Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., author comment AC2
https://doi.org/10.5194/hess-2021-271-AC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

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-AC2, 2021

We thank Reviewer 2 for her/his comments and suggestions that will highly strengthen the quality of the study. In the following, we show our responses directly below the Reviewer's comments in bold font.

This manuscript was challenging to assess. The transferability of model parameters calibrated at gauged locations to ungauged locations using a regionalization approach where parameters are estimated using catchment properties able to estimated at the ungauged catchment is, in many ways, well-worn territory, as the authors also note (L45-58). Much of the discussion in Section 5.2 also points to results that are consistent with previous studies. In my opinion, there has been inconsistent success demonstrated in previous studies as to the utility of this approach and the results presented here are no different than previous studies have found.

The question then is both whether the approach presented here represents such a difference from past studies as to be a substantial departure from past practices that it would be of value to report the results and that the study area and catchments are sufficient to make broader conclusions about this potential new approach.

From what I am able to understand about the approach and the catchments, neither of these meet the criteria so as to make a substantial and broader contribution to our understanding of why or how we might improve on regionalization approaches for parameter estimation at ungauged locations.

We thank the reviewer for the general assessment and her/his questions arising on the novelty of this paper. In the following, we clarify the novelty that we believe we had in this paper to our best knowledge, that has not been explored by previous studies. As the reviewer already mentioned, most of the regionalization approaches are proposed and tested with large samples of gauged catchments, only a few also discussed the regionalization using fewer catchments, e.g. 10 catchments in Wagener and Wheater (2006). However, in Ethiopia dense networks of gauging stations are lacking, a situation very similar to other developing countries and regions. Therefore, the aim of our study is not to present an entirely new method, but instead, we want to adapt (see discussions in the following) the commonly used regionalization approach such that we can use it to help the understanding of water balances in the Ethiopian Rift Valley Basin with the reality of low data availability. Arising from our

adaptions, the novelties of our approach are as follows:

i) As the reviewer noted towards L45-58, we discussed the general approaches that have been used for predictions in ungauged basins. However, we showed a different approach by analyzing the impact of using different parameter sets for regionalization. Other than the typical approach of using the best-calibrated parameters of the gauged catchments (e.g. Wagener and Wheater, 2006), we extract three possible parameter sets for regionalization. Although previous work already considered multiple similar parameter sets for regionalization (Livneh and Lettenmaier, 2013), to our knowledge the differences of using the best-calibrated parameter, the best parameter set in the validation period, and the most stable parameter set considering their performance in calibration and validation period have not yet been explored. Using a spatial split sample test, we show that the best parameter sets of the validation provide better estimates of regionalized parameters than the commonly used best-calibrated parameters.

ii) We express the uncertainty of model parameters going along with using a low sample of catchments for regionalization: Due to the low number of gauged catchments in the Ethiopian Rift Valley Basin, the relationships between model parameters and catchment attributes consist of just 14 points. Therefore, the resulting regionalization can be expected to remain uncertain. We quantify its uncertainty by applying the spatial split sample test 14 times leaving out once each of the catchments and therefore obtain 14 regionalization parameter sets that express the uncertainty of regionalized model parameters in our data-sparse region. We are aware of the possibility to regionalize hydrological signatures but recent work showed that their information content is limited (Addor et al., 2018) and that their regionalization should go along with considering discharge observation uncertainties (Westerberg et al., 2016). Nevertheless, to our knowledge, there has not been an uncertainty quantification of regionalization parameters in a data-sparse region like ours. Other than regionalized signatures, our regionalized model parameters will allow to run the model with climate projections.

iii) We show that applying our model with input data derived from global products (MSWEP, GLEAM) can provide acceptable discharge simulations for both gauged and ungauged catchments. In addition, we show the difference of this approach than the previous ones by the application of global data products to provide acceptable regional model in data-sparse regions. The acceptable simulation results that we obtain in the gauged catchments and with the spatial split-sample test indicate that global products can be used as model inputs to provide reasonable simulations in data-sparse regions. In addition, our regionalized parameters are distinct to those of globally regionalized parameter sets such as the HBV parameters (Beck et al., 2016) indicating that even if only sparse data is available, they can improve regional hydrological simulations.

