Final answer to review comments on manuscript originally entitled "Reducing soil moisture measurement scale mismatch to improve surface energy flux estimation" by Iwema et al. submitted to Hydrology and Earth System Sciences

We thank the Editor for his work and guidance. We address the comments of each referee below. At the end of this answer a marked up version of the revised manuscript is included. For clarity we first address the main issues mentioned by the Editor, based on comments by the three Referees:

1. Consider restructuring the manuscript and improving parts of the manuscript like the discussion of the literature (incomplete as detailed by reviewer #2).

- We have moved Section 2.3 "Soil moisture data comparison methodology" forward; now it is Section 2.2.
- Figure 1, Figure 8a, and Figure 13 (numbering as in the original manuscript) were removed. Figure 3 was moved from the main part of the manuscript to the newly created supplemental material, more precisely Supplement 1.
- We now use the Root Mean Squared Error (RMSE) of latent heat flux as principle validation metric. We have moved Figure 8b, presenting the RMSE of the evaporative fraction, to Supplement 2. We now mention the evaporative fraction analysis only briefly in Section 3.3 of the new manuscript. We produced a figure equivalent to Figure 8b, showing the relative improvement in latent heat flux RMSE against the relative improvement in soil moisture RMSE to the main part of the manuscript. This provides a more consistent structure. We have rephrased sentences in Section 3.3 "Validation", now 3.3 "Validation of the single-objective calibration against eddy-covariance observations", in accordance with this change.
- We have moved Section 3.4 "Were calibrated parameter values physically feasible?" to the new Supplement 3. This allowed us to present a more coherent story without any unnecessary distractions. We refer to this supplement from Appendix 2, which itself was Section 3.6 in the old manuscript ("To which parameters were soil moisture and latent heat flux most sensitive?").
- We have completely rephrased Section 3.5 (now 3.4) "Two-objective calibration against soil moisture and latent heat flux" to provide a clearer interpretation of the results, after comments by Referee #1.
- We have removed Section 3.7 "Could JULES model structure explain the limited improvement in surface energy flux estimation?" and Section 3.8 "Discussion", after comments by Referee #1. We have allocated the paragraphs from these sections to Section 3.1, 3.2, 3.3, and 3.4. This has improved the structure and readability of our manuscript.

2. The link between results and conclusions is questioned by the reviewers. In particular, the weak coupling between ET and soil moisture is not so obvious from the results.

- We have revised our conclusions after comments by all three referees. We don't conclude any longer that JULES has weak soil moisture – evapotranspiration coupling. We attribute the lack of differences between calibrating against the two different soil moisture observations products (PS and CRNS) to (1) the limited effect of calibrating soil parameters on evapotranspiration, (2) the implications of spatio-temporal stability theory, (3) data quality properties of the soil moisture observation techniques, and (4) the self-adjusting behaviour of the wilting point and critical point soil moisture parameters after calibrations.

3. Some topics require additional attention: the role of deeper soil moisture and the rooting depth, and also the quality of the different types of soil moisture measurements.

- We have addressed the issues of the value of deeper soil moisture and rooting depth in our answer to the comments by Referee #2. We have included discussions on these issues in our "Results and Discussion" section.
- We have addressed the quality of different types of soil moisture measurements (uncertainties in Cosmic-Ray Neutron Sensor and point-scale data), as raised by Referee #3, in our "Results and Discussion" section.

In addition to these major points, we have improved the phrasing in some sentences not mentioned by the referees. We have removed a few paragraphs (detailed in the answers to specific comments by the referees) to decrease the size of our manuscript and to improve its readability. We have changed one reference, to Gupta et al. (1998), which was on page 13, line 38 of the original manuscript, to Gupta et al. (1999). Two other references were found missing in the literature list and were therefore added (Gupta et al., 2009; Finkelstein and Sims, 2001).

Finally, we have changed both the title and hypothesis after comments by Referee #2 Ryan Teuling. The title is now:

"Land surface model performance using cosmic-ray and point scale soil moisture measurements for calibration"

The new hypothesis is:

"Reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations"

Sincerely,

Joost Iwema and co-authors

Answer to interactive comment by Anonymous Referee #1

We would like to thank the Referee for carefully reviewing the manuscript. A point-by-point reply to the comments is provided below with original 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. We have addressed the issues mentioned above by the referee to increase the readability. This issue was also raised by the other two referees. The issue with Figure 7 is explained in the 'specific comments'.
All three Referees disagreed with our conclusion that the coupling between soil moisture and latent heat flux in LSM JULES is weak. We agree that our use of the word 'coupling' was not appropriate as also pointed out by the other referees. Our interpretation is that the results obtained in our study are mainly a consequence of calibrating soil parameters, rather than the soil moisture – evapotranspiration coupling in JULES. Other reasons playing a role are the implications of the spatio-temporal stability of soil moisture measurements, the quality of the soil moisture data, and the self-adjusting behaviour of the wilting point and critical point soil moisture.

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 have addressed the issues regarding the clarity and inconsistency of our conclusions in the revised version of the manuscript as pointed out in detail by the reviewer's comments.

COMMENT: *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.*

ANSWER: We have adopted the structure suggested by the referee.

COMMENT: *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.

ANSWER: We agree with the referee's comment. We have created a section "3. Results and Discussion". Section 3.7 and Section 3.8 were removed from the revised manuscript. The different paragraphs were assigned to Section 3.2, 3.3, and 3.4 of the revised manuscript.

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 have adopted the title proposed by Referee #2 Dr. Teuling: "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. In the revised version of the manuscript we have reduced the number of abbreviations. We have however kept certain abbreviations that are commonly known in the land surface modelling community and in the soil moisture measurement community. Another reason to keep these abbreviations is that writing them out would increase the size of the manuscript considerably and labels in figures would not fit with a readable font. These include “PS” (point-scale), “CRNS” (Cosmic-Ray Neutron Sensor), “LSM” (Land Surface Model), “SM” (soil moisture), and “JULES” (Joint UK Land Environment Simulator). We have removed the abbreviations “EC” (eddy-covariance) and “OF” (objective function).

SPECIFIC COMMENTS

COMMENT:

Page 1 line...: 9: "sometimes" is too unprecise. Rephrase e.g. "can be calibrated" 10: EC, PS, LSM, CRNS, JULES - How can the abbreviations in the abstract be reduced? 12: Is there a term “soil-evapotranspiration”? I suggest “soil and evapotranspiration” or “soil-evaporation”. 17: "CRNS calibrations" – What is actually meant is “LSM calibration” or something alike. 30: "atmospheric circulation" - In the UK? Be more specific here or do not mention it. 30: "Because", is "because" naturally a start of the sentence?

ANSWER:

Page 1 line 9: We have replaced “sometimes” with “can be calibrated”.
Line 10: We have removed all abbreviations except “JULES” and “USA” from the abstract. We kept “JULES” because this is how the model is generally referred to by the land surface modelling community. We assume the abbreviation “USA” is commonly understood.
Line 12: The term “soil-evapotranspiration” has been replaced with “soil and evapotranspiration”.
Line 17: We have changed “We found that simulated surface energy partitioning did not differ substantially between the PS and CRNS calibrations” to “We found that simulated surface energy partitioning did not differ substantially between both calibration strategies”.
Line 30: The paragraph on line 28 of page 1 to line 2 of page 2 has been changed to:
“Land Surface Models (LSMs) solve the surface mass (including water), energy, and momentum balances to provide the weather and climate prediction models with lower boundary conditions. The land surface has been shown to play an important role in global atmospheric circulation (Koster et al., 2004). Because the soil moisture state and surface fluxes are so closely connected, it is important to accurately simulate these simultaneously (Henderson-Sellers et al., 1996; Richter et al., 2004; Seneviratne et al., 2010; Dirmeyer, 2011; Dirmeyer et al., 2013).”

COMMENT:

Page 2 line...: 1: "believed" - It is for THESE authors important or it is not. Maybe, why is it important. There is no "believe" in science. 4: "processes." Reference missing. 4: Remove "however". no added value. 14-15. Unclear, specify or rephrase. 22: "in-situ soil moisture" – I suggest to be consistent throughout the paper.in-situ or point scale is the same. So I suggest using one phrasing only. 24: soil moisture “IS” spatially and add reference e.g. Qu et al. 2015 in GRL. 26-28: "EC footprint average soil moisture" or "LSM grid cell" average soil moisture depends on the objective. As is, it is confusing. Rephrase to be clear. 31: Is this the research question? There is another research question on the next page. Either both should be allocated together or only one objective / research question is needed. As is, it is confusing.

ANSWER:

Page 2 line 1, please see answer to previous comment (page 1 line 30).
Line 4: The references in the next sentence (line 4-5) have been added to this sentence and the sentence has been changed to reduce the size of the manuscript (Lines 3-5 of the revised manuscript):

"The increasing complexity of land surface models over the past decades (Sellers, 1997; Seneviratne et al., 2010) has brought with it an increasing number of parameters, with values not easily defined with in-situ measurements because of scale mismatch"
Line 14-15: We rephrased to:
"In an effort to develop global hydrometeorological monitoring and prediction capabilities (Wood et al., 2011), hydrological models and LSMs are now increasingly being applied at the finer `hyper-resolution scale' with grid cells of about 1 km^2."
Line 22: We respectfully disagree with the Referee on this point, because Point Scale soil moisture sensors, Cosmic-Ray Neutron Sensor soil moisture sensors, and Eddy-Covariance sensors are all in-situ; they are on site. We use Point Scale soil moisture sensors to refer to those placed within the soil, and which represent a footprint of a ~4 dm^3 only.
Line 24: "Soil moisture can be spatially non-uniform" has been changed to: "Soil moisture is spatially non-uniform" and the suggested reference will be added.
Lines 26-28 Have been rephrased to (lines 21-22 of the revised manuscript): "Therefore, soil moisture measurements best (i.e. most effectively) representing the soil below the eddy-covariance tower's footprint should be used when the performance of a land surface model is evaluated".
Line 31: The question phrased here is our research question. On page 3, lines 22-23 of the original manuscript we phrased a hypothesis for our research question. We have rephrased this text to: "In order to answer our research question (page 3, line 8 of the revised manuscript) we hypothesise that *reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations*."
Please note that we have changed our hypothesis to a null-hypothesis as proposed by Referee #2.
We have moved the hypothesis to immediately after the research question.

COMMENT:
Page 3 line...: 5 "usually assumed" - this is rather a fact due to soil heterogeneity. 10: One sentence paragraphs should be avoided. Also the use of "which" as often as it is used in the present manuscript, should be avoided. In English, short sentences are much better understood and much clearer. One sentence is preferably 1-2 lines only. 12: add the German CRNS network (Baatz et al 2015) 13: Repetition. 16: "similar" - be more specific. Similar is least informative and makes room for interpretation. 17: which – split the here sentence. 21: which - rephrase to "These sites" also look up the difference between which and that. Is the information you give with "which" really necessary? Then put it in one single sentence as it is worth a single sentence. 27: that – could that be removed here? 30: "Before our modelling exercise" ... This sentence should be moved up within the paragraph. First things first. 30-33: A shorter sentences are desirable.. It can be two or more sentences.

ANSWER:
Page 3, line 5: Due to comments by Referee #2 and Referee #3, we have kept this sentence as it is, but is was moved up (page 2, lines 24-27), to before the research question. The other two referees argued that based on temporal stability theory more soil moisture observation points in space do not necessarily yield better surface energy fluxes.
Line 10: This one sentence paragraph has been combined with the following paragraph.
Line 12: The reference to Baatz et al. (2015) has been added.
Line 13: This sentence has been removed and the information has been added to the first sentence of this paragraph instead: "The CRNS (Zreda et al., 2008) is an above-ground passive sensor which utilises natural cosmic-ray neutron radiation to estimate soil moisture content in the top 10-70 cm. The sensor's footprint area has a radius of about 100 to 300 m surrounding the above-ground sensor (Figure 2; Desilets and Zreda, 2013; Kohli et al., 2015)."
Line 16: Changed to: "Franz et al. (2012) showed soil moisture estimated from CRNS neutron counts differed less than 20% from the average of a co-located point scale soil moisture sensor network at a site in Arizona."

Line 17: We have rephrased this sentence to: "Unlike wireless point scale sensor networks, both the GPS and CRNS technology provide an integrated soil moisture measurement over the entire support volume (Larson et al., 2008; Zreda et al., 2008)."
Line 21: We have changed to: "Twelve of these sites provided sufficient …"
Line 26-27: We have rephrased to: "We did this by calibrating parameters of the Joint UK Land Environment Simulator (JULES; Best et al., 2011) against Point scale and Cosmic-Ray Neutron data separately."
Line 30-33: We have moved the sentences on lines 30-33 to the beginning of the paragraph, lines 31-33 of the revised manuscript.

COMMENT:
Page 4 line...: 10: "used data" - Specify "data". 10: remove brackets and specify e.g. "the upper first 30 cm". 15: Remove "Similar". 18: "CRNS integrated soil moisture" ... rephrase for better understanding e.g. soil moisture integrated over depth from CRNS soil moisture, hereafter referred to as CRNS soil moisture retrieval. 23: Split sentence, remove while. 31: remove which, split sentence 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:
Page 4 line 10: We have specified "data" as follows: "We used point scale soil moisture data from the soil layers up to 30 cm depth only for consistency among all sites. Our main objective was to investigate the difference in information content due to two soil moisture measurement techniques' different horizontal scales in relation to the eddy-covariance footprint, rather than to compare the measurement techniques themselves."
Line 15: "Similar" has been removed.
Line 18: We moved this sentence to the Section "Soil moisture data comparison methodology" and rephrased the last lines of that section to: "We compared PS soil moisture with CRNS soil moisture values computed from vertically homogeneous soil moisture values obtained from the observed neutron counts using the COsmic-ray Soil Moisture Interaction Code (COSMIC; Shuttleworth et al., 2013)."
Line 23: This sentence was split and the second part was rephrased to: "We used data version 3.4 (Goulden et al., 2015) for the three California Climate Gradients sites."
Line 31: We have split the sentence and removed "while" as suggested by the reviewer.
Line 32: 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; Table a) percent gap for each site filled with this method and a figure (Figure a) showing how large the gaps were. This figure shows the percentage missing hours for each gap size category. If the bars for one variable at one site would be summed, this would yield the total percentage of missing hours. Overall, average gaps vary among sites between near zero to 15%. Gaps larger than 15 days occurred at:

- UM: downward shortwave and downward longwave radiation (9% of time)
- DC: net radiation (4% of time)
- KE: downward longwave radiation (4% of time)
- SR: downward longwave radiation (8% of time)
- CS: downward shortwave radiation (3% of time)
- MM: pressure (2% of time)
- AR: temperature, wind speed, and relative humidity (2-4% of time)
- WR: precipitation and pressure (3% of time)

- MO: precipitation (2% of time)

Summarising, large gaps (over 15 days) occurred for no more than 10% of time at any site and occurred at most sites for longwave and shortwave radiation only. Precipitation is likely to be the primary gap factor affecting soil moisture but its occurrence was minimal. Nevertheless, we noted this issue in the manuscript (Section 3.3 of the revised manuscript, see below).
In this study, we used the most regular freely-available data sets to the best of our knowledge to ensure reproducibility. We noticed that outcomes for sites with few large data gaps did overall not differ from the outcomes on sites with more large data gaps. Therefore, we argue that although the data gaps may introduce some uncertainties in the model results, our major conclusions remain unchanged. We discuss the most important large data gaps in Section 3.3:

"A month long gap in longwave and shortwave downward model forcing data occurred in April 2012. This might have affected the results for this site."

and in Section 3.4 of the revised manuscript:
"An issue that could have had effect on our results is the occurrence of gaps in the model forcing data. Gaps in observed meteorological data are often inevitable and must be filled for a land surface model to be run. Percentages of missing hours differed between 0% and 15% at our sites. Gaps larger than fifteen days filled with the moving window gap-filling procedure occurred mainly for downward shortwave and/or downward longwave radiation; at sites UM, KE, SR, DC, TR, and CS. At site WR a data gap of 29 days in precipitation and air pressure occurred. Especially the gap in precipitation data can have negatively affected the model results at this site."

Please note that the gap filling script was actually designed to fill gaps up to 30 days and not 31; we have changed this detail in the new version of the manuscript (page 5, line 24). Precipitation was not gap-filled; missing points were set to zero instead. Finally, we found that Table A1.2 contained errors; we have corrected this in the new version of the manuscript (page 37).

Table a: Percentage of missing hourly model forcing records. In the middle column the mean over all input variables is shown along with the range over all input variables. In the right column the length of the simulation period is shown.

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

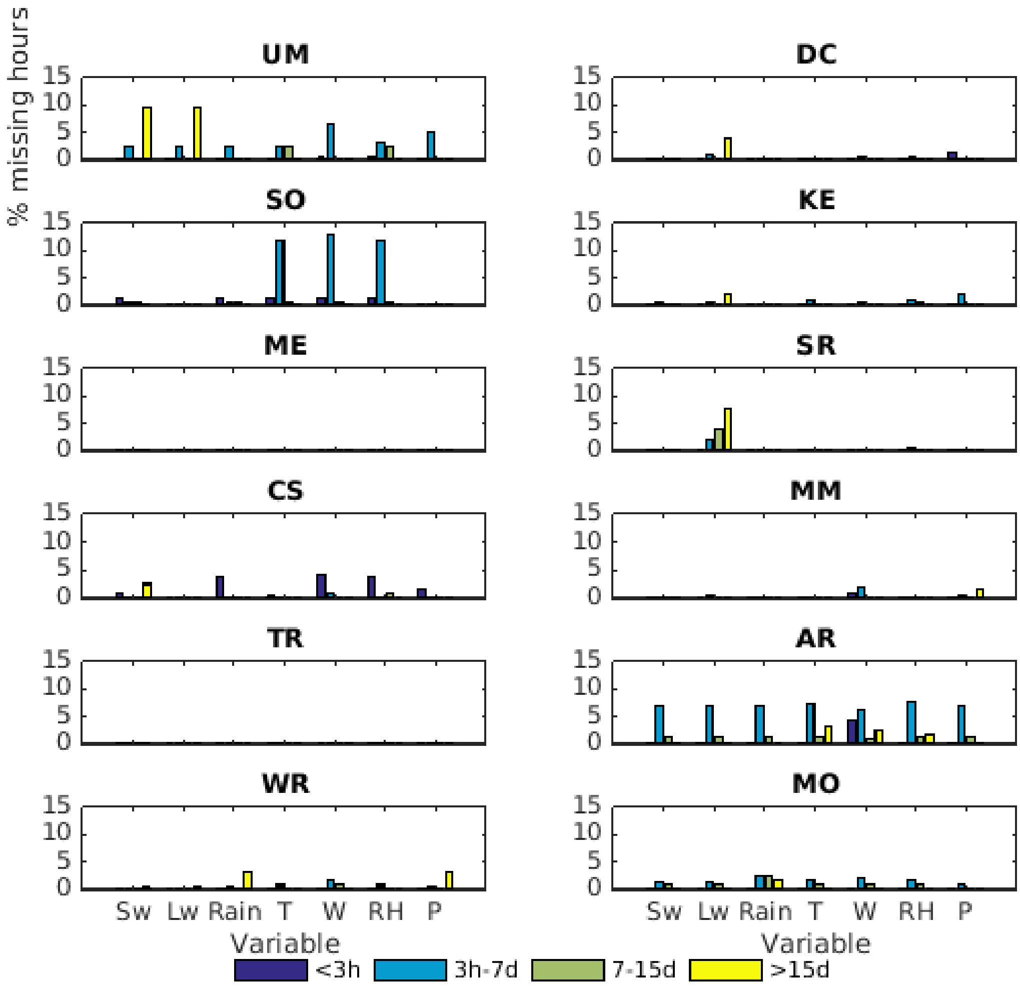


Figure a: Missing hours as percentage of the total hours in the calibration/validation periods. The statistics are shown per forcing variable. Sw is downward shortwave radiation, Lw is downward longwave radiation, Rain is precipitation, T is air temperature, W is wind speed, RH is relative humidity, and P is atmospheric pressure. Please note that for convenience the label Lw is also used for site DC; instead of net radiation. The four categories split the missing hours between gaps of less than 3 hours, of between 3 hours and 7 days, of between 7 days and 15 days, and gaps larger than 15 days.

COMMENT:

Page 7 line...: 8: Rather "Calibration Approaches". Where is the "two-objective calibration" in the methods section? 10-12: Sentence too long. Rephrase into at least two sentences. 12: Add reference (Shuttleworth et al . 2013) 14-15: This reads difficult. Either write saturated hydraulic conductivity instead of sathh or use the symbol introduced in Eq. 5. Same for all other symbols throughout the manuscript. 22-25: Shorter sentences. 27: This this.... rephrase

ANSWER:

Page 7 line 8: Title has been changed to “Calibration and validation approaches”. We have combined this section with the next section (“Validation approach”) and now introduce the “two-objective calibration” in this section.

Lines 10-12: The sentence has been rephrased to: “We chose to calibrate simulated neutron counts against CRNS observed neutron counts using COSMIC (Shuttleworth et al., 2013) to translate simulated soil moisture profiles into equivalent neutron counts.”

Please note that the reference to Shuttleworth et al. (2013) was already provided in Section 2.3 of the old manuscript (Section 2.2 in the new manuscript).

Lines 14-15: We have rephrased: "These included Mualem-Van Genuchten shape parameter b, the water entry pressure parameter (sathh), and the saturation hydraulic conductivity (K_{sat}). The critical point (θ_{crit}) and wilting point (θ_{wilt}) soil moisture content parameters from the evapotranspiration limitation factor were also calibrated."
Lines 22-25: Sentence has been split as follows: "They were constructed by computing the minimum and maximum parameter values for the entire soil texture triangle (based on Wösten PTF). Three organic matter contents were taken into consideration, yielding three triangles. Cay percentages above 70% were excluded to avoid extreme values for parameter b especially."
Line 27: The word "thus" has been removed: This yielded different critical soil moisture content ranges in terms of soil moisture content (m^3m^{-3}) for different sites.

COMMENT:
Page 12 line ...: 32: Remove "It should be noted however that".

ANSWER:
Page 12 line 32: This sentence has been changed to: "Especially for the saturation hydraulic conductivity, the parameter calibration range was non-linear, with a factor 1000 difference between the upper and lower boundaries."
This entire section has been moved to the newly created supplemental material, Supplement 3.

COMMENT:
Page 13 line...: 6: I suggest to put "and also containing stones" to the end of sentence or another place and remove brackets. 7: Here you write sathh out. Much better. However, the sentence may be moved to the discussion section. 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 13 line 6: Sentence has been changed to: "At the KE site, the soil was reported to be coarser than actually reported by the HWSD and to contain stones."
Line 7: We have renamed the Discussion section and rearranged sections to increase the readability. Therefore we could keep this sentence within the section it was before.
Line 13: This issue was also raised by Referee #3. We have decided to move the entire section to the supplemental material, Supplement 3. We refer to it from Appendix 2, which now contains what was previously Section 3.6 "To which parameters were soil moisture and latent heat flux most sensitive?" The reason is this section presents more a quality check of the calibration results than it is a contribution to answering our research question, as mentioned by Referee. We now discuss the most plausible reasons only. We quote the new text here:

"Several reasons have been proposed why apparently non-physically realistic parameter values were found, of which we discuss the most plausible ones here. Richards' Equation applicability at larger horizontal scales is questionable for multiple reasons (Beven and Germann, 2013). The foremost reason is hydraulic continuity, upon which Richards' Equation is based, does usually not occur in the field over distances of multiple metres. This is due to horizontal and vertical soil heterogeneity (e.g. macro-pores, soil layering in organic matter and particle sizes, stones) not (properly) accounted for in the model. The form of Richards' Equation implemented in JULES does not take these natural features into account and therefore effective parameters can be expected to differ from values representing a soil with a homogeneous matrix only. That means the process represented by the equation is not the same as the processes occurring in the field soil. Therefore the parameters' roles in JULES differ from their theoretically assumed roles, and so different effective parameter values can be expected. Explicitly taking into account heterogeneity issues like macro-pores could improve JULES' performance (Rahman and Rosolem, 2017), but doing so was beyond the scope of our study. The vertical discretisation of the soil (layers of 10, 25, 65, and 200 cm) may not be suitable for solving Richards' Equation. Layer thicknesses of no more than one centimetre near the surface are required to accurately simulate water fluxes. Insufficiently fine discretisation at both field (~eddy-covariance footprint) and hydrological catchment scale may not yield realistic water fluxes using realistic parameter values. (Smirnova et al., 1997; Lee and Abriola, 1999; Van Dam and Feddes, 2000; Downer and Ogden, 2004; Beven and Germann, 2013). Testing the effects of this issue on our results, which would entail changing the vertical discretisation of the soil column in JULES, was beyond the scope of our study. Finally, due to the simplifications of natural processes inherent to models, physically realistic parameters often provide unrealistic results (Gupta et al., 1998)."

Line 35: In the new version of the manuscript we have introduced the two-objective calibration in Chapter 2 Data and Methods instead.

COMMENT:

Page 14: line...: 5: comma before respectively. 7: was instead of "could have been" 7: "similar" - quantify or remove. 11: Suggestion: Move "to have obtained such automatic improvement" to the end of sentence 16: "improved soil moisture and latent heat flux". If calibration is done for both, both should be improved, no? 18: What does EF stand for? Too many abbreviations. Also WO: : :etc. There should be a way to distinguish places/sites from variables from acronyms. 19: "coincidence" – Not really, is it? 20: "Similar" - be precise. 21: "this" means what? context not totally clear. Rephrase. 27: "Generally quite weak" - Reference in literature? I cannot quite follow here. 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 5, line 7, line 11: Section 3.5 (now Section 3.4) has been completely rephrased. These errors were therefore removed.
Line 16: Please note that in two-objective calibration simultaneous improvement in both is not necessarily expected (Gupta et al., 1998).
Line 18: EF stands for evaporative fraction. Please note that in order to improve the structure of our manuscript we do not discuss evaporative fraction in this section any longer. We use evaporative fraction as a secondary validation metric and mention it only briefly by referring to the newly created Supplement 3 from Section 3.3. We have mentioned in Section 2.1 that the full site names are given in current Figure 1 (was Figure 2 in the original manuscript). Writing out the site names all the time would increase the size of the manuscript considerably and would necessarily make labels in the figures unreadably small.
Line 19: Due to the complete rephrasing of the manuscript this is now discussed in the following way, which we hope clarifies:
"Scatterplots of normalised RMSE of latent heat flux (LE) and normalised RMSE of soil moisture for all sites are shown in Figure 9 and Figure 10 for the PS and CRNS two-objective calibration strategies

respectively. The RMSE values were normalised with respect to the default solutions, which therefore have normalised RMSE values of 1 (-). The black dots represent individual model runs and the default model run for each site is represented by a red cross The single-objective calibration solution of each single-objective calibration solution is shown as a blue triangle and the compromise solution of each two-objective calibration is shown as a green triangle. These two figures showed that the differences between the compromise solutions of the two calibration strategies observed for SR, DC, and UM in Figure 8, were less meaningful than analysis of Figure 8 only showed.

These findings are based on the shapes of the black point clouds in Figure 9 and Figure 10. We first look at the left edges of the black point clouds. When these edges are close to vertical, a small deterioration in RMSE-SM or RMSE-N (e.g. less than 0.05 (-)) would yield a large deterioration in normalised RMSE-LE (e.g. greater than 0.2 (-)). We observed this for three calibrations with PS soil moisture (SO, AR, and WR; indicated with pink line for WR) and in all these three cases, a less than 0.05 (-) change in normalised RMSE-SM would have yielded worse simulated latent heat flux than with the default parameter set. Seven calibrations with CRNS neutron counts (UM, DC, SO, KE, SR, CS, and WR; indicated with pink line for WR) showed a close to vertical edge. In four cases (UM, DC, KE, and CS) this would have yielded worse simulated latent heat flux than with the default parameter set. A negative slope for this side of the point cloud means an improvement in soil moisture estimation would mean a deterioration in latent heat flux estimation. We observed such negative slope for four calibrations with PS soil moisture (CS, MM, TR, and MO; indicated with continuous black line for MO) and for two calibrations with CRNS neutron counts."

Line 20: Please see our answer to the comment on line 19 for specification.
Line 21: Please see our answer to the comment on line 19. We rephrased this paragraph completely to better clarify.
Line 27 and line 28: Due to the rephrasing of the entire section we have removed the discussion on sites where strong coupling is expected. We did refer to sites in climatic transition zones, where soil moisture provides a first order control on evapotranspiration (Seneviratne et al. 2010). We here cite sentences summarising the findings from Section 3.4 of the revised manuscript:
"In summary, the two-objective calibrations against soil moisture (or neutron counts) and latent heat flux showed fewer substantial differences between calibration with PS soil moisture and calibration with CRNS neutron counts. These results indicate that the differences between both *single*-objective calibration strategies, which showed an advantage for CRNS observations, were possibly not as substantial as it seemed at first."

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. 15: "Weak coupling..." Can the forcing data be the reason here? I disagree, see comments to Figure. 16: "CRNS-N/LE" - rephrase and make clear. 20-21: Suggestions: Split sentence into two. 28: "root zone soil moisture" is where exactly and calculated how? 32: split sentence at "while" and remove "while". Also at other instances in the manuscript, this makes sence.

ANSWER:
Page 15 line 4: Latent heat flux and evaporative fraction performance were computed over day time values (Page 8 line 16 of the original manuscript) only in all analyses except when we plotted monthly mean diurnal cycles in Figure 9 of the original manuscript. The only instance when night time values were used was to plot the monthly mean diurnal cycles shown in Figure 9 (of the original manuscript). We have made this clearer by adding the following remark: "(night time data included

but not used for calibration as discussed previously)” to Section 3.3, page 12, lines 29-30 of the revised manuscript.
Line 15: Please note that this section was removed from the revised manuscript. Our discussions and conclusions changed and therefore the mentioning of the “weak coupling” was removed.
Line 16: We now refer to the two different two-objective calibration approaches as: “calibrations with PS soil moisture” and “calibrations with CRNS soil moisture”, within the context that makes clear we are talking about the two-objective calibrations.
Line 20-21: We have rephrased the sentence and moved it to Section 3.3 of the revised manuscript: “Another issue, which occurred for instance for the PS calibration at SO and for the CRNS calibrations at site AR, was that while surface energy flux estimation improved for a certain period, it deteriorated for another period (data not shown).”
Please note we removed the mentioning of site SR in this sentence because we now use latent heat flux as principle objective function. Latent heat flux yielded different improvement than evaporative fraction in this case.
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.
We now discuss the implications of self-adjusting behaviour of the wilting point and critical point soil moisture in Section 3.3.
Line 32: In our effort to reduce non-essential discussion in the manuscript we removed this sentence.

