Hydrol. Earth Syst. Sci. Discuss., 9, 11199–11225, 2012
www.hydrol-earth-syst-sci-discuss.net/9/11199/2012/
doi:10.5194/hessd-9-11199-2012

Hydrology and Earth System Sciences Discussions

A Bayesian joint probability post-processor for reducing errors and quantifying uncertainty in monthly streamflow predictions

P. Pokhrel, D. E. Robertson, and Q. J. Wang

CSIRO Land and Water, Graham Road, Highett, Victoria, Australia

Received: 7 September 2012 – Accepted: 16 September 2012 – Published: 2 October 2012

Correspondence to: P. Pokhrel (prafullapok@gmail.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Hydrological post-processors refer here to statistical models that are applied to hydrological model predictions to further reduce prediction errors and to quantify remaining uncertainty. For streamflow predictions, post-processors are generally applied to daily or sub-daily time scales. For many applications such as seasonal streamflow forecasting and water resources assessment, monthly volumes of streamflows are of primary interest. While it is possible to aggregate post-processed daily or sub-daily predictions to monthly time scales, the monthly volumes so produced may not have the least errors achievable and may not be reliable in uncertainty distributions. Post-processing directly at the monthly time scale is likely to be more effective. In this study, we investigate the use of a Bayesian joint probability modelling approach to directly post-process model predictions of monthly streamflow volumes. 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.

1 Introduction

Streamflow predictions from a hydrological model can be used for wide range of applications including flood forecasting at short time scales to long term assessments of water resources. Model predictions are subject to errors originating from various sources including input data, calibration data, model structure and parameters. The model is usually calibrated prior to its application to compensate for these errors, thus reducing uncertainty in the predictions. However, a model being a simplified representation of a system will always contain uncertainty in its predictions (Gupta et al., 2005). Post-processors are statistical models that are applied to model predictions to further reduce errors and to quantify uncertainty in the streamflow predictions (Seo et al., 2006).

HESSD

9, 11199–11225, 2012

Post-processors can reduce errors through elimination of systematic bias and/or by reduction of "short memory" or transient errors (Pagano et al., 2011). The former is generally achieved by using simple statistical approaches like quantile mapping or regression (Hashino et al., 2007; Shi et al., 2008), while the latter by prediction updating (Lekkas et al., 2001; Moraweitz et al., 2011). The prediction updating techniques exploit persistence of residuals to correct for errors using linear or non-linear auto-regressive models (WMO, 1992; Shamseldin and O'Connor, 1999; Xiong and O'Connor, 2002; Pagano et al., 2011). Streamflow predictions, even after bias correction and prediction updating contain errors that cannot be eliminated and information on prediction uncertainty is useful for decision makers who use the predictions. Post-processors are generally designed to provide an estimate of the total "lumped" uncertainty in the predictions by constructing statistical models of errors based on model predictions and historical observations (e.g. Krzysztofowicz, 1999; Engeland et al., 2005; Montanari and Grossi, 2008).

In hydrology, post-processors have been mostly used for short-term streamflow or river height forecasting. The examples include Bayesian Forecasting System (BFS; Krzysztofowicz, 1999, 2002; Reggiani and Weerts, 2008), the US National Weather Service (NWS) post-processor (Seo et al., 2006), the General Linear Model Post-Processor (Zhao et al., 2011), the Meta Gaussian post-processor (Montanari and Grossi, 2008) and others. They range in complexity from the NWS post-processor that adopts a fairly simple auto-regressive error structure (Seo et al., 2006), to BFS that uses complex parameterization scheme based on meta-Gaussian distributions. Some are primarily intended for uncertainty quantification but also include components for error reduction (e.g. Krzysztofowicz, 1999, 2002).

Methods for parameterization, parameter estimation and calculation of predictive distributions differ among post-processors, although some common features can be found. Most post-processors produce probabilistic predictive distributions of streamflow (or river height) conditioned on model predictions and recent streamflow observations. They generally assume linear dependence among the variates in a transformed

HESSD
9, 11199–11225, 2012

normal space and most use Normal Quantile Transformation (NQT; Krzysztofowicz, 1997, 1999; Todini, 2008; Li et al., 2010) to normalise the variables. All assume the estimated values of the parameters (of the post-processors) to be "true" and ignore the uncertainty in estimating their values (Krzysztofowicz, 1999, 2002; and others). For a complex post-processor like BFS, this (parametric uncertainty) can be substantial (Seo et al., 2006). More importantly, they are all designed to post-process streamflow predictions at daily or sub-daily time scales.

