Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-737-AC1, 2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

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 #1 The manuscript by Zhang et al. proposes to use multilevel regression for large-scale baseflow index (BFI) prediction. Using 596 catchments in Australia, the BFI prediction accuracy of multilevel regression is evaluated against benchmark BFI calculated by baseflow separation methods. It was found that multilevel regression outperforms classic linear regression, with both models using same explanatory variables. In addition, both regression models outperform two calibrated hydrologic models. The results suggest that climate variability should be considered

in order to better understand the effects of explanatory variables on BFI. The topic is pertinent for the scope of HESS. As far as I am aware of, the application of multilevel regression for BFI is novel. However, in my opinion, the present manuscript is not completely convincing. The assumptions involved in the experiments are not discussed, and the results do not sufficiently support the conclusions. In addition, the manuscript is hampered by a lack of clearness in the description of methods and experiments. So my suggestion is to reject the paper. My justifications are listed below.

Response: Thanks for the positive but critical comments on this study. We do appreciate that this reviewer acknowledges the novelty of this study in terms of using multilevel regression for BFI estimation. To address the issue that the reviewer thought the reviewed version was not convincing, we spent last several months to improve the quality of manuscript, including improvement of assumption in Discussions, resharpening the results and conclusions, adding some more details in methods and experiments, and thorough English proofreading (by a professional English edit provider).

1. The implication of using the BFI ensemble mean as "observed" BFI is not discussed. The average value of BFI calculated by four methods ("ensemble mean") is used as benchmark for evaluating the hydrologic models and regression methods. I agree that this seems to be the best choice given that no true observed BFI is available. However, a plot or statistics should be added to show the agreement among the BFI values given by the four methods. Table 3 seems to suggest significant discrepancy among the four products (My interpretation about Table 3 might be wrong, as no explanation is given in text for this table). Note that the discrepancy among the four products could be an indicator of the uncertainty associated with the ensemble mean, and should be taken into account when performing regression. For example, how does the regression residual compare with the uncertainty of ensemble mean?

Response: This is a very good suggestion regarding the uncertainty from the four methods It is indeed that there are large difference between the four methods, as shown in the following new Figure 1 (the number will be renamed in next version).

In order to eliminate uncertainties from the four algorithms, it is necessary to use the ensemble mean of the four method estimates as the benchmark (denoted as 'the observed BFI') to evaluate multilevel regression and linear regression approaches. We also add one scatter plot to show the BFI derived from four methods.

New Figure 1. Baseflow index derived from four non-tracer baseflow separation methods and ensemble mean value. LH, UKIH, CM, ECK and Mean are the baseflow index estimated from Lyne-Hollick (Lyne and Hollick, 1979), UKIH (Gustard et al., 1992), Chapman-Maxwell (Chapman and Maxwell, 1996), Eckhardt (Eckhardt, 2005) methods. To show more information on the statistics, we add Table 3 (Thanks for picking up that the explanation for Table 3 has not been provided) to show regression approaches (linear regression and multilevel regression) performances when compared to each BFI method. We also use standard deviation to present uncertainty of ensemble mean to evaluate the regression approaches (linear and multilevel).

2. The calibration of the two hydrologic models A major conclusion is that hydrological models overestimate baseflow in Australian catchments. In this study, the two models are calibrated using streamflow. What if the models are calibrated using the BFI as objective function? In that case, would they still be outperformed by the regression models? I missed from my reading why the SIMHYD and Xinanjiang models are selected as representatives of various hydrologic models. It is mentioned in the manuscript and also my understanding that the Xinanjiang model is widely used for humid regions, and it might not be suitable for all Australia catchments. It would be helpful to include results of the calibration goodness-of-fit.