COMMENT:
Page 16 line ...: 4: Again the relatively weak coupling is discussed. Avoid repetition. Restructure. 5: "this suggests that how" - rephrase 9: Parameters should not move up or down after calibration. Variables and model states may be variable. "move closer" to what? 10: Which implications? Please name them and argue why. 14: "did not differ substantially". I disagree. They did differ. 17: "could be used"... you can put "should" instead. "we might have found worse fits". This is speculations. Please give reason or remove speculations. One could state: “You might have found better fits.” 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 4: We agree with the reviewer that there are too many repetitions. We have restructured our manuscript to remove these. The discussion of the “weak coupling” was removed after reassessing our discussions and conclusions.
Line 5: We rephrased lines 4-6 and moved them to Section 3.3, page 12, lines 4-8**: “**One factor causing some of these limited improvements was that, when mean simulated root zone soil moisture (weighted with root density) increased after calibration, the values of the wilting point and critical point soil moisture parameters moved along (data not shown), yielding similar soil moisture stress. This happened for both calibration strategies at site KE, and for the calibration against CRNS neutron counts at sites MO and TR. This could relate to the limited value of simulated absolute soil

moisture for surface energy flux estimation in land surface models (Dirmeyer et al., 2000; Koster et al., 2009).**"**
Line 9: We have rephrased and split the sentence. We have moved this sentence to Section 3.3: "However, we also found the space between wilting point values and critical point values to decrease after calibration. This occurred when the standard deviation of the simulated soil moisture time series decreased due to calibration."
Line 10: We have removed this from the manuscript because it was non-essential information.
Line 14: We think there was no substantial difference between calibrating against PS and against CRNS data with respect to latent heat flux. We completely revised our discussion of the two-objective calibrations in Section 3.5 of the original manuscript (now Section 3.4) to make this clearer. We now start by showing differences between the compromise solutions of both two-objective calibration strategies were limited (Figure 12 of the original manuscript, now Figure 8). We then discuss the left edges of the point clouds in Figure 10 and Figure 11 of the original manuscript (now Figure 9 and Figure 10) and the lower edges of these point clouds. Based on the differences found for the single-objective calibrations and for the compromise solutions of the two-objective calibrations, we summarise in Section 3.4:
"In summary, the two-objective calibrations against soil moisture (or neutron counts) and latent heat flux showed fewer substantial differences between calibration with PS soil moisture and calibration with CRNS neutron counts. These results indicate that the differences between both *single*-objective calibration strategies, which showed an advantage for CRNS observations, were possibly not as substantial as it seemed at first."

We also explain:
"The spatio-temporal stability theory, which implies limited spatial variability in surface energy fluxes, could be one explanation for this (Vachaud et al., 1985; Teuling et al., 2006; Mittelbach and Seneviratne, 2012; Albertson and Montaldo, 2003). Another factor that has possibly played a role is that the spatial scale advantage of the CRNS was masked out by the possibly lower quality of its measurements (see also Section 3.1). Different hydrogen pools than soil moisture affect the neutron measurements at various temporal resolutions (e.g. rainfall interception; Baroni and Oswald, 2015). However, PS sensors also have their limitations. Different (electromagnetic) PS sensors have different designs and properties, which affect the data quality (Robinson et al., 2008; Blonquist et al., 2005). Soil type and soil specific calibration also affect the accuracy and precision of the PS data. For instance, the relationship between electrical permittivity and soil moisture content is strong for quartz rich soils, but less accurate for clay soils (e.g. Ishida et al., 2000; Robinson et al., 2008)."

Line 17: We have removed lines 16-18 because they are a repetition of page 11, lines 6-12. Those lines include the requested explanation:
"The RMSE values reduced relatively more for the CRNS calibration (70% on average over the twelve sites) than for the PS calibration (55% on average). The calibration method could possibly explain this. CRNS calibration was against observed neutron counts, while PS calibration was against observed soil moisture contents. Because neutron counts have an inverse relationship with soil moisture content, the PS calibration was possibly governed by avoiding larger errors occurring during a few brief soil moisture peaks. While focussing on getting the fitting for those peaks right, the PS calibration neglected the smaller errors during dry periods. This would then result in relatively smaller decrease in RMSE values than for the CRNS calibrations because those were fitted with heavier weights to the drier periods."
Line 21: We have rephrased this because our discussions and conclusions have changed. We refer to our answer to the comment on page 16, line 14. This sentence was rephrased and moved to the end of Section 3.3 of the revised manuscript:
"Previous research has also indicated that soil moisture alone is insufficient to estimate soil hydraulic parameters (Vereecken et al., 2008; 2015)."

COMMENT:
Table 2: Add symbols from the equations and use symbols within the text.

ANSWER:
Table 2: Equation symbols are now used in Table 2 and in the text body of the manuscript.

COMMENT:
Figure 1: I suggest removing the figure. The scale mismatch can be pointed out in a single sentence to save space for result figures e.g. The model grid cell size is 1km, the EC footprint is between 100m2 to 1km2 (put here diameter instead), the CRNS footprint is 300m in diameter, the PS soil moisture sensor measures few dm3.

ANSWER:
Figure 1: This figure has been removed. We think the scale mismatch is clear without further textual explanation. It is clear from page 2 lines 17-21 and page 3 lines 18-20 of the revised manuscript.

COMMENT:
Figure 2: This Figure is very informative and necessary. Put Figure 2a and 2b instead of "upper".

ANSWER:
Figure 2: Labels 2a and 2b have been included in the figure and we have referred to these as such in the figure caption.

COMMENT:
Figure 3: Informative, you may want to keep it, however consider removing.

ANSWER:
Figure 3: This figure has been moved to the newly created Supplement 1.

COMMENT:
Figure 4: Here, MO looks like it has a big discrepancy / bias in PS to CRNS soil moisture. Why is this the case / is this in the manuscript? Put the legend (PS, CRNS) beside or below the figure, not in the first subplot. Minimum of Y axis is missing.

ANSWER:
Figure 4: The clear difference between observed PS soil moisture and CRNS soil moisture retrieval shown in Figure 4 was discussed on page 9 line 4-5. The sites were put in order from small to large difference between the two soil moisture products in Figure 4 and in Figure 5. We discussed reasons for the differences between the two soil moisture products (observations/retrievals) in Section 3.1.2 and Section 3.1.3. We were unfortunately not able to identify why the two soil moisture products differed. We were not able to identify a specific reason for the large difference between the two soil moisture products at the Mozark (MO) site.
We have placed the legend below the figure and added minimum values at the vertical axes of the soil moisture plots. We have mentioned the abbreviation “SM” in the figure caption. We have also added the following sentence to the caption: “Daily precipitation is also shown here for each site.”

COMMENT:

Figure 5: Informative. Keep it. CRNS observations are neutron flux. Soil moisture is retrieved from CRNS neutron flux observations. It should be rather something like CRNS soil moisture retrieval than CRNS soil moisture observation. Again, SM is abbreviated in your figure as SM, but not within the paper or at few instances. In general, abbreviations should be reduced.

ANSWER:

Figure 5: We respectfully disagree with the reviewer that CRNS soil moisture is less an observation than PS soil moisture. CRNS soil moisture values are obtained using a relationship (in our case COSMIC) with neutron counts, just like PS soil moisture values are obtained from a relationship with for instance electromagnetic wavelength. We will therefore keep the terminology as is. We have mentioned the abbreviation “SM” in the figure caption.

COMMENT:

Figure 6: What is the added value to Figure 5, because JULES is calibrated based on the data used in Figure 5. I suggest to merge Figures into one with e.g. Figure 5a and 5b.

ANSWER:

Figure 6: The authors would like to argue that there is added value of Figure 6 to Figure 5 (numbering in the original manuscript). Figure 5 shows how similar or different the two soil moisture products are at the different sites. Figure 6 shows how successful calibration was, i.e. how much the Root Mean Squared Error between simulated and observed/retrieved soil moisture was reduced by calibration. Calibration was not as successful in all instances. The relative improvement in Root Mean Squared Error differed between calibration against PS soil moisture and CRNS soil moisture retrieval and also differed between sites. We have decided not to combine Figure 6 with Figure 5 because we think this would increase confusion between their respective purposes (data analyses, versus effect of calibration). Moreover, they are discussed in two separate section.

COMMENT:

Figure 7:Add grid lines. 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. Also you plot a 3 year time series with hourly values. It will be much easier to read, with more information and will have even more meaning if you average over days or months. So far, it is a lot of variability, which is clear beforehand if hourly values are plotted over the course of 3 years. You may consider merging Figure 7 and Figure 4.

ANSWER:

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 b of this answer combines both plots to highlight this fact (notice default run line is omitted as it is irrelevant to this comparison). We decided to keep hourly values because they represent the soil moisture dynamics, which are important with respect to surface energy flux simulation (e.g. Teuling et al., 2009). We have added grid lines.

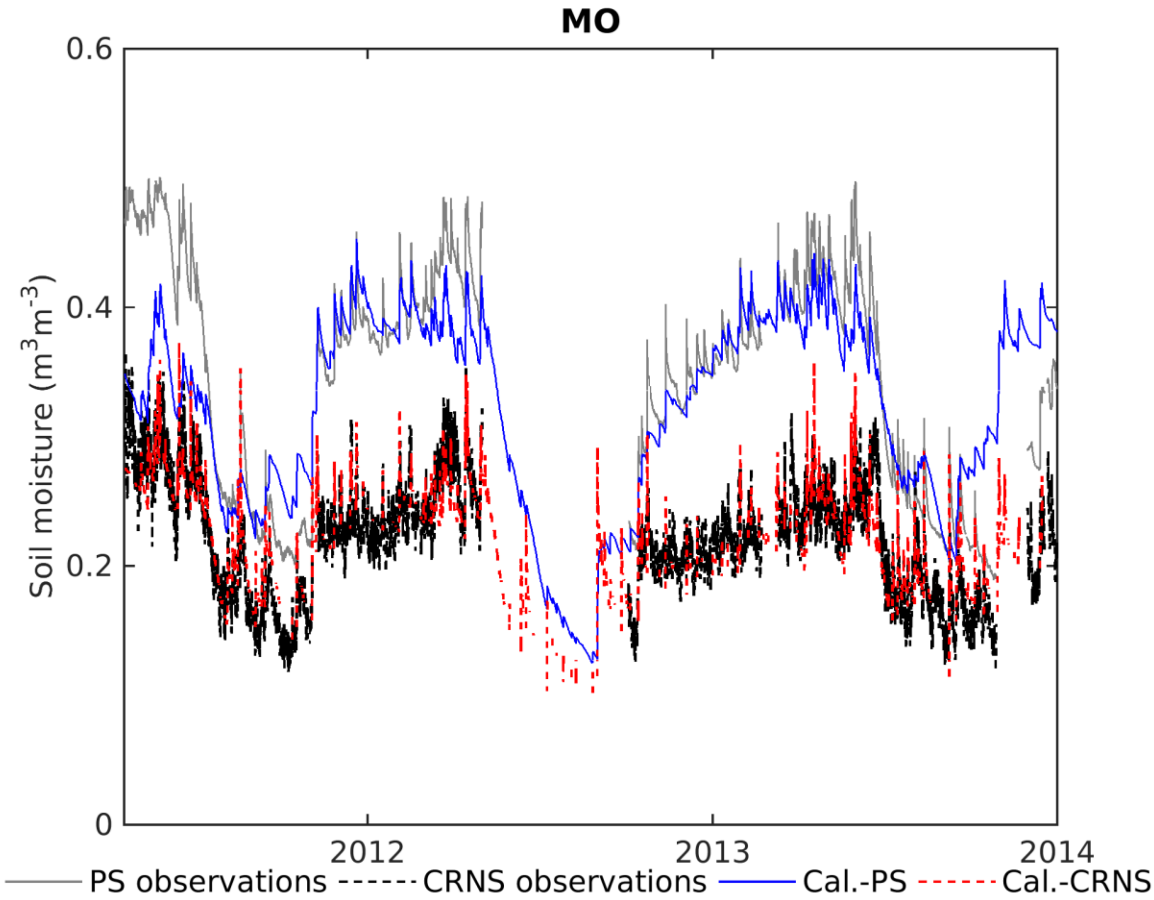


Figure b: Hourly observed (PS and CRNS) soil moisture time series and simulated soil moisture time series after calibration against PS and CRNS soil moisture for site MO.

COMMENT:

Figure 8: The indices a) and b) are missing in the Figure. Axis labels at wrong place. Consider merging with Figure 10,11,12. The data may also fit into one table.

ANSWER:

Figure 8: We have removed Figure 8a in our effort to remove the number of figures. We decided to keep Figure 8b (now Figure 6) as is rather than as a table because we think it is easier to understand the results from a figure. Moreover, we think it is clearer to keep the figures from the single-objective calibrations separate from those of the two-objective calibrations. Please note we now discuss the RMSE of latent heat flux instead of RMSE of evaporative fraction in the manuscript. The new Figure 6 therefore shows the percentual change in RMSE of latent heat flux vs. the percentual change in RMSE of soil moisture

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 reviewer is right that WR and MO show a strong change from default to single-objective calibrated. We therefore changed our conclusion with respect to the soil moisture – evapotranspiration coupling as mentioned in earlier answers.

COMMENT:
Figure 13: Labels and legend are overlapping.

ANSWER:
Figure 13: Has been removed in our effort to reduce the number of figures.

Answer to interactive comment by Referee #2 Dr. Teuling

The authors thank Dr. Teuling for providing comments and suggestions, which will help us to improve our manuscript. Comments are addressed individually.

1. General comments

COMMENT:
The manuscript by Iwema et al. addresses the use of new Cosmic-Ray Neutron Sensor (CRNS) data in land surface model calibration. By using CRNS and in situ sensor data from 12 sites across the U.S., the authors systematically investigate to which degree calibration against this novel data leads to an improvement in the simulation of observed surface fluxes. Unfortunately the results do not show a clear improvement of using CRNS data over in situ data, but this does not in any way affect the quality of the research. Overall, I have a positive impression of the manuscript which I believe would make a good contribution to HESS, but it will need improvements on several aspects. In particular, some of the main conclusions do not seem to follow from the results (a problem also identified by the other referee), the authors should make clear that the in situ data is not used to its full potential so the study is not a clean comparison between datasets but rather a comparison of scale, and the authors seem to have been somewhat selective in the selection of references to identify the knowledge gap. These issues are discussed in more detail below. However I believe these comments can be addressed by minor (mostly textual) revisions.

ANSWER:
We thank the Referee for his effort and are happy to hear he acknowledges the relevance of our study to the HESS community. We have revised the conclusions, so they do follow from the results, as also provided in the answers to the other two referees, and detailed below. We agree the point scale soil moisture sensor data was not used to its' full potential and have made this clearer in the Results and Discussion section. The literature review on the knowledge gap has been improved as suggested by the Referee.

COMMENT:
A first remark concerns the title, which does not seem to reflect the contents of the manuscript. The study really looks at calibration using soil moisture observations at different scales, and it does not look at measurement scales as such. I suggest a title along the lines of "Land surface model performance using cosmic-ray and in situ soil moisture data for calibration". Also, the use of "reducing" is seemingly at odds with the results, which does not show an improvement when using observations at the "model" scale.

ANSWER:
We thank the Referee for helping to improve our title. We have changed the title to:
'Land surface model performance using cosmic-ray and point scale soil moisture measurements for calibration'.

COMMENT:
As was also pointed out by the other referee, the conclusion that JULES has a weak coupling between soil moisture and ET does not follow from the results. In fact, JULES has a strong coupling by definition, and the coupling in reality can only be less strong. The fact that ET estimates do not

improve with improvements in soil moisture is likely because ET is not so sensitive to changes in soil parameters, even for different climate conditions. This is for instance shown in a paper I wrote in 2009 (Teuling et al., 2009), in which I investigated effects of soil parameters on soil moisture and ET. In short, effects of soil parameters on soil moisture are generally large, but effects on ET are small. This is primarily caused by the main effect of soil parameters which is a shift in the mean, rather than the dynamics (this is also consistent with the strong contribution of bias to MSD as reported by the authors). This should be discussed better.

ANSWER:
We thank the Referee for providing this comment. We have changed the conclusions by addressing that the effect of soil parameters on absolute soil moisture is large, but the effects on evapotranspiration are small, as also shown by Teuling et al. (2009). We incorporated this in the Conclusions section:
"… revealed that differences between calibrations with these two soil moisture observation methods did not yield substantially different surface energy flux estimations. These outcomes did not provide sufficient evidence to reject our null-hypothesis "*reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations*". … Another factor, related to this and that contributed to this result, was that calibrating soil parameters had mostly an effect on absolute soil moisture values rather than soil moisture dynamics. Soil moisture dynamics have a greater effect on surface energy flux simulation in land surface models than absolute soil moisture values do."

COMMENT:
Related to this is the question why rooting depth was not optimised along with the soil parameters. If the model rooting depth does not reflect the actual root profile, optimization of soil parameters along will not lead to a better estimation of available soil water. These choices should be better motivated.

ANSWER:
We used site specific data in order to prescribe rooting depth for the analysed sites (as mentioned in original manuscript Page 7, line 32 – Page 8, line 2). The decision was made to avoid prescribing more than one rooting depth parameter for individual sites given that rooting depth in JULES is defined per Plant Functional Type. We recognise this is a limitation of our study.

COMMENT:
In making a case for the validity of their research question, the authors miss out on an important body of literature on spatio-temporal characteristics of soil moisture fields. It has been shown by numerous studies that while soil moisture generally shows a large spatial variability, individual points maintain their rank while the mean changes. This insight started with the "classical" Vachaud et al. study, but numerous other studies (for instance Teuling et al., 2006, Mittelbach and Seneviratne, 2012, among others) have reported similar behaviour. This behaviour implies a relatively small spatial variability of fluxes, which was explained from a theoretical perspective by Albertson and Montaldo (2003) and explored Teuling and Troch (2005), among others. In effect, based on these studies, the hypothesis could also be formulated more neutral in the form of a null-hypothesis: "A reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations..". The results could subsequently be interpreted as insufficient evidence to reject this hypothesis. In any case the introduction should be changed to include a discussion on the relation between point-scale and large-scale soil moisture. This can then also be used to explain the reference to Franz et al. (2012), which now conflicts with the hypothesis (if CRNS and in-situ soil moisture have already been reported to be similar, how can the authors still expect to find a difference in fluxes?).

These comments do not make the research any less relevant, but the phrasing of the hypothesis should be in line with the "state-of-the-art", and not a convenient selection thereof.

ANSWER:
We thank the Referee for this comment. We have included a literature review on spatio-temporal characteristics of soil moisture fields and the effects of these characteristics on evapotranspiration simulation in the Introduction:
"On the other hand, past research has shown soil moisture measured at only one or a few points within an area of similar size as an eddy-covariance footprint, can have similar value to surface energy flux simulation as soil moisture measured at a larger scale (e.g. Vachaud et al., 1985; Teuling et al., 2006; Mittelbach and Seneviratne, 2012). These studies showed that points within a soil moisture observation network keep their rank with respect to the mean soil moisture (anomaly), i.e. they either under –or overestimate the mean (anomaly), so-called spatio-temporal stability. The physical principle behind the spatio-temporal stability theory is that different time variant and hydrological processes either create or destroy spatial soil moisture variability whereas time invariant land surface characteristics induce an effective offset in the spatial variability (Albertson and Montaldo, 2003; Teuling and Troch, 2005; Vanderlinden et al., 2012). When soil moisture reaches values below the critical point (i.e. transpiration becomes moisture limited), spatial variability in soil moisture and fluxes is reduced (Teuling and Troch, 2005). Soil moisture dynamics was found to be a small portion of total soil moisture variability (Mittelbach and Seneviratne, 2012), while having a greater effect on surface energy fluxes than absolute soil moisture in land surface models (Dirmeyer et al., 1999; Teuling et al., 2009). The implication of the spatio-temporal stability theory therefore is that the spatial scale mismatch issue might have limited implications to surface energy flux simulation."

We have, based on the suggested literature, changed our hypothesis to the more neutral null-hypothesis and moved it to immediately after the research question:
"Based on the spatio-temporal stability theory, we phrased the following hypothesis for our research question: *reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations.*"
In the Results and Discussion section, subsection 3.4 "Two-objective calibration against soil moisture and latent heat flux" of the revised manuscript we discuss the implications of this knowledge on our results:
"In summary, the two-objective calibrations against soil moisture (or neutron counts) and latent heat flux showed fewer substantial differences between calibration with PS soil moisture and calibration with CRNS neutron counts. These results indicate that the differences between both *single*-objective calibration strategies, which showed an advantage for CRNS observations, were possibly not as substantial as it seemed at first. The spatio-temporal stability theory, which implies limited spatial variability in surface energy fluxes, could be one explanation for this (Vachaud et al., 1985; Teuling et al., 2006; Mittelbach and Seneviratne, 2012; Albertson and Montaldo, 2003)."

In the Conclusions we come back to this. Based on our results we conclude we found insufficient evidence to reject this null-hypothesis.
"These outcomes did not provide sufficient evidence to reject our null-hypothesis "*reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations*". The spatio-temporal stability theory in soil moisture, on which our hypothesis was based, can possibly explain the limited differences in surface energy flux estimation. This theory implies that spatial variability in surface fluxes is relatively limited within an eddy-covariance tower footprint. Therefore simulated surface energy fluxes after calibration against point scale soil moisture data would not necessarily be better than simulated surface energy fluxes after calibrations against CRNS soil moisture data. "

In addition, please notice that Franz et al. (2012) compared a network of eighteen point-scale soil moisture sensor profiles within a single CRNS footprint. Their goal was to validate the cosmic-ray neutron sensor soil moisture against such network, whereas we assessed differences between a single or a few Point Scale profile(s) and the cosmic-ray neutron sensor, which has a larger footprint. At five sites, the data available to us represented the average of multiple profiles (e.g. six at Santa Rita; Table A1.1). Due to a comment by Referee #1, we changed page 3, lines 15-16 to (page 3, lines 24-25 of the revised manuscript): "Franz et al. (2012) showed soil moisture estimated from CRNS neutron counts differed less than 20% from the average of a co-located point scale soil moisture sensor network at a site in Arizona."

COMMENT:

In applying CRNS and PS soil moisture products, the authors only consider the shallower PS soil moisture observation in order to comply with the CRNS observation depth. While this makes sense given the goal of the study (scale mismatch and not a comparison of observation techniques), it is not sufficiently recognized that this introduces an unfair disadvantage to the PS data. There might be important information in deeper PS observations, in particular during soil moisture-limiting conditions, that is not being considered in this study. It is thus crucial to make a clear distinction between the scale aspect and the observation technique, and acknowledge in the discussion in the discussion that PS soil moisture might give better results when all available observations are used. A better alternative would be to redo the analysis using all observation depths in addition to using only the shallow observations, but this would likely require a substantial amount of work.

ANSWER:

We thank the reviewer for this comment. We agree that not using the deeper soil moisture observations can have introduced disadvantages to the point-scale measurement in the comparison. Notice, however, that most Ameriflux sites provided only publicly available data from shallow sensors with exception of WR (down to 50 cm) and MO (down to 100 cm), respectively. Understanding the impact of deeper soil moisture dynamics is indeed very interesting to be investigated but it is beyond the original scope of our study which focusses on understanding more directly the impact of horizontal footprint. At the end of the Introduction, we included the following sentence:

"We emphasise that we compared the two different soil moisture measurement techniques' scales and not the techniques as such."

We have included a statement in Section 2.1 "Calibration and Validation data: PS, CRNS, and eddy-covariance data" of the revised manuscript:

"We used point scale soil moisture data from the soil layers up to 30 cm depth only for consistency among all sites. Our main objective was to investigate the difference in information content due to two soil moisture measurement techniques' different horizontal scales in relation to the eddy-covariance footprint, rather than to compare the measurement techniques themselves."

In the Results and Discussion, subsection 3.5 of the revised manuscript, we mention:

". In this study we used PS sensor data in the upper 30 cm of the soil only, even though data from deeper sensors was publicly available at two sites (WR and MO). This choice can have provided a disadvantage to the PS data in our comparison, especially at sites with deeper roots and during soil moisture limiting conditions in the upper soil. Investigating the role of deeper roots was however beyond the scope of this study. Our goal was to compare the effects of the difference in horizontal footprint on latent heat flux simulation after using the two different measurement techniques to calibrate model parameters."

2. Detailed comments

COMMENT:

Page 1, Line 27: it is not the coupling between soil moisture and flux partitioning that is strong (this is similar across climates), but the soil moisture control on temperature at seasonal timescales.

ANSWER:

We have changed the phrasing as suggested.

COMMENT:

Page 3, Line 15: If the CRNS data is known to give similar values as the in situ data, how can this be consistent with the hypothesis posed later on? Please change the line of reasoning, but also take into account my previous comments on missing references on soil moisture temporal stability.

ANSWER:

Please also see our reply to the comment about Franz et al. (2012) in the 'general comments'. Please notice that Franz et al. (2012) compared a network of eighteen point-scale soil moisture sensor profiles within a single CRNS footprint. Their goal was to validate the cosmic-ray neutron sensor soil moisture against such network, whereas we assessed differences between a single or a few Point Scale profile(s) and the cosmic-ray neutron sensor, which has a larger footprint. At five sites, the data available to us represented the average of multiple profiles (e.g. six at Santa Rita; Table A1.1). Due to a comment by Referee #1, we changed page 3, lines 15-16 to (page 3, lines 24-25 of the revised manuscript): "Franz et al. (2012) showed soil moisture estimated from CRNS neutron counts differed less than 20% from the average of a co-located point scale soil moisture sensor network at a site in Arizona."

COMMENT:

Page 5, Line 28: Richards equation -> Richards' equation

ANSWER:

We have changed as suggested.

COMMENT:

Page 13, Line 13: "We could think of..." ->All arguments are backed up by references, so it might be better to say "Several reasons have been proposed..."

ANSWER:

We have rephrased as suggested.

COMMENT:

Figures: Generally ok, but I agree with the comments of the other referee that the number can be significantly reduced without sacrificing the main message.

ANSWER:

We have reduced the number of figures in line with the suggestions made by Anonymous Referee #1. Figure 1 was removed, Figure 3 was moved to the newly created Supplement 1. Figure 8a was removed (8b was kept but replaced with a figure showing the change in RMSE of latent heat flux vs. change in RMSE of soil moisture and is now Figure 6). The figure showing the evaporative fraction has been moved to Supplement 2. Figure 13 was removed. Figure A1.1 was moved to Supplement 2 and Figure A2.3 was moved to Supplement 3.

Answer to interactive comment by Anonymous Referee #3

We would like to thank the Referee for carefully reviewing the manuscript. We thank the Referee for his comments on the structure of our manuscript, specification and quantifications, and the English language. We have addressed these issues in the revised version of the manuscript. Below, each comment is addressed individually.

GENERAL COMMENTS

COMMENT:

The manuscript is quite well written and clear, even though some parts should be reduced and summarized. The topic is surely of interest for the HESS readership as cosmic-ray probes represent a new technology for ground measuring soil moisture over large areas. Therefore, we need to assess the impact of this new technology for improving land surface modelling. The paper describes the calibration of JULES land surface model with PS and CRNS at different sites in US. The manuscript is well conceived and applied over a large number of sites thus obtaining reliable and robust results. However, I mostly agree with the comments of previous reviewers, and particularly I believe that several aspects should be improved/changed before the publication. I reported below a list of the general comments to be addressed with also the specification of their relevance.

ANSWER:
We thank the referee for his positive comments regarding the relevance of our work to the HESS community.

COMMENT:
1) MAJOR: I found some of the explanations/justification of the results given in the paper quite weak. They appear to me as speculations, not supported by the performed analyses and results. For instance, I refer to:
(A) The comparison between PS and CRNS soil moisture data (section 3.1.3) shows that soil moisture timeseries are quite similar. The authors expected better performances at homogeneous sites but it was not the case. As shown in "temporal stability" papers (see also Teuling report), PS measurements are usually very well correlated with large scale measurements. Therefore, I expect good correlations. In my opinion, the good (bad) performances are mostly related to the good (bad) quality of soil moisture observations that may be affected by a number of factors (e.g., soil texture, sensor malfunctioning, ...). Therefore, theoretically I could expect that CRNS are better than PS measurements, but due to measurement uncertainties and errors, the larger support scale of CRNS is masked out by the (likely) lower quality of their measurements.
This important aspect, i.e., the quality of soil moisture observations, should be carefully addressed in the paper.

ANSWER:
We thank the Referee for this valuable comment. As also suggested by Referee #2, we have included a literature review on spatio-temporal characteristics of soil moisture fields and the effects of these characteristics on evapotranspiration simulation in the Introduction. The text in the Introduction has been rearranged to accommodate these changes.
In the Results and Discussion section we discuss the implications of this knowledge on our results and we address this issue again in the Conclusions.
We have, based on the suggested literature, changed our hypothesis to the more neutral null-hypothesis (as suggested by Referee #2) and moved it to immediately after the research question:

“Based on the spatio-temporal stability theory, we phrased the following hypothesis for our research question: *reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations.*”
Based on our results we conclude we found insufficient evidence to reject this null-hypothesis.
We have included a discussion on the impact of the quality of observations on the soil moisture signal in the revised version of the manuscript. In Section 3.4 of the revised manuscript:
“Another factor that has possibly played a role is that the spatial scale advantage of the CRNS was masked out by the possibly lower quality of its measurements (see also Section 3.1). Different hydrogen pools than soil moisture affect the neutron measurements at various temporal resolutions (e.g. rainfall interception; Baroni and Oswald, 2015). However, PS sensors also have their limitations. Different (electromagnetic) PS sensors have different designs and properties, which affect the data quality (Robinson et al., 2008; Blonquist et al., 2005). Soil type and soil specific calibration also affect the accuracy and precision of the PS data. For instance, the relationship between electrical permittivity and soil moisture content is strong for quartz rich soils, but less accurate for clay soils (e.g. Ishida et al., 2000; Robinson et al., 2008).”

COMMENT:
(B) The authors attributed the low differences in estimating surface energy fluxes when PS and CRNS measurements are considered to the weak coupling in JULES between soil moisture and evapotranspiration. Actually, I do not believe it is the case, but it should not be a speculation. It should be tested. If the authors want to give this message, they should demonstrate that with a different model or land surface scheme the differences are higher, and likely CRNS is better that PS data (as expected at the beginning). Therefore, I suggest changing the conclusions or, better, implementing an additional LSM and demonstrate the results through a scientifically sound approach.