In Section 5.2 the reviewer is referring to two studies by Goshime et al. (2020) and Abebe et al. (2010). "Much of the discussion in Section 5.2 also points to results that are consistent with previous studies. In my opinion, there has been inconsistent success demonstrated in previous studies as to the utility of this approach and the results presented here are no different than previous studies have found."

We thank the reviewer for these points. In L405-406, we tried to show the consistency of highly sensitive parameters (β, F_C, and L_P) to the previous study in the region (Goshime et al., 2020). In addition, in line L407-409 we discussed

the interaction between parameters that could cause the insensitivity of some model parameters. However, we believe these studies are substantially different from ours considering the above-mentioned (i-iii). We will clarify this in the revised manuscript and include the respective literature.

My recommendation is based on a number of what I see as serious methodological and evaluation questions as well as a highly complimentary presentation of a limited application of the approach to only a small number of catchments. I describe these issues in more detail below. If the manuscript does receive a recommendation other than Reject, I also offer additional minor and editorial comments that the authors need to consider in their revision.

(1) Broader contribution of the work

(1a) The use of weighted least squares (L200), although not necessarily a substantial advance, is what I believe to be the novel aspect of the study. Perhaps if this were emphasized more in the introduction and concentrated in more detail with the existing studies, it might become more clear that this is a more substantial contribution that the impression I was left with. Otherwise, this being mentioned is more detail so later in the discussion paper (in the methods) contributes to this point being lost. I would also be more explicit as to how this work differs from Wagener and Wheater (2006) and the follow on studies that have cited that paper.

Thank you. We will make sure to emphasize more on the weighted least square and mention existing relevant literature in the introduction. The difference of our approach to Wagener and Wheater, (2006) and following studies (Lane et al., 2021; Singh et al., 2014) is that we extract three possible parameter sets to show the difference of regionalization performance derived from using the best-calibrated parameter, the best parameter set in the validation period, and the most stable parameter sets as we described more in (i) above and in subsection 3.3 of the manuscript.

(1b) I do not agree that this work is novel because these approaches have only been applied in data-rich regions (L56-58). In my opinion, the reason these methods have been applied in data-rich areas is to test the limits of these approaches. Even then, the results have certainly been mixed. Certainly, you could have chosen a more data-rich area to test this approach and then removed stream gauges to understand the effects of gauges on the performance of the method.

Thank you for this point. This sentence will be clarified; we will make sure to provide more detail about the observed hydro-climatic data limitation in our region. Using the available global data sets, we showed the possibility and reliability of a regional modeling approach including a quantification of the uncertainty that remains due to the data limitations of our study region. We will clarify this, too, in the revised version of the paper.

(1c) Linked to Comment 1b, it is difficult to make broader conclusive statements about the utility of this approach when only 16 (or 14) catchments are being used. Either way, for a regionalization study, 14-16 gauges is a very limited number. I realize that 2 catchments were removed because they were poor performing, which reduced the number of catchments to 14. I am not sure if removing these 2 catchments was the correct thing to do here; are they poorly performing because the underlying model is not a good representation?

Thank you. As we stated in L83-88 there are low numbers of gauged catchments in the Ethiopian Rift Valley Basin, and this comment is consistent with the

response provided above on the low sample of catchments for regionalization in (ii). As stated by the reviewer in (1c) we removed 2 catchments. We mentioned in L188-190 that the reason for their removal was their poor performance due to the fast flow processes and the occurrence of wetlands immediately above the gauge in catchments #06 and #12, respectively. The model structure does not consider these processes and would therefore provide unrealistic results. We will clarify this in the revised version of the paper.

Were these locations removed just to improve your own study results? It seemed as though there was not a solid technical reason to remove these gauges from the study.

