

Hydrol. Earth Syst. Sci. Discuss., author comment AC1
<https://doi.org/10.5194/hess-2021-271-AC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Tesfalem Abraham et al.

Author comment on "Quantifying the Regional Water Balance of the Ethiopian Rift Valley Lake Basin Using an Uncertainty Estimation Framework" by Tesfalem Abraham et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-271-AC1>, 2021

We thank Reviewer 1 for the thoughtful suggestions and constructive comments that we believe highly strengthening the quality of our study. We structured our responses directly below each reviewer comment in bold font.

This is a review of "Quantifying the Regional Water Balance of the Ethiopian Rift Valley Lake Basin Using an Uncertainty Estimation Framework" by Abraham et al. The paper presents a method to predict streamflow on ungauged basins in a region of Ethiopia. The method first uses behavioral parameter sets to identify parameters that are to be used in the regression analysis (for identifying relationships between parameters and catchment descriptors). A few parameter selection methods are proposed, which are then tested in regionalization using a leave-one-out approach. The authors then extend the study to include an analysis on the streamflow elasticity to better understand how changes in precipitation can affect variations in streamflow.

I found this paper to be informative about the study region, and I think there is potential to this paper. However, I think that there is a fair amount of work left before it can be considered for publication in HESS. Here are my concerns:

We thank Reviewer #1 for understanding the potential ideas that we tried to explain throughout the manuscript. We will incorporate all the concerns arising from the review to complete the manuscript.

1. The literature review is quite outdated. Many seminal papers are presented in this paper, but a lot of work has been done in the past few years that could be used to set a clearer context to this study. For example, see Guo et al. 2021 for an up-to-date review on regionalization approaches across the globe. This will also help contextualize the claim in lines 57-59:

Guo, Y., Zhang, Y., Zhang, L. and Wang, Z., 2021. Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. Wiley Interdisciplinary Reviews: Water, 8(1), p.e1487.

We agree and will add more recent literature on regionalization to explain in more detail the research gaps that we address in our study including the study by Guo et al. (2021). We will clarify that, other than previous regionalization studies, we consider three possible sets of model parameters and evaluate them

for their adequateness for regionalization, and use the results of multiple spatial split-sample tests to quantify the uncertainty that goes along when regionalizing parameter sets from a low number of catchments. In addition, we tried to show that globally available precipitation and potential evapotranspiration data sets can be used for regional modeling studies.

2. Lines 77-89: This section is more of a description of the method. I suggest the authors better define the problem they are trying to solve and provide clear objectives. The way they are presented in the paper, the objectives are not clear to me.

Thank you, we will modify this paragraph, to make sure that our research problems and objectives are clearly stated.

3. Lines 106-107: This sentence is unclear. Which global parameter sets? Which climate forcings? Please be specific. I suggest removing this "overview" paragraph and focus on a step-by-step description of the methodology. The steps can then refer to the figure to see where they fit in.

The Reviewer is referring to the statement "We apply global parameter sets and climatic forcings for...". We will clarify it by defining what we mean by global parameter sets and climatic forcing at this point.

Also, we agree with the comment and will remove the overview paragraph and provide a detailed description of the overall methodology. We will make sure that Fig 2. is referenced in the appropriate location in this section.

4. Figure 1: I think there are steps missing here. For example, how does the calibration fit into this process? I assume it is in the parameter estimation step, where the best set from the behavioral sets are identified, but that should be clarified. And the parameters are only computed on the gauged basins (obviously) but is there also a validation step?

We assume the reviewer is referring to Figure 2 because Figure 1 mentioned above is only a description of the study area. We will make sure to include the missing steps mentioned by reviewer #1. We will update the flow chart by adding the best parameter estimation steps from calibration, validation, and stable parameter sets in the "Parameter estimation in gauged catchments" box. In addition, we will clarify the best parameters selected from the validation steps.

5. Section 3.1: This section presents the data and catchment properties. I would suggest moving this to the "study area" section, since it deals with the data and properties of the study area.

Thank you – we will move that section accordingly.

6. Also, it would be important to state why the data are only available until 2007. Perhaps there was a decision to close gauging stations, etc., but for the reader it feels as if the study was completed on data that has not been updated in the past 14 years.

