Results and Discussion" or "3. Results; 4. Discussion; 5. Conclusion". Please, stick to either one but do not mix.
The topic is complex and in general well addressed, but a new structure will increase readability and will make writing the paper more easy.

ANSWERS: We will revise the structure of our manuscript appropriately to improve its readability and clarity. We thank the referee for his/her suggestions.

COMMENT: The title reflects the content. However, I would suggest a modification of the title to e.g. "Improved land surface processes by calibration with cosmic ray soil moisture measurements at the model scale".

ANSWER: We thank the reviewer for making this point which was also raised by Referee #2. We are considering the following new title: ”Land surface model performance using cosmic-ray and point scale soil moisture data for calibration”.

COMMENT: The abstract is concise and summarizes the paper well. It may be modified if conclusions are changed.

ANSWER: We appreciate that the Referee found our abstract to be concise and a good summary of our manuscript

COMMENT: In general, there is a large number of abbreviations (e.g. PS), symbols (in formulas), short names (e.g. smcrit). This makes the paper very difficult to follow. It is necessary to use either abbreviations, or symbols also in the text, ommit short names and in general write the names out more often. This paper almost needs a List of Abbreviations. Please, reduce them.

ANSWER: This is a valid point raised by the referee. Whenever possible, we will reduce the number of abbreviations and/or include a List of Abbreviations for clarity. Notice that some abbreviations are already widely accepted by the scientific community (e.g., LSM, CRNS) and we believe in those cases, the use of them are justified.

SPECIFIC COMMENTS
The referee has made a number of suggestions to significantly improve the quality of the manuscript, in particular regarding to its structure, clarity, and use of language. We thank the Referee for his comments and will make appropriate changes in the revised version. Below, we focus on the main comments that concern with the methodology and conclusions drawn.

COMMENT:
Page 4, line 32: "More than 31 days were gap filled" using average diurnal pattern. This sounds like a really high uncertainty. Is this the case for precipitation, too? Is the high uncertainty reflected in the results? If so, where? Is it feasible to mark this in the Figures? Is it feasible to remove these periods from the calibration period? How much of modeled periods was filled with diurnal patterns?

ANSWER:
We thank the reviewer for raising this point. This was also mentioned by reviewer #3. We provide a table below summarising the average (average of the seven forcing variables) percent gap for each site. Precipitation was not gap-filled; missing points were set to zero instead. Overall, average gaps vary among sites between near zero to 15%. As pointed out by the reviewer, this can introduce some uncertainty in the analysis and we will highlight data when describing the dataset used in this study. We will also mentioned in the text the original gap period in the data.

Site	Percentage missing hours filled: mean (range)	Size of time series in years
UM	7 (2-12)	1.5
DC	1 (0-5)	3.7
SO	7 (0-15)	3.8
KE	1 (0-3)	4.6
ME	0 (0-0.1)	1.6
SR	2 (0-14)	3.6
CS	2 (0-5)	3.7
MM	1 (0-2)	2.7
TR	0.1 (0-0.2)	3.6
AR	10 (8-14)	2.5
WR	2 (0-4)	2.6
MO	3 (1-7)	2.7

COMMENT:
Page 13, line 13: I suggest to discuss these reasons. Actually, it is worth investigating each of the points to either accept them or rule them out. Mentioning all of these points / reasons is not getting the manuscript closer to the objective. 35: The "two-objective calibration" was not mentioned before. However, it is good to have it.

ANSWER:
Page 14, line 13: We thank the referee for his/her comments. Note, this issue was also raised by Referee #3 which suggested for removal. We re-assess this section of the manuscript for the revised version of the manuscript.

COMMENT:
Page 14, line 19: "coincidence" – Not really, is it?

ANSWER:
Page 14, line 19: We understand that the word “coincidence” may not have been chosen appropriately in this context. We explain our observation better in the revised version of the manuscript. What we meant to say is that there is: A small deterioration in PS soil moisture or CRNS neutron simulation could yield a much worse latent heat flux simulation, even of the same order as the default run. This happened for example for the CRNS calibration at SR (Figure 11), where a deterioration of only about 0.05 in the normalised RMSE-neutron count could have yielded the same normalised RMSE-LE as the default run (1). This implies a large (e.g. 0.8) improvement in soil moisture (in terms of in normalised error) would not necessarily mean an improvement in latent heat flux.

