A Bayesian Joint Probability Post-Processor for Reducing Errors and Quantifying Uncertainty in Monthly Streamflow Predictions

Prafulla Pokhrel, David E. Robertson and Q.J. Wang

We would like to thank the reviewer for the assessment on the paper and constructive suggestions/comments. We have revised the manuscript, incorporating most of the comments made by the reviewer. To help in the assessment of our revision, reviewer's comments and our specific response to those comments are included.

Response to comments made by Reviewer # 2:

General comments: The paper is well written, and has good structure, figures are informative and have very good quality. The topic is interesting and has practical significance. Overall, I recommend the paper for publication, although I would suggest authors consider several minor additions.

Response: We thank the reviewer for the positive assessment that the paper is well written and the recommendation for publication with several minor additions.

Comment 1: Firstly, the methods presented in the paper are complex and perhaps less accessible to a wide body of readers who may not be familiar with bayesian methods. Perhaps it would be good to add a short paragraph describing basics of the bayesian inference in not-so-technical language.

Response: We have now provided a more detailed information as well as a simplified equation on the Bayesian inference scheme. We believe that it will provide some introductory material to readers not familiar with the topic.

The section 3.2 with added description on Bayesian inference now reads as –
"The BJP modelling approach assumes that a set of predictands y(2), and their predictors y(1) follow a joint multivariate normal distribution in a transformed space. Normalization of the variables is achieved by using the log-sinh transformation (Wang et al 2012). The log-sinh transformation replaces the previously used Yeo-Johnson transformation (Yeo and Johnson 2000, Wang et al 2009, Wang and Robertson 2011). Although both have data normalization and variance stabilization properties, the log-sinh has been shown to outperform the Box-Cox based Yeo-Johnson transformation when applied to catchments with highly skewed data (Wang et al 2011). The posterior distribution of the parameters $p(\theta|Y_{OBS})$, including mean, variance and transformation parameters for each variable and a correlation matrix for the multivariate normal distribution, is estimated using a Bayesian inference (equation 1).

$$p(\theta|Y_{OBS}) \propto p(\theta) \cdot p(Y_{OBS}|\theta) \quad (1)$$

where $\mathrm{Y_{OBS}}$ contains the historical data of both predictor y(1) and predictand y(2) variables used for model inference, and θ is the parameter vector. $p(\theta)$ is the prior distribution of the parameters of the multivariate normal distribution, representing any information available before the use of historical data YOBS. $p(Y_{OBS}|\theta)$ is the likelihood function defining the probability of observing the historical data given the model and the parameters. The posterior parameter distribution is approximated by 1000 sets of parameters sampled using a MCMC method.

The posterior predictive density for a new event is given by equation (2). Here, we present a brief introduction of the BJP modelling approach, Details of the method for the numerical evaluation of Eq. (1), (2) and the prior distribution of the parameters can be found in Wang et al. (2009) and Wang and Robertson (2011).

$$f\big(y(2)\big|y(1)\big) = p(y(2)|y(1); Y_{OBS}) = \int p(y(2)|y(1), \theta) \cdot p(\theta|Y_{OBS}) \cdot d\theta \qquad (2)”$$

Comment 2: Secondly, it would be good to demonstrate not just that the method works for the purpose intended, but also how its performance compares to simpler methods of error and bias correction, such as linear regression or quantile remapping.

Response: A linear regression model generally assumes that the residual errors have constant variance. This causes problem when one wants to quantify prediction uncertainty, as larger events tend to have much larger variances in prediction uncertainty than smaller events. For this reason, we can rule out the use of linear regression as directly applied to the predictors and predictands.

Having said above, the BJP method is essentially the same as the use of linear regression as applied to transformed predictors and predictands. There are only some subtle differences in the way parameters are estimated (but BJP can handle missing data and zero flow problem). For this reason, we have not made a comparison with the use of linear regression as applied to transformed predictors and predictands.

We do not feel quantile mapping is any simpler than the use of parametric transformations to normalise data and stablise variance. An advantage of the BJP method is that it allows for uncertainty in the transformation parameters.

Comment 3: Thirdly, authors stress the method’s skill in reducing simulation bias. Obviously, the post-processor reduces biases/errors when they are large. However, equally teaching are cases when the method is not skilful. Perhaps a plot of the gain in RMSEP expressed as a percentage of original RMSEP would be better in expressing postprocessor skill. Also, it would be interesting to see how performance of the postprocessor varies due to other factors, such as error variance.