Please see the comment above.

(2) Methodological and evaluation concerns

(2a) I missed where non-linear regressions are being used in conjunction with weighted (linear) least squares (L200)? I see later in L218 that the non-linear regression is discussed but with not much justification or explanation as to why this is the case.

The form of Equation 10 looks like the form of a regression equation when the regression was performed in log space and then transformed back to normal space. In other words, the logs of the response and predictor variables were taken to linearize the relation between them (to better ensure the assumption of a linear relation for the regression) and then the regression was performed on the log-transformed variables.

Of course, an additive model in log space is a multiplicative model in normal space. So to get the values back to normal space, Equation 10 is what the regression equation looks like when the additive linear model is re-transformed back to normal space.

Seeing that you do not mention anywhere that you performed the regression on the logs of the response and predictor variables, I am not understanding why you would apply the non-linear equation shown in Equation 10 for this reason. More justification is then needed for the application of Equation 10 to the data.

We explained this already in L216-220 of the submitted manuscript, however, our elaborations may not be clear enough. We will clarify the misunderstanding and try to give more explanation about the non-linear regression option we propose. In this study, we apply the weighted linear regression between the catchment properties (independent variables) and model parameters (response variable). We choose the linear regression for the correlating multiple catchment properties with the model parameter as shown in Table S1. However, we chose a non-linear regression equation on the normal scale (not log-scale) if there is only one catchment property correlating with a model parameter. For instance in Table S1, we can see that L_P is correlated only with Elevation in such cases we applied the non-linear regression.

In addition, to increase the representation of more identifiable catchments, we applied a weighted regression on the normal scale. We are aware of the possibility to do non-linear regression on the log-transformed scale however, the correlation coefficient of L_P with Elevation is superior on the normal scale than the log-transformed ones that can provide a better regression model on the normal scale. Furthermore, we will make sure to include more discussion about the relationships between the catchment properties and model parameters that form the regression model.

(2b) Keep in mind NSE values less than 0 have the interpretation that the mean of the

data is a better model than the model being proposed (in this case, the regionalization model is worse than simply using the mean of the data as the model). NSE values less than 0.5 are likely poor fits and those less than 0.25 are approaching the case where would have been better off using the mean of the observed data instead of the regionalization approach. You make the statement on L383 that "79% of the catchments had a NSE > 0"; however, I do not believe this is a statement that puts the method in a positive light. Surely you could find a simpler model (even the drainage area ratio, perhaps) that would achieve the same success as having 80% of the model results better than using the mean of the data. The reverse of the statement on L383 means that 3 catchments (20%) of the 14 catchments did have an NSE < 0 using this regionalization method. How would one in practice guarantee that they were applying the regionalization to an ungauged location where the method would not provide a worse estimate than the mean of the data?

It is correct that our regionalization approach resulted in low performance in few evaluation catchments. However, in the study, we focus on the challenges in regionalization in data-limited conditions showing the applicability of global forcing data for the regionalization of data sparse-regions considering the resulting uncertainties. Using the best validation parameters, the regression model does not perform well in three catchments as discussed in L383. However, the median value of NSE for the 14 catchments is 0.56 that we believe is a sufficient performance in regionalization that started with an NSE threshold of 0.5 and above (Fig. 5b). Therefore, we believe our approach provides a basis for regional model estimation and uncertainty quantification for low catchment numbers in the data-sparse regions.

Furthermore, our objective is not to create a new regionalization technique however, we try to introduce regionalization methods that can be adapted to data-sparse regions by using global datasets. Poor performances of parameter regionalization are also reported by previous studies. A study showing regionalization of HBV parameters using the 10 most similar donor catchments has resulted in a median daily NSE value of -0.02 and monthly NSE of 0.17 in the 1113 evaluation catchments globally (Beck et al., 2016). Other studies also showed poor performance of hydrologic signatures during multiple regression. For example, a study by Zhang et al., (2018) has performed with a NSE value of 0.16 for the multiple regression of slope. They also found an NSE value of 0.06 while regionalizing the slope of the flow duration curve using a log-transformed multiple linear regression in the leave-one-out approach. The same study has also shown NSE < 0 performance of signature regionalization using a hydrologic model (SIMHYD) on 605 catchments in Australia. Therefore, the abovementioned difficulties in regression of model parameters coupled with the use of global data products in a data-sparse region can be expected to result in considerable uncertainty. However, our approach provides good reason to assume that acceptable median NSE value can be obtained despite low catchment numbers.