This is a good point. We explain that the streamflow data we collect from the Ministry of Water Irrigation and Energy (MoWIE) in Ethiopia ends in 2007. After this period, administrative structural change was made and individual basin authorities are given the mandate to collect and manage the data. Due to such changes, a long period (2008-2015) of data is still not available for the user, and the available ones from 2016 up to now are of poor quality than the ones we choose for our study period (1995-2007).

Therefore, we will make sure to update this information.

7. Lines 137-140: Please state clearly which properties and data are used as catchment descriptors here. These sentences as-is are pretty vague.

We will add more detail to the section to clarify the catchment descriptors that are used.

8. Lines 211-213: Indicate that CV is the standard deviation divided by the mean.

We will define CV as the standard deviation divided by the mean.

9. The methodology presented in section 3.5 seems biased, in my opinion. At this stage, towards lines 230-235, the authors explain that the three regression models (trained on parameters coming from the calibration period, validation period and "stable" parameters) are verified on the validation period only. This is problematic, because at this stage I could foresee that the regression model trained on the "best validation" parameter set would probably be the best during verification. This is because the hydroclimatic conditions play a major role in the ability to regionalize in the first place (as stated a bit further in the paper). So parameter sets that are "good" on this period, are probably going to be better in regionalization than parameters trained on other periods, simply because the hydroclimatic conditions are more similar by default (given the proximity of the catchments). I think that to even the field, the same process should have been completed by testing the regression models on the calibration period and the full period as well, to complete the experiment design. I would be fairly confident that the regression models would perform best on their corresponding training period. I suggest the authors include this analysis in a revised version to be able to analyze this aspect and contextualize the claim that one regression model is "better" than the others. This can be done by updating figure 5b, where we can see the effect I am referring to.

Thank you for these helpful suggestions. It is correct that the procedure we applied to evaluate the regionalization performance was done only for the validation phase by considering this period to be a prediction phase. We recognize that the hydroclimatic conditions of each evaluation period could potentially affect their performance. We will therefore repeat the evaluation reserving one independent time slot different from the currently used calibration and validation periods to independently evaluate the three different regression models. We will carry out additional evaluations for the calibration period and for the whole simulation period as well to compare the performances in each period.

We will also update Figure 5b, by including the evaluation of regionalization performance in the calibration periods and for the whole simulation periods.

10. Figure 4: It is important to note that the small distributions of 5, 8, 13 and 15 are caused by the fact that they barely hit 0.5 NSE, meaning that only a few parameter sets are even allowed in this analysis. Whereas catchments with higher NSEs have many more parameter sets that are above 0.5. Perhaps one approach would have been to keep only the top 0.1 NSE from the maximum or something similar. Why keep parameter sets that have NSE values of 0.5 if some runs give 0.7 / 0.8 NSE? It seems that these are less "behavioral" than those at 0.7 if the maximum is 0.75. Perhaps keeping a fixed range vs their maximum value would allow for a better comparison.

Thank you for this important point. As stated by reviewer #1, we used behavioral parameter sets resulting a NSE above 0.5 resulting in different remaining numbers of parameter sets. For catchments #5, #8, #13, and #15, 741, 25, 593, and 2,877 parameter sets remain after applying the threshold. With 593

parameter sets and larger, we have reason to assume that the distributions of model parameters and NSE values of catchments #5, #13, and #15 are not biased by a low sample size despite being just slightly above the 0.5 NSE level. We will provide this information to the revised manuscript and add to the discussion of the revised manuscript that the distributions derived from the remaining parameter sets of catchment #8 might be biased by a low sample.

However, we prefer to remain with the threshold-based separation of behavioral and non-behavioral parameter sets as a threshold-based approach helps to explicitly state under which minimum performance requirements, i.e. $NSE \geq 0.5$, our regionalization by the CV-weighted regression was conducted.

11. Following comment #10, line 275-276: "The parameters in these catchments remained insensitive" should be revised. It is not that they are insensitive. It is that the only few parameter sets with $NSE > 0.5$ had to have those parameter values.

We will make sure to revise as suggested.

12. Figure 5 is extremely vague for me, I am not sure what I am looking at even after reading the text, legend and caption a few times. Please consider displaying in another fashion or providing a more detailed interpretation.