Response: Thanks for the suggestion. However, the main point is to evaluate how hydrological models that is calibrated against daily streamflow time series data are good for BFI estimates. We clearly demonstrate that traditional calibrated hydrological models overestimate baseflow for Australian catchments. It is not appropriate to use a single BFI value to calibrate hydrological model The two hydrological SIMHYD and Xinanjiang models are widely and successfully used for runoff time series pre-

diction in Australian catchments (Zhang and Chiew, 2009; Zhang et al., 2009; Li et al., 2009; Li et al., 2012; Zhou et al., 2013; Zhou et al., 2015, etc). It is noted that we used an updated version of SIMHYD and Xinanjiang that use remote LAI data together with climate forcing to improve overall performance for predicting runoff time series in Australia catchments. Therefore, evaluation of these kinds of models provide a great opportunity to check reliability of hydrological modelling for estimating BFI. We will include calibration results in terms of NSE and bias in the next version. References: Li H, Zhang YQ, Vaze J, Wang B, (2012). Separating effects of vegetation change and climate variability using hydrological modelling and sensitivity-based approaches. Journal of Hydrology, 420-421, 403-418. Li HX, Zhang YQ, Chiew F, Xu S, (2009). Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index, Journal of Hydrology, 370, 155-162, DOI: 10.1016/j.jhydrol.2009.03.003. Zhang YQ, Chiew FHS, (2009). Relative merits of three different methods for runoff predictions in ungauged catchments, Water Resources Research, DOI: 10.1029/2008WR007504. Zhang YQ, Chiew FHS, Zhang L, Li H, (2009). Use of remotely sensed actual evapotranspiration to improve rainfall-runoff modelling in southeast Australia, Journal of Hydrometeorology, DOI: 10.1175/2009JHM1061.1. Zhou YC, Zhang YQ, Vaze J, Lane P, Xu SG, (2015). Impact of bushfire and climate variability on streamflow from forested catchments in southeast Australia. Hydrological Sciences Journal (DOI: 10.1080/02626667.2014.961923). Zhou YC, Zhang YQ, Vaze J, Lane P, Xu S, (2013). Improving runoff estimates using remote sensing vegetation data for bushfire impacted catchments, Agricultural Forest and Meteorology, 182-183,332-341.

3.1. The description of the multilevel regression method needs to be improved. This is related to comment # 5. I also made suggestions in specific comments in the hope to improve the clarity of this section. What is the physical meaning in the BFI context of the correlation coefficient (rho) between alpha and beta? I missed from my reading a discussion about rho in the results. The definition in Eq (6) is contradictory to line 284. Line 284 says rho is a between-group correlation, but it is defined in Eq. (6) as

the correlation of intercept and beta within a same group.

Response: We added more information about the multilevel regression approach in method section. The details were found below: “(i.e., geology, climate, and vegetation), the flow routing would be controlled by several effects, the dominate factors cause the varying hydrological processes of catchments at large scales,” in lines 253-255. Physical meanings of alpha and beta are intercept and slope that reflects hydrological processes in a catchment influenced by the hydrological backgrounds (i.e., routing pathway). This productivity presents the relationship among a target catchment and surrounding ones. Herein, the rho is the correlation coefficient between different groups. In Eq (6), the represents the correlation coefficient between group levels and is cross groups.

3.2. Please provide more information about how intercepts and slopes/coefficients are estimated, the software package you used or scripts developed, along with other relevant algorithm configurations/settings so that interested readers can follow up your work. The equations in section 3.3 are all developed for the univariate regression case. Since multiple explanatory variables are used, I suggest including a matrix-vector equation in which the slope is a vector.

Response: we add more information about the multilevel regression method in method section. The details can be found below: “Generally, this approach is a two-stage regression, estimating the effects for each in stage one (within-group), and fitting group effects on group-level predictors in stage two (Gelman and Hill, 2006).” in lines 287-289. We also add the specifics on how to perform the multilevel approach in this study. The details were found below: “Herein, the “lmer” function in R package of “arm” was used to perform the multilevel regression.” in lines 311-312. In Equation (3), X represents the multiple variables (catchment attributes), the specifics of intercept and slope were displayed in the next Equation (4) to (8). In this matrix-vector, the slope β is included in those Equations.