(2c) In calculating the NSE based on the actual values of flow, what were the range of flow values? If no attempt to balance the weight of the high and low flows in the NSE calculation, the NSE itself would be most affected by the fit of the model at the highest flows, and thus the NSE may only be a reflection of how well the parameters are estimating flows for the largest flows. For example, a difference of 0.1 cms and 5 cms would be a poor fit but if your high flow values are large (on the order of 100s or 1000s of cms) a difference of 4.9 cms would register as an excellent fit for NSE and this fit – simply by the numerical calculation of the NSE - would swamp any of the fits at the low flows since the differences squared would be so much less. Would it not be better to compute the NSE on the logs of the streamflows? Or at least split the flows into high, low, and mid flows so that these issues of scale are not affecting the interpretation of fit?

Thank you for these valuable points. We did not attempt to balance the weight of the high and low flows in this study. Throughout the 14 catchments, we have highly variable ranges of flow. For instance in catchment #05 flow ranges from 0 m^3/s to 700 m^3/s. For the revised manuscript, we will explore if the use of a more balanced logNSE would improve our results.

(2d) There are no regression equations provided or regression diagnostics for the equations so that one could assess whether these are valid regression equations with statistically significant explanatory variables. To use these regression equations in prediction mode and calculate uncertainty and prediction intervals (which is done in Section 4.2), the behavior of the regression equations must adhere to the properties of a linear regression (statistically significant explanatory variables, homoscedastic residuals, uncorrelated and normally-distributed residuals, and uncorrelated explanatory variables).

We will make sure to provide the regression equation with their prediction intervals. Since several weighted regression equations were derived (i.e. 14 regression equations for every nine parameters), we will provide them with their R^2 in the supplement in an extra xlsx file.

(2e) In Equation 8, the weights are described as 1/CV (the reciprocal of the CV; L213). I was having difficulty understanding this. The CV = standard deviation / mean; the reciprocal is then mean / standard deviation. The weights in a weighted least squares regression are, ideally, 1 / variance. How then were you able to achieve a weight equal to 1 / variance by using the inverse of the CV? This needs to be clarified in more detail so the reader can follow along.

In the weighted regression procedure, higher weights are assigned for more identifiable catchments by considering their performance and variability during parameter estimation. In our approach, given behavioral parameter sets, different catchments showed different parameter variability. Using a weight 1/CV or 1/variance, both cases would result in a similar result. For a parameter, introducing a constant (the mean) into the regression will not change the relative weights to each catchment since the scaling factor is the same, meaning that the regression will remain the same. However, by doing this there is an advantage to compare the weights of a catchment for different parameters because using 1/CV removes the influence of magnitude and units of a parameter.

(2f) For insensitive parameters (Figure 4), such as Mmaxeas, it seems it would be advantageous to incorporate this knowledge somehow into your regionalization scheme, although it would be unclear how this would hold up for ungauged locations. On L432-433, the statement is made "Our study shows the insensitivity of model parameters to be related to catchment properties." I am not sure how that can be. If a parameter is insensitive to model calibration, then it would have no preference for the value; therefore, why would one expect this parameter to be estimable or predictable? Would it not be better to just simply randomly generate a value for this parameter from a uniform distribution of values given the parameter range in Table 3?

Then in an ungauged location, how would one be able to predict whether this was a catchment that was insensitive to the parameter Mmaxeas or if it was one of the 3 catchments (figure 4) that was highly sensitive to this parameter?