For many hydrological applications, such as seasonal streamflow forecasting, water resources and climate change assessments, monthly streamflow volumes are of primary interest. 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 are reliable in uncertainty distributions and have the least errors achievable. It is likely much more effective to apply post-processing directly at the monthly time scale, where pre-processed monthly volumes may come either from aggregating daily model outputs or simply from monthly models.

In this study, we investigate the use a Bayesian joint probability (BJP) modelling approach to post-process model predictions of monthly streamflow volumes. The BJP method was originally developed for forecasting seasonal streamflows in Australia (Wang et al., 2009). Here we apply it for bias correction, prediction updating and uncertainty quantification of monthly streamflow volumes generated from a monthly water balance model. The BJP method uses a parametric transformation to normalise data and stabilise variance. It allows for parameter uncertainty in the post-processor, and this can be important when dealing with monthly variables, which have far fewer data points than daily variables. In this study, we assess three formulations of the BJP post-processor in their ability to reduce error and quantify uncertainty.

The paper is structured as follows. Section 2 describes the catchments and data used in the study. Section 3 presents the hydrological model used and the formulations of the BJP post-processor. Evaluation of the post-processor is given in Sect. 4 and followed by discussions in Sect. 5. Conclusions are drawn in Sect. 6.

HESSD
9, 11199–11225, 2012

2 Study area and data

We test BJP post-processor in 18 catchments located in Queensland, Victoria (including one at the border with New South Wales) and Tasmania (see Fig. 1). The Victorian catchments are further divided into 3 regions; upper Murray, central Victoria and southern Victoria (see Table 1).

The catchments range in size from 127 to 36 000 km^2. The Queensland catchments are the largest in size and experience grassland to semi-arid type of climate, characterized by low rainfall and high evapotranspiration. The mean annual rainfall is less than 600 mm and the catchments are dry during the austral winter. In contrast, the Victorian catchments experience a temperate climate, with higher rainfall (617–1400 mm) occurring during the austral winter and spring. The Tasmanian catchments experience temperate oceanic climate and are the wettest with mean annual rainfall in excess of 1900 mm. The Tasmanian catchments are wet throughout the year.

We use observed monthly streamflow data obtained from various water resource management agencies and the Bureau of Meteorology, Australia. For most catchments, with the exception of some in Queensland and Victoria, the data are available from 1950 to 2008 (see Table 1). The monthly catchment average rainfall and potential evapotranspiration for each catchment are calculated from a 5 km gridded dataset available from the Australian Water Availability Project (AWAP; Jones et al., 2009).

3 Methods

In each catchment, we calibrate parameters of a hydrologic water balance model and generate predictions. The "raw" predictions generated by the model contain errors that are unreconciled during calibration process. The BJP post-processor aims to reduce such errors and quantify uncertainty. This section describes the process of generating streamflow predictions and their subsequent post-processing.

HESSD
9, 11199–11225, 2012

3.1 Generation of streamflow predictions using a hydrological model

We use a monthly model known as WAPABA (Water Partition and Balance; Wang et al., 2011) model to generate streamflow predictions. WAPABA is a modified version of the Budyko framework model (Zhang et al., 2008) and consists of two storages and five parameters. The model uses consumption curves to partition water into different components based on the availability of water (supply) and demand. WAPABA has been tested in 331 catchments in Australia and has demonstrated ability to perform well (Wang et al., 2011).

We calibrate WAPABA using the shuffled complex evolution search method (SCE; Duan et al., 1994) for a period of five years. Prior to every model run we allow a five year warm up period to reduce model sensitivity to state initialization errors. We use a scalarized multi-objective measure consisting of a uniformly weighted average of the Nash-Sutcliffe efficiency coefficient (Nash and Sutcliffe, 1970), the Nash-Sutcliffe efficiency of log transformed flows, the Pearson correlation coefficient and a symmetric measure of bias (Wang et al., 2011). Finally, we use the calibrated parameters to produce streamflow predictions using the observed rainfall.