Response: A problem with including RMSEP as percentage of original RMSEP is that it tends to unnecessarily highlight errors that are of very small magnitudes. However, we do appreciate reviewer’s comment that cases when the post-processor is not skilful are also important.

We therefore have added an extra figure in the manuscript that shows post-processor’s performance across small and large error values and have included a discussion that clearly mentions conditions where the post-processor will not be effective.

In addition, the added plots also demonstrate how the performance of post-processor varies against RMSEP error values (This conveys similar message as the plotting against the error variance, except that it is measured in probability space).

We have added following figure and discussion in section 5 of the manuscript –

“In general, the reduction of errors by the post-processor does not necessarily depend upon the magnitude of errors and occur for small as well as large errors (figure 9). However, reductions of errors are not possible in all situations, as illustrated by the points lying in the 1:1 lines. As with all the statistical methods, the effectiveness of the BJP post-processor depends upon the correlation between predictand and predictors, stationarity in relationship (between predictors and predictands) and persistence in the error structure that allow for prediction updating. The post-processor is not effective in situations where none of these occur, this seem to be the case for many points lying in the 1:1 lines, most prominent among them being the high RMSEP error values (>0.25) corresponding to predictions in Nillahcootie (see figures 9 and 3). However, more importantly the BJP post-processor is able to preserve skill (not degrade performance) of WAPABA prediction even when error correction is not possible."

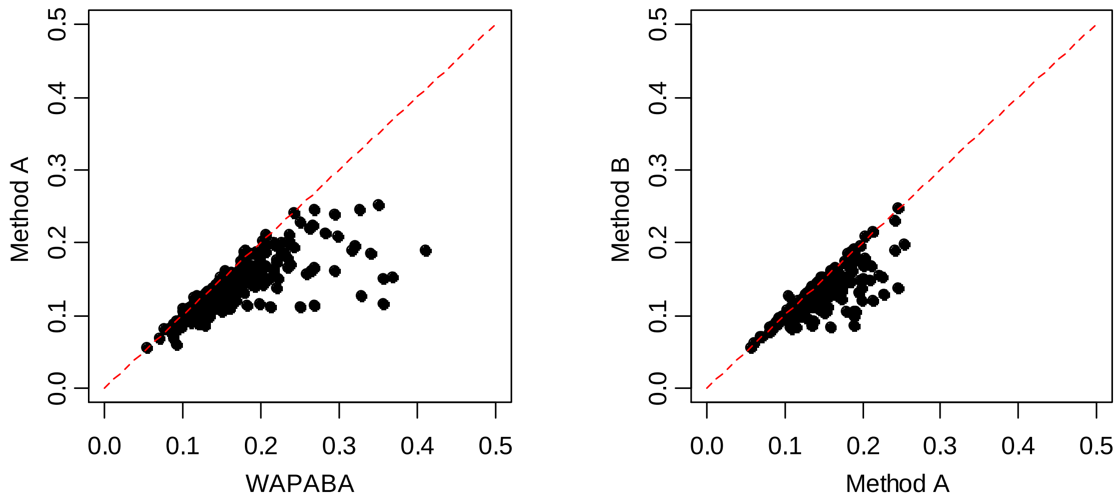


Figure 9: Scatter plots of RMSEP values; (left) WAPABA predictions vs. Method A, (right) Method A vs. Method B.

Comment 4: Lastly, the abstract revolves around the difference between post-processing on the daily and monthly basis, suggesting that the paper is about contrasting these two, which it isn't. The abstract should be rephrased to reflect the contents of the paper appropriately.

Response: We have changed the abstract, which now reads –

"Hydrologic model predictions are often biased and subject to heteroscedastic errors originating from various sources including data, model structure and parameter calibration. Statistical post-processors are applied to reduce such errors and quantify uncertainty in the predictions. In this study, we investigate a statistical post-processor based on Bayesian Joint Probability (BJP) modelling approach to reduce errors and quantify uncertainty in streamflow predictions generated from a monthly water balance model. The BJP post-processor reduces errors through elimination of systematic bias and through transient errors updating. It uses a parametric transformation to normalise data and stabilise variance and allows for parameter uncertainty in the post-processor. We apply the BJP post-processor to 18 catchments located in eastern Australia and demonstrate its effectiveness in reducing prediction errors and quantifying prediction uncertainty."