ANSWER:
We thank the Referee for this comment. This is a point that was also raised by the other two referees. 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 the comments by Reviewer #2 that the results obtained in our study are a consequence of calibrating soil parameters, rather than the soil moisture – evapotranspiration coupling in JULES. Other factors yielding limited effects were the implications of spatio-temporal stability theory, data quality properties of the soil moisture observation techniques, and the self-adjusting behaviour of the wilting point and critical point soil moisture parameters after calibrations. We here cite the main part of our Conclusions section:

“…revealed that differences between calibrations with these two soil moisture observation methods did not yield substantially different surface energy flux estimations. These outcomes did overall not provide sufficient evidence to reject our null-hypothesis “*reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations*”. The spatio-temporal stability theory in soil moisture, on which our hypothesis was based, can possibly explain the limited differences in surface energy flux estimation. This theory implies that spatial variability in surface fluxes is relatively limited within an eddy-covariance tower footprint. Therefore simulated surface energy fluxes after calibration against point scale soil moisture data would not necessarily be better than simulated surface energy fluxes after calibrations against CRNS soil moisture data. Another factor, related to this and that contributed to this result, was that calibrating soil parameters had mostly an effect on absolute soil moisture values rather than soil moisture dynamics. Soil moisture dynamics have a greater effect on surface energy flux simulation in land surface models than absolute soil moisture values do. Related to this we observed that after calibration the wilting point and critical point soil moisture parameter values adjusted themselves in a similar way as the root zone soil moisture did, yielding similar soil moisture control on transpiration despite changes in soil moisture values. In other cases we found calibration against soil moisture to improve surface energy fluxes during certain periods, but to deteriorate surface energy fluxes during other periods, yielding similar overall performance. Yet in other cases evapotranspiration was not limited by soil moisture stress. The potential scale advantage of the Cosmic-Ray Neutron Sensor was possibly masked out by the possibly lower measurement quality of this sensor because other hydrogen pools than soil moisture affect the neutron count observations. Future use of CRNS soil moisture data could however benefit from improved knowledge on the effects of additional hydrogen pools (e.g. Baroni and Oswald, 2015) and of the sensor footprint (Köhli et al., 2015). In this study, our results are conditioned to a single land surface model (JULES). For additional understanding of the importance of both PS and CRNS measurements for simulated surface energy quantities can be extended to other models in the future."

COMMENT:
(C) The range of reasons reported at page 13, lines 13-34 are only speculations. I suggest removing.

ANSWER:
We thank the referee for his/her comments. Note, this issue was also raised by Referee #1. We have decided to move the entire section to the newly created Supplement 3. We refer to it from Appendix 2, which now contains the text from what was previously Section 3.6. The reason is this section presents more a quality check of the calibration results than it is a contribution to answering our research question, as mentioned by Referee #1. We now discuss the most plausible reasons only. We would like to note that speculation is allowed when discussing (parameter value) results. We quote the new text here:

"Several reasons have been proposed why apparently non-physically realistic parameter values were found, of which we discuss the most plausible ones here. Richards' Equation applicability at larger horizontal scales is questionable for multiple reasons (Beven and Germann, 2013). The foremost reason is hydraulic continuity, upon which Richards' Equation is based, does usually not occur in the field over distances of multiple metres. This is due to horizontal and vertical soil heterogeneity (e.g. macro-pores, soil layering in organic matter and particle sizes, stones) not (properly) accounted for in the model. The form of Richards' Equation implemented in JULES does not take these natural features into account and therefore effective parameters can be expected to differ from values representing a soil with a homogeneous matrix only. That means the process represented by the equation is not the same as the processes occurring in the field soil. Therefore the parameters' roles in JULES differ from their theoretically assumed roles, and so different effective parameter values can be expected. Explicitly taking into account heterogeneity issues like macro-pores could improve JULES' performance (Rahman and Rosolem, 2016), but doing so was beyond the scope of our study. The vertical discretisation of the soil (layers of 10, 25, 65, and 200 cm) is not suitable for solving Richards' Equation. Layer thicknesses of no more than one centimetre near the surface are required to accurately simulate water fluxes. Insufficiently fine discretisation at both field (~eddy-covariance footprint) and hydrological catchment scale may not yield realistic water fluxes using realistic parameter values. (Smirnova et al., 1997; Lee and Abriola, 1999; Van Dam and Feddes, 2000; Downer and Ogden, 2004; Beven and Germann, 2013). Testing the effects of this issue on our results, which would entail changing the vertical discretisation of the soil column in JULES, was beyond the scope of our study. Finally, due to the simplifications of natural processes inherent to models, physically realistic parameters often provide unrealistic results (Gupta et al., 1998)."

COMMENT:
2) MAJOR: I found quite strange that by using the default parameter values performs the same than using the parameter values calibrated on soil moisture data in terms of evaporation fraction (EF) estimation. Even though soil moisture data were of low quality, or the coupling between soil moisture and EF is weak, soil moisture observations represent local data that should give some information to the model. Therefore, I expected better results with respect to the default parameterization. What happens if JULES is calibrated on EF data? How the corresponding modelled soil moisture data compare with PS and CRNS observations? By looking at the results reported in the paper, it seems that using soil moisture observations is needless if we have the purpose of improving land surface modelling. I suggest the authors to improve the discussion and the analysis of the results.

ANSWER:
This is an important point raised by the reviewer and it is in line with Referee #2's comments. It is correct that default parameters in certain cases yielded better surface energy fluxes than after calibration against soil moisture data. This is not an error in our results, but actually an outcome of our study. This corresponds with previous research (Dirmeyer et al 1999; 2000; Koster et al., 2009; Teuling et al., 2009), which showed absolute soil moisture has limited effect on surface energy partitioning, as mentioned by Referee #2. Model soil moisture dynamics have higher impact on surface energy flux simulation. Calibrating soil parameters against soil moisture observations mostly affects the simulated absolute soil moisture values and less so simulated soil moisture dynamics. Our results confirm this. With this reasoning we conclude, in the revised version of our manuscript, that, among other reasons, the results obtained in our study are a consequence of calibrating soil parameters, rather than the soil moisture – evapotranspiration coupling in JULES.
In the main part of revised manuscript we evaluate latent heat flux instead of evaporative fraction. This has made our manuscript more readable. We moved Figure 8 (original manuscript numbering) to Supplement 2 and replaced it with a figure representing the percentage change in RMSE of latent heat flux vs. the percentage change in RMSE of soil moisture. The new figure contain the plots showing the percentage change only, while the bar plots showing the absolute values of RMSE of latent heat flux were removed. It is beyond the scope of our research to do a two-objective calibration including evaporative fraction rather than latent heat flux, in combination with soil moisture.
We have improved our discussion of the results, especially the two-objective calibrations (Section 3.4 in the revised manuscript). We rephrased this section completely to improve the clarity of our analysis, as suggested by Referee #1. We do not think soil moisture data are needless in the context of improving land surface models. Our results showed limited differences between the two soil moisture data products in improving surface energy fluxes and showed simulated surface energy fluxes do not necessarily improve after calibration against soil moisture data. This is in agreement with previous research (Vereecken et al., 2008). Moreover, Gupta et al. (1998) showed a single-objective function is usually insufficient to obtain effective parameter values that yield improved states and fluxes. Our two-objective calibration generally yielded improvement in latent heat fluxes and soil moisture, showing the value of soil moisture. The value of absolute soil moisture to improve surface energy fluxes is limited. New research should therefore focus on using soil moisture data in such way that their main value, which is in the anomalies and dynamics, is optimally used.

COMMENT:

3) MODERATE: I found the description of the results with too many details in several parts of the text (e.g., section 3.1.1, page 9 lines 14-28, page 12, lines 3-18). I suggest not discussing the results for each site, but trying to summarize the most important findings and to focus the discussion on these results.

ANSWER:
We thank the Referee for this comment. We have made our discussions in Section 3.1.1, Section 3.1.2, and Section 3.3 more concise. The subsection headers under Section 3.1 have been removed, yielding one section. The text of former Section 3.1.1 has been reduced to:
"In Figure 2 comparison between PS and CRNS soil moisture time series shows that the seasonal trends of the two soil moisture observation products were similar. The two soil moisture products however also differed from each other, in distinct ways at different sites. PS soil moisture observations were systematically higher than CRNS soil moisture observations at eight of twelve sites. At three sites (DC, SO, CS) PS soil moisture dried down quicker than CRNS soil moisture, while at ME the opposite behaviour was observed. At KE, MM, TR, and MO PS showed a higher seasonality signal (up to 50% higher) than CRNS. Peaks in PS soil moisture were at three sites (UM, KE, TR) up to twice as high as in CRNS soil moisture. In addition, the CRNS data appears noisier than the PS data, which is an effect of inherent randomness in neutron radiation reaching the CRNS sensor element. This effect was more pronounced for lower neutron intensity."

The first paragraph of former Section 3.1.2 (page 9 lines 14-28) has been reduced to:
"The differences seen in Figure 2 are also summarised in Figure 3, which shows a gradual site to site increase in RMSD between observed PS and CRNS and soil moisture data series (RMSD-SM_{obs})."

In Section 3.3, page 12 lines 3-18 of the original manuscript have been changed to:

"To see if validation results would be different if only those periods during which soil moisture stress occurred were evaluated, we also computed the RMSE values (not shown) over these periods only. During these periods plant water uptake limitation factor β was below one (data not shown). Only at site SO did we see somewhat better performance during these periods compared to the original validation period, after both PS and CRNS calibration.

We explored trends between relative improvement in surface energy flux estimation and soil wetness, precipitation, vegetation type, vegetation height, and soil characteristics. No clear trends were discovered. The only feature that could be distinguished was that for the two sites with a bare soil tile (DC and SR) PS and CRNS calibration improved EF and LE. Rooting depth did also not explain relative improvement in surface energy flux estimation. Finally, larger differences between the two observed soil moisture time series did generally yield more different simulated surface energy flux time series, with a clear exception for site UM, where the two calibration approaches yielded more than 20% difference in change in RMSE-LE.

In Figure 7 the monthly mean diurnal latent heat cycles of four sites (UM, SR, WR, and MO) are shown (night time data included but not used for calibration as discussed previously). Both calibration strategies yielded overestimation in March and April and underestimation from June to August at site UM. A month long gap in longwave and shortwave downward model forcing data occurred in April 2012. This might have affected the results for this site. At SR, both PS and CRNS calibration improved latent heat flux during periods of low evapotranspiration, while during the other periods only the CRNS calibrations yielded better results. AT WR calibration against PS soil moisture yielded overestimation of latent heat flux, while CRNS calibration yielded too low latent heat flux. The PS calibration at MO yielded latent heat flux underestimation, whereas CRNS calibration did not change LE substantially."

COMMENT:
In the specific comments, I added some corrections and suggestions that should be implemented. On this basis, I believe the paper deserves to be published only after a major revision.

ANSWER:
The authors appreciate that the Referee believes our paper deserves to be published after a major revision.

Specific comments

COMMENT
P2, L26: The sentence "past research indicates ...wetting and drying periods" is too vague. At least, references should be included. However, I note that it is still an open issue to fully understand in which conditions soil moisture variability is higher. For instance, it is not the same if absolute or anomaly soil moisture values are analysed (see e.g., Mittelbach and Seneviratne, 2012; doi:10.5194/hess-16-2169-2012).

ANSWER:
We have removed this sentence in our effort to reduce the size of the manuscript. We do agree with the referee that this is still an open issue.

COMMENT:
P3, L8-10: The sentence is incomplete (only a single sensor is needed for?), please check.

ANSWER:
We have rephrased this sentence: "Newer soil moisture sensor techniques, for instance one which makes use of the Global Positioning System (GPS; Larson et al 2008, 2010), and the Cosmic-Ray Neutron Sensor (CRNS; Zreda et al., 2008) have the advantage that their installation requires less time and work effort because only a single above ground sensor is needed."

COMMENT:
P4, L32: The gap-filling of 30 days seems to me a very large window. Does it affect the results? Some tests should be made.

ANSWER:
We thank the reviewer for raising this point. This was also mentioned by Reviewer #1. We here provide the same answer as to Reviewer #1 and have included the same table and the same figure as in the answer to Reviewer #1. Table a (below) summarises the average (average of the seven forcing variables) percent gap for each site filled with this method and a Figure a shows how large the gaps were. This figure shows the percentage missing hours for each gap size category. If the bars for one variable at one site would be summed, this would yield the total percentage of missing hours.
Overall, average gaps vary among sites between near zero to 15%. Gaps larger than 15 days occurred at:

- UM: downward shortwave and downward longwave radiation (9% of time)
- DC: net radiation (4% of time)
- KE: downward longwave radiation (4% of time)

- SR: downward longwave radiation (8% of time)
- CS: downward shortwave radiation (3% of time)
- MM: pressure (2% of time)
- AR: temperature, wind speed, and relative humidity (2-4% of time)
- WR: precipitation and pressure (3% of time)
- MO: precipitation (2% of time)

Summarising, large gaps (over 15 days) occurred for no more than 10% of time at any site and occurred at most sites for longwave and shortwave radiation only. Precipitation is likely to be the primary gap factor affecting soil moisture but its occurrence was minimal. Nevertheless, we noted this issue in the manuscript (Section 3.3 of the revised manuscript, see below).

In this study, we used the most regular freely-available data sets to the best of our knowledge to ensure reproducibility. We noticed that outcomes for sites with few large data gaps did overall not differ from the outcomes on sites with more large data gaps. Therefore, we argue that although the data gaps may introduce some uncertainties in the model results, our major conclusions remain unchanged. We discuss the most important large data gaps in Section 3.3:

"A month long gap in longwave and shortwave downward model forcing data occurred in April 2012. This might have affected the results for this site."

and in Section 3.4 of the revised manuscript:

"An issue that could have had effect on our results is the occurrence of gaps in the model forcing data. Gaps in observed meteorological data are often inevitable and must be filled for a land surface model to be run. Percentages of missing hours differed between 0% and 15% at our sites. Gaps larger than fifteen days filled with the moving window gap-filling procedure occurred mainly for downward shortwave and/or downward longwave radiation; at sites UM, KE, SR, DC, TR, and CS. At site WR a data gap of 29 days in precipitation and air pressure occurred. Especially the gap in precipitation data can have negatively affected the model results at this site."

Please note that the gap filling script was actually designed to fill gaps up to 30 days and not 31; we have changed this detail in the new version of the manuscript (page 5, line 24). Precipitation was not gap-filled; missing points were set to zero instead. Finally, we found that Table A1.2 contained errors; we have corrected this in the new version of the manuscript (page 37).

Table a: Percentage of missing hourly model forcing records. In the middle column the mean over all input variables is shown along with the range over all input variables. In the right column the length of the simulation period is shown.

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

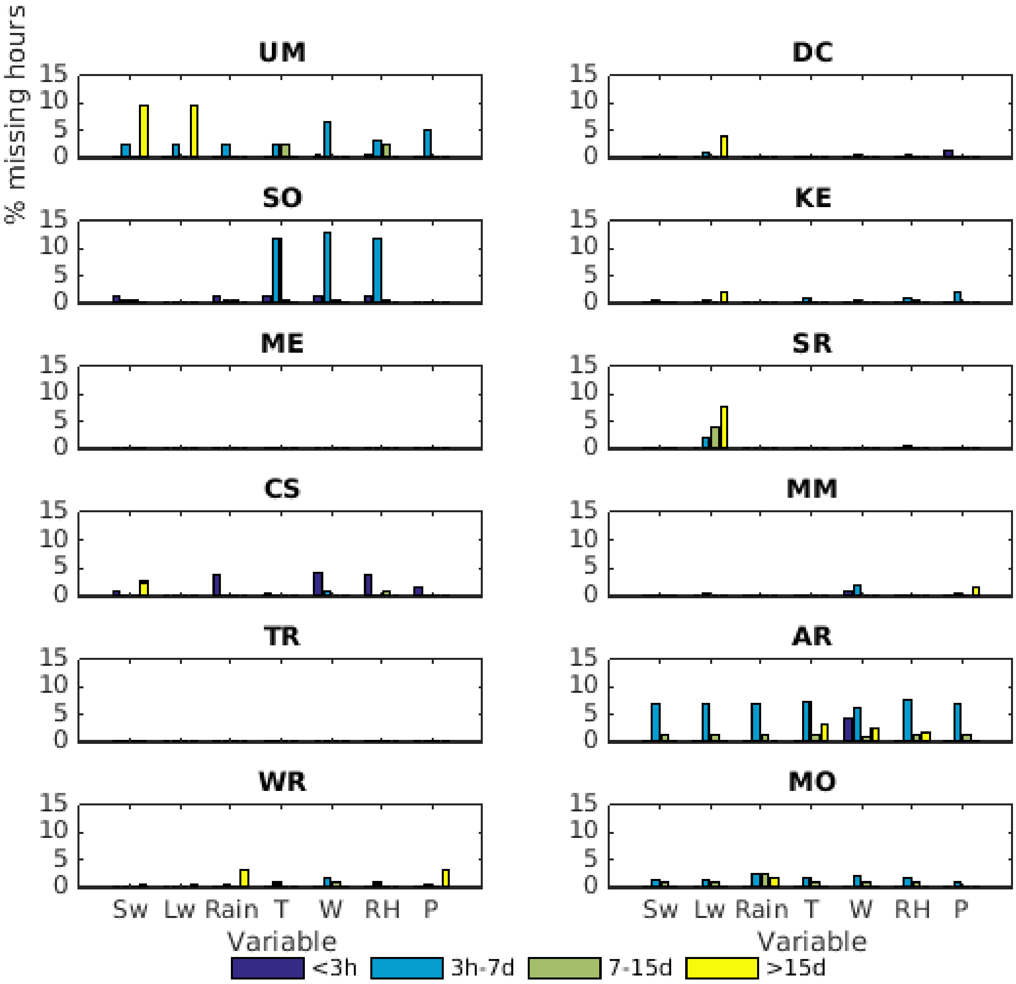


Figure a: Missing hours as percentage of the total hours in the calibration/validation periods. The statistics are shown per forcing variable. Sw is downward shortwave radiation, Lw is downward longwave radiation, Rain is precipitation, T is air temperature, W is wind speed, RH is relative humidity, and P is atmospheric pressure. Please note that for convenience the label Lw is also used for site DC; instead of net radiation. The four categories split the missing hours between gaps of less than 3 hours, of between 3 hours and 7 days, of between 7 days and 15 days, and gaps larger than 15 days.

COMMENT:
P5, L6-10: It should be better to insert an equation here.

ANSWER:
We have inserted the equation. Please notice this section has been moved up (now Section 2.2) as suggested by Referee #1.

COMMENT:
P9, L20: The larger differences between PS and CRNS at wetter sites are expected. Higher is soil moisture, higher will be the differences.

ANSWER:
We removed the discussion on the correlations with wetness conditions (first three sentences of this paragraph).

COMMENT:
P10, L5: Figure A1 should be Table A1?

ANSWER:
Correct, we have changed this cross-reference.

COMMENT:
P10, L18-22: The difference in sensing depth might be the cause of some of the differences between PS and CRNS. However, it could be checked with specific analysis. Otherwise, I suggest removing.

ANSWER:
We have removed this text.

COMMENT:
P10, L29: It is obvious that after the calibration on soil moisture data the RMSE values will reduce. The model is tested with the data used for calibration.

ANSWER:
We have rephrased this line: “The degree to which the objective function (RMSE-SM; Figure 6) values decreased differed between sites …”

COMMENT:
P11, L1-2: The use of RMSE for calibration reduces the mean error between modelled and observed data. Therefore, the effect of having less extreme peaks and valleys is due to the selected objective function.

ANSWER:
We have rephrased: “This was due to the selected objective function (RMSE), which reduces the mean error between modelled and observed data.”

COMMENT:
P11, L24-28: From here it is not clear the number of sites over which an improvement in EF estimation is obtained. 11 sites (P11, 22) or 12 sites (P11, L28). Check also later in the text (e.g., P13, L36).

ANSWER:
We rephrased this section to make it better understandable and readable. Please also notice that we are now using latent heat as principle objective function instead of EF. We rephrased page 11, lines 24-28 of the original manuscript to:

“Calibration against PS soil moisture improved RMSE-LE at six sites (UM, KE, DC, SR, SO, and ME), while calibration against CRNS neutron counts improved RMSE-LE at eight sites (KE, DC, SR, ME, SO, AR, MM, and WR). At five sites RMSE-LE improved after calibration against both PS soil moisture and CRNS neutron counts. Calibration yielded lower RMSE-LE after calibration against CRNS neutron counts than after calibration against PS soil moisture at all but three sites (UM, KE, and WR). Figure 6 also shows that RMSE-SM decreased substantially less (i.e. >20% difference) after calibration against PS soil moisture than after calibration against CRNS neutron counts (i.e. >20% difference between both strategies on the horizontal axis of Figure 6, which occurred at six sites). At five of these six sites the relative change in surface energy flux performance was also smaller (sites MO, AR, TR, ME, and SR). This indicates that it cannot be excluded that further improvement in soil moisture simulation after calibration against PS data could have yielded better surface energy fluxes."

We have removed the sentence on page 13 line 36 and completely rephrased this section to make it better understandable.

COMMENT:
P12, L3-6: It is not clear to me what the authors want to demonstrate with this analysis. Please clarify.

ANSWER:
We rephrased the entire paragraph to clarify this. Please note we have removed Figure 13 from the manuscript in our effort to reduce the number of figures.

"To see if validation results would be different if only those periods during which soil moisture stress occurred were evaluated, we also analysed the performance over these periods only (data not shown). During these periods plant water uptake limitation factor β was below one (data not shown). Only at site SO did we see somewhat better performance during these periods compared to the original validation period, after both PS and CRNS calibration."

COMMENT
P12, L32: Change "on the edges" with "within the edges".

ANSWER:
We changed as suggested.

COMMENT:
P15, L9-12: Not clear to me how the "multiplier" values are used. Please clarify.

ANSWER:
Please see Table 2, columns 4 and 5.

COMMENT:
P16, L4-11: As mentioned above, I found not scientifically sound to attribute the low performances in term of EF improving to the weak coupling of JULES. Moreover, it's not clear to me the discussion of the value of absolute soil moisture with respect to anomalies.

ANSWER:
Please see our answer to General Comment 1B. We removed Section 3.8 "Discussion" from the revised manuscript. Discussion was moved to Section 3.2:

"Previous research (e.g. Teuling et al., 2009) has shown calibrating soil parameters has a large effect on simulated absolute soil moisture values (bias), but substantially less on soil moisture seasonality and dynamics. Our finding supports this."

Section 3.3:

“One factor causing some of these limited improvements was that, when mean simulated root zone soil moisture (weighted with root density) increased after calibration, the values of the wilting point and critical point soil moisture parameters moved along (data not shown), yielding similar soil moisture stress. This happed for both calibrations at site KE, and for the calibration against CRNS neutron counts at sites MO and TR. This could relate to the limited value of simulated absolute soil moisture for surface energy flux estimation in land surface models (Dirmeyer et al., 2000; Koster et al., 2009). However, we also found the distance between wilting point values and critical point values to decrease after calibration. This occurred with a simultaneous decrease in standard deviation of the simulated soil moisture. The self-adjusting behaviour of the wilting point and critical point parameters was also indicated by parameter sensitivity analysis (Appendix 3), which showed soil moisture was substantially more sensitive to a change in critical point value than latent heat flux.”,

and:
“In summary, the single-objective calibrations against CRNS neutron counts yielded larger improvements in simulated surface energy fluxes than single-objective calibrations against PS soil moisture (Figure 6). Improvements in surface energy flux estimation were however substantial for four calibrations against PS soil moisture and five calibrations against CRNS neutron counts. Limited improvements in surface energy flux estimation after calibration could partly be attributed to the limited value of absolute soil moisture to estimate surface energy fluxes with land surface models. This seems reasonable because calibration mostly affected absolute soil moisture (Figure 4). This result corresponds with earlier research that showed model soil moisture dynamics and seasonality have a larger effect on surface energy flux simulation than absolute soil moisture (e.g. Teuling et al., 2009; Dirmeyer et al., 1999). Previous research has also indicated that soil moisture alone is insufficient to estimate soil hydraulic parameters (Vereecken et al., 2008; 2015).”

And to Section 3.4:
“In summary, the two-objective calibrations against soil moisture (or neutron counts) and latent heat flux showed fewer substantial differences between calibration with PS soil moisture and calibration with CRNS neutron counts. These results indicate that the differences between both *single*-objective calibration strategies, which showed an advantage for CRNS observations, were possibly not as substantial as it seemed at first. The spatio-temporal stability theory, which implies limited spatial variability in surface energy fluxes, could be one explanation for this (Vachaud et al., 1985; Teuling et al., 2006; Mittelbach and Seneviratne, 2012; Albertson and Montaldo, 2003). Another factor that has possibly played a role is that the spatial scale advantage of the CRNS was masked out by the possibly lower quality of its measurements (see also Section 3.1). Different hydrogen pools than soil moisture affect the neutron measurements at various temporal resolutions (e.g. rainfall interception; Baroni and Oswald, 2015). However, PS sensors also have their limitations. Different (electromagnetic) PS sensors have different designs and properties, which affect the data quality (Robinson et al., 2008; Blonquist et al., 2005). Soil type and soil specific calibration also affect the accuracy and precision of the PS data. For instance, the relationship between electrical permittivity and soil moisture content is strong for quartz rich soils, but less accurate for clay soils (e.g. Ishida et al., 2000; Robinson et al., 2008).”

COMMENT:
Figure 8: Labels a) and b) are missing.

ANSWER:
We have removed Figure 8a and therefore the labels are not being used any longer. Please note Figure 8b was moved to the newly created Supplement 2.

NEW REFERENCES

Albertson, J. D., and Montaldo, N.: Temporal dynamics of soil moisture variability: 1. Theoretical basis, Water Resour. Res., 39, 1274, doi:10.1029/2002WR001616, 2003.

Baatz, R., Bogena, H. R., Hendricks Franssen, H.-J, Huisman, J.A., Montzka, C., and Vereecken, H.: An empirical vegetation correction for soil water content quantification using cosmic ray probes, Water Resour. Res., 51, 2030-2046, doi:10.1002/2014WR016443, 2015.

Baroni, G., and Oswald, S.E.: A scaling approach for the assessment of biomass changes and rainfall interception using cosmic-ray neutron sensing. J. Hydrol., 525, 264-276, doi:10.1016/j.hydrol.2015.03.053, 2015.

Blonquist, J. M., Jones, S. B., and Robinson, D. A.: Standardizing characterization of electromagnetic water content sensors: Part 2. Evaluation of seven sensing systems. Vadose Zone J., 4, 1059-1069.

Desilets, D., and Zreda, M.: Footprint diameter for a cosmic-ray soil moisture probe: Theory and Monte Carlo simulations. Water Resour. Res., 49, 3566-3575, doi:10.1002/wrcr.20187, 2013.

Dirmeyer, P. A., Dolman, A. J., and Sato, N.: The pilot phase of the Global Soil Wetness Project. Bull. Amer. Meteor. Soc., 80, 851-878, 1999.

Dirmeyer, P. A., Zeng, F. J., Ducharne, A., Morrill, J. C., and Koster, R. D.: The Sensitivity of Surface Fluxes to Soil Water Content in Three Land Surface Schemes. J. Hydromteor., 1, 121-134, 2000.

Downer, C. W., and Ogden, F. L.: Appropriate vertical discretization of Richards' equation for two-dimensional watershed-scale modelling. Hydrol. Process., 18, 1-22, 2004.

Famiglietti, J. S., Dongryeol, R., Berg, A. A., Rodell, M., and Jackson, T. J.: Field observations of soil moisture across scales. Water Resourc. Res, 44, W01423, doi:10.1029/2006WR005804, 2008.

Finkelstein, P. L., and Sims, P. F.: Sampling error in eddy correlation flux measurements. J. Geophys. Res., 106(D4), 3503-3509, 2001.

Gupta, H. V., Bastidas, L. A., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. L.: Parameter estimation of a land surface scheme using multicriteria methods. J. Geophys. Res., 104(D16), 19491-19503, 1999.

Gupta, H. V., King, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377, 80-91, 2009.

Ishida, T., and Makino, T., and Wang, C. J., 2000. Dielectric-relaxation spectroscopy of kaolinite, montmorillonite, allophane, and imogolite under moist conditions. Clays Clay Miner, 48, 75-84.

Lee, D. H., and Abriola, L. M.: Use of Richards equation in land surface parameterizations. J. Geophys. Res., 104(D22), 27519-27526, 1999.

Mittelbach, H. and Seneviratne, S. I.: A new perspective on the spatio-temporal variability of soil moisture: temporal dynamics versus time-invariant contributions. Hydrol. Earth Syst. Sci., 16, 2169-2179, doi:10.5194/hess-16-2169-2012, 2012.

W. Qu, H. R. Bogena, J. A. Huisman, J. Vanderborght, M. Schuh, E. Priesack, H. Vereecken, Predicting subgrid variability of soil water content from basic soil information, Geophysical Research Letters, 2015, 42, 1-8.

Rahman, M, and Rosolem, R.: The effects of chalk representation in land surface modelling. Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-244, accepted, 2017.

Sellers, P. J., and Hall, F. G.: FIFE in 1992: Results, scientific gains, and future research directions. J. Geophys. Res., 97(D17), 19091-19109, doi:10.1029/92JD02173, 1992.

Smirnova, T. G., Brown, J. M., and Benjamin, S. G.: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. Monthly Weather Review, 125, 1871-1884, 1997.

Teuling, A. J., and P. A. Troch: Improved understanding of soil moisture variability dynamics. Geophys. Res. Lett., 32, L05404, doi:10.1029/2004GL021935, 2005.

Teuling, A. J., Uilenhoet, R., Hupet, F., Van Loon, E. E., and Troch, P. A.: Estimating spatial mean root-zone soil moisture from point-scale observations. Hydrol. Earth Syst. Sci., 10, 755-767, 2006.

Teuling, A. J., Uilenhoet, R., Van den Hurk, B., and Seneviratne, S. I.: Parameter Sensitivity in LSMs: An Analysis Using Stochastic Soil Moisture Models and ELDAS Soil Parameters. J. Hydrometeorol., 10(3), 751-765, doi:10.1175/2008JHM1033.1, 2009

Vachaud, G., Passerat de Silans, P., Balabanis, P., Vauclin, M.: Temporal stability of spatially measured soil water probability density function, Soil Sci. Soc. Am. J., 49, 822–828, 1985.

Van Dam, J. C., and Feddes, R. A.: Numerical simulation of infiltration, evaporation, and shallow groundwater levels with the Richards equation. J. Hydrol., 233, 72-85.

Vanderlinden, K., H. Vereecken, H. Hardelauf, M. Herbst, G. Martínez, M. H. Cosh, and Pachepsky, Y. A.: Temporal Stability of Soil Water Contents: A Review of Data and Analyses. Vadose Zone J., 11., doi:10.2136/vzj2011.0178, 2012.