COMMENT:
Page 14, line 28: What is the reasons that at these sites strong coupling is expected? At what time? Is the calibration done during the time of strong coupling? Very unclear how this conclusion is reached. I suggest an individual point of discussion. This is also the point where I cannot follow the conclusion drawn. It would be a major setback of JULES which needs to be justified much stronger. Above all, the results in the Figures show a meaningful difference in ET due to calibration / calibration method.

ANSWER:
Page 14, line 28: As mentioned above, we will improve the discussion to appropriately address the issue of consequence of soil parameters calibration rather than strong/weak coupling as originally stated in the manuscript.

COMMENT:
Page 15 line...: 4: "day time only" Why suddenly "day time only"? Is this throughout the manuscript the case or just here? LE measurements are difficult in the night. Diurnal cycles are difficult to obtain. Only modeled LE at times of observed ET should be compared, because there is no observations at other times. I do not see this here.

ANSWER:
Page 15 line 4: This is an important point raised by the referee. Latent heat flux and evaporative fraction performance were computed over day time values only (Page 8 line 16). Notice that night time values were used to plot the monthly mean diurnal cycles shown in Figure 9, but did not make any contribution to the computation of the validation metrics (RMSEs). We will make this clearer where we present the results and in the figure captions.

COMMENT:
Page 15, line 28: "root zone soil moisture" is where exactly and calculated how?

ANSWER:
Page 15, line 28: JULES does not provide a root zone soil moisture output weighted by the presence of roots that can be directly compared with the values of the wilting point and critical point soil moisture. Therefore we computed an estimation of root zone soil moisture by computing the soil moisture stress weighting factor for each layer at each time step (Equation 3). We then computed the relative contribution from each layer to the total root zone soil moisture stress factor. This relative contribution is a function of the root density and soil thickness. To finally obtain the weighted root zone soil moisture we multiplied the JULES soil moisture of each layer with the weighting factor.

COMMENT:
Page 16, line 14: "did not differ substantially". I disagree. They did differ.

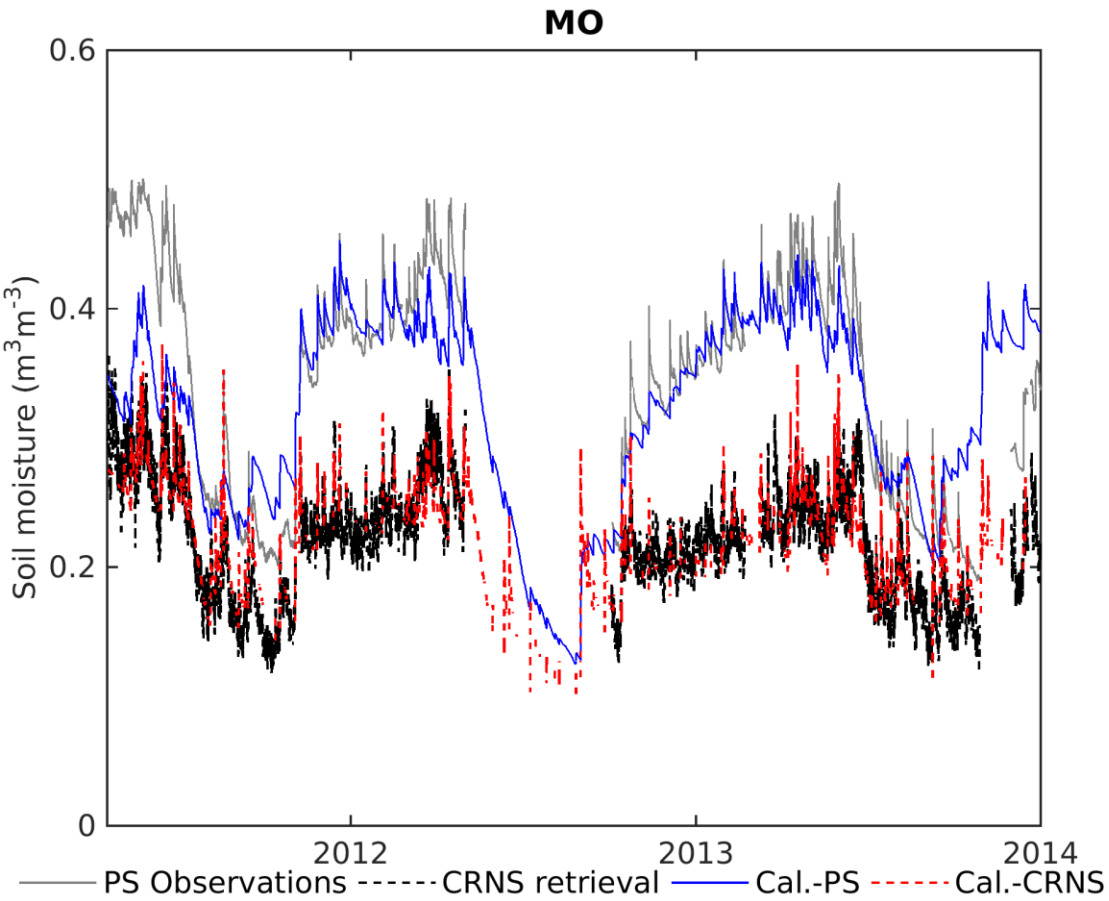
ANSWER:
Page 16, line 14: We argue there was no substantial difference between calibrating against PS and against CRNS data with respect to latent heat flux across sites. We base this on the results from the two-objective calibrations against soil moisture and latent heat and against neutron counts and latent heat (Figure 12). Only at SR (CRNS better), DC (CRNS better), TR (PS better), and UM (PS better) did we observe a substantial (larger than 10%) difference between PS and CRNS calibration, based on the compromise solutions from the two-objective calibrations. To increase readability, we consider using latent heat as principle validation metric instead of evaporative fraction. The results for evaporative fraction (figure 8) would then be moved to the appendix, while Figure A2.1 would become the new Figure 8 (and a scatter plot like current Figure 8b will be added.