Specific comments: -
Comment 5: p.11200 line3 ".. further reduce: : :" further compared to what?

Response: No longer relevant

Comment 6: - p. 11202 line 12 "..reliable in uncertainty distribution.." what does that mean, exactly?

Response: By reliable we mean that the post-processed probability distributions are statistically consistent with the observed frequency. The definition of reliability is included in section 4.2.1. In addition we have also re-phrased the sentence to make it more understandable. The sentence now reads –

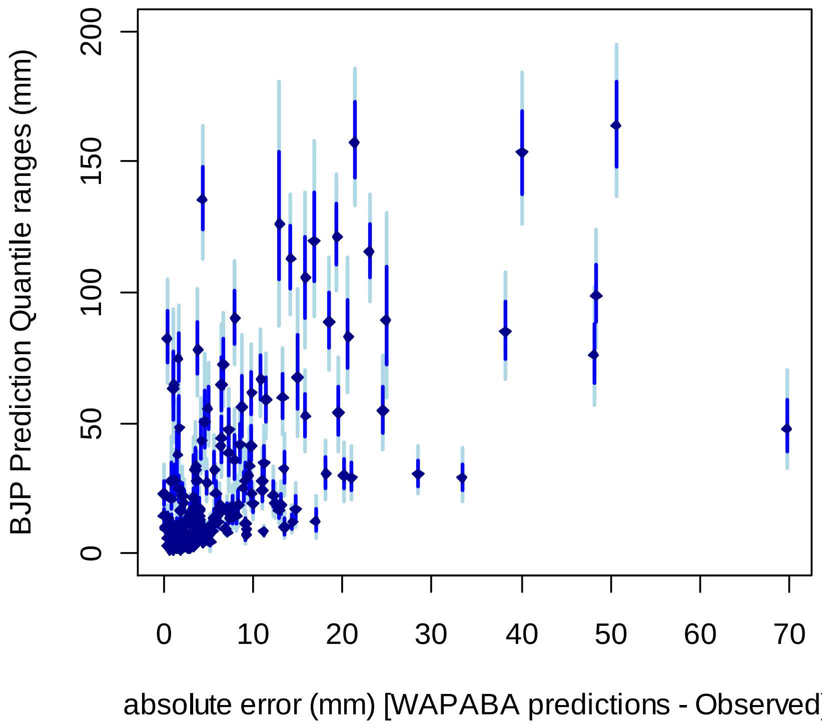
"While daily predictions from daily models may be post-processed at the daily time scale and then aggregated to monthly, there is no guarantee that the monthly volumes so produced have reliable uncertainty distributions and least errors achievable."

Comment 7: - p. 11203, line 7 "..grassland to semi-arid type of climate". Maybe other, more informative term than grassland could be used.

Response: The word 'grassland' omitted from the manuscript.

Comment 8: - p. 11219 fig.2 perhaps it would be interesting to plot too the BJP quantile ranges against the simulation error?

Response: The figure below shows BJP quantile ranges vs. the absolute error *[abs(WAPABS simulation –BJP prediction]*. The figure does not add anything new to the manuscript so we do not include in the revised manuscript.



Comment 9: - p 11212 line 20. The authors write that in case of the use of post-processor for predictions based on rainfall forecast, the post-processor would increase uncertainty spread to account for uncertainty in forecast rainfall. This is difficult to conceptualise. Perhaps a procedure for such application could be outlined in a couple of sentences?

Response: The manuscript now includes a description of the procedure where the post-processor could be used on rainfall forecasts. We have modified and added a passage, which reads –

"However, the post-processor is equally applicable in the real world applications using rainfall forecast ensembles. In such case the hydrologic model could be forced with rainfall forecast ensembles, to create streamflow forecast ensembles. The streamflow forecast ensembles could then be post-processed to reduce errors and further quantify hydrologic uncertainty in the streamflow forecast (Seo et al. 2006). An alternative approach would be to force hydrologic model using the mean of rainfall forecast ensembles. Then train the post-processor on the deterministic streamflow forecast produced by the hydrologic model and finally post-process the deterministic forecast to reduce error and quantify the total uncertainty (Pokhrel et al 2012). "