3.2 Statistical post-processing

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, including mean, variance and transformation parameters for each variable and a correlation matrix for the multivariate

HESSD
9, 11199–11225, 2012

normal distribution, is estimated using a Bayesian inference. The sampling of the posterior parameter distribution is done by using a Markov chain Monte Carlo method. The posterior predictive density for a new event is given by:

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

where Y_{OBS} contains the historical data of both predictor and predictand variables used for model inference, and θ is the parameter vector. Details of the method for the numerical evaluation of Eq. (1) can be found in Wang et al. (2009) and Wang and Robertson (2011).

To apply BJP as a post-processing tool we implement three methodologies with different combinations of the predictors.

Method A: method A represents the simplest case where only WAPABA predictions is applied as the predictor ($y(1)$ in Eq. 1) and the observed streamflows as the predictand ($y(2)$ in Eq. 1). This combination is to achieve two post-processing objectives, correction of systematic bias and quantification of uncertainty. The bias correction is achieved through the regression property embedded within the BJP modelling approach (see Wang et al., 2009).

Method B: for method B, we add a second predictor over that used for method A. We add streamflow data observed one month previously. The inclusion of lagged streamflow observations is to add auto-regressive component to the post-processor and allow prediction updating. This method reduces errors through correction of systematic bias as well as prediction updating and quantifies uncertainty.

Method C: for method C, we introduce a third predictor, the WAPABA model outputs simulated in the previous month. This inclusion is to further improve the prediction updating ability of the post-processor by utilizing the persistence in the simulated time series.

HESSD
9, 11199–11225, 2012

For each method, we first train the post-processor using the historically observed data. To account for seasonal effects, we establish 12 different models for each month of the year. For each month, the post-processed probabilistic predictions are generated using a "leave-one-out" cross validation procedure. This consists of sampling the parameters using all but the year of interest and then generating predictions for the "left out" year. The cross validation period in most catchments period is about 59 yr (1950–2008).

Figure 2 is an example of the post-processed predictions generated by the BJP post-processor. This example is to provide the reader an appreciation of how the post-processed predictions from the BJP post-processor may look like. A detail evaluation of the post-processor, with respect to the post-processing qualities, will be presented in Sect. 4. The example is drawn from Lake Eildon in central Victoria and shows $[0.1, 0.25, 0.5, 0.75, 0.9]$ quantiles and observed streamflow values plotted chronologically. In this case the post-processed predictions do not show any obvious trend with time and the widths of the quantile intervals seem to cover the expected number of the observed values.

4 Results

In this section we assess the quality of the probabilistic predictions generated by using the three methods and evaluate how effective the BJP post-processor is in reducing errors and quantifying uncertainty.

4.1 Reduction of error

We assess the ability of BJP post-processor to reduce errors by using a measure of accuracy called Root Mean Squared Error in Probability (RMSEP; Wang and Robertson, 2011). RMSEP (Eq. 2) measures error in a probability space. An advantage of RMSEP over the more commonly used mean squared error or root mean squared error is that, it places equal emphasis on errors obtained at all events rather than on a few large

HESSD
9, 11199–11225, 2012

errors occurring at large events.

$$\text{RMSEP} = \left[\frac{1}{n}\sum_{t=1}^{n}(F_{\text{CLI}}(y^t) - F_{\text{CLI}}(y^t_{\text{OBS}}))^2\right]^{\frac{1}{2}} \quad (2)$$

where, y^t and y^t_{OBS} are the predictions and observations at $t = 1, 2, \ldots, n$ events, respectively. The predictions can be either WAPABA simulations or the medians of post-processed distributions. F_{CLI} is the cumulative historical distribution, and $F_{\text{CLI}}(y)$ is the non-exceedance probability.

4.1.1 Performance of the WAPABA model

The RMSEP error values of the WAPABA predictions are shown in Fig. 3. Each row in the figure corresponds to a catchment and each column to a month. In general, except for river Cape and in Queensland, Thompson and O'Shannassy Reservoirs in southern Victoria, the RMSEP values are relatively higher in drier months or months when the catchments just start to get wet. This occurs during August–October in Queensland, May–March in upper Murray, January–May in central Victoria and February–March in Tasmania.