Could you simplify your regionalization by only regionalizing sensitive parameters and then assigning a random, uniformly distributed value to the insensitive parameters?

Thank you for this helpful remark. As shown in Figure 4, M_{MAXBAS} showed

insensitivity in all catchments except #05 and #08that are sensitive towards the lower values. For the estimation of parameters in the ungauged catchments, we will incorporate the reviewer's suggestion by generating random values for the insensitive parameters in their parameter range to test if they can improve the regional model evaluation.

Throughout line L423-432, we discussed already the interaction between catchment properties and model parameters. For instance, the insensitivity of K_1 and K_2 is directly attributed to the small drainage area and slope of catchments in #08 and #10. Whereby the increase in K_1 and K_2 may not affect the outflow condition due to the resulting less soil moisture in the upper and lower reservoirs. We also provided an example for this in L427-432. Therefore, our statement in L432-433 refers to the influence of catchment properties on the model parameter identifiability.

(2g) The use of the word "stable" parameter set is not very clear. The definition of the "stable" parameter set is the set of parameters that "shows the smallest difference between the calibration and validation NSE". But this does not consider also picking the parameter set with the highest NSE as well.

Thank you for these points. We will clarify that stable parameters are those that are showing the smallest difference between the calibration and validation NSE as we defined in Eq. 6. We tried to answer the concerns of the reviewer (but not in detail), by picking the most stable parameter set by considering their performance in calibration and validation period in L82-83. We will make sure to explain better that we have already picked parameter set with the highest NSE while selecting stable parameter sets.

It also does not explain how the validation period has a higher NSE than the calibration period for some catchments.

This comment is consistent with our response to question (3a) below.

Lastly, how does this criteria help in determining the best parameter set for regionalization? What is the benefit of transferability when you have a "stable" set of parameters at one location? In other words, what would be the guarantee that a parameter set will work well at another location just because it is "stable"?

Our regional model was tested for parameters derived from the calibration, validation, and the most stable parameter sets. This approach produces three regional models for our study region that would increase the chance to choose the best parameter set for regionalization. In addition, our approach produces a reliable model by reducing the uncertainty that could be propagating from using single parameter sets.

Concerning the transferability of stable parameter sets, our entire procedure shows the possibility to produce a spatially evaluated robust regional model.

(2h) Section 3.5: I am not understanding the validation and selection of the parameter sets (L229-233). From what I could understand, the parameter sets are tested on the validation phase and in leave-one-out mode. What is the leave-one-out method not sufficient itself to assess the performance of the method? Also, it would seem a longer period of calibration (one that includes both of what you term the calibration and validation periods) provide better parameter estimates? I am not understanding why the leave-one-out approach to measure uncertainty is not enough to evaluate the approach?

In L229-233 we present the regional models derived from using catchment properties and model parameters. However, in our split sample test, that we apply before the regionalization, we use three types of best parameters (from calibration, validation, and the most stable parameters sets). We then used these three parameter sets to produce three regression models in the leave-one-out procedure, and select the one performing best (as shown in Figure 5b). Using parameters exclusively from a (even longer) calibration period goes along with the risk of regionalizing over-fitted parameters, which is shown by our analysis that identifies the best validation parameter set at the superior set for regionalization. We will clarify this in the revised version of the manuscript in addition with a stricter evaluation of the regional model following the remarks of reviewer #1.

It is also unclear in the methods when calibration and validation are used. You could use Figure 2 to clarify this. From my reading, in Figure 2, you could modify the box "regional regression" to read "regional regression using calibration parameters" and then "evaluation of the regression procedure using validation and leave one out". Although, as I note, I do understand why the validation and leave-one-out are both used.

Regarding the calibration and validation periods, we mentioned in Section 3.3., L180-181, that the calibration period is set from 1995–2002 and the validation period is from 2003–2007. In addition, we will make sure to mention this in Figure 2. We will also modify the box to read "regional regression using calibration parameters" and then "evaluation of the regression procedure using validation and leave one out".