Thank you. We will prepare Figure 5 again in a more readable format and clarify it in the caption.

13. Lines 297 – 315: I think these results should be provided with some sort of note that they are strongly dependent on the available dataset and that they must be taken with a grain of salt for the abovementioned reasons: 1- Not a lot of training data; 2- some catchments have a large spread of possible values due to having a $NSE \gg 0.5$, whereas others have NSE barely above 0.5, which plays on the identifiability of parameter sets.

We will rewrite this paragraph to clarify that our results are only based on available streamflow data and on the global forcing datasets.

14. Figure 8: Here the CVs are not clear to me. Why do 2 neighboring catchments have similar elasticities, but have CVs that range from essentially 0 to 180% ? a CV of 180 means that the standard deviation is 1.8x the average, so I am supposing that the precipitation is extremely low there? and neighboring catchments are very different in this regard? Please provide a bit more guidance to clarify.

Thank you! These are important questions. The reviewer is referring to the extreme differences that we found for the wettest year CVs resulting in neighboring catchments. The differences in CV are found for the ungauged catchment numbers 22 and 28 and gauged catchment #05 in Figure 8. As we explained in lines 440-443, the precipitation in the ungauged catchments number 22 and 28 is very low compared to gauged catchment #05 as shown in Table S2 and Table 2. In addition, the catchments properties in the ungauged regions are highly different that influences the 14 sets of model parameters derived from the regionalization that increased the resulting CV values. We see extreme differences in annual precipitation values of 928.9 mm, and 631.7 mm in the ungauged catchments number 22 and 28 respectively that is highly different from 1319.6 mm in the gauged catchment #05 (Table S2 and Table 2). Such extreme differences in precipitation together with other catchment properties will cause high variability of the regionalized discharge even in the neighboring catchments. In addition to precipitation variability, this effect is linked with the remoteness of the most of gauged catchments used to establish

the regional model as we noted in lines 440-443. We will expand our elaborations accordingly in the revised version of the manuscript.

15. Line 361: "With an average decrease of 0.40% from calibration to validation..." What exactly does this represent? 0.40% of the NSE value? Of another error metric? Please specify.

Thank you. We meant the change of the average NSE values from calibration to validation by "an average decrease of 0.40 % from calibration to validation" it is. We will make sure to clarify this.

16. Lines 367-369. Please revise following my comment #9.

Will do.

17. Lines 379-388: This section is restating the results. It would be beneficial to restructure the text to focus on the lessons learned from the experiment and dig deeper into the results to explain them and find links with the literature. this entire paragraph (lines 379-389) only has one such sentence of interest, the last one that compares to the Beck et al. 2016 study.

Thank you for the suggestion! We will update this paragraph accordingly making sure to cite relevant literature supporting our experiment.

18. Lines 414-415: Was it really well identified, or is it simply that most parameter sets were barely able to provide 0.5 NSE (and looking at figure 3, it would seem that catchment #15 did not attain 0.5 at all in calibration)?

With 2,877 parameter sets with $NSE \geq 0.5$, we believe that the derived distributions are representative and that our conclusions on the identifiability of the model parameters for catchment #15 are valid. Please see our response to comment 10 of this review.

19. Line 442: "regress" à regression?

Yes.

20. Lines 484-485: This sentence kind of pops up from nowhere and has little relevance to the rest of the paper. I would suggest removing it.

We will remove the sentence.

Finally, the paper should be proofread by a professional English speaker as there are quite a lot of syntax errors which sometime distract from the content. In a similar vein, it is more typical to use a neutral and objective writing style to "depersonalize" the text. Instead of writing "we use the HBV model...", try to use "the HBV model was used...". I am unsure of the official HESS policy on this, but it is good practice.

Thank you. We will perform a professional language check. In addition, we will re-check the HESS policy for the writing style. Both styles are quite often used in different scientific journals and we picked the preference of using the first person pronoun.

Reference

Guo, Y., Zhang, Y., Zhang, L. and Wang, Z.: Regionalization of hydrological

modeling for predicting streamflow in ungauged catchments: A comprehensive review, Wiley Interdiscip. Rev. Water, 8(1), 1–32, doi:10.1002/wat2.1487, 2021.