This suggests an inability of the model to properly characterize low flows and to capture the change in catchment dynamics from being dry to getting wet. There can be various reasons for this, including, for example, non-consistency in the data between the calibration and evaluation period (with the calibration period being either wetter or direr), the choice of objective function for calibration, inadequate model structure or a combination of these. While it might be interesting to investigate the causes of poor model performance in these catchments from a model diagnostic point of view, this is beyond the scope of this study. Here we only focus on evaluating whether the errors can be reduced by the post-processor.

HESSD
9, 11199–11225, 2012

4.1.2 Method A: bias reduction

Figure 4a shows the differences in RMSEP error values between the WAPABA predictions and that produced from method A (WAPABA prediction – method A). The values are colour coded with blue indicating the reductions in RMSEP error values and red indicating increases.

In general, the result shows that method A effectively reduces systematic bias present in WAPABA predictions. This is manifested as reductions in RMSEP error values over the 18 catchments. The reductions in RMSEP roughly follow the error patterns seen in Fig. 3. In most cases the differences in RMSEP values are either positive or zero, indicating that the post-processor either reduces errors or preserves (does not degrade) performance of the WAPABA predictions. The highest reductions in RMSEP values occur in Lake Eildon and Goulbourn Weir of central Victoria.

4.1.3 Method B: prediction updating

Figure 4b shows the benefit of prediction updating by assimilating the recent streamflow observations (method B). We use the difference between method A and B (method A – method B) to indicate any further reductions in errors achieved by prediction updating. As in previous case blue indicates reductions in errors and red indicates increases.

The figure shows further reductions in RMSEP values after bias correction (method A). The reductions occur in most of the catchments. The reductions in errors are governed by whether the errors are present after bias correction and the persistence in the streamflow observation data. For example, the WAPABA predictions in River Cape of Queensland and Lake Nillahcootie of central Victoria (Fig. 3) show the presence of substantially large error values in initial few months even after bias correction, but cannot be corrected due to the lack of persistence in the errors. In the upper Murray region, central Victoria and southern Victoria, reductions occur in most of the catchments, and in some catchments (such as Cairn Curran Reservoir), it is greater than that achieved through bias correction. In Tasmanian catchments, the reductions are negligible.

HESSD
9, 11199–11225, 2012

4.1.4 Method C: prediction updating using WAPABA lagged simulation

Figure 4c shows additional benefits achieved by assimilating "lagged" streamflow simulation. The difference is measured relative to method B (method B – method C) such that positive (blue) values indicate further reductions in RMSEP error values over that achieved by B. The result shows that the benefits of adding lag-1 WAPABA streamflow tend to be negligible in most catchments and seasons. Although some reductions in RMSEP error values can be observed in Maroondah reservoir (in southern Victoria), for the months of February, March and May in other catchments it is close to zero.

4.2 Quantification of uncertainty

The post-processor should be able to quantify the uncertainty in predictions. As a measure of the ability to quantify uncertainty, we assess if the probabilistic predictions generated by the post-processor are reliable and robust. We assess the predictions generated using all three methods in 18 catchments, but present results for Lake Eildon using method B as a general representation.

4.2.1 Assessment of reliability

Reliability refers to "statistical consistency" of the predictive probability distributions with the observed frequency of the events (Toth et al., 2003; Robertson et al., 2012). In this study we use PIT (probability integral transform) uniform probability plots (Wang et al., 2009; Wang and Robertson, 2011) to assess the overall reliability of the post-processed predictive distributions. We choose PIT uniform probability plots over other methods because they are more suited to smaller sample sizes (Wang et al., 2009).

The PIT of the observed value is given as, $\pi^t = F^t(y^t_{\mathrm{OBS}})$, where, $F^t(y^t_{\mathrm{OBS}})$ is the non-exceedance probability of the observed streamflow in the predictive distribution. The predictive distributions are said to be reliable if the PIT values are distributed uniformly. To check uniformity, we plot PIT values corresponding to each event in a uniform

HESSD
9, 11199–11225, 2012

probability plot (Wang et al., 2009; Wang and Robertson, 2011). A close alignment of the values to 1 : 1 indicates uniformity and therefore reliable distributions. Deviations from the 1 : 1 line indicate if the predictive distributions are too low, high or if the uncertainty spreads are too wide or narrow. The details on how to interpret the PIT plots can be found in Thyer et al. (2009), Wang et al. (2009) and Wang and Robertson (2011).

