Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-737-AC2, 2018

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Comparing hydrological modelling, linear and multilevel regression approaches for predicting baseflow index for 596 catchments across Australia" *by* Junlong Zhang et al.

Junlong Zhang et al.

junlongzhangcq@hotmail.com

Received and published: 14 May 2018

Anonymous Referee #2 This study compares two hydrologic models to two regression approaches for estimating base flow index (BFI) index at a large number of catchments across different climatic regions in Australia. The benchmark BFI is the mean of four recession flow approaches to estimating BFI. The study focuses on the important topic of estimating BFI at ungaged sites and provides a contribution to the literature in introducing a new approach to BFI estimationâËŸAËĞTmulti-level regression approaches, which are common in other fields but less often applied in hydrology.

General comments: The manuscript is has appealing and informative figures and includes a thoughtful cross validation of the approaches. My first main comment is that the clarity and presentation could be improved, both in terms of the writing and language of the paper, as well as the comparisons between models. More explanation for which comparisons are made and why, as well as some general editing would help make the study more clear (see specific comments for some suggestions). One important language clarification is the use of "cross-level interactions". It does not seem that interactions are used in the regression models based on the equations and I suspect this is a terminology issue. Perhaps "correlation" is a more appropriate term?

Response: Thanks very much for the favourable and constructive comments on our present study. We followed your suggestions to add more explanation and did thorough edits on the manuscript as you can see from the responses to comments for both reviewers. One aim of this study is to introduce a new approach to improve the prediction of baseflow index across a large scale. To get this message cross, we compare the multilevel approach to two other approaches including classic linear regression and hydrological approaches for predicting BFI. It is clear that the multilevel approach outperforms the other two. The cross-level interactions used here is to indicate the relationships and correlations between different climate zones. This terminology is used to describe the complexity of those relationships. To clarify the issue raised, the terminology is further clarified as "...the complexity of correlations between different backgrounds" in line 492. We conducted many revisions to improve manuscript quality, including thorough proofreading and other concerns raised.

Clarification of the motivation between the approach for comparing the hydrologic and regression models would be helpful. Based on figure 5, it appears that the hydrologic models selected were not at all useful for predicting BFI. Are these models frequently used for this estimation of BFI? If not, might there be more appropriate models to provide a comparison to the regression approaches? Also, it would be good to introduce BFI duration curves before showing them as a way to compare approaches in Figure

4. I also wondered why the multilevel regression wasn't included in Figure 4. It would be helpful to the reader also to have figure 5 to match the structure of Figures 6 and 7.

Response: we agree that more clarification is required. The reason to compare with the classic linear regression is that they are widely used approaches for hydrological signature predictions and it can be easily applied. In terms of hydrological modelling, it is a major tool estimating runoff time series. However, it is not clear how good is hydrological modelling for predicting BFI. Figure 5 clearly demonstrates that hydrological modelling perform very poorly in terms of BFI prediction. What we want to alert hydrological community that it is not reliable to use traditional calibrated hydrological models for baseflow separation. Adding the BFI duration curves from regression approaches into Figure 4 is a very good suggestion. In the following figure, we include duration curves from all three approaches (two hydrological models, linear regression, and multi-level regression).

Figure 4. BFI duration curves from benchmark, regression and hydrological modelling approaches. In terms of introduction of BFI duration curves, we would introduce them before we show them as a way to compare the approaches. Good suggestion!

Next, in the comparison of traditional to multi-level regression models, it would be helpful to directly state more explicitly that the approach from equation (1) isn't used in the comparison in the study. (Though it might be worth doing that just to see it as a comparison, especially given the similarity of the regional coefficients shown in figure 8?) It was surprising that, for the multi-level model, NSE and bias was almost identical in the leave-one-out cross validation as when the model was fit for the whole set of catchments. Can you provide some possible explanations for this, especially given the large degradation for the traditional model?

Response: We added the clarification of this explanation, the details "The approach from equation (1) isn't used in the comparison in the study owning to its limitation." are shown in lines 241-242. As it was mentioned in Method section, the multilevel

approach considers the relationships between the catchment attributes and around it, this consideration is closer to the catchment hydrological processes, and therefore its degradation from model calibration to model validation is much smaller than that for the classic linear regression approach. However, we didn't clear show the degree of this merit. To clear demonstrate the benefit of using the Multi-level regression, we added bar plots to show the degradation for the traditional to multi-level regression models. The description is also provided as follows: New Figure 2. The difference of NSE and Bias between calibration and validation for linear regression approach ((a) and (c)) and multilevel regression approach ((b) and (d)). New Figure 2 summarises the degradation from calibration to validation for the two regression approaches. First, multilevel regression is very stable and there are no noticeable degradation across all climate regimes. Second, there exists strong degradation for linear regression, and the degradation is much stronger for the whole dataset than for the sub datasets (i.e. data from different climate regimes). Third the degradation for the linear regression is stronger in arid and tropics climate regimes than in Equiseasonal and Winter rainfall climate regimes.

