Editor’s comments:

Dear authors,

We have received the reports from two reviewers on your revised manuscript. They both appreciate your efforts.

However, Reviewer #2 points out that several of his/her major comments have not been addressed with sufficient depth and in sufficient detail. I largely concur with this view.

The main points are the following:

(1) the previous concern that the manuscript seems more like a technical report than a scientific paper has not yet been convincingly addressed in the revised version. What are the knowledge gaps and science questions to be addressed in your analysis? What is the novelty of that? These points remain very vague and unclear. Yet, they should be the core of any scientific paper.

(2) the manuscript lacks balance in the reporting and analysis of results with respect to the expectations raised in the introduction

(3) some parts of the analysis remain rather superficial and need to be developed in more depth, such as the relationships of attributes with model performance.

These points, as in detail described by Reviewer #2, need to be substantially and convincingly developed before this manuscript could potentially be considered for publication in HESS.

I thus return the manuscript to you with a request for such an in-depth major revision. The revised version will be sent out for a further round of reviews.

I am looking forward to a revised version of your manuscript

Best regards,

Markus Hrachowitz

Response:

Dear editor,

Thank you for your comments. We have revised the introduction to better present the research gap that motivated our study. Specifically, we have incorporated explicit mention of the challenge posed by parameter estimation difficulties, as highlighted by Fisher and Koven (2020), within the introduction (lines 67-77, tracked version). Moreover, we have streamlined the introduction to better present the specific problems our research aims to tackle.

Additionally, we have expanded our analysis by adding Sections 5(c) and 5(d). In Section 5(c), we have provided a detailed comparison of the simulation skill between the VIC and Noah-MP models (at lines 581-614 and illustrated in Figures 10, 11, S11-S13). This comparative analysis encompasses baseline, post-calibration, and post-regionalization scenarios. In Section 5(d), we compared the performance between post-regionalization and post-calibration runs for each model (at lines 626-641 and demonstrated in Figure 12) (For further details, please refer to our response to Reviewer 2 and the manuscript). These results should provide guidance for selecting the most appropriate model depending on the hydrological condition being analyzed.

Fisher, R.A. and Koven, C.D.: Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. Journal of Advances in Modeling Earth Systems, 12(4), p.e2018MS001453, 2020.

Reviewer 2's comments:

Dear editor, dear authors,

I'm sorry to say that my concerns and critical comments (from the first review round) have only been partially addressed. My first comment was related to the formulation of the research gaps and the scientific novelty of the research objectives. This part still needs to be substantially improved. The revision includes two new paragraphs indicating that one of the new contributions is the application of two models addressing structural model uncertainty. I wonder to what extent and what precisely is a new contribution here. There are numerous papers examining the uncertainty and its reduction using multi-model ensembles (e.g. Her et al, 2019 among many others), so it is important to review existing studies on this topic along with the formulation of what new research hypothesis is tested in the manuscript. The second

new paragraph indicates that the selected models are widely accepted models. I agree. Still there are again numerous comparisons that include these models (e.g. Liu et al, 2020 among many others), so it is important to highlight what are still the open research gaps and what new understanding can be obtained in the current study. From the revisions made, I still do not clearly see what the research gaps are and how the current study adds to a new scientific understanding/knowledge. The calibration of a model is more a technical question than a scientific objective of a research paper.

A large part of the results refers to the regionalisation. The scientific objective of this part (again, research gaps and novelty) are not covered in the Introduction in its current form. There are studies linking the performance of regionalisation methods to the climate or physiographic attributes so it will be interesting to link the current objectives and results with previous studies. Moreover this part seems to be more larger and significant than the calibration part, but many results are just presented in the supplement.

The final remark is related to the interpretation of the correlations between calibrated model efficiency and physiographic attributes. In my opinion most of the correlations are rather weak (even statistically significant) and some more detailed attribution to the processes (and how these are represented in the models) will be an interesting extension of the analysis.

References

Her, Y., Yoo, SH., Cho, J. et al. Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. Sci Rep 9, 4974 (2019). https://doi.org/10.1038/s41598-019-41334-7

Liu et al (2020) https://doi.org/10.1016/j.jhydrol.2019.124534

Response:

Dear reviewer,

Thank you for your comments. We have revised the introduction to better present the research gap that motivated our study. Our experience is that calibration of models such as those we implemented at continental or subcontinental scales has been hampered by the difficulty of parameter estimation. This point is in fact made by the by Fisher and Koven (2020), which identifies the problem (stated slightly differently) as one of three outstanding

“grand challenges” in land surface modeling. This is the main motivation of our research, and we now so state explicitly in our introduction (lines 67-77, tracked version). Additionally, we have refined the entire introduction to enhance its clarity and focus more directly on the specific problems our research addresses.

Regarding the results interpretation, we expanded our analysis in Section 5(c)Comparison of VIC and Noah-MP simulation skill (at lines 581-614 and Figures 10, 11, S11-S13). We compared the skills between two models for baseline, post-calibration and post-regionalization. We explored where the basins VIC does better than Noah-MP and vice versa. We also investigated the skills characteristics when we divided the study region into four categories: coastal snow dominated basins, coastal rain dominated basins, interior wet basins and interior dry basins. We found that when partitioned into hydroclimatic categories, VIC outperforms Noah-MP in all but interior dry basins following regionalization, where Noah-MP is better. We further evaluated the performance of the two models after regionalization in simulating annual average flows ,flood flows (POT3), and low flows. We found that VIC outperforms Noah-MP in simulating annual mean streamflow and flood simulations in most cases. Conversely, Noah-MP performs better for low flows.

We added Section 5(d) Comparison of Post-Regionalization and Post-Calibration Performance (at lines 626-641 and Figure 12). In this section, we analyzed the performance differences between the regionalized and calibrated runs for each model. We found that post-regionalization, both VIC and Noah-MP performance declines in comparison with the calibrated run, with declines more pronounced for VIC. The performance degradation is greatest in interior dry basins for both models.

These findings provide insights for choosing the most suitable model based on specific hydrological scenarios, thus enhancing the accuracy of predictions and improving water resource management.

Fisher, R.A. and Koven, C.D.: Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. Journal of Advances in Modeling Earth Systems, 12(4), p.e2018MS001453, 2020.