Land surface model performance using cosmic-ray and point scale soil moisture measurements for calibration Reducing soil moisture measurement scale mismatch to improve surface energy flux estimation

Joost Iwema[1], Rafael Rosolem[1,2], Mostaquimur Rahman[1], Eleanor Blyth[3], Thorsten Wagener[1,2]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK
[2]Cabot Institute, University of Bristol, Bristol, UK
[3]Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK

Correspondence to: Joost Iwema (joost.iwema@bristol.ac.uk)

Abstract. At the so-called hyper-resolution scale (i.e. grid cells of 1 km^2) Land land Surface surface Mmodel (LSM) parameters are sometimescan be calibrated with Eddy-Covariance (EC) flux data and Point point-sScale (PS) soil moisture data. However, measurement scales of Eddy-Covariance and point-scalePS data differ substantially. In our study, we investigated the impact of reducing the scale mismatch between surface energy flux data and soil moisture data observations by replacing point-scalePS soil moisture data with observations derived from Cosmic-Ray Neutron Sensors (CRNS) made at larger spatial scales. Five soil and -evapotranspiration parameters of the Joint UK Land Environment Simulator (JULES) were calibrated against point-scalePS and Cosmic-Ray Neutron SensorCRNS soil moisture data separately. We calibrated the model for twelve sites in the USA representing a range of climatic, soil, and vegetation conditions. The improvement in surface energylatent heat partitioning flux estimation for the two calibration solutions was assessed by comparing to Eddy-Covariance flux data and to a version of JULES runs simulations with default parameter values. We found that simulated latent heat fluxessurface energy partitioning did not differ substantially between both calibration strategiesthe espite their differences in actual soil moisture observations. This was mainly caused by the limited effect of calibrating soil parameters on soil moisture dynamics and surface energy fluxes. Other factors that played a role were limited spatial variability in surface fluxes implied by soil moisture spatio-temporal stability, and data quality issues.We concluded that potential differences due to distinct spatial scales represented by the point-scalePS and Cosmic-Ray Neutron SensorCRNS soil moisture sensor techniques were substantially undermined by the weak coupling between soil moisture and evapotranspiration within JULES.

1 Introduction

The land surface water and energy balances are coupled through the process of evapotranspiration. Soil moisture is one of the main water reservoirs near the land surface and can hence importantly control the surface water and energy balances. Soil moisture provides a first order (i.e. direct) control on evapotranspiration when there is insufficient water to meet the evaporative demand (Manabe et al., 1969; Budyko, 1956; Seneviratne et al., 2010). An indirect effect of soil moisture on surface energy flux partitioning is for instance the damping effect on soil and air temperature, which in its turn affects humidity, evapotranspiration, boundary-layer stability, and in some cases precipitation (Seneviratne et al., 2010). The coupling betweencontrol of soil moisture and on temperature at seasonal scalessurface energy flux partitioning is especially strong in transitional climate regions (Koster 2004).

Land Surface Models (LSMs) solve the surface mass (including water), energy, and momentum balances to provide the weather and climate prediction models with lower boundary conditions. The land surface has been shown to play an important role in global atmospheric circulation (Koster et al., 2004). Because the soil moisture state and surface fluxes are so closely connected, it is believed important to accurately simulate these simultaneously (Henderson-Sellers et al., 1996; Richter et al., 2004; Seneviratne et al., 2010; Dirmeyer, 2011; Dirmeyer et al., 2013).

Land surface models have become more complex over the past decades (Sellers, 1997; Seneviratne et al., 2010). The first land surface models included only limited representations of soil hydrological and biophysical processes. The `second-generation' models included the main physical processes within vegetation and in soil, and the later developed `third generation' models included plant physiological processes like photosynthesis and carbon assimilation (Seneviratne et al., 2010). Theeis increasing complexity of land surface models over the past decades (Sellers, 1997; Seneviratne et al., 2010) has brought with it an increasing number of parameters, with values not easily defined with in-situ measurements because of scale mismatch,. e.g.For instance, stomatal resistance measured at leaf level is not the same as canopy stomatal resistance needed for LSMs (Blyth et al., 1993). Soil hydraulic parameter values (e.g. soil hydraulic conductivity) are often obtained from laboratory experiments on soil cores for cubic centimetres to cubic decimetres. The soil properties and processes at this scale can however differ from those at the LSM grid cell sizes, which are often as large as hundreds to tens of thousands of square kilometres (Pitman 2003). Due to the different governing processes upscaling the soil hydraulic properties from soil core scale to field scale is non-trivial (Vinnikov 1996; Crow et al., 2012).

In an effort to develop global hydrometeorological monitoring and prediction capabilities (Wood et al., 2011), hydrological models and LSMs are now increasingly being applied at the finer `hyper-resolution scale' with grid cells of about 1 km^2.Following discussion on how to move forward with hydrological models and LSMs in an effort to develop models that cover all land surface of the globe (Wood et al., 2011; Beven and Cloke, 2012), models are increasingly being applied at the finer `hyper-resolution scale' with grid cells of about 1 km^2. Typically, parameters are calibrated and validated at this hyper-resolution against in-situ measurements from sources such as eEddy-Ccovariance (EC) flux towers and point-scale (PS) soil moisture sensors (e.g. Time Domain Transmissivity; TDT and Time Domain Reflectometry; TDR) (Stockli et al., 2008; Richter et al., 2004; Blyth et al., 2010; Blyth et al., 2011; Rosolem et al., 2013). Such calibration/validation data has become more widely available at hyperresolution scale (Baldocchi, 2001; Smith, 2012). However, the horizontal footprints of different measurement techniques' vary from each other: EC Eddy-covariance surface energy flux data represent a downwind footprint of 100 m^2 to 1 km^2, while in-situ soil moisture probes link to much smaller surface areas by representing a support volume (soil volume represented by sensor; Blöschl and Sivapalan, 1995) of ~4 dm^3 only (Running et al., 1999; Kurc and Small, 2007;

Vivoni et al., 2008). Soil moisture can beis spatially non-uniform within the EC eddy-covariance footprint due to heterogeneity in soil properties, vegetation, and topography. How variable the soil moisture content is under certain heterogeneous conditions depends on the wetness conditions.; It is however not yet entirely clear under which conditions variability is highest and differs for temporal mean soil moisture and anomalies (Famigliettti et al., 2008; Mittelbach and Seneviratne, 2012)past research indicates the variability to be highest during soil wetting and drying periods. Therefore, soil moisture measurements best (i.e. most effectively) representing the soil below the eddy-covariance tower's footprint should be used when the performance of a land surface model is evaluated. If soil moisture is measured at only a single or few locations with limited support volume, like with PS sensors, the measured soil moisture content might be different from the effective soil moisture state that controls the surface exchange processes. It is therefore often assumed that soil moisture measured at a scale closer to the footprint of an eddy-covariance tower is more informative than a single or a couple of PS sensor profiles for studying land-surface processes and constraining model parameters at scales of $>\sim 100$ m^2 (Robinson et al., 2008). This poses a potential scale mismatch issue when a single or a few PS sensors are used, as depicted in Figure 1. On the other hand, past research has shown soil moisture measured at only one or a few points within an area of similar size as an eddy-covariance footprint, can have similar value to surface energy flux simulation as soil moisture measured at a larger scale (e.g. Vachaud et al., 1985; Teuling et al., 2006; Mittelbach and Seneviratne, 2012). These studies showed that points within a soil moisture observation network keep their rank with respect to the mean soil moisture (anomaly), i.e. they either under –or overestimate the mean (anomaly), so-called spatio-temporal stability. The physical principle behind the spatio-temporal stability theory is that different time variant and hydrological processes either create or destroy spatial soil moisture variability whereas time invariant land surface characteristics induce an effective offset in the spatial variability (Albertson and Montaldo, 2003; Teuling and Troch, 2005; Vanderlinden et al., 2012). When soil moisture reaches values below the critical point (i.e. transpiration becomes moisture limited), spatial variability in soil moisture and fluxes is reduced (Teuling and Troch, 2005). Soil moisture dynamics was found to be a small portion of total soil moisture variability (Mittelbach and Seneviratne, 2012), while having a greater effect on surface energy fluxes than absolute soil moisture in land surface models (Dirmeyer et al., 1999; Teuling et al., 2009). The implication of the spatio-temporal stability theory therefore is that the spatial scale mismatch issue might have limited implications to surface energy flux simulation. The question which then rises is,

d*oes reduced observation scale mismatch improve LSM energy partitioning flux estimationes?*

Based on the spatio-temporal stability theory, we phrased the following hypothesis for our research question:

reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations.

In recent years, new soil moisture measurement techniques have been developed that have, compared to in-situpoint-scale soil moisture sensing techniques (TDT and TDR), a reduced scale mismatch with EC eddy-covariance surface energy flux measurements. Improvement in wireless technology and remote data collection technology have made the development of PS soil moisture sensor networks more feasible (Cardell-Oliver, 2005; Ritsema et al., 2009; Trubilowicz, 2009; Bogena et al., 2010; Robinson et al., 2008). ~~It is usually assumed that, due to spatial variability in soil physical characteristics, a network of PS sensor profiles is more informative than a single or a couple of PS sensor profiles for studying land-surface processes and constraining model parameters at scales of >~100 m^2 (Robinson et al., 2008).~~

Newer soil moisture sensor techniques, for instance one which makes use of the Global Positioning System (GPS; Larson et al 2008, 2010), and the Cosmic-Ray Neutron Sensor (CRNS; Zreda et al., 2008) have the advantage that their installation requires less time and work effort because only a single above ground sensor is needed~~have the advantage that only a single, above ground sensor is needed, which is easier to install than a PS network~~.

The CRNS (Zreda et al., 2008) is an above-ground passive sensor which utilises natural cosmic-ray neutron radiation to estimate soil moisture content in the top 10-70 cm. The sensor's footprint area has a radius of about 100 to 300 m surrounding the above-ground sensor (~~(Figure 2;~~ Desilets and Zreda, 2013; Köhli et al., 2015). Networks of CRNS have been established in various countries (e.g. Cosmic-ray Soil Moisture Observation System; COSMOS (Zreda et al., 2012), COSMOS-UK (Evans et al., 2016), the Australian National Cosmic-Ray Soil Moisture Monitoring Facility CosmOZ (Hawdon et al., 2004), and TERrestrial ENvironmental Observatoria; TERENO (Baatz et al., 2015)). Franz et al. (2012) showed soil moisture estimated from CRNS neutron counts differed less than 20% from the average of a co-located point scale soil moisture sensor network at a site in Arizona.

Unlike wireless point scale sensor networks, both the GPS and CRNS technology provide an integrated soil moisture measurement over the entire support volume (Larson et al., 2008; Zreda et al., 2008). We chose to answer our research question using the CRNS technology because the COSMOS network provides publicly available data for multiple years at a range of sites co-located with Ameriflux/FLUXNET ~~EC~~ eddy-covariance towers (ORNL-DAAC, 2015). ~~We used t~~Twelve of these sites ~~which~~ provided sufficient LSM forcing data, PS soil moisture data, CRNS data, and ~~EC~~ eddy-covariance LE and sensible heat flux (H) data. ~~We phrased the following hypothesis for our research question: .~~

Before our modelling exercise we first compared the PS and CRNS data. The outcomes of this data analysis were mainly used to see if the results from the calibration and validation yielded larger differences in surface energy flux estimation at sites where the two soil moisture observation products showed higher deviation from each other. To investigate our research question we made the LSM simulated soil moisture content match the observed PS or CRNS data as closely as possible. We did this by calibrating parameters of the Joint UK Land Environment Simulator (JULES; Best et al., 2011) against point-scale and Cosmic-Ray Neutron data separately. We subsequently validated the results against ~~EC~~eddy-covariance observed data over the same periods. To assess the change in soil moisture and surface energy fluxes after calibration we compared the calibrated runs against a default run with parameter values computed from a widely used soil properties database. We emphasise that we compared the two different soil moisture measurement techniques' scales and not the techniques as such. ~~Before our modelling exercise we first compared the PS and CRNS data. The outcomes of this data analysis were mainly used to see if the results from the calibration and validation yielded larger differences in surface energy flux estimation at sites where the two soil moisture observation products showed higher deviation from each other.~~

2 Data and Methods

2.1 Calibration and Validation data: PS, CRNS, and ~~EC~~ eddy-covariance data

PS soil moisture and CRNS neutron count data from twelve Ameriflux/COSMOS sites were used (~~Figure 2~~Figure 1; full COSMOS site names are shown in this figure). These twelve sites covered eight of twenty Ecoclimatic domains of the US National Ecological Observatory Network (NEON; www.neonscience.org) (~~Figure 2~~Figure 1). These twelve sites hence represent a variety of climates and land cover types, but also different soil types (Table 1~~; for full site names see Appendix 1, Table A1.1~~).

Hourly PS data for nine sites were obtained from the publicly available Ameriflux Level 2 data source (ORNL). Data for the three California Climate Gradients sites (DC, CS, and SO) were obtained at http://www.ess.uci.edu/~california/ (data version 3.4; Goulden et al., 2015). The number of PS profiles, the installation depths, and sensor types differed between the twelve study sites (Table A1.1 in Appendix 1). We used point scale soil moisture data from the soil layers up to 30 cm depth only for consistency among all sites. There were only two sites reporting soil moisture data at greater depths: WR at ~~(~~50 cm~~)~~, and MO at ~~(~~100 cm). Our main objective was to investigate the difference in information content due to two soil moisture measurement techniques' different horizontal scales in relation to the eddy-covariance footprint, rather than to compare the measurement techniques themselves. Quality control was applied to filter out spurious and unrealistic data points due to sensor errors. The PS data was then interpolated to the JULES soil layer on which the model was calibrated.

Hourly CRNS neutron count data were obtained from the COSMOS network website (www.cosmos.hwr.arizona.edu). Corrections were applied as by Zreda et al. (2012). Water vapour corrections (Rosolem et al., 2013) were applied with respect to dry air (Bogena et al., 2013). ~~Similar~~ The quality control approach used for the PS analysis was also applied to CRNS neutron count data series to remove unrealistic points. Snow cover periods were also removed for the analysis. A 5-hour moving average window was applied to the observed CRNS neutron counts (following Shuttleworth et al., 2013; Rosolem et al., 2014). ~~We compared PS soil moisture with CRNS integrated soil moisture computed from vertically homogeneous soil moisture values obtained from the observed neutron counts using the COsmic-ray Soil Moisture Interaction Code (COSMIC; Shuttleworth et al., 2013).~~

We used the exact same data for the soil moisture data comparison as for the calibration (providing that both PS and CRNS were available in the same period). As validation data, latent heat (LE) ~~and sensible heat (H)~~ flux hourly data from Ameriflux Level 2 data source was used for nine sites. We used data version 3.4 (Goulden et al., 2015)~~while~~ for the three California Climate Gradients sites ~~data version 3.4 (Goulden et al., 2015)~~. Quality control was applied to the LE and H flux data to remove outliers and unrealistic data points.

2.2 Soil moisture data comparison methodology

To compare PS and CRNS soil moisture data we computed the Mean Squared Deviation (MSD) and its decomposition (Gupta et al., 2009) into:

(1) The squared difference between the means (structural bias)

(2) The squared difference between the standard deviations (indicates different seasonality)

(3) A term relating to the coefficient of correlation (indicates different dynamics)

Which yields the following equation:

$$MSD = (\mu_{PS} - \mu_{CRNS})^2 + (\sigma_{PS} - \sigma_{CRNS})^2 + 2 \cdot \sigma_{PS} - \sigma_{CRNS} \cdot (1 - r)$$

Where μ ($m^3 m^{-3}$) is the observed mean, σ ($m^3 m^{-3}$) is standard deviation, and r (-) is the coefficient of correlation.

We however scaled the relative contributions of the three MSD components to the Root Mean Squared Deviation (RMSD) instead of MSD, to keep units of $m^3 m^3$. We then ranked the sites from lowest to highest RMSD. According to our hypothesis, we would expect to see a larger difference in simulated surface energy fluxes after calibration when the two soil moisture time series differ most. We compared PS soil moisture with CRNS soil moisture values computed from vertically homogeneous soil moisture values obtained from the observed neutron counts using the COsmic-ray SoiI Moisture Interaction Code (COSMIC; Shuttleworth et al., 2013). ~~In reality heterogeneous profiles are common across different sites and conditions.~~

2.~~2~~3 JULES forcing data and initial conditions

JULES requires precipitation, air temperature, atmospheric pressure, wind speed, specific humidity, downward shortwave radiation, and downward longwave radiation as meteorological forcing data. Quality controlled hourly data was obtained from Ameriflux Level 2 and the three California Climate Gradients sites. At some of the sites however, certain specific forcing data was not available from Ameriflux and hence data from different sources ~~was~~ were used (Table A1.2 of Appendix 1).

Model input data was gap-filled following Rosolem et al. (2010) because JULES requires continuous time series except for precipitation where gaps were set to zero. For all sites gaps smaller than 3 hours were filled by linear interpolation.~~,~~ ~~while~~ ~~l~~Larger gaps of up to 30~~1~~ days were gap-filled using the average diurnal pattern of the preceding and following 15 days. In addition, some remaining gaps in Downward Shortwave Radiation and Downward Longwave Radiation at Wind River (WR) were filled using linear least squares relationships with NLDAS-2 data (http://disc.sci.gsfc.nasa.gov/uui/datasets?keywords=NLDAS). At Soaproot (SO) and Coastal Sage (CS) data gaps in the atmospheric pressure time series were filled with NLDAS data. At ~~Tonzi Ranch (TR)~~CS NLDAS data were also used to gap fill air temperature. Gap filling at Santa Rita Creosote was done with data from the nearby Sahuarita site followed by the gap-filling procedure described above.

~~**2.3 Soil moisture data comparison methodology**~~

~~To compare PS and CRNS soil moisture data we computed the Mean Squared Deviation (MSD) and its decomposition (Gupta et al., 2009) into:~~

~~(1) The squared difference between the means (structural bias)~~

~~(2) The squared difference between the standard deviations (indicates different seasonality)~~

(3) A term relating to the coefficient of correlation (r; indicates different dynamics): $2 \cdot \sigma_{ps} \cdot \sigma_{crns} \cdot (1-r)$

We however scaled the relative contributions of the three MSD components to the Root Mean Squared Deviation (RMSD) instead of MSD, to keep units of m^3m^{-3}. We then ranked the sites from lowest to highest RMSD. According to our hypothesis, we would expect to see a larger difference in simulated surface energy fluxes after calibration when the two soil moisture time series differ most. The CRNS soil moisture values computed with COSMIC represent the integrated signal computed from vertically homogeneous profiles, whereas in reality heterogeneous profiles are common across different sites and conditions.

2.4 Calibration and validation methodology

2.4.1 Joint UK Land Environment Simulator (JULES)

We used the Joint UK Land Environment Simulator (JULES; Best et al., 2011; Clark et al., 2011) in this study. JULES can be coupled as lower boundary condition to the UK Met Office Unified Model (Cullen, 1993). Within JULES choices can be made (e.g. canopy radiation model type) and certain modules (e.g. vegetation dynamics) can be switched on or off to operate at different levels of complexity. In addition, we chose the UK Variable resolution configuration (UKV) because it is the standard setting when JULES is run coupled with the UK Met Office Unified Model. The UKV land grid cell size is 1 km by 1 km. However, our study focused on JULES standalone simulations at the 12 grid points located at the sites investigated. The UKV setting employs the multi-layer canopy radiation module with surface heat capacity and snow beneath the canopy, the single canopy layer `big leaf' approach for leaf-level photosynthesis (which computes radiation absorption with Beer's law). Soil heat conductivity was calculated using the approach of Dharssi et al. (2009). We used the default JULES-UKV soil layering (Supplement 1, Figure S1.1Figure 3). The hydraulic bottom boundary condition in JULES is free drainage.

JULES computes the transport of water through the soil using a finite difference representation of the Richards' equation. The vertical fluxes are computed with the Buckingham-Darcy equation. JULES-UKV uses the Mualem-Van genuchten (Van Genuchten, 1980; Mualem, 1983) soil water retention equations. The Van Genuchten equation calculates soil water content θ (m^3m^{-3}) from soil hydraulic pressure head ψ (m):

$$\frac{\theta-\theta_{res}}{\theta_{sat}-\theta_{res}} = \frac{1}{[1+(\alpha\Psi)^n]^{1-1/n}}, \quad (1)$$

with shape parameter n (-), α (m^{-1}) representing the inverse of the water entry pressure, θ_{res} (or smres; m^3m^{-3}) is the empirical residual soil moisture content (without physical meaning), and θ_{sat} (or smsat; m^3m^{-3}) is the saturated soil moisture content. In JULES parameter n is defined as b = 1/(n-1) and sathh=α^{-1} (m).

The Mualem equation computes the unsaturated hydraulic conductivity K:

$$K = K_{sat}\frac{\theta-\theta_{res}}{\theta_{sat}-\theta_{res}}\left[1-(1-\frac{\theta-\theta_{res}}{\theta_{sat}-\theta_{res}}^{1/(1-\frac{1}{n})})^{1-\frac{1}{n}}\right]^2, \quad (2)$$

where K_{sat} (or satcon; $mm \cdot s^{-1}$) is the saturated hydraulic water conductivity.

The values of the Mualem-Van Genuchten parameters need to be defined by the user for each grid cell/point based on soil characteristics.

In JULES soil moisture directly interacts with transpiration (through root water uptake) and bare soil evaporation as described hereafter (see also Figure S~~A~~1.~~1~~ 2 of ~~Appendix~~ Supplement 1). JULES first computes the potential photosynthesis, which is a function of three limiting factors: Rubisco limitation, radiation limitation, and photosynthetic product transport limitation. The potential photosynthesis is multiplied with a soil moisture reduction factor to obtain the actual photosynthesis. To obtain this soil moisture reduction factor the model first computes a limiting factor for each layer:

$$\beta_i = \begin{cases} 1, & \theta_i \geq \theta_{crit} \\ \frac{(\theta_i - \theta_{wilt})}{(\theta_{crit} - \theta_{wilt})}, & \theta_{wilt} < \theta_i < \theta_{crit}, \\ 0, & x \leq \theta_{wilt} \end{cases} \tag{3}$$

where θ_i is the unfrozen soil moisture content in layer i, θ_{crit} is the critical point soil moisture content below which soil moisture is limiting the root water uptake (matrix water potential -330 cm in JULES), and θ_{wilt} is the wilting point soil moisture content below which no root water uptake occurs (-15000 cm in JULES). These reduction factors are multiplied with the root density in the layer. These weighted reduction factors are then summed to obtain the root zone soil moisture reduction factor. From the actual photosynthesis the plant stomatal conductance is computed. Separately the bare soil surface conductance, which is a function of the soil moisture content in the upper soil layer and the critical soil moisture, is computed. The surface conductance is then computed as a function of the stomatal conductance and the bare soil surface conductance.

The potential evapotranspiration is also calculated separately. This variable is multiplied with the saturated land fraction to compute the free water evaporation (e.g. lake and canopy evaporation). The rest of the potential evapotranspiration is multiplied with the surface conductance to obtain the bare soil evaporation + plant transpiration. Together these fluxes are the actual evapotranspiration (water) or latent heat flux (energy). The amount of water drawn from the top soil layer through bare soil evaporation depends on the bare soil surface conductance. The distribution of the root water uptake between the layers depends on the weighted soil moisture limitation factor for each layer. The water extraction from the soil in its turn directly affects the soil moisture content in the different layers at the start of the next time step. These soil moisture contents then affect the soil moisture redistribution, surface runoff, and deep drainage.

JULES-UKV also requires a number of initial conditions: the amount of unfrozen water and snow stored on the surface (on canopy and on soil surface; set to zero in this study), snow properties (set to JULES-UKV defaults), the surface temperature (set to the air temperature of the hour before the first simulation time step), the soil temperature of each layer (set to the soil temperature from Ameriflux data the hour before initial time step) the soil water content in each layer (set to the soil water content from the PS observed moisture content of the hour before the first simulation time step and applied homogeneously throughout the profile). Soil moisture was spun up by running a maximum of five cycles and stopped when soil moisture convergence was lower than or equal to 10% compared to the previous cycle.

2.4.2 Calibration and validation approaches

At each site we calibrated JULES against observed PS observed soil moisture and against CRNS observed neutron counts respectively. We chose to calibrate simulated neutron counts against CRNS observed neutron counts using COSMIC

(Shuttleworth et al., 2013) to translate simulated soil moisture profiles into equivalent neutron counts.. ~~We chose to calibrate simulated neutron counts against CRNS observed neutron counts because it allowed us to take into consideration vertical heterogeneity in modelled soil moisture by computing neutron counts from modelled soil moisture profiles using COSMIC.~~

We computed the Root Mean Squared Error (RMSE) between simulated and observed hourly time series. To better match the observed soil moisture/neutron count time series, we calibrated five JULES parameters that influence the model soil moisture state (Figure S1.2 of Supplement 2 These included ~~three~~ Mualem-Van Genuchten shape parameter~~s~~: b, the water entry pressure parameter (sathh), and the saturation hydraulic conductivity (K_{sat}).~~satcon~~ The critical point (θ_{crit}) and wilting point (θ_{wilt})~~and~~ soil moisture content parameters from the evapotranspiration limitation factor were also calibrated~~limitation factor β parameters sm_{crit} and sm_{wilt}~~. We chose these parameters because they are, in theory, directly linked to the movement of moisture in the soil and to the effects of soil moisture on transpiration in JULES.

To assess the effects of calibration on soil moisture and surface energy flux simulation we compared the calibrated solutions against a default run at each site. The parameter values for the default case were derived from soil properties (percentages clay, loam, and organic matter) reported by the Harmonised World Soil Database (HWSD; FAO, 2009) for each of the twelve sites. These properties were used in the Wösten Pedotransfer function (~~Wösten PTF;~~ Wösten et al., 1997) to obtain values for b, sathh, and K_{sat}~~satcon~~. Parameter values for θ_{crit} and θ_{wilt}~~smwilt and smcrit~~ were subsequently obtained with the Van Genuchten formula.

The parameter calibration ranges were the same for all sites (Table 2). They were constructed by computing the minimum and maximum parameter values for the entire soil texture triangle (based on Wösten Pedotranser Function). Three organic matter contents were taken into consideration, yielding three triangles. Cay percentages above 70% were excluded to avoid extreme values for parameter b especially.~~They were constructed by computing the minimum and maximum parameter values for the entire soil texture triangle (based on Wösten PTF), while also taking into consideration three organic matter contents (yielding three triangles), but excluding clay percentages above 70% to avoid extreme values for parameter b especially.~~ The range for θ_{crit}~~smcrit~~ was set to 10-90 % of the saturated soil moisture content. The saturated soil moisture content was computed as a function of the dry soil bulk density (bd): θ_{sat}~~smsat~~ =1 – bd / 2.65 (Brady and Weil, 1996). Soil bulk density values obtained from the COSMOS network were used. This ~~thus~~ yielded different θ_{crit}~~smcrit~~ ranges in terms of soil moisture content (m^3m^{-3}) for different sites. The residual soil moisture content (defined implicitly in JULES) was set to zero because the Wösten Pedotransfer Function~~FT~~ does not consider it.

JULES' remaining two ancillary parameters; the soil heat conductivity and soil heat capacity, were for each site computed as a function of soil properties (HWSD) with De Vries' (1963) method. The bare soil albedo was set constant at 0.38 (-) for all sites. Plant Functional Type ~~(PFT)~~ parameters were set to JULES defaults except for the e-folding rooting depth (depth above which 86 % of plant roots are present) and the canopy height, at sites where more specific information was available from Ameriflux/COSMOS site information or from site specific literature.

We calibrated JULES using the BORG Multi-Objective Algorithm (BORG-MOEA or BORG; Hadka and Reed, 2013). This calibration tool was designed for multi-objective problems but also works for single-objective calibration. BORG employs multiple optimisation algorithms simultaneously to obtain convergence while also keeping the searched parameter space wide. The algorithm measures progress with the epsilon-progress technique, which uses the objective function~~ (OF)~~ space divided in boxes with sides of size epsilon. Epsilon is a user defined value for each objective function (we used epsilon values of 0.001 m^3m^{-3} for PS and 1 cph for CRNS, Kollat et al., 2012). Only if a new solution resides inside a box with a better objective function value, BORG considers it is progress. If no progress was obtained after 200 runs, the algorithm had stagnated. In this case the BORG algorithm triggers a restart, which consists of (among other techniques) changing population size to maintain a diverse population and to escape local optima. We used a maximum number of 3,000 runs and an initial population size of 100 runs.

~~2.4.3 Validation approach~~

As ~~main~~ validation metric we chose RMSE between the observed and simulated latent heat flux (LE). ~~Additionally, we computed the RMSE between the observed and simulated Evaporative Fraction EF = LE / (LE + H). The EF tells how the net surface radiation is partitioned between all LE and H fluxes.~~ ~~Additionally, we computed the RMSE between observed and simulated LE.~~ Because data quality issues with ~~EC~~ eddy-covariance data are often observed during ~~at~~ night time (Goulden et al., 1996; 2006; 2012; Aubinet et al., 2010), we computed these metrics over day time hours only. We defined day time as downward shortwave radiation > 20 Wm^{-2}. To avoid extreme RMSE-EF values, we used hours with both observed and simulated ~~H and~~ LE values ≥ 1 Wm^{-2} only. Otherwise a few hours with small LE or H values would have dominated the RMSE values, while this would probably have been due to forcing or ~~EC~~ eddy-covariance data inconsistencies and would not relate to soil moisture temporal variability.

~~This was not completely surprising because~~It is known that single-objective calibration is often insufficient to constrain parameters to simulate different states and fluxes well (Gupta et al., 199~~8~~9). In addition, Vereecken et al. (~~2015~~2008) argued that calibrating soil hydraulic parameters against soil moisture only does not guarantee better surface energy fluxes. To find out whether it was actually feasible to expect better surface energy flux simulation when soil moisture was improved we performed calibrations where we optimised the model for two objectives simultaneously. We employed the BORG algorithm to simultaneously optimise the RMSE of (day time) latent leat flux~~ RMSE~~ and the RMSE of all day soil moisture~~ RMSE~~ (using PS soil moisture and CRNS neutron counts separately). ~~Both optimisation experiments are defined as PS/LE and CRNS/LE.~~We analysed the trade-off between the two objective functions. We computed the compromise solution for each two-objective-calibration. Plotted within the normalised two-objective solution space, the compromise solution is the model run which has the smallest distance to the origin. This means no other solution can be obtained that yields a better approximation for one objective function without deteriorating the other.

3 Results and Discussion

3.1 Soil moisture Ddata analyses

3.1.1 Time series analyses

In Figure 4 Figure 2 comparison between PS and CRNS soil moisture time series shows that the seasonal trends of the two soil moisture observation products were similar. The two soil moisture products however also differed from each other, in distinct ways at different sites. PS soil moisture observations were systematically higher than CRNS soil moisture observations at eight of twelve sites. At three sites (DC, SO, CS) PS soil moisture dried down quicker than CRNS soil moisture, while at ME the opposite behaviour was observed. At KE, MM, TR, and MO PS showed a higher seasonality signal (up to 50% higher) than CRNS. Peaks in PS soil moisture were at three sites (UM, KE, TR) up to twice as high as in CRNS soil moisture. In addition, the CRNS data appears noisier than the PS data, which is an effect of inherent randomness in neutron radiation reaching the CRNS sensor element (Zreda et al., 2012). This effect was more pronounced for lower neutron intensity.