My second main comment is regarding the development of the regression questions. How were these models fit? It was interesting, as noted in the paper, that a number of the parameters appear very close to zero (Figure 8). Did you check to see if the variables included in the regression were statistically significant? In the discussion it is noted that "P and ETP have the most significant effects on BFI" but was this was not reported in the results.

Response: We add more description on model development in method section. The text now includes "Herein, the "lmer" function in R package of "arm" (https://cran.r-project.org/web/packages/arm/index.html) was used to perform the multilevel regression." in lines 311-312. We performed the normalisation (Eq (10)) before building the model, which make sure that data meet the assumptions of regression.

It would be a good idea to check that regression assumptions (normality, constant

variance and independence of model residuals) are met for both models and report how well these assumptions are met. Also checking for influential observations could be good. It seems as though a few outliers in the traditional regression might have a lot of influence and be related to the reduced performance for these models. Finally, it might be helpful to add a table comparing NSE and bias for the regression approaches across regions and for full sample vs leave-one-out cross validation (to make it easier than trying to compare across the figures). Also, calculation of % bias relative to the BFI index could provide useful context for what the bias reflects relative to the value of BFI itself.

Response: Yes, these assumptions were checked and both models met the assumptions. To minimise the influence from outliers, we did normalisation first, and then applied the regressions. We add a difference NSE and Bias figure (Figure 5, as showed previously) to compare the performance of using full dataset and using sub datasets. It is clear that the linear regression shows much poorer results when using the full dataset than using the sub datasets. However, the multilevel regression overcomes this issue.

Specific comments: Title: Possible to make it more informative?

Response: Thanks for the suggestion. We thought the current one should be a best we being informative

Abstract: Line 37-39 “Our study indicates the multilevel regression approach should be used for predicting large-scale baseflow index such as Australian continent where sufficient catchment predictors are available.” The word “should” makes it a very strong statement. It might be best to tone it down some, such as: “could improve upon predictions of large-scale baseflow index compared to the other methods studied”

Response: We changed this statement to “Our study indicates the multilevel regression approach could improve upon predictions of large-scale baseflow index such as Australian continent where sufficient catchment predictors are available” in lines 37-39.

Line 43-47 These are very helpful. Maybe add in more of the information to the abstract?

Response: We added that information in abstract, the details were found below: “The two calibrated and regionalised hydrological models perform similarly poorly in predicting BFI with a Nash-Sutcliffe Efficiency (NSE) of -8.44$\sim$-2.58 and an absolute percent bias (Bias) of 81$\sim$146 (overestimates baseflow); the classic linear regression is intermediate with the NSE of 0.57 and bias of 25; and the introducing multilevel regression approach is best with the NSE of 0.75 and bias of 19.” in lines 33-37.

Line 51: “important indicator of catchment hydrogeological characteristic” a bit unclear. Removing “important indicator of” might make it more clear: “catchment hydrogeological characteristic”.

Response: Done.

Line 71: where do ensembles come in? “ensemble estimates from the non-tracer methods at gauged catchments”

Response: We add more details. The text now says “The details about ensemble estimates are descripted in section 3.1 Baseflow separation algorithm” in lines 72-73.

Line 81: probably “literature” rather than “literatures”

Response: Done as suggested.

Line 84: A little confused about the ensemble piece and the non-tracer designation

Response: To clarify the confusing, the text now says “This study use ensemble mean of BFI as a benchmark to evaluate the BFI predictions from two hydrological models (SIMHYD and Xinanjiang models) ” in lines 84-86.

Line 111 “studies” rather than “literatures”

Response: Done

Line 118: remove the word “the”

Response: Done.

Line 131 Grammar: “There are 596 catchments selected across Australia for assessing the three methods”

Response: It is changed to “We selected 596 catchments across Australia to assess three methods (hydrological modelling, linear regression and multilevel regression) for their skills for predicting BFI” in lines 133-135.

Line 191: More detail here about how “Figure 2 demonstrates how the recession constant is estimated” would be helpful.

Response: We add details for this recession constant. In lines 198-200, the text say “The method plots dQ/dt against Q with the 5% lower envelope, which represents the slowest recession rate (Figure 3).”.

Line 192: perhaps “empirical BFI” or “benchmark BFI” would be more appropriate as it is not an observed quantity

Response: Thanks, we will change it to “benchmark BFI”.

Line 194-195: “either to evaluate the two hydrological models for BFI prediction, or to build” I think you mean to both evaluate and to build?

Response: We rewrote this sentence to “The observed BFI was used to evaluate the two hydrological models in BFI prediction, but to build the linear and multilevel regression approaches together with the catchment attributes.” in lines 203-206.

Line 231: missing the word “of”

Response: Done.

Line 234: “slope”, “X represents the”

Response: Done.

Line 239 confusing “one level reflects hydrological background should be introduced”

Response: This statement has revised to “hydrological backgrounds should be introduced” in lines 258-259.

Line 252 define n; do you mean i here? “j is catchment in each climate zone”

Response: Done.