(3) For these methodological reasons given in (2), there are a number of questions related to the results and interpretations:

(3a) Figure 3 shows that the model performs better in the validation phase for some catchments, which is quite puzzling. Why would parameters perform better under validation rather than calibration for some catchments? I believe this needs to be explained thoroughly, unless I am not understanding the methods, in which case, this needs to be better explained in the methods.

Thank you for this point. We have used behavioral parameter ranges (parRANGE) during calibration with NSE ≥ 0.5, from which we select the best validation parameter. Therefore, from these samples, there will be a possibility of one best parameter set which can perform better than the best-calibrated parameter sets in some catchments. However, we agree with the concerns of the reviewer that, on the normal calibration and validation, using a single parameter set, the best parameters of the calibration period will eventually show superior performance than the validation period. We will add this critical point to the discussion of the revised paper.

(3b) The sections on elasticity and uncertainties would need to be evaluated after the comments in (2) are addressed, as I am not sure the methods themselves were applied in a manner consistent with the assumptions of linear regression nor am I certain the nonlinear regression was needed because a log-transform of the data did not appear to be used.

We will make sure to update this section based on the comment in (2). The comment regarding the selection of the regression options is consistent with the response in (2a).

(3c) Figure 5a: Please add a 1-to-1 line on the figure so that the reader can determine for

themselves how much worse the regionalization method performs. By presenting the x and y-axes at different starting locations, it gives the impression that the methods are somewhat similar, unless the reader looks carefully at the axes values.

Thank you. We will make sure to add a 1-1 line.

(3d) The conclusions discuss how identifiable parameters are able to be reasonably well reproduced but one cannot know a priori which parameters are identifiable at an ungauged location. How would one be able to apply this conclusion in practice then, when a leave-one-out approach is not possible? How would one know which parameters are sensitive and insensitive and at which catchments are there exceptions? Otherwise, this proposed method does not seem to very useful in practice.

Thank you for this remark, which we think, is a misunderstanding. Every regionalization study relies on a set of donor catchments where discharge observations are available. For those catchments, model parameters have to be obtained by inverse parameter estimation, during which parameter sensitivities can be obtained. If well-identifiable, more reliable regional relationships can be obtained for those parameters using the most dominant catchment attributes. Either large or low samples of catchments, a leave-one-out procedure should always be possible, too. We will re-phrase this part of the conclusions for clarification.

(4) The data statement is inadequate. Please note the EGU data policy: https://www.hydrology-and-earth-system-sciences.net/policies/data_policy.html. Having the streamflow data "available upon request" is not consistent with the EGU data policy. If the data are not publicly accessible, a detailed statement as to why this is the case must be stated. Otherwise, the data needs to be placed in a public repository and cited.

Thank you! We obtained streamflow data from the Ethiopian Ministry of Water Irrigation and Energy (MoWIE) by formal request, and they do not allow sharing the data among 3rd parties. In case if anyone wants, this data can be acquired through a formal request. So, we will make sure to add a detailed statement of why data is not publicly available.

Minor Comments:

L152: There should be a clear statement here that these 3 parameters are also calibrated, much like it is stated in line 169. Consider modifying it to read "Three calibrated parameters…"

We will modify this as stated in L152.

Table 2: The headings are not formatted for easy readability and cut off mid-word.

We will make sure to update the headings in a more readable format.

Figure 3 - add the abbreviations Cal, Val, and Stb to the caption.

We will add them.

Line 361: Decrease of 0.4% in what?

Thank you. We meant the change of the average NSE values from calibration to validation. We will make sure to clarify this.

Line 367: What model was 3 regressions? Were there not 6 parameters to estimate via regression? Or do you mean there were 3 explanatory variables in each regression model? If the equations were shown, this would help clarify the number of regressions.