At UM PS was peakier (peak height up to 0.10 $m^3 \cdot m^{-3}$ compared to <0.05 $m^3 \cdot m^{-3}$). At DC PS dry down was more gradual compared to CRNS and reached soil moisture contents at least 0.01 $m^3 m^3$ above PS SM. At SO PS dried down quicker and was then systematically drier than CRNS. At KE PS was peakier (peak height up to 0.20 $m^3 \cdot m^{-3}$ compared to <0.10 $m^3 \cdot m^{-3}$ for CRNS) and systematically wetter. ME was characterised by a slower dry down during summer for PS compared to CRNS. At SR PS was systematically higher (~0.25 $m^3 \cdot m^{-3}$) than CRNS. At CS PS SM dried down quicker and then stayed at a systematically wetter soil moisture content (0.1 $m^3 m^{-3}$ higher) than CRNS soil moisture. At MM the seasonality was the main difference, with PS being up to 0.1 $m^3 m^{-3}$ drier in summer and 0.05 $m^3 m^{-3}$ wetter during winter. At TR PS was peakier and during wet periods (winter) systematically higher (up to 0.2m $m^3 m^3$) than CRNS. At AR PS was mostly wetter. Also at WR PS was systematically wetter (up to 0.10 $m^3 \cdot m^{-3}$). During spring PS dried down slower than CRNS. At MO PS was wetter as well (up to 0.20 $m^3 \cdot m^{-3}$) but also showed a higher seasonality (up to 1.5 times larger amplitude) than CRNS.

PS SM observations were systematically higher than CRNS SM observations at eight of twelve sites. In addition, the CRNS data showed high frequency variation, especially clear at SO, MM, TR, WR, and MO. This was an effect of inherent randomness in neutron radiation reaching the CRNS sensor element. This effect was more pronounced for lower neutron intensity, when there was relatively more hydrogen prevalence in the surroundings of the CRNS.

3.1.2 Similarity metrics

The differences seen in Figure 4 Figure 2 are also summarised in Figure 5Figure 3, which shows a gradual site to site increase in RMSD between observed PS and CRNS and soil moisture data series ($RMSD\text{-}SM_{obs}$). However, between WR and MO there is a sudden ~60% increase in $RMSD\text{-}SM_{obs}$. The systematic bias mentioned in the soil moisture time series analysis for SR reflects here in the large (~90%) contribution from the difference in mean soil moisture. The difference in seasonality mentioned in the time series analyses for MM reflects in the large contribution (~50%) from seasonality (i.e. standard deviation). The higher soil moisture content for PS at TR during winter shows up in the relatively large contribution (~60%)

~~from seasonality. The slower dry down at DC is reflected by the high contribution (~70%) from dynamics (coefficient of correlation).~~ Overall, bias contributed to 50% or more of the total error at seven out of twelve sites.

Additional analyses indicated that differences between the two soil moisture estimates could not be clearly related to any physical site characteristics besides the mean soil wetness.

~~Scatter plots of RMSD-SM_{obs} against mean and standard deviation of observed PS and CRNS soil moisture content (not shown) revealed that the difference between PS and CRNS soil moisture content is larger at wetter sites. A similar plot, with RMSD-SM_{obs} against mean and standard deviation of precipitation (also not shown) showed a weaker correlation. No correlation was found between MSE-SM_{obs} decomposition and wetness conditions.~~ Dominant vegetation type seemed not to have an effect on the similarity between the two soil moisture data products: both forested and grass sites~~sites~~ included those with relatively small RMSD-SM_{obs} ~~(UM, SO, ME)~~ and those with relatively high RMSD-SM_{obs} ~~(MM, WR, MO)~~. ~~The same holds for grass sites (KE, TR, AR).~~ Only bare/shrub covered sites (DC, SR, CS) were all below the sites' average RMSD-SM_{obs} of 0.05~~49~~ m^3m^{-3}. These four sites were however also relatively dry. No correlation between dominant land cover and decomposition was found. The grassy sites had the most even distributions between the three decomposition terms. Soil type and soil bulk density were also investigated for correlations with RMSD-SM_{obs}, but no trends were discovered.

Larger differences between PS and CRNS soil moisture could be expected at sites with more heterogeneous soil or vegetation. Static satellite photos of the sites from the COSMOS website did not indicate systematically more heterogeneous conditions at the sites where PS and CRNS soil moisture differed more. Site info (e.g. topography, presence of rocks) from COSMOS and Ameriflux did also not clearly show more horizontally heterogeneous soil properties for those sites.

~~Concluding, the differences between the two soil moisture products could not be clearly related to physical site characteristics besides the mean soil wetness.~~

~~3.1.3 Discussion of data analyses results~~

The fact that the soil moisture time series of PS and CRNS differed from each other in various ways could be related to a number of issues. First, PS sensor types and numbers of sensors differed between the sites (~~Figure~~ Table A1 of Appendix A). Second~~ly~~, the installation methods may also have been different between the sites. ~~Thirdly~~Finally, the exact installation locations of the PS sensors may in certain cases have been for instance next to a macropore, or near roots, while at other sites they were coincidently installed in a homogeneous soil patch.

The presence of hydrogen pools other than soil moisture (e.g. biomass, intercepted water, and water in litter layer) also affect the observed CRNS neutron count. Because different hydrogen pools are more present at certain sites than at others, the uncertainty on neutron count observations varies between sites. The results did however not show effects of land cover and soil properties on the similarity between the two soil moisture products. Another factor is that, the quality of the calibration of COSMIC could possibly be different for different sites. Finally, at multiple sites, the PS and CRNS soil moisture time series were quite similar. This could be expected at rather homogeneous sites. Moreover, Köhli~~Kohli~~ et al. (2015) suggested the

CRNS footprint to be around 150-200 m instead of 300 m as reported by Desilets and Zreda (2013). In that case the differences between the two soil moisture observation techniques could be smaller than initially thought.

~~**3.1.4 Limitations of the data comparison methodology**~~

We derived vertically constant CRNS soil moisture values from observed neutron counts with COSMIC. This method contains inherent uncertainty because in reality soil moisture is often not vertically homogeneous. ~~Comparing these derived CRNS soil moisture values with PS soil moisture data integrated to certain layers (0-10cm and 10-35cm) is therefore a deviation from field conditions. PS installation depths differed between sites and CRNS measurement depth varied between sites and over time. The comparison might therefore have been `more valid' at some sites than at others.~~

The different lengths (between sites) of the time series used can have affected the metrics and sites' ranking. However, we used at least one year of data at all sites.

For these reasons these data analyses should be seen as a first simple comparison of the two data types only. The outcomes of the calibration (against PS and CRNS) and validation provide insight in the effects of the differences between the two soil moisture products on JULES' surface energy flux partitioning and latent heat flux simulation.

3.2 Single objective ~~C~~calibration against soil moisture observations

~~Calibration reduced the objective function (RMSE-SM) values in all cases (Figure 6~~). ~~However, t~~The degree to which the objective function (RMSE-SM; ~~–~~ Figure 4) values decreased ~~this happened~~ differed between sites and the two calibration strategies (PS or ~~/~~CRNS), with decreases of 21% (AR-PS) to 93% (UM-CRNS). While the errors of the default runs existed mostly of systematic bias, after calibration the difference in dynamics was the largest source of uncertainty and in 16 out of 24 cases this contribution actually increased in absolute terms. ~~This happened because peaks and valleys became less extreme after calibration. Higher peaks yield higher correlation (with scatter plot slope > 1) because the relatively higher peaks within a time series differ more clearly from smaller (noise) peaks (e.g. SR-PS, WR in Figure 7). When the peaks are then reduced by calibration, the high peaks do not stand out as much anymore and the correlation decreases.~~The increase in difference in dynamics was due to the selected objective function (RMSE), which reduces the mean error between modelled and observed data. Previous research (e.g. Teuling et al., 2009) has shown calibrating soil parameters has a large effect on simulated absolute soil moisture values (bias), but substantially less on soil moisture seasonality and dynamics. Our finding supports this.

The RMSE values reduced relatively more for the CRNS~~PS~~ calibration (70% on average over the twelve sites) than for the ~~CRNS~~ PS calibration (55% on average). The calibration method could possibly explain this. CRNS calibration was against observed neutron counts, while PS calibration was against observed soil moisture contents. Because neutron counts have an inverse relationship with soil moisture content, the PS calibration was possibly governed by avoiding larger errors occurring during a few brief soil moisture peaks. While focussing on getting the fitting for those peaks right, the PS calibration neglected the smaller errors during dry periods. This would then result in relatively smaller decrease in RMSE values than for the CRNS calibrations because those were fitted with heavier weights to the drier periods.

~~Figure 6~~ Figure 4 also shows that the relative improvement was not systematically lower or higher for sites with higher similarity between the two observed soil moisture time series (actually the largest improvement was for the CRNS calibration at UM). Therefore, it appears that ~~T~~the quality of the default runs was hence ~~maybe simply determined~~predominated by the quality of the chosen default parameter values.

~~Figure 7~~Figure 5 shows soil moisture time series for four selected sites: UM was chosen because PS and CRNS soil moisture were most similar there, SR was a site with moderate difference, and at WR and MO PS and CRNS soil moisture were most different. ~~The~~ Simulated soil moisture ~~time series approached the observations better in all cases,~~improved especially at UM and WR, where the default runs overestimated soil moisture contents. ~~The s~~Simulated soil moisture dynamics became more similar to both~~at WR approached those of~~ PS and CRNS observed soil moisture dynamics~~closely~~, even though~~while thes~~e observed soil moisture dynamics differed from each other substantially at sites WR and MO.

3.3 Validation of the single-objective calibrations against eddy-covariance observations

While calibration errors decreased for soil moisture, ~~the~~ latent heat flux~~Evaporative Fraction~~ estimation improved for ~~eleven~~ fourteen out of twenty-four calibrations (~~Figure 8a~~Figure 6) ~~and we observed similar results for latent heat flux (Figure A2.1 of Appendix 2)~~. This ~~basically~~ means that an improvement in simulated soil moisture did not necessarily lead to better estimation of surface energy fluxes. Calibration against PS soil moisture improved RMSE-LE at six sites (UM, KE, DC, SR, SO, and ME), while calibration against CRNS neutron counts improved RMSE-LE at eight sites (KE, DC, SR, ME, SO, AR, MM, and WR). At five sites RMSE-LE improved after calibration against both PS soil moisture and CRNS neutron counts. Calibration yielded lower RMSE-LE after calibration against CRNS neutron counts than after calibration against PS soil moisture at all but three sites (UM, KE, and WR). Figure 6 also shows that RMSE-SM decreased substantially less (i.e. >20% difference) after calibration against PS soil moisture than after calibration against CRNS neutron counts (i.e. >20% difference between both strategies on the horizontal axis of Figure 6, which occurred at six sites). At five of these six sites the relative change in surface energy flux performance was also smaller (sites MO, AR, TR, ME, and SR). This indicates that further improvement in soil moisture simulation after calibration against PS data could have yielded better surface energy fluxes.

~~In fact, the magnitude of those improvements (EF: 3% to 30%; LE: 1% to 37%) were comparably to the deterioration from other calibration cases (EF: 2% to 28%; LE: 1% to 29%). The results also showed no systematically greater improvement across all sites for either of the two calibration strategies. Compared to the default case, CRNS calibration yielded lower RMSE-EF at four sites while PS calibration yielded better EF at eight sites~~.

In ~~Figure 8~~Figure 6 ~~b~~ 10% change in latent heat flux ~~EF~~ estimation was chosen to distinguish substantial change from non-substantial change, derived from the approximate error in eddy-covariance~~EC~~ ~~sensible heat flux~~ data (Sellers and Hall, 1992; Finkelstein and Sims, 2001). ~~It shows that i~~In just ~~three~~ four cases for PS calibration and ~~three~~ five cases for CRNS calibration the improvement in latent heat flux was actually substantial. Using this threshold also reveals that for calibrations with a more than 60% change in RMSE-SM, RMSE-LE did not change substantially in six cases. This again shows that a change in simulated soil moisture did not necessarily mean a substantial change in surface energy flux simulation. Analysis of the RMSE of evaporative fraction (EF=LE/(LE+H)), which shows the ability to simulate surface energy partitioning, yielded similar overall results as our analysis of the RMSE of latent heat flux (Supplement 2, Figure S2.1).

One factor causing some of these limited improvements was that, when mean simulated root zone soil moisture (weighted with root density) increased after calibration, the values of the wilting point and critical point soil moisture parameters moved along (data not shown), yielding similar soil moisture stress. This happed for both calibration strategies at site KE, and for the calibration against CRNS neutron counts at sites MO and TR. This could relate to the limited value of simulated absolute soil moisture for surface energy flux estimation in land surface models (Dirmeyer et al.et al., 2000; Koster et al., 2009). However, we also found the distance between wilting point values and critical point values to decrease after calibration. This occurred with a simultaneous decrease in standard deviation of the simulated soil moisture. The self-adjusting behaviour of the wilting point and critical point parameters was also indicated by parameter sensitivity analysis (Appendix 3), which showed soil moisture was substantially more sensitive to a change in critical point value than latent heat flux.

Another issue, which occurred for instance for the PS calibration at SO and for the CRNS calibrations at site AR, was that while surface energy flux estimation improved for a certain period, it deteriorated for another period (data not shown). A third cause of limited improvement in surface energy flux estimation occurred for example at site ME (data not shown). Soil moisture stress at ME was relatively limited (beta mostly above 0.6) and during periods of soil moisture stress, the latent heat flux was dominated by different stress factors.One cause was that even when root zone soil moisture changed considerably (not shown), soil moisture stress changed little (self-adjusting behaviour, Figure 13), and hence there was no considerable change in surface energy partitioning. We found this to be the main reason for KE, MO-CRNS, and TR-CRNS.

At WR and CS, PS calibration improved EF most, while at these same sites CRNS calibration did not yield improvement. Finally, larger differences between the two observed soil moisture time series did yield more different simulated surface energy flux time series at some sites (MO and WR compared to UM and DC) (e.g. UM compared to MO) while at other sites it did not (e.g. TR and AR compared to UM and DC)..

To see if validation results would be different if only those periods during which soil moisture stress occurred were evaluated, we also analysed the performance over these periods only (data not shown). At sites DC, KE, SR, CS, MM, and AR integrated root zone soil moisture content ($SM_{root\ zone}$) During these periods plant water uptake limitation factor β waswas all the time (>93%, not shown) below the critical point soil moisture (smcrit)below one (data not shown). The surface energy partitioning time series and RMSE values of the other sites were additionally analysed over periods with $SM_{root\ zone}$ below the critical point. Only at site SO did we see somewhat better performance during these periods compared to periods with $SM_{root\ zone}$>smcritthe original validation period, after both PS and CRNS calibration.

We explored trends between relative improvement in surface energy flux estimation and soil wetness, precipitation, vegetation type, vegetation height, and soil characteristics. No clear trends were discovered. The only feature that could be distinguished was that for the two sites with a bare soil tile (DC and SR) PS and CRNS calibration improved EF and LE. Rooting depth did also not explain relative improvement in surface energy flux estimation. Finally, larger differences between the two observed soil moisture time series did generally yield more different simulated surface energy flux time series, with a clear exception for site UM, where the two calibration approaches yielded more than 20% difference in change in RMSE-LE.

In ~~Figure 9~~ Figure 7 the monthly mean diurnal latent heat cycles of four sites (UM, SR, WR, and MO) are shown (night time data included but not used for calibration as discussed previously). Both calibration strategies yielded overestimation in March and April and underestimation from June to August at site UM. A month long gap in longwave and shortwave downward model forcing data occurred in April 2012. This might have affected the results for this site.~~The PS calibration for site UM yielded, compared to the default run, a higher overestimation of latent heat flux during winter and spring, while in summer and autumn no large change occurred. The CRNS calibration yielded a higher underestimation during summer.~~ At SR, both PS and CRNS calibration improved latent heat flux during periods of low evapotranspiration, while during the other periods only the CRNS calibrations yielded better results. AT WR calibration against PS soil moisture yielded overestimation of ~~overestimation of LE~~ latent heat flux~~decreased after PS calibration in all seasons~~, while CRNS calibration yielded too ~~high~~ low ~~LE,~~latent heat flux~~ especially during summer~~. The PS calibration at MO yielded ~~LE~~ latent heat flux underestimation~~ during summer while it improved during winter~~, whereas CRNS calibration did not change LE substantially~~ during any season~~.

In summary, the single-objective calibrations against CRNS neutron counts yielded larger improvements in simulated surface energy fluxes than single-objective calibrations against PS soil moisture (Figure 6). Improvements in surface energy flux estimation were however substantial for four calibrations against PS soil moisture and five calibrations against CRNS neutron counts. Limited improvements in surface energy flux estimation after calibration could partly be attributed to the limited value of absolute soil moisture to estimate surface energy fluxes with land surface models. This seems reasonable because calibration mostly affected absolute soil moisture (Figure 4). This result corresponds with earlier research that showed model soil moisture dynamics and seasonality have a larger effect on surface energy flux simulation than absolute soil moisture (e.g. Teuling et al., 2009; Dirmeyer et al., 1999). Previous research has also indicated that soil moisture alone is insufficient to estimate soil hydraulic parameters (Vereecken et al., 2008; 2015). ~~Vereecken et al. (2015) commented that even improved land surface models in combination with better soil moisture observations do not necessarily yield correct latent and sensible heat flux estimation.~~ Our findings support this for some sites~~.~~

To better understand these implications, the two-objective simulations against soil moisture and latent heat flux simultaneously (discussed in the next section) provides further insight in these results.

~~3.4 Were calibrated parameter values physically feasible?~~

~~If parameter values obtained after calibration were not physically feasible (e.g. representing a sandy soil while there was a clay soil) then, if model structure is assumed to represent biophysical processes sufficiently well, that could yield undesirable results. We investigated whether the parameter values obtained were `realistic' by seeing if they were substantially different from the soil conditions reported by HWSD (used to obtain default values), COSMOS, and Ameriflux. We analysed the calibrated parameter values from the single-objective calibrations for each site with the parallel coordinate plots in Figure A2.3 of Appendix 2. Calibration yielded values far from default for sensitive parameter b in two cases (MM-PS and TR-PS) only.~~

The saturation hydraulic conductivity (satcon) took on substantially different values in eight cases. The critical point and wilting point soil moisture parameters assumed values far from the defaults in three and four cases respectively. The wilting point multiplier, which was actually calibrated, took a wide range of values, but the actual wilting point parameter stayed closer to the default values in most cases.

Although the parameter calibration range of sathh was rather wide compared to the range of soil types from HWSD at the twelve sites, the values were on, or very near towithin the boundaries at all sites for the default runs and in thirteen cases after the calibrations. The saturated hydraulic conductivity was on within the edges for 22 runs. It should be noted however that, especially for the saturated hydraulic conductivity, the parameter calibration range was non-linear, with a factor 1000 difference between the upper and lower boundaries.

COSMOS/Ameriflux soil information (Table 1) for UM suggests sandier conditions, which would yield higher saturation hydraulic conductivity, which was indeed obtained after PS calibration. At SO calibration yielded higher values for the saturation hydraulic conductivity and a lower value for b, which is in line with the thick organic layer occurring there. At the KE site, the soil was reported to be coarser (and also containing stones) than actually reported by the HWSD. This would mean higher saturation hydraulic conductivity, but a higher values was not obtained after calibration. At SR and MM the soil was reported to be of finer texture than the data from HWSD. Lower satcon values than the defaults were however not obtained. At TR a clay hardpan was found at 30-40 cm, which would impede drainage and calibration could hence be expected to yield lower satcon. This did however not happen after our calibrations. AR and WR were reported to be sandy, in contrast with the loam soils reported by HWSD. Higher satcon values were found after CRNS calibration at AR and after both calibrations at WR.

We could think of a range of reasonSeveral reasonss have been proposed why apparently non-physically realistic parameter values were found, for instance:

- Horizontal and vertical soil heterogeneity (e.g. macro-pores, soil layering in organic matter and particle sizes, stones) was not (properly) accounted for in the model
- The vertical discretisation of the soil (layers of 10, 25, 65, and 200 cm) is not suitable for solving Richards' Equation (Beven and Germann, 2013)
- Richards' Equation applicability at larger horizontal scales is questionable for multiple reasons (Beven and Germann, 2013).
- JULES' bottom boundary condition (free drainage) may not have been suitable in all cases. We used Figure 3 from Fan et al. (2013) to get an idea of the water table depth. This did not indicate the presence of shallow groundwater tables (< 5m), which would have made free drainage unsuitable as bottom boundary condition. However, the resolution of the used map was rather coarse for our purpose. Moreover, other conditions, such as perched groundwater tables and shallow hard rock could be present and impede free drainage.
- The wilting point and critical point soil moisture contents were computed as a function of soil properties only; pressure heads of -16000 cm and -300 cm respectively were used regardless of vegetation type. More complex assumptions are however made in more complex models. For instance, a common soil water stress function, similar to the one used in JULES, is the Feddes function (Feddes et al., 1978). This function can take into account two different critical soil moisture content values; one for high potential transpiration and one for low potential transpiration. Moreover, the critical soil moisture pressure (-300 cm in JULES) can take different values for

Formatted: Normal, Space After: 0 pt, Line spacing: single, No bullets or numbering

~~different vegetation types. Taylor and Ashcroft (1972) for instance reported values between -150 cm and -15000 cm for crops and grass. The results presented in Section 4.3 indicated these two parameters might govern the soil wetness importantly in JULES.~~

- ~~Physically realistic parameters often provide unrealistic results (Gupta et al., 1998).~~

3.~~5~~ 4 Two-objective calibration against soil moisture and latent heat flux

The results from the two-objective calibrations suggest that differences in latent heat flux estimation improvement between compromise solutions for both two-objective calibration strategies were limited (Figure 8). Only at three sites were these differences more than 5% (at sites SR, DC, and UM). Improvements were substantial (according to our 10% threshold for improvement in RMSE of latent heat flux) for six calibrations with PS soil moisture and for eight calibrations with CRNS soil moisture.

Scatterplots of normalised RMSE of latent heat flux (LE) and normalised RMSE of soil moisture for all sites are shown in Figure 9 and Figure 10 for the PS and CRNS two-objective calibration strategies respectively. The RMSE values were normalised with respect to the default solutions, which therefore have normalised RMSE values of 1 (-). The black dots represent individual model runs and the default model run for each site is represented by a red cross The single-objective calibration solution of each single-objective calibration solution is shown as a blue triangle and the compromise solution of each two-objective calibration is shown as a green triangle. These two figures showed that the differences between the compromise solutions of the two calibration strategies observed for SR, DC, and UM in Figure 8, were less meaningful than analysis of Figure 8 only showed.

These findings are based on the shapes of the black point clouds in Figure 9 and Figure 10. We first look at the left edges of the black point clouds. When these edges are close to vertical, a small deterioration in RMSE-SM or RMSE-N (e.g. less than 0.05 (-)) would yield a large deterioration in normalised RMSE-LE (e.g. greater than 0.2 (-)). We observed this for three calibrations with PS soil moisture (SO, AR, and WR; indicated with pink line for WR) and in all these three cases, a less than 0.05 (-) change in normalised RMSE-SM would have yielded worse simulated latent heat flux than with the default parameter set. Seven calibrations with CRNS neutron counts (UM, DC, SO, KE, SR, CS, and WR; indicated with pink line for WR) showed a close to vertical edge. In four cases (UM, DC, KE, and CS) this would have yielded worse simulated latent heat flux than with the default parameter set. A negative slope for this side of the point cloud means an improvement in soil moisture estimation would mean a deterioration in latent heat flux estimation. We observed such negative slope for four calibrations with PS soil moisture (CS, MM, TR, and MO; indicated with continuous black line for MO) and for two calibrations with CRNS neutron counts.

Next, we look at the lower edges of the point clouds in Figure 9 and Figure 10. When this edge is horizontal, this means good latent heat flux estimation can be obtained for a wide range of soil moisture estimation performances (for instance site ME, indicated with pink lines in both figures). When this edge has a negative slope, this means the best latent heat flux estimation is obtained for worse soil moisture (for instance site TR, indicated with pink lines in both figures). We observed these two features for all two-objective calibrations. Single-objective calibration against latent heat flux would however only have necessarily yielded worse soil moisture than the default parameter set for six calibrations with respect to PS soil moisture (sites UM, KE, SR, TR, AR and WR) and two calibrations with respect to CRNS neutron counts (sites TR and AR). The implication of these results is that the quality of latent heat flux simulation did not depend strongly on the quality of soil moisture simulation.

In summary, the two-objective calibrations against soil moisture (or neutron counts) and latent heat flux showed fewer substantial differences between calibration with PS soil moisture and calibration with CRNS neutron counts. These results indicate that the differences between both *single*-objective calibration strategies, which showed an advantage for CRNS observations, were possibly not as substantial as it seemed at first. The spatio-temporal stability theory, which implies limited spatial variability in surface energy fluxes, could be one explanation for this (Vachaud et al., 1985; Teuling et al., 2006; Mittelbach and Seneviratne, 2012; Albertson and Montaldo, 2003). Another factor that has possibly played a role is that the spatial scale advantage of the CRNS was masked out by the possibly lower quality of its measurements (see also Section 3.1). Different hydrogen pools than soil moisture affect the neutron measurements at various temporal resolutions (e.g. rainfall interception; Baroni and Oswald, 2015). However, PS sensors also have their limitations. Different (electromagnetic) PS sensors have different designs and properties, which affect the data quality (Robinson et al., 2008; Blonquist et al., 2005). Soil type and soil specific calibration also affect the accuracy and precision of the PS data. For instance, the relationship between electrical permittivity and soil moisture content is strong for quartz rich soils, but less accurate for clay soils (e.g. Ishida et al., 2000; Robinson et al., 2008). An issue that could have had effect on our results is the occurrence of gaps in the model forcing data. Gaps in observed meteorological data are often inevitable and must be filled for a land surface model to be run. Percentages of missing hours differed between 0% and 15% at our sites. Gaps larger than fifteen days filled with the moving window gap-filling procedure occurred mainly for downward shortwave and/or downward longwave radiation; at sites UM, KE, SR, DC, TR, and CS. At site WR a data gap of 29 days in precipitation and air pressure occurred. Especially the gap in precipitation data can have negatively affected the model results at this site. In this study we used PS sensor data in the upper 30 cm of the soil only, even though data from deeper sensors was publicly available at two sites (WR and MO). This choice can have provided a disadvantage to the PS data in our comparison, especially at sites with deeper roots and during soil moisture limiting conditions in the upper soil. Investigating the role of deeper roots was however beyond the scope of this study. Our goal was to compare the effects of the difference in horizontal footprint on latent heat flux simulation after using the two different measurement techniques to calibrate model parameters.

The single-objective calibration against PS and CRNS soil moisture did in thirteen of twenty-four cases not yield better surface energy fluxes and in only six cases was improvement substantial. This was not completely surprising because single-objective calibration is often insufficient to constrain parameters to simulate different states and fluxes well (Gupta et al., 1998). In addition, Vereecken et al. (2015) argued that calibrating soil hydraulic parameters against soil moisture only does not guarantee better surface energy fluxes. To find out whether it was actually feasible to expect better surface energy flux simulation when soil moisture was improved we performed calibrations where we optimised the model for two objectives simultaneously. We employed the BORG algorithm to simultaneously optimise daytime latent leat flux RMSE and soil moisture RMSE (using PS and CRNS separately). Both optimisation experiments are defined as PS/LE and CRNS/LE.

The results from the PS/LE and CRNS/LE calibrations are shown in Figure 10 and Figure 11 respectively. From the results we learned that for five of the 12 PS calibrations and four of the 12 CRNS calibrations a substantially better LE estimation than obtained with the single-objective calibration could have been obtained while maintaining a similar soil moisture error (black points, representing all two-objective calibration runs, were present right below the single objective calibration solutions).

The shapes of the Pareto fronts did not suggest automatic improvement in LE estimation with improved soil moisture estimation except in a few cases (e.g. DC in higher RMSE-SM range and SO). To have obtained such automatic improvement, the lower edges of the point clouds should have shown positive correlations.

We plotted the compromise solutions (the runs which did relatively well for both RMSE-SM/N and RMSE-LE) on top (green triangles) to see if a two-objective calibration would have led to improved soil moisture and latent heat flux. This was the case for all PS/LE compromise solutions except at TR and for all CRNS/LE compromise solutions except at UM. This suggests that two-objective calibrations could have yielded improved soil moisture and latent heat flux. This is also clear from Figure 13, where the relative improvements for the compromise solutions with respect to the default solutions are shown.

The large differences in EF/LE simulation between the PS and CRNS single-objective calibrations at WR and MO seem, based on Figure 10 and Figure 11, mostly a coincidence of which best solution was chosen. At WR a CRNS-calibration solution with the same neutron count/soil moisture RMSE but better RMSE-LE could have been obtained, yielding LE performance similar to the PS-calibration. At MO this was true for the PS calibration. At other sites where we saw differences between PS and CRNS (e.g. MM), the Pareto fronts of Figure 10 and Figure 11 suggest these differences could as well have been quite limited. Figure 12 supports this conclusion because at all sites except SR (where the CRNS compromise solution was clearly better with respect to LE) differences in LE simulation between the compromise solutions were similar between PS/LE and CRNS/LE calibration.

Our results (Figure 10 and Figure 11; especially the rather horizontal lower edges of the point clouds in those figures), suggest that the coupling between soil moisture and latent heat flux was generally quite weak in JULES, even at sites where such coupling is expected to be relatively strong (e.g. DC, SR, KE, AR). Moreover, differences between surface energy flux simulation of PS and CRNS calibrations were minimal.

3.6 To which parameters were soil moisture and latent heat flux most sensitive?

We investigated if a change in some parameters had a relatively small effect on soil moisture, while at the same time having a large effect on latent heat flux. In such case calibration could yield better soil moisture, but the parameter value might be inappropriate for latent heat flux. The inverse (influential on SM but not on LE) could also occur. We explored this by performing a sensitivity analysis with Morris' method (Morris, 1991) as implemented in the SAFE Toolbox (Pianosi et al., 2015) on the exact same parameter value ranges as used during the calibration. We computed the sensitivity indices (mean and standard deviations of the elementary effects) on the RMSE values (i.e. our Objective Functions; OFs) of simulated vs. observed PS soil moisture, simulated vs. observed CRNS neutron counts, and simulated vs. observed latent heat flux (day time only).

The results (shown for four sites in Figure A2.2 of Appendix 2) were consistent across most sites: all three OFs were most sensitive to changes in parameter b and least sensitive to the wilting point multiplier. Finch and Haria (2006) also found JULES parameter b to be most influential on soil moisture and latent heat, at a UK chalk site. The critical point soil moisture content was influential with respect to soil moisture / neutron counts but had at all sites less effect on latent heat flux. The lack of effect from the wilting point soil moisture content can probably be attributed to the use of the multiplier, despite being a common approach (Prihodko et al., 2008; Rosolem et al., 2012); a certain multiplier value could be good in combination with a certain value for the critical point but not for a different critical point value.