Figure 5 shows the PIT uniform probability plots of the post-processed predictions generated for the months of February and July in Lake Eildon. The dotted inclined lines depict the Kolmogorov 5 % significance band. The PIT values in the plots align quite uniformly along the diagonal 1 : 1 line (solid inclined line) and are well within the significance band. This suggests that the post-processed predictive distributions are over all reliable and the width of uncertainty intervals are of appropriate spread (not too wide or narrow). The result is similar for all the months in Lake Eildon (figures not included).

4.2.2 Assessment of robustness

Robustness refers to "conditional reliability" of the predictive distributions over time and event size. To measure the robustness of the predictive distributions against time, we plot PIT values chronologically and analyse the plot for the presence of any trends or patterns. The distributions are robust (over time) if the PIT values are distributed uniformly. Any existing trends or patterns indicate the presence of systematic errors in the distributions (Wang et al., 2009; Wang and Robertson, 2011).

Figure 6 (top row) shows the PIT values plotted chronologically for February and July. The PIT values tend to be distributed randomly against time, devoid of any trends or patterns indicating that distributions are robust. In fact, this was the case for all the months in Lake Eildon (figure not included).

To measure robustness of the post-processed predictions against flow magnitudes, we plot post-processed prediction quantiles and the observed streamflow values against the medians of the predictions. As in previous case we analyse the plot to

HESSD
9, 11199–11225, 2012

detect presence of any trends or patterns. Figure 6 (bottom row) shows the post-processed quantiles plotted against event magnitude.

The figure shows that the quantiles increase with event sizes and the medians are consistent with the observed flows. The observed flows are scattered randomly over the medians, suggesting that the post-processed quantiles are robust with respect to event magnitudes. The plots also show that the width of the uncertainty intervals are of appropriate spread for all the event size.

This verification approach is applied for all post-processing method (A, B and C) for all the catchments, for each month. In general, the results are consistent to the results obtained in Lake Eildon.

5 Discussion

The results show that large bias can occur in predictions despite calibrating WAPABA using a multicriteria objective function that includes a symmetric measure of bias. This is not surprising because maximization of the scalarized function is a result of compromise between four objective functions and does not necessarily lead to removal of systematic bias in all catchments and in all months. The presence of bias is especially high in Lake Eildon. The BJP post-processor eliminates bias in the predictions effectively, resulting in bias close to zero throughout the year. This can be better appreciated in Fig. 7 which shows monthly percentage bias obtained by WAPABA predictions and its elimination by method A.

Furthermore, it is interesting to note that the bias correction is not just due to linear changes in slope or intercepts but also due to non-linear changes as illustrated by Fig. 8a, b. The figures demonstrate non-linear compensations to WAPABA predictions by the BJP post-processor. The log-sinh transformation in combination with the BJP model parameter inference, allow for the non-linear corrections of errors thus allowing for corrections of conditional as well as unconditional biases.

HESSD
9, 11199–11225, 2012

Our results show that further error reductions can be possible through prediction updating. This contradicts the assumptions made by Li et al. (2011), who assume that persistence in error structure at monthly time step is negligible. However, we note that the results tend to be catchment specific. In our case the improvements are mostly seen in catchments that have substantial streamflow contribution from the slow responding mechanisms (resulting in longer memory) in the catchment. This seems to be the case in upper Murray, central Victoria and southern catchments, where significant reductions in errors can be observed. The two catchments in Tasmania and the one in Queensland and central Victoria have shorter catchment "memory" with the streamflow being dominated by fast responding runoff processes and therefore the benefits of prediction updating is negligible.

We acknowledge that the rainfall forecast uncertainty represents a major source of uncertainty in streamflow forecast (Krzysztofowicz, 1999; Kuczera et al., 2006). In this study, however, we run the water balance model in a simulation mode. Therefore the total uncertainty quantified by the post-processor is the "lumped" combination of the hydrologic uncertainty, rainfall measurement uncertainty, the streamflow measurement uncertainty and the uncertainty in inferring the values of the parameters of the BJP post-processor. However, the post-processor is equally applicable in the real world applications using rainfall forecasts. In such case the uncertainty spread quantified by the post-processor will be wider to reflect the uncertainty in forecasting rainfall.

