Response to Editor's comments

Dear Editor,

We thank you for the comments raised on the submitted manuscript. We have accommodated the suggested modifications as outlined below. Our responses to the comments follow the comments made (repeated in quotation marks), followed by our responses in bold. Below is a list point-by-point response to the comments raised on the previously submitted manuscript.

"What information can be expected from an averaged soil moisture (in lateral space, over depth and over time) in general?"

- **Whereas soil moisture has a short correlation length, averaging soil moisture over a profile can yield valuable information concerning month-to-month variability in soil moisture and the associated seasonal cycle. A study which largely inspired the analysis for this paper, presented related information from a number of sensors at various soil depths. See for example the study by Yuan and Quiring (2017)**

"What added information or reliable estimate can be derived from the respective sources? Or in other words what is the trivial part which can be gained by much simpler means like the water balance or distributed linear stores etc.?"

- **Assuming "respective sources" in the above comment refers to the data sources (e.g., modelled, satellite, satellite-based modelled data sources used) then we anticipate that: 1) Testing the ability of the different sources of data to capture the soil moisture dynamics as driven by known (in some cases quantified) as well as unknown (and therefore also unquantified) drivers of the fluctuations in soil moisture at the study area and study sites. 2) We acknowledge that the water balance/distributed linear stores approach can provide a simplified way of capturing average seasonal effects and rainfall-soil moisture relationships. The more complex land surface model (coupled land surface-atmospheric model), that is presented here, capture some of the key processes which drive fluctuations in soil moisture such as vegetation. This presents an opportunity to investigate if the modelled drivers can yield patterns that are representative of the SM over time. For practical purposes: if a simple model (e.g., water balance model) is more accurate than the more complex models and satellite observations then one will use the simpler model. One of the goals of the paper is to evaluate if these data sources, and more specifically the process based model CCAM-CABLE output is reliable or accurate for the specific area (latitude/longitude) having a specific climate condition (semi-arid) and ecosystem (savanna). We attempt to establish if the models in particular CCAM-CABLE can be used at all to simulate local and regional soil moisture conditions for the selected resolution (spatial and temporal). We conclude on page 29 (paragraph 785-790) with regard to using satellite derived SM or process based models (i.e., CCAM-CABLE). Regarding that all these models are limited in their assumption about the drivers of soil moisture, the evaluation output can potentially inform their fitness for purpose and potential avenues for improvement.**

"Given your expertise and previous publications on the method, I suspect that allowing a slightly more clear focus of the manuscript towards the actual topic of the special issue would be beneficiary to both - the manuscript itself and the contribution to the special issue."

- **We thank the Editor for the recommended additions, in order to improve the manuscript with this regard and align it with the special's issue actual topic, the standardised soil moisture index (SSI) is computed and used to analyse spatial and temporal as well as inter-model variations of the simulated soil moisture. The calculation demonstrate a way of establishing a link between the soil moisture states and topography, and hence landscapes. This additional discussion on landscapes is described in Sect. 2.4.4 and the results are presented in Sect. 3.2.**

"You summarise that the coherence of CCAM-CABLE and GLEAM results indicate "that the key physical processes that drive soil moisture in CCAM-CABLE and GLEAM, at the surface and root zone, lead to an appreciable degree of mutual information." Since mutual information is well-defined in information theory, I would expect some more detailed reference and calculus what information is shared and conveyed in the different approaches. More importantly, having a strongly seasonal input signal, I would expect quite substantial shared information between monthly averages of soil moisture and rainfall. So what information is added by the models then?"

- **We thank the editor for the suggestion, to address the above comments, in the updated manuscript we compute mutual information using Shannon entropy as described by Kraskov et al. (2004). Detailed calculus and references in relation to the calculation are outlined in Section 2.4.5 of the manuscript.**
- **A calculation of mutual information as part of model evaluation, gives a perspective on the extent to which different soil moisture models captures similar processes on a grid cell as driven by variables such as precipitation and temperature.**
- **A comparison of rainfall and soil moisture signal at Skukuza and Malopeni on Figure 2, reflects that there could be a time lag between the observed soil moisture and accumulated monthly rainfall. Regarding that the models do differ in their assumption about the drivers of soil moisture, the mutual information calculation in this case becomes instrumental in diagnosing the extent to which differences in the respective model outputs are comparable.**

"As far as I understood, your reference data at Skukuza and Malopeni are both in sandy loamy soils. Your gridded soil map however, classifies the area near the stations as silty loam. If the soil types are important (as you claim), how does this affect your results? How does this relate to the covariance structures as estimate of uncertainty, when the covariance for the reference station sites is very low and other references are lacking? Are this topographic effects or patterns in the rainfall estimates or other compartments of the water balance? "

- **We thank the Editor for raising this comment. The gridded soil type data present a 25*25 km^2 grid cell indicating the dominant soil type. The dominant soil type at the grid cell is deemed representative of the entire grid cell thus introducing a soil type homogeneity assumption at a grid cell. The grid cell within which the flux towers are located is silty loam. However, at the flux tower foot print (i.e., 1 km^2) the soil type is found to be predominantly sandy loam. This is now clarified in the manuscript in Sect. 2.1.1 and Sect. 2.1.2 for Skukuza and Malopeni respectively. This however does not affect the interpretation of the mutual information presented at the spatial resolution of 25 km. Furthermore, instead of using the covariance mutual information is now computed using the Shannon entropy as discussed in Section 2.4.4 of the manuscript.**

- **A discussion on spatial patterns of the standardised soil moisture index is added and used to link to link the soil moisture spatial patterns to topographic features at the study region and hence the associated landscapes.**

"Having a strongly seasonal signal and rather limited reference data, I am not quite sure that I understood the value of employing the cross-wavelet analysis. I do not find it very surprising to see a significant annual period and I would not expect any shift in this frequency. What can one learn about the validity of CCAM-CABLE or GLEAM-v3a if there is no ground truthing data involved? What is the frequency spectrum of the reference time series? Maybe even the mean annual soil moisture dynamics could serve as a reference when the inter-annual deviations are specified? Again, my question would link back to the information contained in the different sources under study."