3.7 Could JULES model structure explain the limited improvement in surface energy flux estimation?

We found multiple causes for the lack of improvement in surface energy fluxes after calibration against soil moisture data and the quite weak coupling between JULES soil moisture and latent heat flux seen from the single and two-objective calibrations (PS-SM/LE and CRNS-N/LE). One cause was that even when root zone soil moisture changed considerably (not shown), soil moisture stress changed little (self-adjusting behaviour, Figure 13), and hence there was no considerable change in surface energy partitioning. We found this to be the main reason for KE, MO-CRNS, and TR-CRNS.

Field Code Changed

AR-CRNS, ME-CRNS, SR-PS, and SO-PS however did yield change in soil moisture stress (Figure 13), but other factors limited improvement in EF/LE. AR-CRNS, SR-PS, and SO-PS did yield different EF and LE timeseries, but because performance deteriorated during certain periods but improved during other periods, overall performance did not change. For instance, the CRNS calibration at AR yielded better EF and LE during wet periods with high LE, while during subsequent periods of drying the estimation was worse. EF and LE estimation after PS calibration at SR was better during dry periods while it was worse during wetter periods (monsoon). PS calibration at SO improved EF and LE during dry periods after summer while the simulation was worse in spring. Soil moisture stress at ME was relatively limited (beta mostly above 0.6) and during periods of soil moisture stress, the latent heat flux was dominated by different stress factors.

At UM, CS, MM, WR, and MO, EF and LE were substantially different between single-objective PS and CRNS calibrations. Figure 13 shows that in these cases the soil moisture stress was very different. At UM, CS, and MO, root zone soil moisture was similar (not shown) between the PS and CRNS calibration solutions, causing different stress due to different wilting point and critical point soil moisture, while at MM and MO, the root zone soil moisture actually changed considerably.

These results indicate the relatively large effect of the wilting point and critical point soil moisture parameters on the calibration results. The self-adjusting behaviour can yield similar soil moisture stress for different root zone soil moisture, while in other cases the wilting point and critical point values can be such that simulated soil moisture is close to the observations but root zone soil moisture stress is different. With respect to the two-objective calibrations the self-adjusting behaviour contributed to the possibility of having highly different soil moisture but similar latent heat flux performance.

3.8 Discussion

Our results suggest a relatively weak coupling between soil moisture and evapotranspiration in JULES. In combination with the self-adjusting behaviour of wilting point and critical point soil moisture values this suggests that how soil moisture observations are used to calibrate JULES should be carefully considered beforehand. In the Land Surface Modeling, community it is known that the absolute value of soil moisture content has limited information content for the model (Dirmeyer et al., 2000; Koster et al., 2009). However, we also found the space between wilting point values and critical point values to decrease after calibration. This occurred when the standard deviation of the simulated soil moisture time series decreased due to calibration.However, we found the wilting and critical point soil moisture parameters not only to move up or down after calibration, but also to move closer when the standard deviation in simulated root zone soil moisture content decreased. Our findings might also have implications for data assimilation. When certain critical and wilting point soil moisture parameter values are chosen that do not match the magnitude of the assimilated soil moisture, soil moisture stress could become too low or too high.Our findings might also have implications for data assimilation if certain critical and wilting point soil moisture parameter values are chosen that do not match the magnitude of the assimilated soil moisture.

Finally, the potential differences in terms of contribution to model performance observed between the PS and CRNS sensors are undermined by the weak coupling in the current model structure of JULES. Namely, even at sites where the two observed soil moisture time series differed most, surface energy partitioning simulation did not differ substantially between PS and CRNS calibration. As mentioned in Section 2.3 the calibration approach differed for the two soil moisture products; we calibrated against PS soil moisture and CRNS neutron counts. We chose to calibrate against neutron counts because it is the way in which CRNS data could be used in the context of LSM when the COSMIC operator is used. If we had calibrated against representative CRNS soil moisture values, we might have found worse fits, more similar to those of the PS calibrations.

Previous research has indicated that soil moisture alone is insufficient to estimate soil hydraulic parameters (Vereecken et al., 2008). Vereecken et al. (2015) commented that even improved land surface models in combination with better soil moisture observations do not necessarily yield correct latent and sensible heat flux estimation. Our findings support this.

4 Conclusions

We investigated whether reducing the spatial scale mismatch between the surface energy flux data and soil moisture data could be reduced through the use of Cosmic-Ray Neutron Sensors (CRNS). Five soil- and evapotranspiration parameters of LSM JULES were calibrated against Point Scale (PS) and CRNS soil moisture data separately, for twelve COSMOS/Ameriflux sites with different climate, land cover, and soil properties. Next, at each site, the improvement in latent heat flux surface energy partitioning and latent heat fluxes for the two calibration solutions was assessed by comparing the fit with EC eddy-covariance data and a version of LSM JULES runs with default parameter values based on a widely used soil database. Before the calibrations we compared the observed soil moisture data from the two sensor types. These analyses showed the differences between PS and CRNS soil moisture happened in different ways at the investigated sites. While at certain eight sites there were mainly systematic biases, at three other sites the seasonality was mostre different, while and at one other sites different time series dynamics were the main cause of differences between the two soil moisture observations. We found the difference between the two soil moisture products to be larger at wetter sites.

The single-objective calibration of JULES parameters against point-scale soil moisture and Cosmic-Ray Neutron Sensor neutron counts did not necessarily yield an improvement in model surface energy partitioning and latent heat flux simulation. The analysis of these sSingle-objective (soil moisture) calibrations and multi-objective calibrations (against (1) PS soil moisture and latent heat flux and (2) CRNS neutron counts and latent heat flux) revealed that differences between calibrations with these two soil moisture observation methods did overall not yield substantially different surface energy flux estimations. did not yield substantial differences between PS sensor and CRNS calibrations in simulating latent heat flux at eleven of the twelve sites. Moreover, at sites where the observed soil moisture time series from the two observation techniques diverged more, the differences between the resulting surface energy fluxes were not larger. These outcomes did not provide sufficient evidence to reject our null-hypothesis "*reduced scale mismatch does not lead to LSM flux estimates closer to eddy covariance observations*". The spatio-temporal stability theory in soil moisture, on which our hypothesis was based, can possibly explain the limited differences in surface energy flux estimation. This theory implies that spatial variability in surface fluxes is relatively limited within an eddy-covariance tower footprint. Therefore simulated surface energy fluxes after calibration against point scale soil moisture data would not necessarily be better than simulated surface energy fluxes after calibrations against CRNS soil moisture data. Another factor, related to this and that contributed to this result, was that calibrating soil parameters had mostly an effect on absolute soil moisture values rather than soil moisture dynamics. Soil moisture dynamics have a greater effect on surface energy flux simulation in land surface models than absolute soil moisture values do. Related to this we observed that after calibration the wilting point and critical point soil moisture parameter values adjusted themselves in a similar way as the root zone soil moisture did, yielding similar soil moisture control on transpiration despite changes in soil moisture values. In other cases we found calibration against soil moisture to improve surface energy fluxes during certain periods, but to deteriorate surface energy fluxes during other periods, yielding similar overall performance. Yet in other cases evapotranspiration was not limited by soil moisture stress. The potential scale advantage of the CRNS was possibly masked out by the possibly lower measurement quality of this sensor because other hydrogen pools than soil moisture affect the neutron count observations. Future use of CRNS soil moisture data could however benefit from improved knowledge on the effects of additional hydrogen pools (e.g. Baroni and Oswald, 2015) and of the sensor footprint (Köhli et al., 2015). The assumed benefits of a larger footprint of the Cosmic-Ray Neutron Sensor were not clear to JULES due to its weak coupling behaviour observed in this study. We found a few causes of the lack of improvement: (1) the wilting point and critical point soil moisture moved up or down in a similar way as the root zone soil moisture did, yielding similar transpiration for different soil moisture conditions; (2) calibration against soil moisture improved surface energy fluxes during certain periods, but deteriorated surface energy fluxes during other periods, yielding similar overall performanceIn this study, our results are conditioned to a single land surface model (JULES). For additional understanding of the importance of both PS and CRNS measurements for simulated surface energy quantities can be extended to other models in the future.

~~; (3) soil moisture was not the main surface energy partitioning controlling factor~~.

~~Because our findings indicate that the coupling between model soil moisture and evapotranspiration in JULES is somewhat weak, we recommend to improve the representation of the related land surface processes at sub-kilometre scale; the spatial scale which is consistent with Cosmic-Ray Neutron Sensor and Eddy Covariance flux data.~~

Acknowledgements

This research was supported by the Queen's School of Engineering (University of Bristol) Ph.D. scholarship and the by Engineering and Physical Sciences Research Council (EP/L504919/1). Additional support for this work was also provided by the Natural Environment Research Council (A MUlti-scale Soil moisture-Evapotranspiration Dynamics study (AMUSED); grant number NE/M003086/1). Funding for Ameriflux data resources was provided by the US. Department of Energy's Office of Science. We would like to thank the investigators of the sites used for providing us with site information through personal communication. The authors would like to thank Jos van Dam, referee Ryan Teuling, and two anonymous referees for providing constructive comments.~~.~~ Finally, the authors would like to thank the Editor, Harrie-Jan Hendricks Franssen, for guiding the revision process.

References

Albertson, J. D., and Montaldo, N.: Temporal dynamics of soil moisture variability: 1. Theoretical basis, Water Resour. Res., 39, 1274, doi:10.1029/2002WR001616, 2003.

Anderson, R.G., Goulden, M.L.: Relationship between climate, vegetation, and energy exchange across a montane gradient. J. Geophys. Res., 116, G01026, doi:10.1029/2010JG001476, 2011.

Aubinet, M., C. Feigenwinter, B. Heinesch, C. Bernhofer, E. Canepa, A. Lindroth, L. Montagnani, C. Rebmann, P. Sedlak, and Van Gorsel, E.: Direct advection measurements do not help to solve the night-time CO2 closure problem: Evidence from three different forests, Agric. For. Meteorol., 150(5), 655–664, doi:10.1016/j.agrformet.2010.01.016, 2010.

Baatz, R., Bogena, H. R., Hendricks Franssen, H.-J, Huisman, J.A., Montzka, C., and Vereecken, H.: An empirical vegetation correction for soil water content quantification using cosmic ray probes, Water Resour. Res., 51, 2030-2046, doi:10.1002/2014WR016443, 2015.

Baldocchi, D., Eva Falge, Lianhong Gu, Richard Olson, David Hollinger, Steve Running, Peter Anthoni, Ch Bernhofer, Kenneth Davis, Robert Evans, Jose Fuentes, Allen Goldstein, Gabriel Katul, Beverly Law, Xuhui Lee, Yadvinder Malhi, Tilden Meyers, William Munger, Walt Oechel, K. T. Paw, Kim Pilegaard, H. P. Schmid, Riccardo Valentini, Shashi Verma, Timo Vesala, Kell Wilson, and Steve Wofsy, 2001: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. Bull. Amer. Meteor. Soc., 82, 2415–2434, doi: 10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, 2001

Baroni, G., and Oswald, S.E.: A scaling approach for the assessment of biomass changes and rainfall interception using cosmic-ray neutron sensing. J. Hydrol., 525, 264-276, doi:10.1016/j.hydrol.2015.03.053, 2015.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Menard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description - part 1: Energy and water fluxes. Geoscientific Model Development 4(3), 677-699, 2011.

Beven, K. J., and Cloke, H. L.: Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al., Water Resour. Res., 48, W01801, doi:10.1029/2011WR010982, 2012.

Beven, K., and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49, 3071–3092, doi:10.1002/wrcr.20156, 2013

Bloeschl, G., Sivapalan, M.: Scale issues in hydrological modelling: A review. Hydrological Processes 9 (3-4), 251{290. URL http://dx.doi.org/10.1002/hyp.3360090305, 1995.

Blonquist, J. M., Jones, S. B., and Robinson, D. A.: Standardizing characterization of electromagnetic water content sensors: Part 2. Evaluation of seven sensing systems. Vadose Zone J., 4, 1059-1069.

Blyth, E. M., Dolman, A. J. and Wood, N.: Effective resistance to sensible- and latent-heat flux in heterogeneous terrain. Q.J.R. Meteorol. Soc., 119: 423–442. doi:10.1002/qj.49711951104, 1993.

Blyth, E., Gash, J., and Lloyd, A.: Evaluating the JULES Land Surface Model Energy Fluxes Using FLUXNET Data. Journal of Hydrometeorology, 11, 509-519, 2010.

Blyth, E., Clark, D.B., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M, and Sitch, S.: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. Geosci. Model Dev., 4, 255-269, doi: 10.5194/gmd-4-255-2011, 2011.

Bogena, H.R., J.A. Huisman, H. Meier, U. Rosenbaum, and A. Weuthen: Hybrid wireless underground sensor networks: Quantification of signal attenuation in soil. Vadose Zone J. 8. 3. 755-761, 2009.

Bogena, H. R., Huisman, J. A., Baatz, R., Hendricks Franssen, H. J., Vereecken, H.: Accuracy of the cosmic-ray soil water content probe in humid forest ecosystems: The worst case scenario. Water Resources Research 49, 5778-5791, doi:10.1002/wrcr.20463, 2013.

Brady, N.C., and Weil, R.R.: The nature and properties of soils (11th ed.). Prentice Hall, New York, 1996.

Budyko, M.I.: Тепловой баланс земной поверхности (Heat Balance of the Earth's Surface). Gidrometeoizdat, Leningrad. 255 pp, 1956.

Cardell-Oliver, R., Kranz, M., Smettem, K. and Mayer, K.: A reactive soil moisture sensor network: Design and field evaluation. International journal of distributed sensor networks, 1(2), 149-162, 2005.

Cavanaugh, M.L., Kurc, S.A., Scott, R.L.: Evapotranspiration partitioning in semiarid shrubland ecosystems: a two-site evaluation of soil moisture control on transpiration. Ecohydrology 4 (5), 671e681, 2011.

Chen, X., Rubin, Y., Ma, S., Baldocchi, D.: Observations and stochastic modelling of soil moisture control on evapotranspiration in a Californian oak savanna. Water Resources Research, 44, W08409, doi:10.1029/2007WR006646, 2008.

Clark, D. B., Mercado, L. M. , Sitch, S., Jones, C.D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Menard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Blyth, E., Boucher, O. , Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description - part 2: Carbon fluxes and vegetation dynamics. Geoscientific Model Development 4(3), 677-699, 2011.

Cullen, M. J. P.: The unified forecast/climate model, Meteorol. Mag., 122, 81–94, 1993.

Crow, W.T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., Rosnay, P., Ryu, D., Walker, J. P.: Upscaling sparse groundbased soil moisture observations for the validation of coarse-resolution satellite soil moisture products. Rev. Geophys. 50 (RG2002), 1-20. URL http://dx.doi.org/10.1029/2011RG000372, 2012.

De Vries, D.A.: Thermal properties of soils. In D.A. De Vries, and Afgan, N.H. (eds.): Heatand mass transfer in the biosphere. I. Transfer processes in plant environment, 5-28, Washington, DC: Scripts, 1963.

Desilets, D., and Zreda, M.: Footprint diameter for a cosmic-ray soil moisture probe: Theory and Monte Carlo simulations. Water Resour. Res., 49, 3566-3575, doi:10.1002/wrcr.20187, 2013.

Dharssi, I., Vidale, P. L., Verhoef, A., Macpherson, B., Jones, C., and Best, M.: New soil physical properties implemented in the Unified Model at PS18, Met Office Technical Report 528, Met Office, Exeter, UK, 2009.

Dirmeyer, P.A., Zeng, F.J., Ducharne, A., Morrill, J.C., and Koster, R.D.: The Sensitivity of Surface Fluxes to Soil Water Content in Three Land Surface Schemes. J. Hydrometeor., 1, 121–134, doi: 10.1175/1525-7541(2000)001<0121:TSOSFT>2.0.CO;2, 2000.

Dirmeyer, P. A., Dolman, A. J., and Sato, N.: The pilot phase of the Global Soil Wetness Project. Bull. Amer. Meteor. Soc., 80, 851-878, 1999.

Dirmeyer, P. A., Zeng, F. J., Ducharne, A., Morrill, J. C., and Koster, R. D.: The Sensitivity of Surface Fluxes to Soil Water Content in Three Land Surface Schemes. J. Hydromteor., 1, 121-134, 2000.

Dirmeyer, P.A.: The terrestrial segment of soil moisture-climate coupling. Geophysical Research Letters, 38 (16), 2011.

Dirmeyer, P.A., Jin, Y., Singh, B., Y, X.: Trends in Land-Atmosphere Interactions from CMIP5 Simulations. Journal of Hydrometeorology, 14, 829-849, doi: 10.1175/JHM-D-12-0107.1, 2013.

Downer, C. W., and Ogden, F. L.: Appropriate vertical discretization of Richards’ equation for two-dimensional watershed-scale modelling. Hydrol. Process., 18, 1-22, 2004.

Evans, J. G., Ward, H. C., Blake, J. R., Hewitt, E. J., Morrison, R., Fry, M., Ball, L. A., Doughty, L. C., Libre, J. W., Hitt, O. E., Rylett, D., Ellis, R. J., Warwick, A. C., Brooks, M., Parkes, M. A., Wright, G. M. H., Singer, A. C., Boorman, D. B., and Jenkins, A.: Soil water content in southern England derived from a cosmic-ray soil moisture observing system – COSMOS-UK. Hydrol. Process., doi: 10.1002/hyp.10929, 2016

FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009. Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Fan, Y., Li. H., and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth. 339(6122, 940-943, doi:10.1126/science.1229881, 2013.

Feddes, R.A., Kowalik, P.J., and Zaradny, H.: Simulation of field water use and crop yield. Simulation Monographs. Wageningen: Pudoc, 1978.

Finch, J.W. and Haria, A: The representation of chalk soils in the JULES/MOSES soil hydrology model. Rept., NERC/Centre for Ecology and Hydrology, 2006.

Finkelstein, P. L., and Sims, P. F.: Sampling error in eddy correlation flux measurements. J. Geophys. Res., 106(D4), 3503-3509, 2001.

Franz, T. E., Zreda, M., Rosolem, R., Ferre, T. P. A.: Field validation of a cosmic-ray neutron sensor using a distributed sensor network. Vadose Zone Journal 11 (4), doi:10.2136/vzj2012.0046, 2012.

Goulden, M. L., Munger, J.W., Fan, S.M., Daube, B.C., and Wofsy, S.C.:, Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy, Global Change Biol., 2(3), 169–182, doi:10.1111/j.1365-2486.1996.tb00070.x, 1996.

Goulden, M. L., Winston, G. C., McMillan, A.M.S., Litvak, M. E., Read, E. L., Rocha, A. V., and Elliot, J. R.: An eddy covariance mesonet to measure the effect of forest age on land-atmosphere exchange, Global Change Biol., 12(11), 2146–2162, doi:10.1111/j.1365-2486.2006.01251.x, 2006.

Goulden, M.L., Anderson, R.G., Bales, R.C., Kelly, A.E., Meadows, M., and Winston, G.C.: Evapotranspiration along an elevation gradient in California's Sierra Nevada. Journal of Geophysical Research, 117, G03028, doi:10.1029/2012JG002027, 2012.

Goulden, M.L.: Measurement of Energy, Carbon and Water Exchange Along California Climate Gradients. Website: http://www.ess.uci.edu/~california/, last visited: July 2016. 2015.

Gupta, H.V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research, 34(4), 751-763, 1998.

Gupta, H. V., Bastidas, L. A., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. L.: Parameter estimation of a land surface scheme using multicriteria methods. J. Geophys. Res., 104(D16), 19491-19503, 1999.

Gupta, H. V., King, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377, 80-91, 2009.

Hawdon, A., D. McJannet, and J. Wallace.: Calibration and correction procedures for cosmic-ray neutron soil moisture probes located across Australia, Water Resources Research, 50(6), 5029-5043. doi:10.1002/2013WR015138, 2014.

Hadka, D., and Reed P.: Borg: an auto-adaptive many-objective evolutionary computing framework. Evolut. Compt., 21(2), 231-259, 2013.

Henderson-Sellers, A., McGuffie, K., Pitman, A.J.: The Project for Intercomparison of Land-Surface Parametrization Schemes (PILPS): 1992 to 1995. Clim. Dyn. 12, 849–859, 1996.

Ishida, T., and Makino, T., and Wang, C. J., 2000. Dielectric-relaxation spectroscopy of kaolinite, montmorillonite, allophane, and imogolite under moist conditions. Clays Clay Miner, 48, 75-84.

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 Resources Research 51, doi:10.1002/2015WR017169, 2015.

Kollat, J.B., Reed, P M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrological models meaningful? Water Resour. Res., 48, W03520, doi:10.1029/2011WR011534.

Koster, R.D., Dirmeyer, P.A., Guo, Z.C., Bonan, G., Chan, E., Cox, P., Gordon, C.T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y.C., Taylor, C.M., Verseghy, D., Vasic, R., Xue, Y.K., Yamada, T.: Regions of strong coupling between soil moisture and precipitation. Science 305, 1138–1140, 2004.

Koster, R.D., Guo, Z., Yang, R., Dirmeyer, P.A., Mitchell, K., Puma, M.J.: On the nature of soil moisture in land surface models. J. Climate 22 (16), 4322–4335, 2009.

Kurc, S. A. and E. E. Small: Soil moisture variations and ecosystem-scale fluxes of water and carbon in semiarid grassland and shrubland, Water Resources Research, 43, W06416, doi:10.1029/2006WR005011, 2007.

Larson, K.M., Small, E.E., Gutmann, E.D., Bilich, A.L., Braun, J.J., Zavorotny, V.U.: Use of GPS receivers as a soil moisture network for water cycle studies. Geophysical Research Letters, 35, L24405, doi:10.1029/2008GL036013, 2008.

Larson, K.M., Braun, J.J., Small, E.E., Zavorotny, V.U., Gutmann, E.D., Bilich, A.L.: GPS Multipath and Its Relation to Near-Surface Soil Moisture Content. IEEE J. Sel. Topics in Appl., Earth Obs. And Rem. Sens., 3(1), 91-99, 2010.

Lee, D. H., and Abriola, L. M.: Use of Richards equation in land surface parameterizations. J. Geophys. Res., 104(D22), 27519-27526, 1999.

Manabe, S.: Climate and ocean circulation. Part I: the atmospheric circulation and the hydrology of the earth's surface. Mon. Weather Rev. 97, 739–774, 1969.

Mittelbach, H. and Seneviratne, S. I.: A new perspective on the spatio-temporal variability of soil moisture: temporal dynamics versus time-invariant contributions. Hydrol. Earth Syst. Sci., 16, 2169-2179, doi:10.5194/hess-16-2169-2012, 2012.

Morris, M.D.: Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33(2), doi:10.2307/1269043, 1991.

Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media, Water Resour. Res., 12(3), 513–522, doi:10.1029/WR012i003p00513, 1976.

ORNL-DAAC, 2015. Ameriflux Web Page. http://ameriflux.ornl.gov.

Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for Global Sensitivity Analysis. Environmental Modelling and Software 70, 80-85, doi:10.1016/j.envsoft.2015.04.009, 2015.

Pitman, A.J.: The evolution of, and revolution in, land surface schemes designed for climate models. Int. J. Climatol. 23, 479–510, 2003.

Qu, W., Bogena, H. R., Huisman, J. A., Vanderborght, J., Schuh, M., Priesack, E., Vereecken, H.: Predicting subgrid variability of soil water content from basic soil information, Geophysical Research Letters, 2015, 42, 1-8.

Rahman, M, and Rosolem, R.: The effects of chalk representation in land surface modelling. Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-244, accepted, 2017.

Richter, H., Western, A. W., and Chiew, F. H. S.: The Effect of Soil and Vegetation Parameters in the ECMWF Land Surface Scheme. J. Hydrometeor., 5, 1131–1146, doi: 10.1175/JHM-362.1, 2004.

Ritsema, C.J., Kuipers, H., Kleiboer, L., Van Den Elsen, E., Oostindie, K., Wesseling, J.G., Wolthuis, J.W. and Havinga, P.: A new wireless underground network system for continuous monitoring of soil water contents. Water resources research, 45(4), 2009.

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., Ogden, F., Selker, J., Wendroth, O.: Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. Vadose Zone Journal 7 (1), 358{389. URL http://dx.doi.org/doi:10.2136/vzj2007.0143, 2008.

Rosolem, R., W. J. Shuttleworth, X. Zeng, S. R. Saleska, and T. E. Huxman: Land surface modeling inside the Biosphere 2 tropical rain forest biome, J. Geophys. Res., 115, G04035, doi: 10.1029/2010JG001443, 2010.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Gonalves de Gonalves, L. G., Zeng, X.. Towards a comprehensive approach to parameter estimation in land surface parameterization schemes. Hydrological Processes 27 (14), 2075{2097, 2013b.

Rosolem, R., W. J. Shuttleworth, M. Zreda, T. E. Franz, X. Zeng, and Kurc. S. A.: The Effect of Atmospheric Water Vapor on Neutron Count in the Cosmic-Ray Soil Moisture Observing System. J. Hydrometeor., 14, 1659–1671, doi: 10.1175/JHM-D-12-0120.1.

R Rosolem, T Hoar, A Arellano, JL Anderson, WJ Shuttleworth, X Zeng, TE Franz: Assimilation of near-surface cosmic-ray neutrons improves summertime soil moisture profile estimates at three distinct biomes in the USA. Hydrol. Earth Syst. Sci. Discuss., 11(5), 5515-5558, doi: 10.5194/hessd-11-5515-2014, 2014.

Running, S.W., Baldocchi, D.D., and Turner, D.: A global terrestrial monitoring network, scaling tower fluxes with ecosystem modelling and EOS satellite data. Remote Sensing, 70, 108-127, 1999.

Skott, R.L., Cable, W.L., Hultine, K.R.: The ecohydrologic significance of hydraulic redistribution in a semiarid savanna. Water Resources Research, 44 (2), 1990.

Sellers, P. J., W. J. Shuttleworth, J. L. Dorman, A. Dalcher, and Roberts, J.M.: Calibrating the Simple Biosphere Model for Amazonian tropical forest using field and remote sensing data. Part I: Average calibration with field data. J. Appl. Meteor., 28, 727–759, 1989.

Sellers, P. J., and Hall, F. G.: FIFE in 1992: Results, scientific gains, and future research directions. J. Geophys. Res., 97(D17), 19091-19109, doi:10.1029/92JD02173, 1992.

Sellers, P.J., Dickinson, R.E., Randall, D.A., Betts, A.K., Hall, F., Berry, J., Collatz, G., Denning, A., Mooney, H., Nobre, C., Sato, N., Field, C., Henderson-Sellers, A.: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. Science 275 (5299), 502–509, 1997.

Seneviratne, S. I. and Corti, T. and Davin, E. L. and Hirschi, M. and Jaeger, E. B. and Lehner, I. and Orlowsky, B. and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A review, Earth-Sci. Rev., 99, 125—161, doi="10.1016/j.earscirev.2010.02.004, 2010.

Shuttleworth, J., Rosolem, R., Zreda, M., and Franz, T.: The COsmic-ray Soil Moisture Interaction Code (COSMIC) for use in data assimilation, Hydrol. Earth Syst. Sci., 17, 3205-3217, doi:10.5194/hess-17-3205-2013, 2013.

Smirnova, T. G., Brown, J. M., and Benjamin, S. G.: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. Monthly Weather Review, 125, 1871-1884, 1997.

Smith, A. B., J. P.Walker, A. W.Western, R. I.Young, K. M.Ellett, R. C.Pipunic, R. B.Grayson, L.Siriwardena, F. H. S.Chiew, and H.Richter: The Murrumbidgee soil moisture monitoring network data set, Water Resour. Res., 48, W07701, doi:10.1029/2012WR011976, 2012.

Stöckli, R., Lawrence, D.M., Niu, G.Y., Oleson, K.W., Thornton, P.E., Yang, Z.L., Bonan, G.B., Denning, A.S., Running, S.W.: Use of FLUXNET in the community land model development. J. Geophys. Res 113 (G1), G01025, 2008.

Taylor, S.A., and Ashcroft, G.M.: Physical Edaphology. Freemand and Co., San Francisco, California, 434-435, 1972.

Trubilowicz, J., K. Cai, and M. Weiler 2009, Viability of motes for hydrological measurement, Water Resour. Res., 45, W00D22, doi:10.1029/2008WR007046, 2009.

Teuling, A. J., and P. A. Troch: Improved understanding of soil moisture variability dynamics. Geophys. Res. Lett., 32, L05404, doi:10.1029/2004GL021935, 2005.

Teuling, A. J., Uilenhoet, R., Hupet, F., Van Loon, E. E., and Troch, P. A.: Estimating spatial mean root-zone soil moisture from point-scale observations. Hydrol. Earth Syst. Sci., 10, 755-767, 2006.

Teuling, A. J., Uilenhoet, R., Van den Hurk, B., and Seneviratne, S. I.: Parameter Sensitivity in LSMs: An Analysis Using Stochastic Soil Moisture Models and ELDAS Soil Parameters. J. Hydrometeorol., 10(3), 751-765, doi:10.1175/2008JHM1033.1, 2009

Vachaud, G., Passerat de Silans, P., Balabanis, P., Vauclin, M.: Temporal stability of spatially measured soil water probability density function, Soil Sci. Soc. Am. J., 49, 822–828, 1985.

Van Dam, J. C., and Feddes, R. A.: Numerical simulation of infiltration, evaporation, and shallow groundwater levels with the Richards equation. J. Hydrol., 233, 72-85.

Vanderlinden, K., H. Vereecken, H. Hardelauf, M. Herbst, G. Martínez, M. H. Cosh, and Pachepsky, Y. A.: Temporal Stability of Soil Water Contents: A Review of Data and Analyses. Vadose Zone J., 11., doi:10.2136/vzj2011.0178, 2012.

Van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. America J., 44, 892–898, 1980.

Vereecken, H., Huisman, J.A., Bogena, H., Vanderborght, J., Vrugt, J.A., and Hopmans, J.W.: On the value of soil moisture measurements in vadose zone hydrology: A review. Water Resources Research, 44, doi:10.1029/2008WR006829, 2008.

Vereecken, H., Huisman, J.A., Hendricks Franssen, H.J., Bruggemann, N., Bogena, H.R., Kollet, S., Javaux, M., van der Kruk, J., and Vanderborght, J: Soil hydrology: recent methodological advaces, challenges, and perspectives. Water Resources Research, 51, 10.1002/2014WR016852, 2015.Appendices

Vinnikov, K. Y., Robock, A., Shuang, W., Schlosser, A.: Scales of temporal and spatial variability of midlatitude soil moisture. Journal of Geophyisical Research 101, 7163{7174. URL http://dx.doi.org/10.1029/95JD02753, 1996.