COMMENT:
Page 16, line 21: "Our findings support this." Your findings were that the JULES model is not an "improved land surface model", so the manuscript as is cannot support this. However, I see an impact of soil moisture states on latent heat flux in your model runs. Just have a look on Figure 11, how the RMSE in LE is reduced by calibration at sites DC,SO,KE,SR,WR. Only at few sites, RMSE in LE became higher. Soil moisture seemed to impact latent heat flux in JULES. The authors can be and should be more positive in the results, discussion and conclusion.

ANSWER:
Page 16, line 21: This is a very important point mentioned by the reviewer. We will re-phrase our conclusions on JULES because, as the reviewer mentions, JULES soil moisture does affect simulated surface energy partitioning and latent heat flux. The authors do still not see a substantial difference in surface energy partitioning and latent heat flux simulation when CRNS neutron counts are used instead of PS soil moisture. Our revision will focus on the limited effects of calibrating soil parameters on evapotranspiration as pointed out by Referee #2 based on his previous work (e.g., Teuling et al., 2009). We will also expand our analysis to incorporate discussions on soil moisture data quality from both measurement approaches (as pointed out by Referee #3) and expected behaviour from spatio-temporal stability (as pointed out by Referee #2).

COMMENT:
Figure 7: How can CRNS and PS soil moisture at MO be so different as in Figure 5, but Model results after calibration be so similar as in Figure 7. It seems very strange.

ANSWER:
Figure 7: Please notice that the overall calibration based on PS and CRNS (i.e., top and bottom MO panels in Figure 7) actually yielded different results. The figure on page 6 of this answer combines both plots to highlight this fact (notice default run line is omitted as it is irrelevant to this comparison).



Hourly observed (PS and CRNS) soil moisture time series and simulated soil moisture time series after calibration against PS and CRNS soil moisture for site MO.

COMMENT:
Figure 9: Interestingly WR and MO show a really strong change from Default to Calibrated. How do you get to the conclusion that there is no coupling of LE to soil moisture in JULES?

ANSWER:
Figure 9: The Referee is right that WR and MO show a strong change from default to single-objective calibrated. As mentioned above, we will improve the discussion to appropriately address the issue of consequence of soil parameters calibration rather than strong/weak coupling as originally stated in the manuscript.

REFERENCES
Teuling et al, 2009. Parameter sensitivity in LSMs: An analysis using stochastic soil moisture models and ELDAS soil parameters. J. Hydrometeorol., 10(3), 751-765.