- **We thank the editor for the recommendations. To address this comment, we resorted to using observed data for the cross-wavelet analysis. To achieve this, we gap-filled the in situ observations using the multiple imputations method. This is unavoidable regarding that the cross-wavelet analysis only works with complete datasets. Gap filling was only employed for the Skukuza since it has the least gaps as compared to Malopeni. As a result, the cross-wavelet analysis is now computed between CCAM-CABLE, GLEAM-v3a and in situ observations both at the surface and rootzone, thereby presenting an opportunity to evaluate the time lag and cyclic features of the modelled soil moisture signal. Furthermore. The gapfilled dataset is also used in the simulation of the onset and offset of the wet period analysis. The two analyses mentioned are carried out for the years ranging between 2001 and 2014.**

"Moreover I find it rather difficult to trace the respective reference data throughout the manuscript. Since you argue about using flux station data, I suspect that the water and energy balances can be closed with the available data. However, you only compare the most difficult to evaluate part of the water balance (soil moisture). Would there reside as much information about the respective states and processes in ET, P and Q? How do these compare to the downscaled GCM and GLEAM inputs?"

- **A detailed analysis on ET has been done on a separate manuscript currently under review. Conducting a water balance study would be interesting. However, certain components of the water balance such as drainage and runoff are not measured at the sites.**
- **To maximise the use of the reference datasets, we gapfilled the in situ observations as mentioned above. This gap filled datasets enabled cross-wavelet analysis and the computation of the onset and offset of the wet period which is only possible on a complete dataset. However, the Malopeni datasets would still not feature in these analyses due to large gaps in the data, imputation of such datasets is likely to result in unreliable estimates of soil moisture.**

"I find it rather difficult to extract information from Figure 5 and 6. Since you are using an R2 for evaluation, maybe a respective Observed vs. Modelled dotty plot would be more insightful about the fit? How about further measures of correlation? Figure 6 is particularly difficult to evaluate. The lack of data appears to lead to overly positive rankings of correlation (in Fig. 7)."

- **We thank the Editor for the recommended improvements. Figures 5 and 6 are now represented quantitatively using the R^2 values plots (Fig. 3). However, the dotty plots of Observed vs. Modelled as recommended by the Editor are also presented in Appendix A.**
- **We agree with the Editor that the lack of data at the Malopeni site may lead to high correlations. This is expected if the number of data points that are compared are following a similar pattern.**

"Coming back to the overall manuscript and the proposed major revisions, I also suggest to include the figures along the lines of argumentation. Fig. 1 might benefit from some topographic information instead of politic boundaries. Fig. 2 did not give me much insight, which might be due to my scepticism to the suitability of the employed method. Fig. 3 and 4 might be combined?"

- We thank the Editor for the recommended improvements:

- **Fig. 1 – has been updated as suggested by the Editor, to include topographic information (i.e. altitude).**
- **Fig. 2 – has been moved to the appendix for description purposes.**
- **Fig. 3 and 4 have been combined as recommended by the Editor.**

"With regard to your author comments to the reviews, I found it relatively difficult to grasp your intended lines of revision. I sincerely hope you will find my comments helpful in this respect. Please also note that both reviewers indicated room for improvement especially for scientific quality and scientific significance of the manuscript."

- **A number of revisions has been made on the updated manuscript considering comments from both the reviewers and Editor. These include minor and major changes such as:**
    - **Clarifying the in situ data constraints as raised by the two reviewers (L102-L103), improved descriptions of the passive and active satellite sensors (Section 2.2).**
    - **The description of the CCAM-CABLE model has been improved to address some of the concerns raised by the reviewers (Section 2.3.1); the improvements are contained in L204-216.**
    - **Motivation for using monthly data in the analysis is included in L257-260.**
    - **We cite the reference for Eq.1 in L283-284.**
    - **To maximise the use of reference data, we extended the use of in situ data to the cross-wavelet analysis and the simulation of the onset and offset of the wet period. This has been archived by employing the multiple imputations procedure since these analyses require complete datasets. The description of the multiple imputation procedure is outlined in L243-352.**
    - **The covariance analysis is now replaced with mutual information described in Section 2.4.5 as recommended.**
    - **The standardised soil moisture index is now computed among the models and used to link soil moisture simulations to landscapes.**

"Since our special issue is an inter-journal one which opens the opportunity to publish related data in ESSD, I would like to encourage you to consider publication of the data and scripts either as appendix to your MS or as own data publication. As far I understood, the data you used is partly hosted on publicly available repositories already. However you like to see the HESS data policy for details: https://www.hydrology-and-earth-system-sciences.net/about/data_policy.html"

- **The links to the analysis scripts are now shared**.

Regards

Floyd Khosa