Thank you. We used the three best-estimated parameters from calibration, validation, and stable sets as shown in L192-196. Using these parameters, we derive three regional models for 14 catchments that reproduce the nine HBV parameters as shown in Figure 6. However, in Figure 6 we have presented nine parameters reproduced from using only the best-validated parameters. Therefore, the three-regression models refer to the regional models deriving from using the best parameters of calibration, validation, and stable parameters sets. We will provide the equation of the regression models, please also refer to the reply to comment (2d).

L368: What is the "optimal regional model"? I have not seen this term defined anywhere else in the text?

In this regard, the "optimal regional model" is the best regional model derived from using the best parameters of calibration, validation, and stable sets. The comparison of these optimal regional model performances was shown in Figure 5b, where the NSE of a regional model derived from using the best-validated parameter has shown superiority over the other two. We will make sure to define this term throughout the paper for more clarity.

L369-371: How is the "spatial cross-validation" different from the "Leave-One-Out method"? Only the leave-one-out method was described earlier as a validation method. In L377-378, how could a robust spatial cross-validation be completed with only 14 (or 16) catchments?

Thank you for these points. Throughout this paper, we used these two terms interchangeably and they are not different from each other. Both terms describe the spatial validation of the regionalized model by leaving out one catchment at a time by producing a 14-regression model that quantifies the uncertainty of regionalization as well.

The comment/question referring to L377-378, about robust cross-validation using 14-16 catchment is consistent with our response in (ii) above.

L373: The text states that "this method is more stable and more resilient to errors…" but an explanation would be needed here, as I am not convinced this is the case.

Thank you. We will make sure to explain this more.

L379: Change to read: "A scatter plot of monthly NSE values between parameters estimated from the model calibration and parameters regionalized from the regression equations show…"

Thank you. We will make sure to change this.

Lines 446-452: No evidence is offered to support these points.

We will make sure to add relevant literature that supports these statements.

Reference

Abebe, N. A., Ogden, F. L. and Pradhan, N. R.: Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation, J. Hydrol., 389(3–4), 301–310, doi:10.1016/j.jhydrol.2010.06.007, 2010.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N. and Clark, M. P.: A Ranking of Hydrological Signatures Based on Their Predictability in Space, Water Resour. Res., 54(11), 8792–8812, doi:10.1029/2018WR022606, 2018.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J. and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resour. Res., 52(5), 3599–3622, doi:10.1002/2015WR018247, 2016.

Goshime, D. W., Absi, R., Haile, A. T., Ledésert, B. and Rientjes, T.: Bias-Corrected CHIRP Satellite Rainfall for Water Level Simulation, Lake Ziway, Ethiopia, J. Hydrol. Eng., 25(9), 05020024, doi:10.1061/(asce)he.1943-5584.0001965, 2020.

Lane, R. A., Freer, J. E., Coxon, G. and Wagener, T.: Incorporating Uncertainty Into Multiscale Parameter Regionalization to Evaluate the Performance of Nationally Consistent Parameter Fields for a Hydrological Model, Water Resour. Res., 57(10), e2020WR028393, doi:https://doi.org/10.1029/2020WR028393, 2021.

Livneh, B. and Lettenmaier, D. P.: Regional parameter estimation for the unified land model, Water Resour. Res., 49(1), 100–114, doi:10.1029/2012WR012220, 2013.

Singh, R., Archfield, S. A. and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments - A comparative hydrology approach, J. Hydrol., 517, 985–996, doi:10.1016/j.jhydrol.2014.06.030, 2014.

Wagener, T. and Wheater, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, J. Hydrol., 320(1–2), 132–154, doi:10.1016/j.jhydrol.2005.07.015, 2006.

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A. and Freer, J.: Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resour. Res., 52(3), 1847–1865, doi:10.1002/2015WR017635, 2016.

Zhang, Y., Chiew, F. H. S., Li, M. and Post, D.: Predicting Runoff Signatures Using Regression and Hydrological Modeling Approaches, Water Resour. Res., 54(10), 7859–7878, doi:10.1029/2018WR023325, 2018.