4. Discussions centering around estimated coefficients/slopes are not convincing. While I appreciate the discussion section which connects findings from this study to literature, I do not think the reasoning there is convincing. For the multilevel regression, the estimated intercept and coefficients show only small differences among climate groups. Given the wide error bar (Figure 8), these differences are not statistically significant. The similarity of estimated intercept and slopes suggests that the effects of explanatory variables (including climatic controlling factors) are similar across climate groups. It is risky to conclude that the effects of P, ET, and F vary among climate groups. In terms of understanding the controlling factors in different climate groups, it seems that multilevel regression provides limited advantage over the classic linear regression.

Response: Thanks for the critical comment. We acknowledge that the uncertainty related to the coefficients/slopes. Therefore, we will soften our tone in discussion for the effects of P, ET and F among climate groups. It is indeed it is challenging to use the parameters obtained from the regressions to explain the physical control of baseflow generation.

5. Discussions centering around cross-interaction are unclear to me. Cross-interaction seems to be an important concept and a major strength of multilevel regression. The term is brought up in several places but never clearly defined for the context of BFI prediction. For example, I do not understand why interactions crossing various group levels are primary drivers to influence baseflow processes (line 255). What is the interaction refer to? Interaction between which and which?

Response: We add the term of cross-interaction in next version to illustrate the meaning of this term. The details were found below: “One limitation of the linear regression approach is that it uses constant parameters to predict BFI, and cannot handle cross-interactions at different spatial scales (correlation coefficient between groups) (Qian et al., 2010)” in lines 102-105. The baseflow processes are influenced by the complexity of catchment attributes. However, the catchment is embedded in its geographical

context (i.e., climate). There are many characteristics impact the baseflow processes. Moreover, there are internal connections between catchment attributes which determine the baseflow processes. Considering the influences derived from various catchment attributes and the influences in different backgrounds (i.e., climate zones used this study) can improve prediction of the baseflow. These influences from between-catchment attributes are denoted to cross-interaction. To this end, we introduce multilevel regression to overcome this issue.

6. I like it that the manuscript is well-structured. The presentation quality does not meet the standard of HESS. As mentioned in my comments above, the description of methods and experiments is not complete or precise for reproduction. The manuscript would benefit from a thorough proofreading and English language editing. There are repeated sentences (line 405-410), incorrect reference to figures (e.g., line 448 Figure 9) and equations (line 275), typos, and grammar mistakes.

Response: We appreciate for the critical comments. To make our manuscript meeting HESS quality, we not only did technical corrections/clarification, added more into to discussion, but also thorough English language editing through professional English edit service. We are confident that the next version will meet HESS standard.

Specific comments Line 225, do you mean that in leave-one-out cross validation, the parameters of the models are filled in by either taking the calibrated parameter values in the closest catchment or a combination of parameter values from several basins that are both spatially and hydrologically close?

Response: In leave-one-out cross validation stage, two approaches including the spatial proximity and integrated similarity were used for prediction. For spatial proximity, the parameters calibrated at geographically closest catchment are used for the 'targeted ungauged catchment'; for integrated similarity approach, the parameters calibrated at physically most similar catchment are used for the 'targeted ungauged catchment'. The details are given the in lines 236-240, and the references cited are give in line 237.

Eqn. (2). Given line 247, "we further assumed that the effects of those predictors on BFI vary with climate zones. . .", alpha and beta should be replaced with alpha_j and beta_j, respectively. Then what is the difference between (2) and (3)?

Response: Eqn. (2) is used to predict the BFI in individual climate zone, and one parameter set is used, but Eqn. (3) considers the interactions between each climate zones. In Eqn. (2), the total data is divided into four datasets based on climate zones, and the linear regression model is then built for each zone separately. The α and β are determined by the catchment attributes X in each climate zone, and represent the individual climate zone. For Eqn. (3), the α and β in one climate zone are not only determined by the catchment attributes in one climate zone but also the connections with other climate zones around it. The relationship within- and -between groups is all considered.