Vivoni, E.R., Moreno, H.A., Mascaro, G., Rodriguez, J.C., Watts, C.J., Garatuza-Payan, J. and Scott, R.L.: Observed relation between evapotranspiration and soil moisture in the North American monsoon region. Geophysical Research Letters, 35(22), 2008.

Wood, E.F., et al.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47, W05301, doi:10.1029/2010WR010090, 2011.

Wösten, J. H. M.: Pedotransfer functions to evaluate soil quality. In Gregorich, E. G. and Carter, M. R. (eds.) Soil Quality for Crop Production and Ecosystem Health. Elsevier, Amsterdam. pp. 221–246, 1997.

Zreda, M., Desilets, D., Ferre, T. P. A., Scott, R. L.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic- ray neutrons. Geophysical Research Letters 35, 1-5, doi:10.1029/2008GL035655, 2008.

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., Rosolem, R.: Cosmos: the cosmic-ray soil moisture observing system. Hydrology and Earth System Sciences 16, 4079-4099, doi:10.5194/hess-16-4079-2012, 2012.

Table 1: Site characteristics. Altitude from COSMOS website, land cover percentages from Ameriflux and publications. Harmonised World Soil Database (HWSD) data was used to define default model parameter values (here only soil categories shown).

Site	Altitude (meter above sea level; COSMOS)	Land cover (%)		HWSD dominant soil type	Site soil info (Ameriflux, COSMOS, literature)	Data sources
		Dominant	Remaining			
UM	220	100% broadleaf		Loamy sand	Deep well drained soils	COSMOS, ORNL-DAAC (2015)
DC	1300	46% shrubs	46% bare, 8% needleleaf	Sandy loam		Anderson and Goulden (2012)
SO	1160	63% needleleaf	37% shrubs	Loam	0-10 cm thick organic litter layer, bedrock at 1-2 m	COSMOS, Goulden et al. (2012)
KE	1531	100% C4-grass		Loam	Coarse-loamy, limestone fragments	ORNL-DAAC (2015)
ME	1253	100% needleleaf		Loamy sand	Sandy, minimal organic	COSMOS, ORNL-DAAC (2015)
SR	989	76% bare	24% shrubs	Loam	Silty clay loam	COSMOS, Cavanaugh et al. (2011)
CS	457	100% shrubs		Loam		Anderson and Goulden (2012)
MM	288	100% broadleaf		Loam	Well drained clay loam	COSMOS, ORNL-DAAC (2015)
TR	177	60% C3-grass	40% broadleaf	Loam	Sandy clay loam, clay hardpan at 30-40 cm	COSMOS, ORNL-DAAC (2015), Chen et al. (2008)
AR	314	100% C4-grass		Loam	Sandy	COSMOS, ORNL-DAAC (2015)
WR	371	100% needleleaf trees		Loam	5-10 cm organic layer, silty sand	COSMOS, ORNL-

					DAAC (2015)
MO	219	100% broadleaf trees	Loam	Silty loam	COSMOS, ORNL-DAAC (2015)

Table 2: Calibrated parameter definitions and calibration ranges.

JULES parameter name	Unit	Role	Range	
			Minimum	**Maximum**
b	-	Mualem-Van Genuchten parameter (b=1/(n-1))	0.63	24.43
sathh	m	Mualem-Van Genuchten parameter (sathh=α^{-1})	0.09	28.01
K_{sat}satcon	mm s^{-1}	Mualem saturationed hydraulic conductivity	$3 \cdot 10^{-5}$	$4.3 \cdot 10^{-1}$
smθ_{crit}	m^3m^{-3}	Critical soil moisture content	0.1·θsm$_{sat}$	0.9· θ_{sat}smsat
smθ_{wilt}	m^3m^{-3}	Wilting soil moisture content	0.1·θsm$_{crit}$	0.9· θ_{crit}smcrit

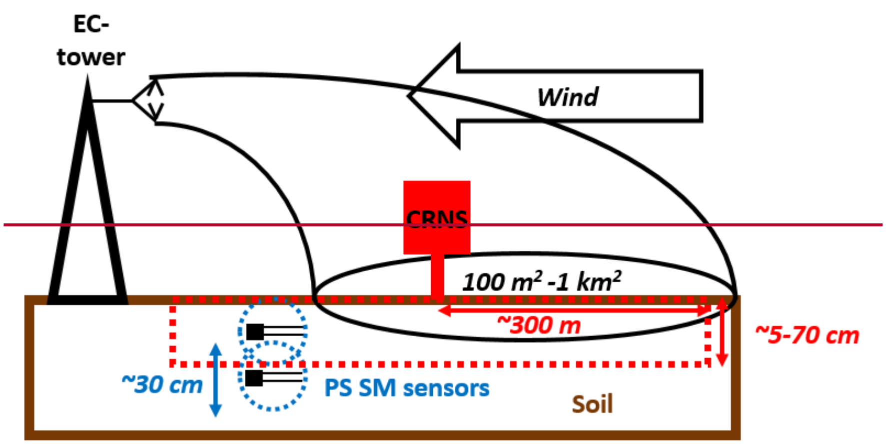


Figure 1: Schematic depicting the scale mismatch between Point Scale (PS) soil moisture sensor footprint (highlighted in dashed blue) and Eddy Covariance tower footprint (highlighted with solid black circle on the ground). The Cosmic Ray Neutron Sensor (CRNS; footprint highlighted in dashed red) may help fill the gap this scale mismatch leaves.

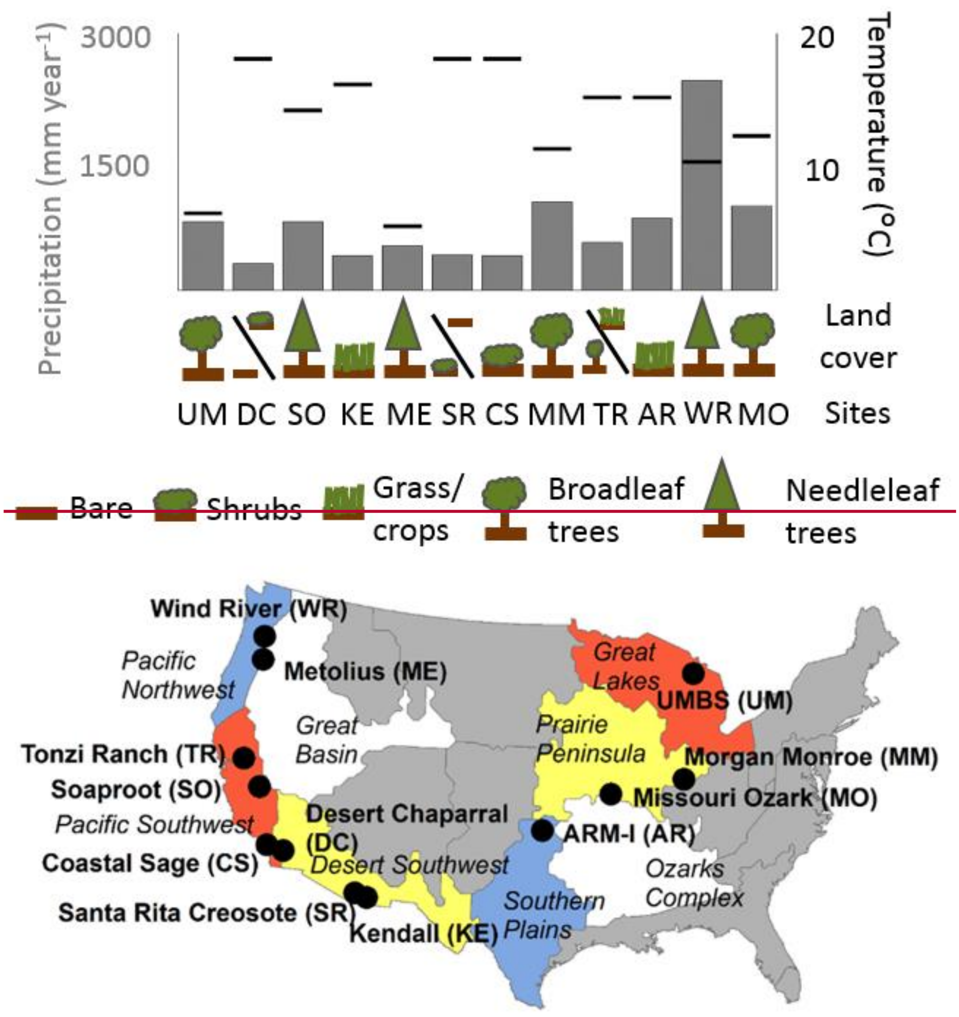
Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript



Precipitation (mm year^{-1})
3000
1500
20
10
Temperature (°C)
Land cover
UM DC SO KE ME SR CS MM TR AR WR MO
Sites
Bare
Shrubs
Grass/ crops
Broadleaf trees
Needleleaf trees
Wind River (WR)
Metolius (ME)
Pacific Northwest
Great Basin
Tonzi Ranch (TR)
Soaproot (SO)
Pacific Southwest
Coastal Sage (CS)
Santa Rita Creosote (SR)
Desert Chaparral (DC)
Desert Southwest
Kendall (KE)
Southern Plains
ARM-I (AR)
Great Lakes
UMBS (UM)
Prairie Peninsula
Morgan Monroe (MM)
Missouri Ozark (MO)
Ozarks Complex

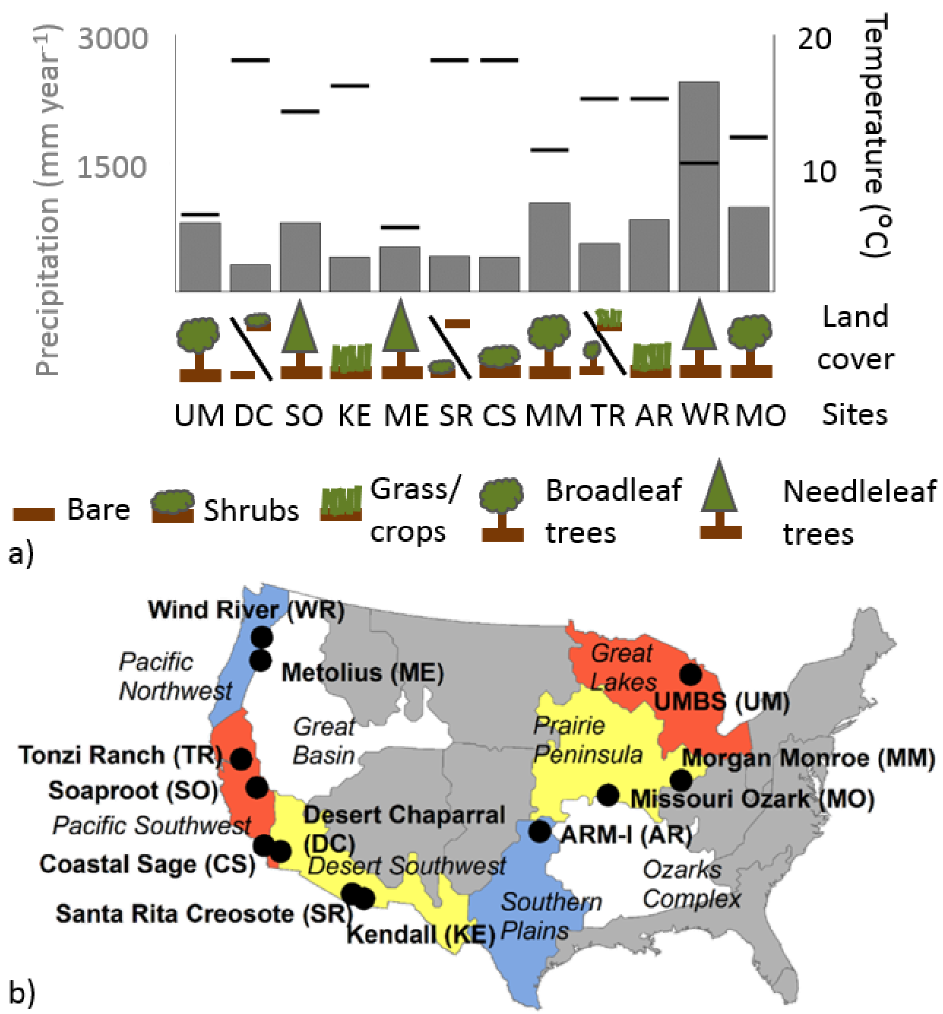
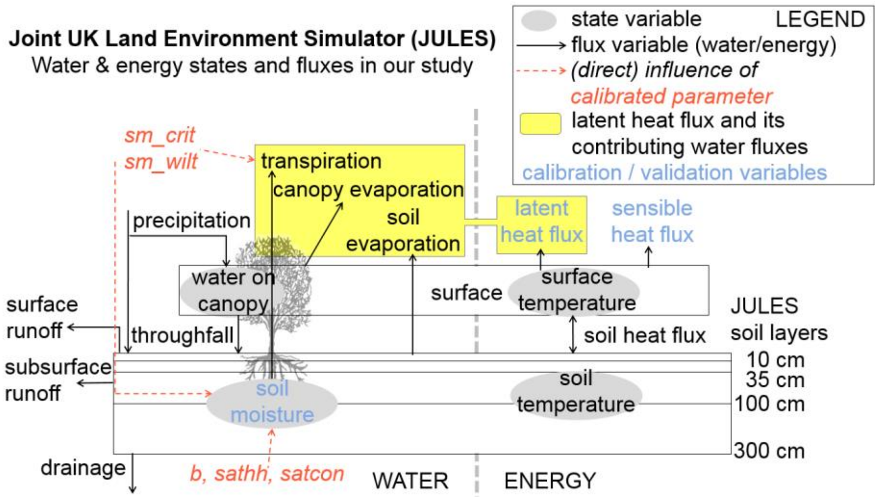
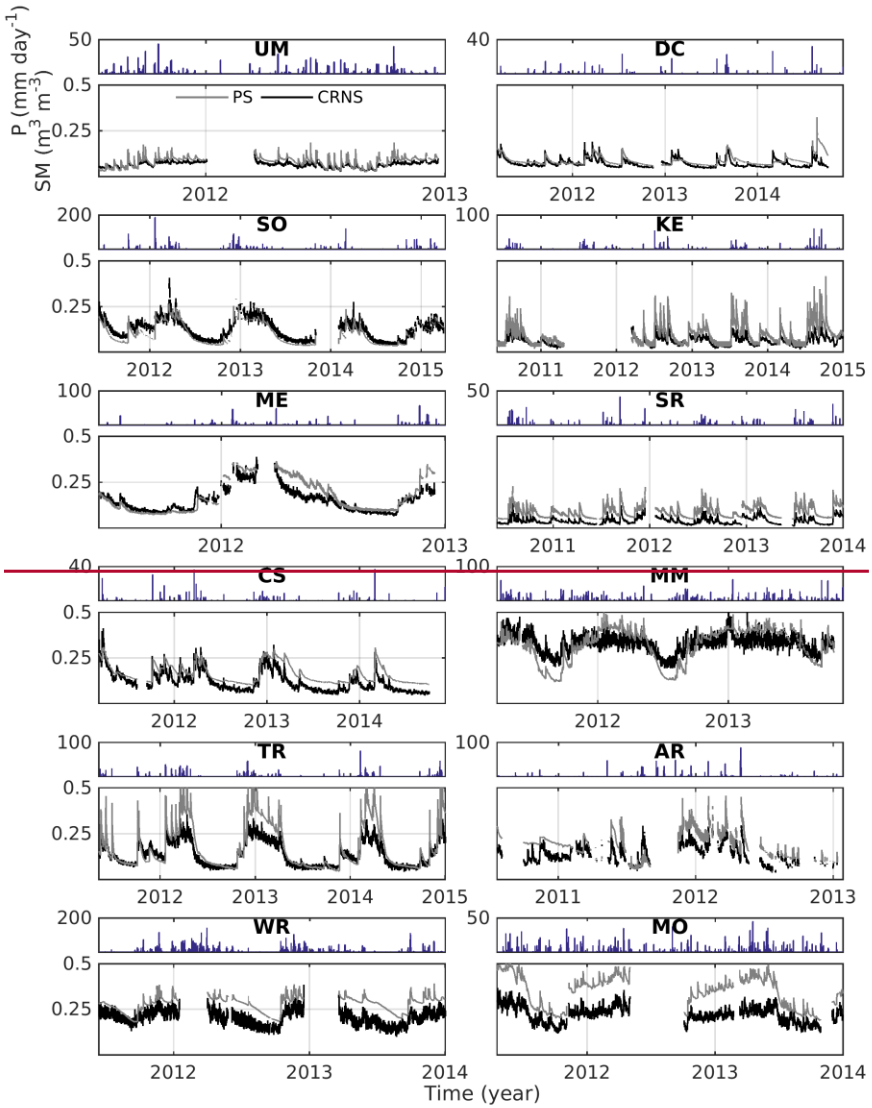


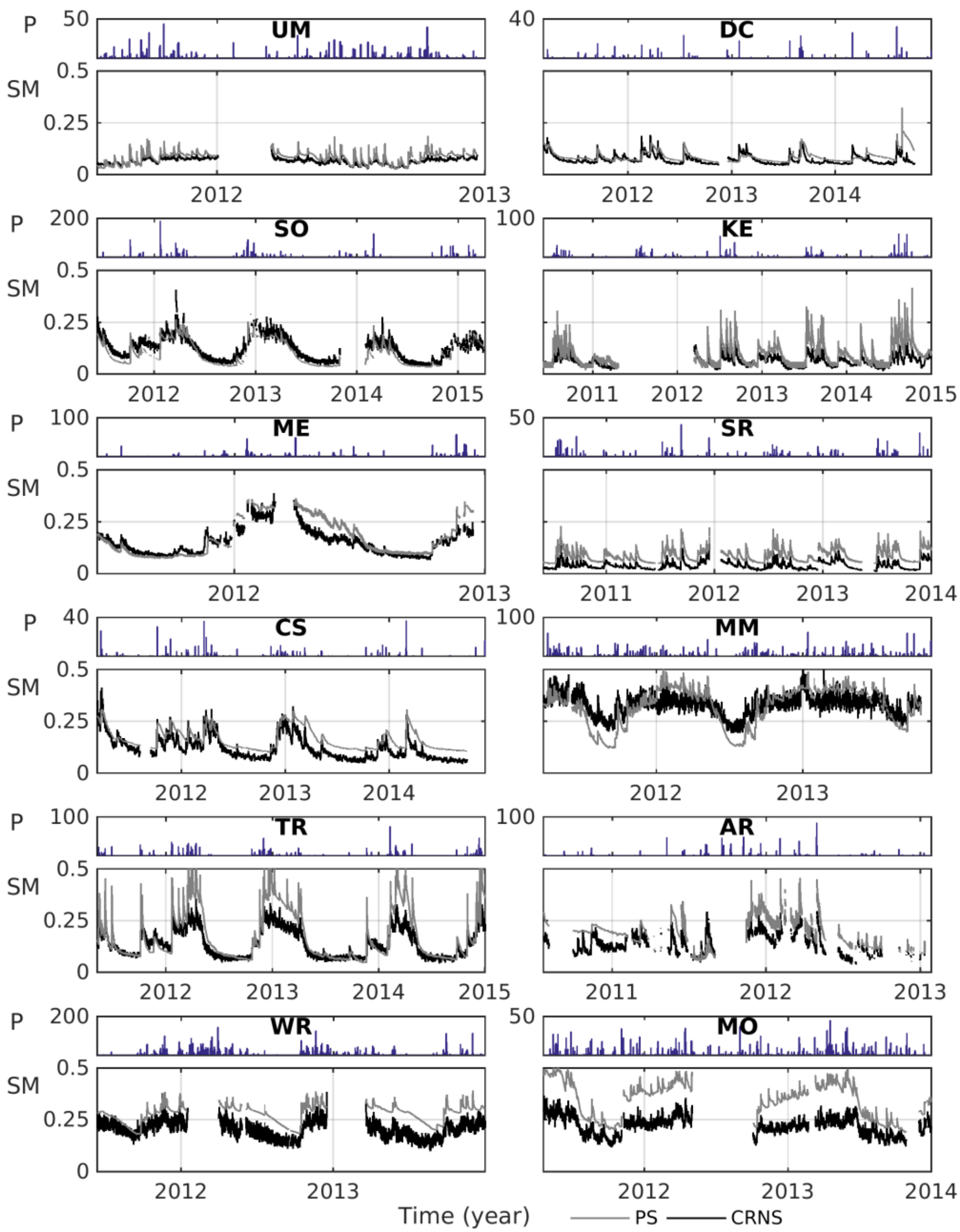
Figure 112: The upper fFigure a shows the yearly mean precipitation, air temperature, and dominant land cover types for the twelve Ameriflux/COSMOS sites used. At sites DC, SR, and TR two different land cover types were shown because they covered similar areas in size. The map below (b) shows the locations of the twelve sites within eight NEON Ecoclimatic Domains. Data sources: COSMOS, ORNL-DAAC (2015), Goulden et al. (2012), Anderson and Goulden (2012), Scott et al. (1990), Chen et al. (2008).



~~**Figure 3: JULES water and energy states and fluxes most relevant to our study. Calibrated parameters and their direct effects shown. Please, refer to Table 2 for detailed description of parameters.**~~



P (mm day^{-1})
SM (m^3 m^{-3})
UM
DC
PS
CRNS
SO
KE
ME
SR
CS
MM
TR
AR
WR
MO
Time (year)



UM
DC
SO
KE
ME
SR
CS
MM
TR
AR
WR
MO
P
SM
Time (year)
PS
CRNS

Figure 224: PS and CRNS observed soil moisture (SM; m^3m^{-3}) time series for the twelve study sites. Notice the PS soil moisture time series have been linearly interpolated from individual measurement depths to the corresponding JULES soil layers. CRNS soil moisture was obtained using COSMIC while assuming vertically homogeneous soil moisture. Daily precipitation (mm day^{-1}) is also shown here for each site.

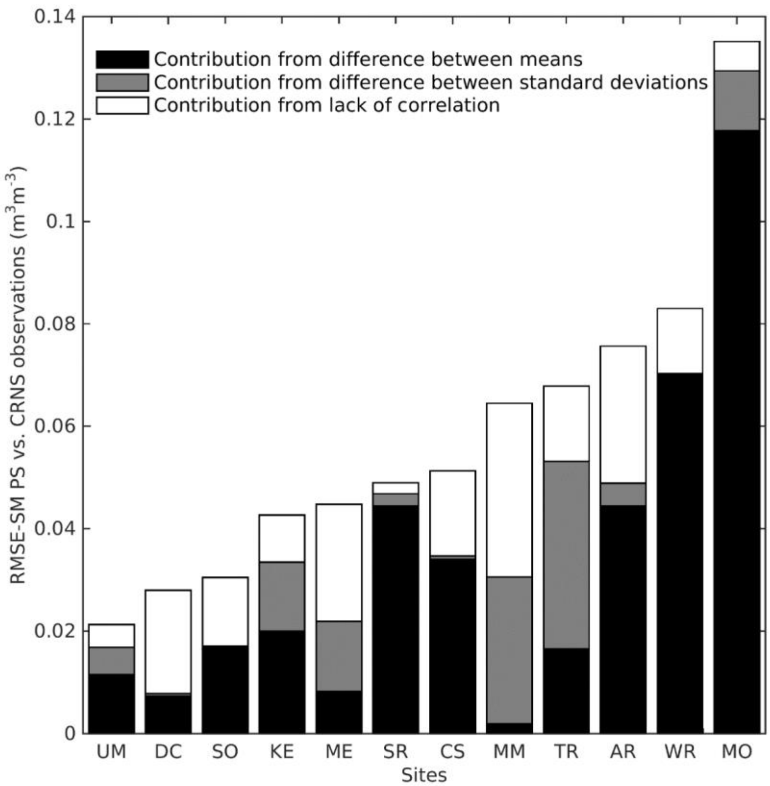


Figure 335: Root Mean Squared Error (RMSE) between observed PS and CRNS soil moisture (SM; m^3m^{-3}). MSE decomposition (Gupta et al., 2009) was calculated and the fractions were then applied to the RMSE values. Sites are ranked from most similar to most different.

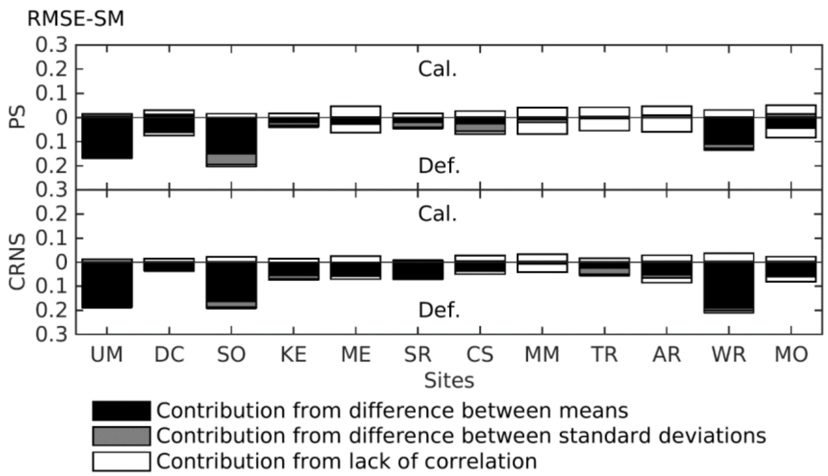
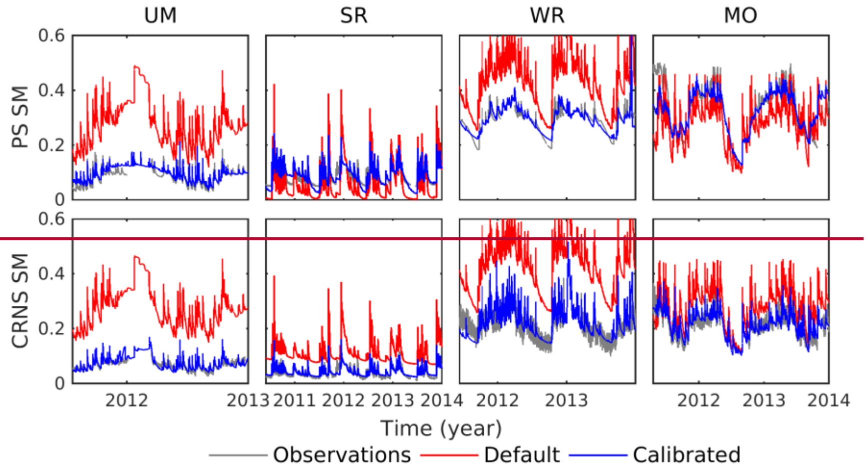


Figure 446: Objective function (Root Mean Square Error; RMSE (m^3m^{-3})) values between observed PS/CRNS soil moisture and JULES simulated soil moisture. For each calibration the RMSE of the default run is shown from the horizontal axis down, and the result after calibration is shown in the upward directions. The different error contributions from the MSE decomposition are shown as stacked bar plots.



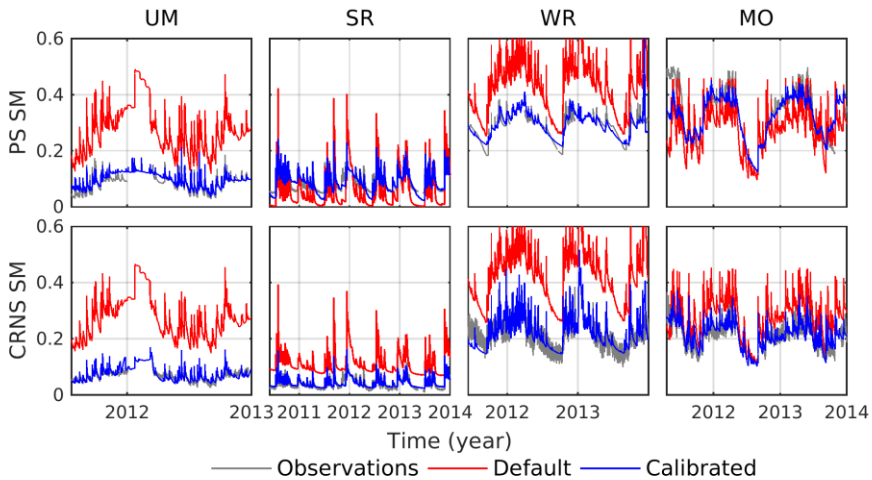
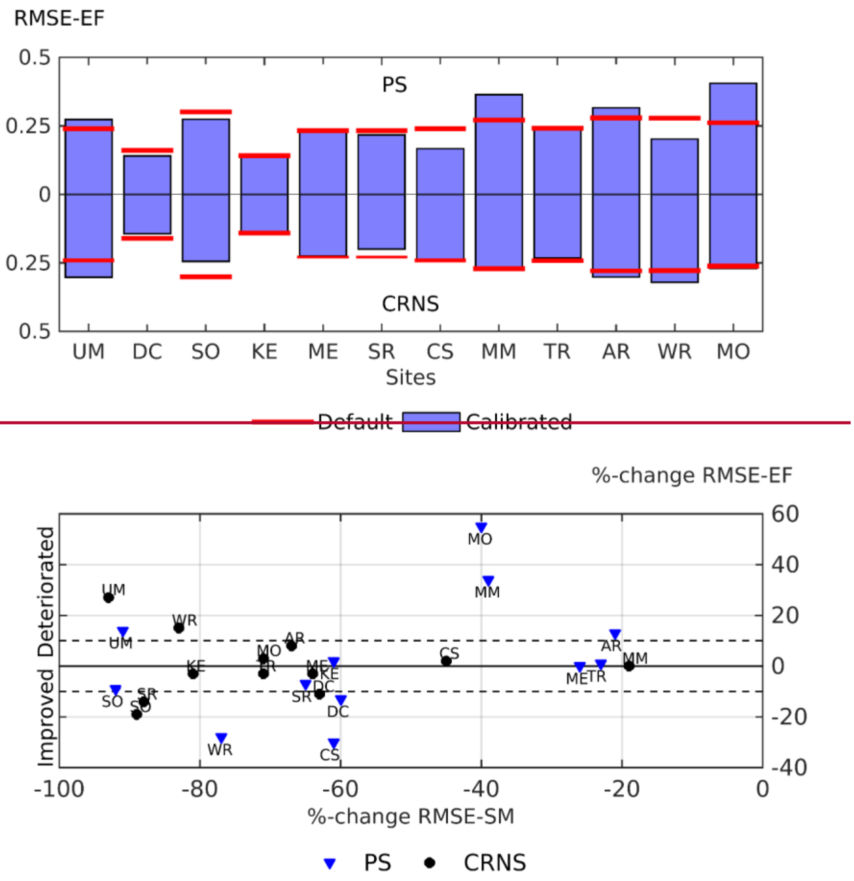


Figure 557: Hourly soil moisture time series of JULES default and PS/CRNS calibrated runs against observations (PS or CRNS) for four of twelve sites.