Line 294: more details about how leave-one-out cross validation works would be helpful (ie, one of the catchments was omitted from fitting of the regressions and then those models used to predict at that catchment, as if the catchment was ungaged”

Response: We add more introduction on this validation approach. In the leave-one-out cross-validation, (1) each catchment is left out in turn, and is purposely treated as “ungauged”; (2) the predictive relationship is then developed using data from the remaining catchments; and (3) the relationship is used to predict the baseflow index for the catchment not used in developing the relationship.

Line 315: might be good to add what a NSE of 0 means or that below 0 indicates that the mean value would provide a better prediction

Response: Good suggestion. Yes, we add that a NSE of 0 and below 9 indicates very poor correspondence.

Line 380 Grammar “large biases to use hydrological models”

Response: We revised this statement to “Our results suggest there are large biases using the hydrological models to predict BFI.” in lines 404-405.

Line 387 Fragment. “This suggests that better estimate streamflow.”

Response: we reshaped and merged this statement to its last sentence, the details were “This suggests that those two hydrological models can better estimate streamflow, rather than its components, such as baseflow” in lines 411-413.

Line 402 Grammar “As studies by Santhi et al. (2008) and PenÌCËĞ a-Arancibia et al. (2010), they”

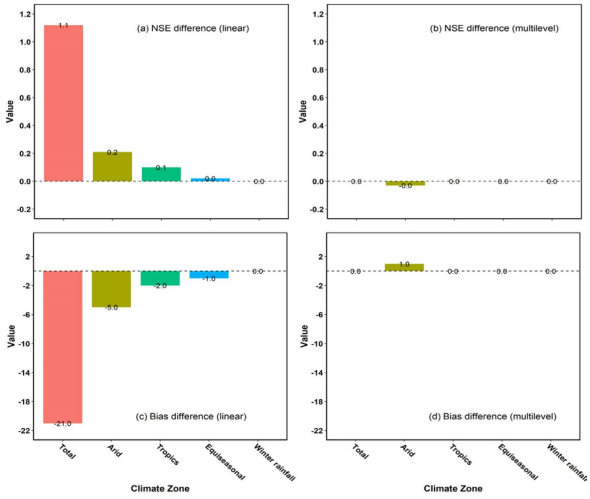
Response: we revised this statement to “As studies by Santhi et al. (2008) and Peña-Arancibia et al. (2010), they have shown that climate attributes can be used to best predictors for recession constant” in lines 438-440.

Line 436 What does this mean? The models don’t have interaction terms so this is somewhat confusing. “When interactions crossing level have been implemented, adding those two factors can greatly improve performance of multilevel regression approach.”

Response: In multilevel regression approach, the interactions between different climate zones have been considered, the details were shown from Eq (3) to Eq (8). Specifically, the model is varying-intercept α and varying-slope β. To reduce this confusion, we revised this statement to “When varying-intercept α and varying-slope β have been implemented, adding those two factors can greatly improve performance of multilevel regression approach.” in lines 472-473.

Figure 9 appears to be missing Response: Done as suggested.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-737, 2017.



New Figure 2. The difference of NSE and Bias between calibration and validation for linear regression approach ((a) and (c)) and multilevel regression approach ((b) and (d)).

Fig. 1.

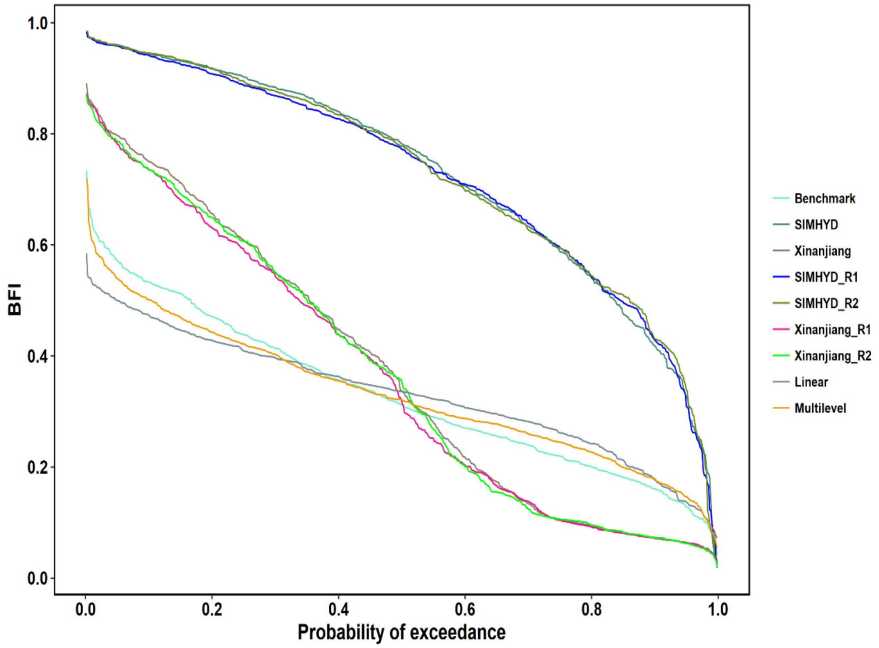


Figure 4. BFI duration curves from benchmark, regression and hydrological modelling approaches.

Fig. 2.