Line 356. Figure 7 shows the validation results, which support the conclusion that multilevel regression outperforms classic linear regression. However, this is not the case for Figure 6, which is based on the data used for regression. Multilevel regression uses more parameters, so it is anticipated that it will fit the data better.

Response: What you are said is true. The main point is for prediction (i.e. the results shown in Figure 7). Another important point the initial version is not clear the degradation from calibration to validation for both linear regression and multilevel regression approaches (as shown in new Figure 2 (this will be renumbered in the next version)).

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.

Line 360. What are possible reasons that the two regression models perform similarly for the winter rainfall climate group?

Response: The biggest difference between classic linear regression and multilevel regression is whether the cross-level interactions are considered. In other words, multilevel regression considers the complexity of hydrologic cycles between catchments at different climates. However, in winter rainfall climate zone, the relationship of hydrological processes between catchments is not so much close (i.e. without the cross-level interactions). Therefore, the performance of multilevel regression is very similar to that of linear regression approach. We add more discussion into Discussion section. The text now includes "When the cross-level interactions are not so strong, the benefit of using multilevel regression approach will be limited. This is the case for winter rainfall climate zone" in lines 430-433.

Line 383 "model structure is more important than parameterisation" The reasoning here is unclear to me. Did the three parameterisation schemes lead to different parameters?

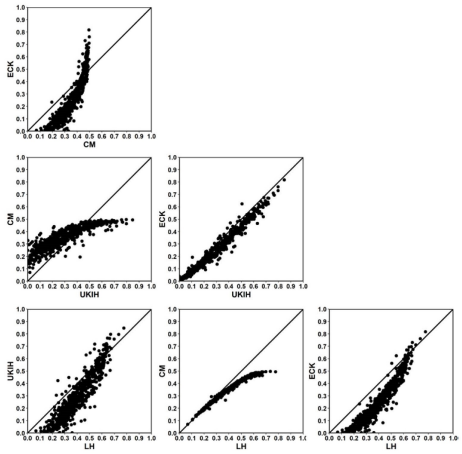
Response: Thanks for the comment. This should be "model structure is more important than parameter regionalisation". The text now says "It seems that model structure has a stronger effect on the outcomes of prediction than parameter regionalisation since the three regionalisation schemes show similar results" in lines 406-408.

Table 3 is not referred to in text. How is the statistics calculated? For example, is the value of 114 calculated by fitting the regression model to UKIH results, or using the regression model fitted to the ensemble mean?

Response: Explanation of Table 3 has been added into text. The statistic is calculated between regression and each of benchmark separation approaches (LH, UKIH, CM,

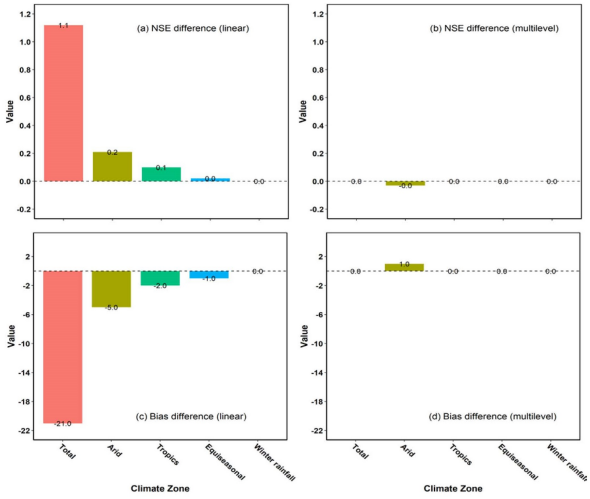
ECK, and ensemble).

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



New Figure 1. Baseflow index derived from four non-tracer baseflow separation methods and ensemble mean value. LH, UKIH, CM, ECK and Mean are the baseflow index estimated from Lyne-Hollick (Lyne and Hollick, 1979), UKIH (Gustard et al., 1992), Chapman-Maxwell (Chapman and Maxwell, 1996), Eckhardt (Eckhardt, 2005) methods

Fig. 1.



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. 2.