RMSE-EF
PS
CRNS
0.5
0.25
0
0.25
0.5
UM DC SO KE ME SR CS MM TR AR WR MO
Sites
Default
Calibrated
%-change RMSE-EF
Improved Deteriorated
60
40
20
0
-20
-40
-100 -80 -60 -40 -20 0
%-change RMSE-SM
PS
CRNS

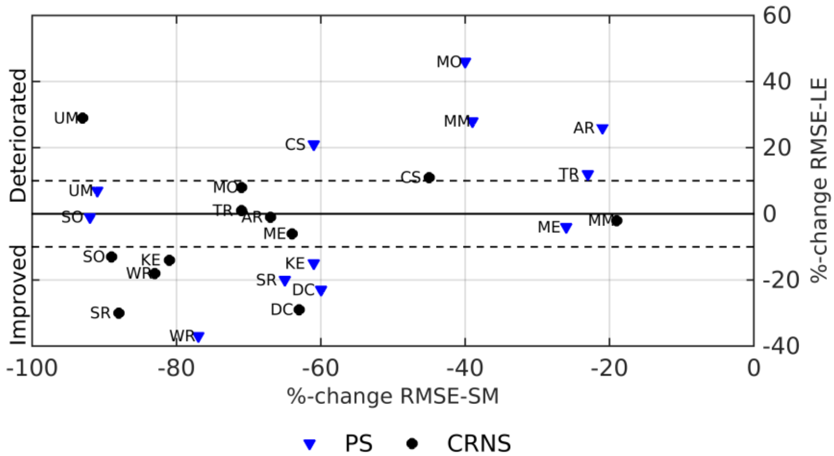


Figure 668: 8a) Relative change in Root Mean Square Error (RMSE) values after single-objective calibration against PS soil moisture and CRNS neutron count data. Change in RMSE between observed and simulated latent heat flux (LE) is plotted against change in RMSE between observed and simulated soil moisture.between observed and simulated Evaporative Fraction EF=LE/(LE+H), per site, for default and calibrated (PS/CRNS) runs. 8b) Relative change in RMSE-EF and RMSE-SM after calibration plotted against each other.

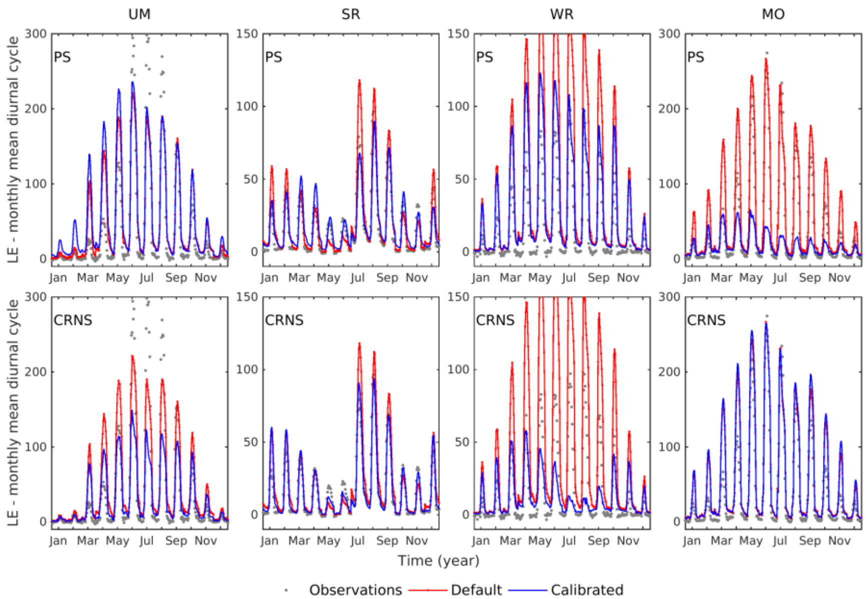


Figure 779: Monthly mean diurnal latent heat flux (LE) cycles. The upper row contains PS calibrated solutions, the lower row shows the CRNS calibrated solutions. Monthly mean diurnal latent heat flux cycles. The upper row contains PS calibrated solutions, the lower row shows the CRNS calibrated solutions.

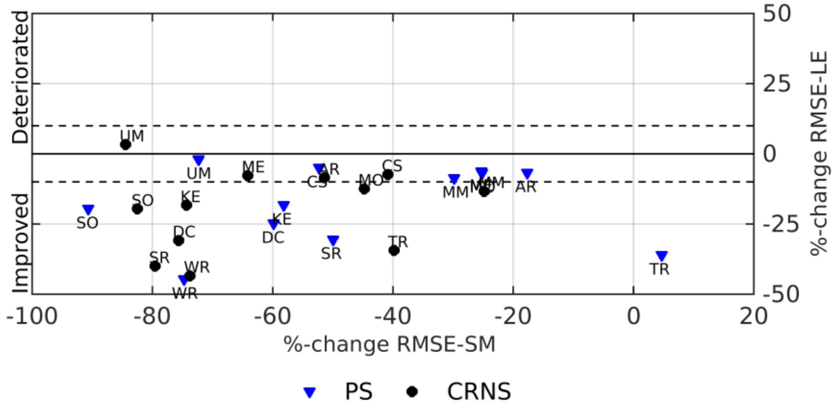
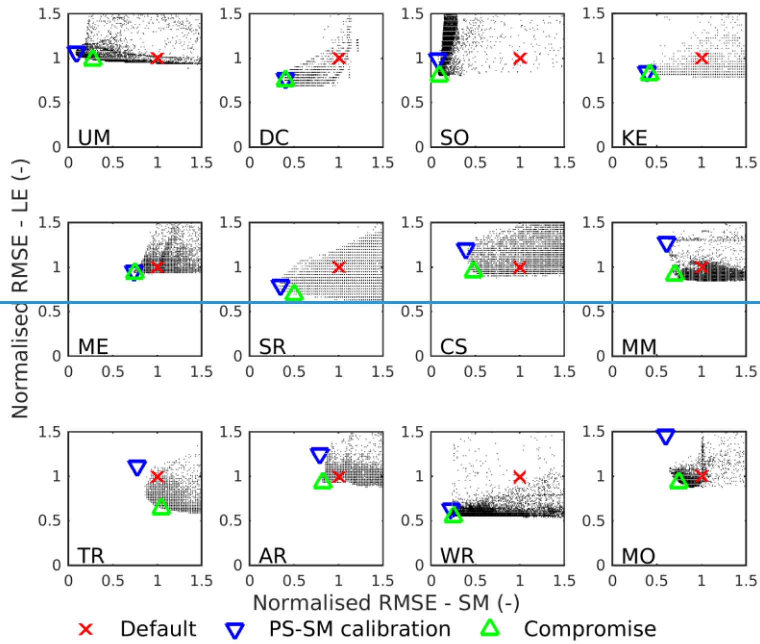


Figure 8: Relative change in Root Mean Square Error (RMSE) values after two-objective calibration against (1) PS soil moisture and latent heat flux and (2) against CRNS neutron count data and latent heat flux. Change in RMSE between observed and simulated latent heat flux (LE) is plotted against change in RMSE between observed and simulated soil moisture.

Formatted: Keep with next



UM
DC
SO
KE
ME
SR
CS
MM
TR
AR
WR
MO
Normalised RMSE - LE (-)
Normalised RMSE - SM (-)
Default
PS-SM calibration
Compromise

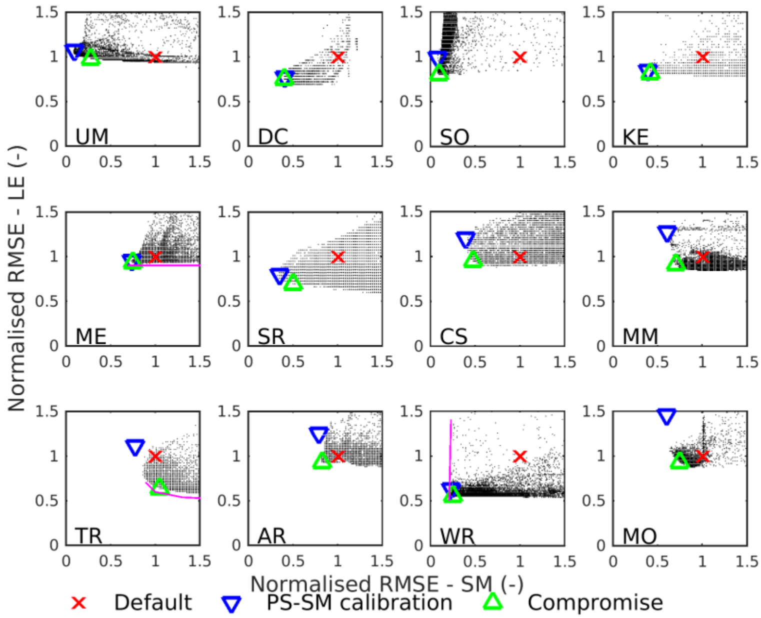
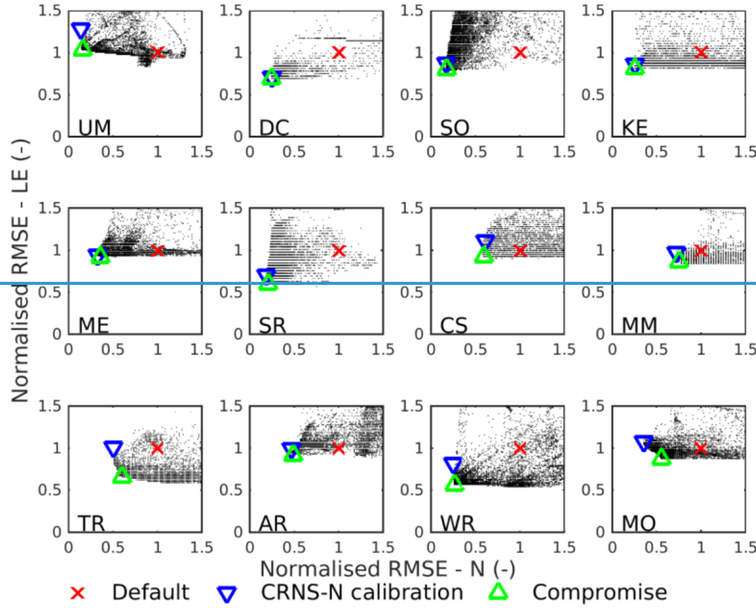


Figure 9: All runs from the two-objective calibrations against PS soil moisture and day time hourly latent heat flux (LE) at each site. Default run, PS-SM single objective calibrated run, and compromise solution runs are shown on top. All values were normalised to the values of the Default run. The plots were zoomed in to the area of 0 to 1.5 times the Default values.

Figure 910: All runs from the two-objective calibrations against PS soil moisture and day time hourly latent heat flux (LE) at each site. Default run, PS-SM single objective calibrated run, and compromise solution runs are shown on top. All values were normalised to the values of the Default run. The plots were zoomed in to the area of 0 to 1.5 times the Default values.

Formatted: Keep with next



UM
DC
SO
KE
ME
SR
CS
MM
TR
AR
WR
MO
Normalised RMSE - LE (-)
Normalised RMSE - N (-)
Default
CRNS-N calibration
Compromise

Formatted: Keep with next

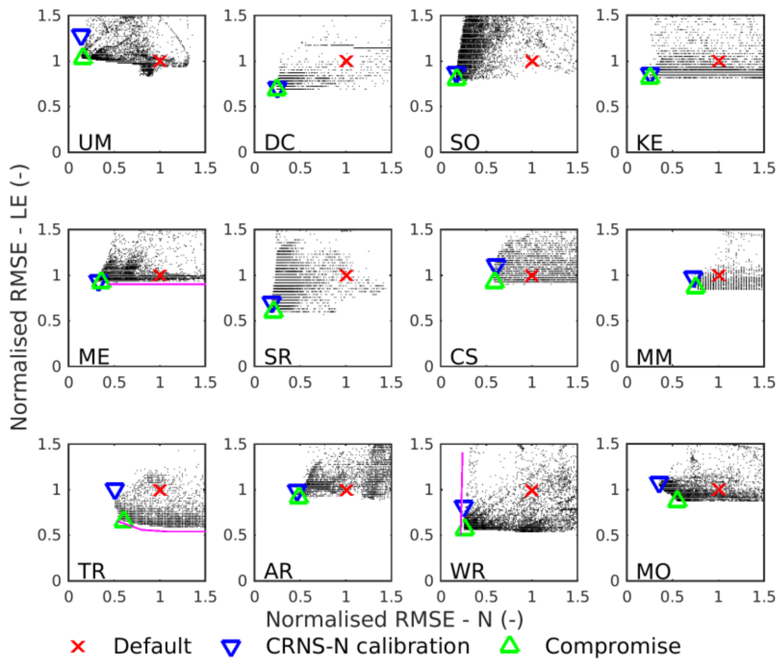


Figure 10: Same as Figure 10 but for the two-objective calibrations against CRNS neutron counts (representing soil moisture) and day time hourly latent heat flux (LE) at each site.

~~Figure 1011: Same as Figure 10 but for the two-objective calibrations against CRNS neutron counts (representing soil moisture) and day time hourly latent heat flux (LE) at each site.~~

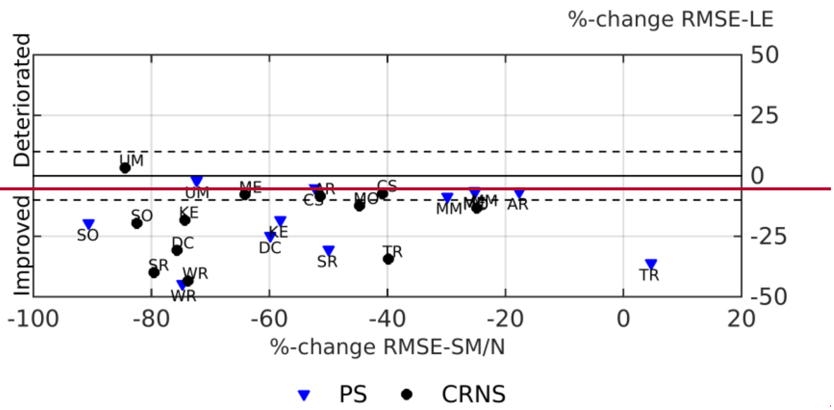


Figure 12: Relative change in performance (RMSE) of multi-objective compromise solutions compared to JULES simulations with default parameter values. The relative change in RMSE of soil moisture (SM; for PS) and neutron counts (N; for CRNS) are shown on the horizontal axis. The relative change in RMSE of latent heat is shown on the vertical axis.

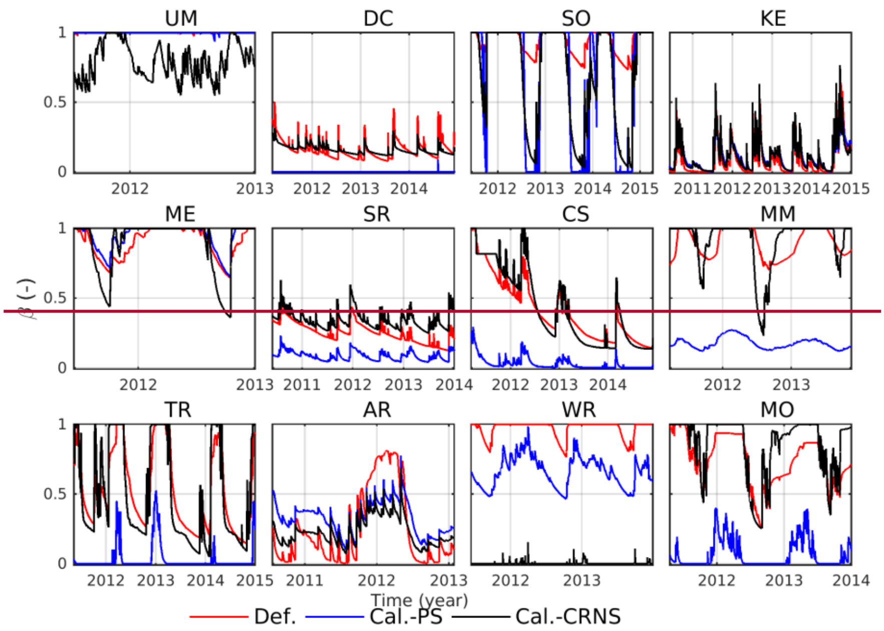


Figure 13: Soil moisture stress factor beta timeseries for all default runs and single objective (SM) calibration solutions.

Appendices

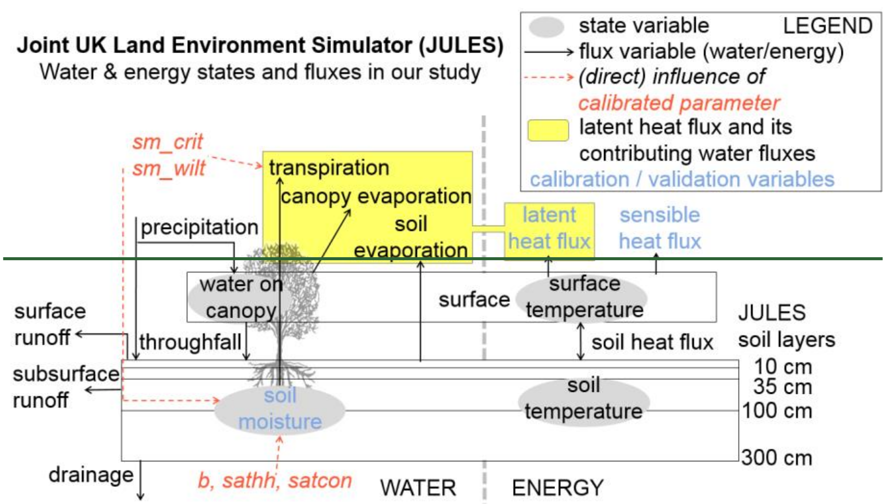
Appendix 1: Additional Data and Methods tables and figures

Table A1.1: PS types, installation depths, number of profiles. In the last column it is shown to which JULES layer the observations were linearly interpolated.

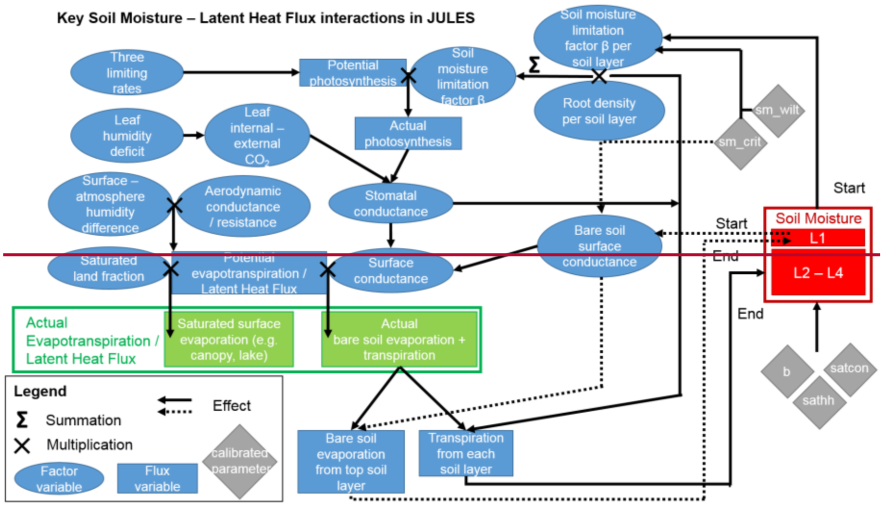
Site	Type	Installation layers (cm below surface)	Number of profiles	Interpolated to JULES layer (cm)
UMBS	Campbell Scientific CS615 / CS616 (reflectometer)	0-30 average	1	0-10
Desert Chaparral UCI	Campbell Scientific CS616 (reflectometer)	0-30 average	4	0-10
Soaproot	Campbell Scientific CS615 (reflectometer)	0-30 average	4	0-10
Kendall	Stevens Water Hydra Probe	5,15	1	0-10
Metolius	Campbell Scientific CS615 (reflectometer)	0-30 average	1	0-10
Santa Rita Creosote	~~TDR~~Time Domain Resistivity	2.5,12.5	6	0-10
Coastal Sage UCI	Campbell Scientific CS616 (reflectometer)	0-30 average	4	0-10
Morgan Monroe	Campbell Scientific CS615 (reflectometer)	0-30 average	1	0-10
Tonzi Range	ThetaProbe (ML2)	0,(20)	2	0-10,(10-35)
ARM-1	Decagon Echo2 EC-20 (capacitance)	10,25	2	10-35
Wind River	Campbell Scientific CS615 (reflectometer)	0-30 average	1	0-10
Mozark	Delta-T	10	1	0-10

Table A1.2: Data used per site for downward longwave radiation (L_{win}), net radiation (R_{net}), and atmospheric pressure (P_{atm}). A is Ameriflux Level 2, COSMOS is COSMOS website, and NLDAS is National Land Data Assimilation Systems Forcing data Phase 2 (http://disc.sci.gsfc.nasa.gov/uui/datasets?keywords=NLDAS).

Site	$L_{w,in}$	R_{net}	P_{atm}
UM	A		A
DC	~~A~~	A	COSMOS~~A~~
SO	NLDAS~~A~~		A
KE		A	~~COSMOS~~A
ME	~~NLDAS~~A		~~COSMOS~~A
SR	A		A
CS	~~A~~NLDAS		COSMOS~~A~~
MM	A~~NLDAS~~		A
TR	NLDAS		~~COSMOS~~A
AR	A		A
WR	A		A
MO	A		A



~~**Figure 3: JULES water and energy states and fluxes most relevant to our study. Calibrated parameters and their direct effects shown. Please, refer to Table 2 for detailed description of parameters.**~~



~~Figure A1.1: Key interactions between soil moisture and latent heat flux in JULES. The parameters calibrated in this study are indicated.~~

~~Appendix 2: Additional Results and Discussion tables and figures~~

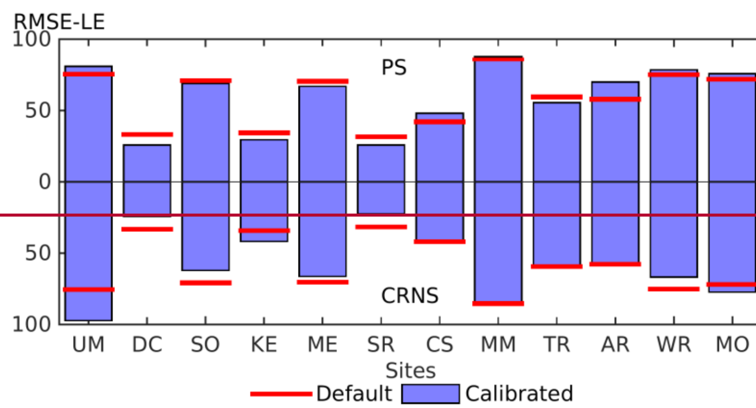


Figure A2.1: Root Mean Squared Error of simulated versus EC observed latent heat flux (w m^{-2}). The metric was computed over hourly day time data only. The upward bars represent PS-calibrated solutions, the downward bars represent CRNS-calibrated solutions. The Default runs are represented by the red horizontal lines, displayed twice for each site, to compare with both calibration strategies. **Appendix 3:**

Appendix 2: Parameter sensitivity analysis

We investigated if a change in some parameters had a relatively small effect on soil moisture, while at the same time having a large effect on latent heat flux. In such case calibration could yield better soil moisture, but the parameter value might be inappropriate for latent heat flux. The inverse (influential on SM but not on LE) could also occur. We explored this by performing a sensitivity analysis with Morris' method (Morris, 1991) as implemented in the SAFE Toolbox (Pianosi et al., 2015) on the exact same parameter value ranges as used during the calibration. We computed the sensitivity indices (mean and standard deviations of the elementary effects) on the RMSE values (i.e. our Objective Functions) of simulated vs. observed PS soil moisture, simulated vs. observed CRNS neutron counts, and simulated vs. observed latent heat flux (day time only).
The results (shown for four sites in Figure A2.2 of Appendix 2) were consistent across most sites: all three Objective Functions were most sensitive to changes in parameter b and least sensitive to the wilting point multiplier. Finch and Haria (2006) also found JULES parameter b to be most influential on soil moisture and latent heat, at a UK chalk site. The critical point soil moisture content was influential with respect to soil moisture / neutron counts but had at all sites less effect on latent heat flux. This could relate to the self-adjusting behaviour of the wilting point and critical point soil moisture described in Section 3.3. The lack of effect from the wilting point soil moisture content can probably be attributed to the use of the multiplier, despite being a common approach (Prihodko et al., 2008; Rosolem et al., 2012); a certain multiplier value could be good in combination with a certain value for the critical point but not for a different critical point value.
We have evaluated the differences between default and calibrated parameters and their degree of physical realism in Supplement 3. If parameter values obtained after calibration were not physically feasible (e.g. representing a sandy soil while there was a clay soil) then, if model structure is assumed to represent biophysical processes sufficiently well, that could yield undesirable results. We wide ranges of parameter values, especially for the water entry pressure (sathh) and the saturation hydraulic conductivity. However, physically realistic parameters also often provide unrealistic results (Gupta et al., 1998).

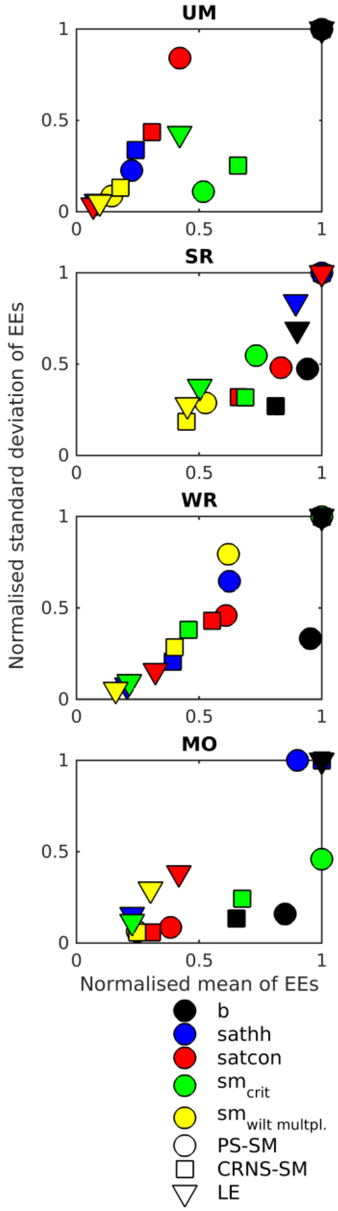

UM
SR
WR
MO
Normalised standard deviation of EEs
Normalised mean of EEs
0
0.5
1
b
sathh
satcon
sm crit
sm wilt multpl.
PS-SM
CRNS-SM
LE

Figure A22.21: Morris sensitivity indices for the five calibrated parameters at four sites: UM, SR, WR, and MO. The three Objective Functions are the RMSEs between simulated and observed PS soil moisture (PS-SM), CRNS neutron counts (CRNS-N), and latent heat flux (LE). The mean elementary effects (EEs), representing the main effects, are displayed along the horizontal axis. The standard deviations of the elementary effects, representing parameter interactions, are shown along the vertical axis. Values were normalised to the most influential parameter for each Objective Function.

~~**solutions**The foremostThis is due to horizontal and vertical soil heterogeneity (e.g. macro-pores, soil layering in organic matter and particle sizes, stones) not (properly) accounted for in the model. The form of Richards' Equation implemented in JULES does not take these natural features into account and therefore effective parameters can be expected to differ from values representing a soil with a homogeneous matrix only. Explicitly taking into account heterogeneity issues like macro-pores could improve JULES' performance (Rahman and Rosolem, 2016), but doing so was beyond the scope of our study.~~

~~The vertical discretisation of the soil (layers of 10, 25, 65, and 200 cm) is not suitable for solving Richards' Equation. Layer thicknesses of no more than one centimetre near the surface are required to accurately simulate water fluxes. Insufficiently fine discretisation at both field (~eddy-covariance footprint) and hydrological catchment scale may not yield realistic water fluxes using realistic parameter values. (Smirnova et al., 1997; Lee and Abriola, 1999; Van Dam and Feddes, 2000; Downer and Ogden, 2004; Beven and Germann, 2013). Testing the effects of this issue on our results, which would entail changing the vertical discretisation of the soil column in JULES, was beyond the scope of our study.~~

~~due to the simplifications of natural processes inherent to models,~~

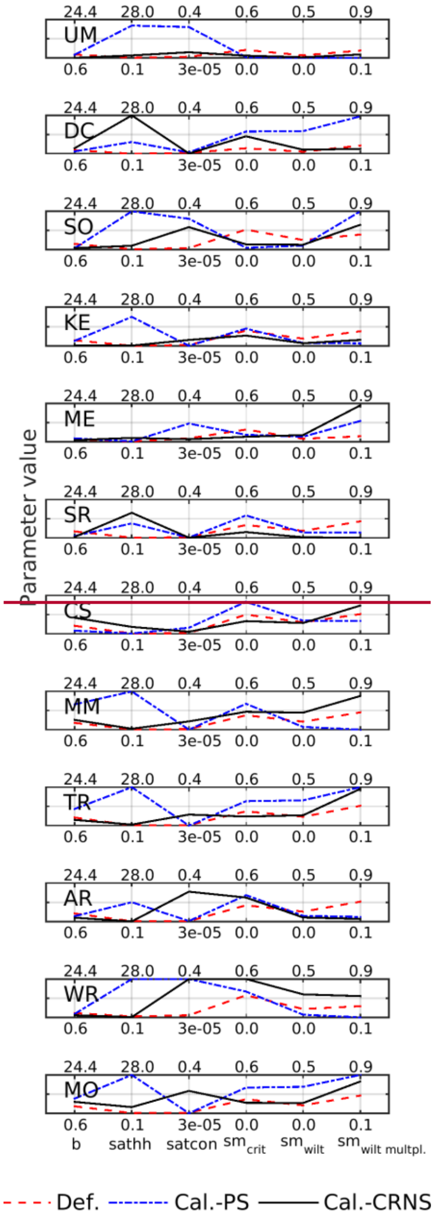

24.4 28.0 0.4 0.6 0.5 0.9
UM
DC
SO
KE
ME
SR
CS
MM
TR
AR
WR
MO
0.6 0.1 3e-05 0.0 0.0 0.1
Parameter value
b
sathh
satcon
sm_{crit}
sm_{wilt}
$sm_{wilt\ multpl.}$
Def.
Cal.-PS
Cal.-CRNS

Figure A2.3: Parameter values of the Default, PS-calibrated, and CRNS-calibrated solutions for each site shown in parallel coordinate plots. Parameter b is Mualem-Van Genuchten parameter 1/(n-1), sathh is Mualem-Van Genuchten parameter α^{-1}, satcon is the saturated soil hydraulic conductivity in the Mualem function, sm_{crit} is the critical point soil moisture content, sm_{wilt} is the wilting point soil moisture content, and is the multiplier (multiplied with sm_{crit}) which was actually calibrated instead of sm_{wilt}. The upper and lower bound values of the calibration ranges are shown for each parameter and were the same for all sites