6 Summary and conclusions

In this study, we present a statistical post-processor capable of reducing errors and quantifying uncertainty in monthly streamflow predictions. The statistical post-processor is based on the BJP modelling approach (Wang et al., 2009). The BJP post-processor is applied to 18 catchments in Australia and its ability to reduce errors, through reductions of systematic bias and prediction updating, and to quantify uncertainty in the monthly streamflow predictions is assessed.

HESSD
9, 11199–11225, 2012

The study shows that the BJP post-processor is capable of improving the accuracy of the streamflow predictions by reducing systematic bias in most of the catchments. In many cases reduction of bias is achieved by means of non-linear relationship between model predictions and the observed streamflow values. The post-processor also demonstrates its useful property in preserving the accuracy (does not increase error) of predictions when bias correction is not possible.

Prediction updating through the assimilation of recent streamflows by the post-processor results in further reductions in RMSEP error values over that achieved by bias correction alone, and is most effective for catchments showing stronger persistence in the prediction errors. Benefits of prediction updating using additional information from the water balance model simulation seem to be very marginal and do not justify the added complexity of introducing another predictor to the post-processor.

The BJP post-processor is capable of generating probabilistic predictions that are overall reliable. The uncertainty quantified by the processor is of appropriate spread. The post-processed predictive distributions are robust with respect to time and event magnitude.

Acknowledgements. This research has been supported by the Water Information Research and Development Alliance between the Australian Bureau of Meteorology and CSIRO Water for a Healthy Country Flagship, the South Eastern Australian Climate Initiative, and the CSIRO OCE Science Leadership Scheme. Streamflow and GIS data were provided by the Murray–Darling Basin Authority, Melbourne Water, HydroTasmania, Goulburn-Murray Water, the Australian Bureau of Meteorology and the Queensland Department of Environment and Resource Management. We would also like to acknowledge James C. Bennett and Roger Hughes for their useful suggestions and help in editing the manuscript.

References

Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol., 158, 265–284, 1994.

HESSD
9, 11199–11225, 2012

Engeland, K., Xu, C. Y., and Gottschalk, L.: Assessing uncertainties in a conceptual water balance model using Bayesian methodology, Hydrol. Sci. J., 50, 45–63, 2005.
Gupta, H. V., Beven, K. J., and Wagener, T.: Model calibration and uncertainty estimation, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G. and McDonnell, J. J., John Wiley & Sons Ltd, Chichester, 1–17, 2005.
Hashino, T., Bradley, A. A., and Schwartz, S. S.: Verification of probabilistic streamflow forecasts, IIHR Report No. 427, IIHR-Hydroscience & Engineering and Dept. of Civil and Environ. Eng., The Univ. of Iowa, Iowa City, IA, 125 pp., 2002.
Jones, D. A., Wang, W., and Fawcett, R.: High-quality spatial climate data-sets for Australia, Aust. Meteorol. Oceanogr. J., 58, 233–248, 2009.
Krzysztofowicz, R.: Transformation and normalization of variates with specified distributions, J. Hydrol., 197, 286–292, 1997.
Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, Water Resour. Res., 35, 2739–2750, 1999.
Krzysztofowicz, R.: Bayesian system for probabilistic river stage forecasting, J. Hydrol., 268, 16–40, 2002.
Kuczera, G., Kavetski, D., Franks, S., and Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters, J. Hydrol., 331, 161–177, doi:10.1016/j.jhydrol.2006.05.010, 2006.
Lekkas, D. F., Lees, M. J., and Imrie, C. E.: Improved nonlinear transfer function and neural network methods of flow routing for real-time forecasting, J. Hydroinform., 3, 153–164, 2001.
Li, L., Xia, J., Xu, C.-Y., and Singh, V. P.: Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models, J. Hydrol., 390, 210–221, 2010.
Li, L., Xu, C.-Y., Xia, J., Engeland, K., and Reggiani, P.: Uncertainty estimates by Bayesian method with likelihood of AR(1) plus Normal model and AR(1) plus multi-normal model in different time-scales hydrological models, J. Hydrol., 406, 54–65, doi:10.1016/j.jhydrol.2011.05.052, 2011.
Montanari, A. and Grossi, G.: Estimating the uncertainty of hydrologic forecasts: a statistical approach, Water Resour. Res., 44, W00B08, doi:10.1029/2008WR006897, 2008.
Morawietz, M., Xu, C.-Y., and Gottschalk, L.: Reliability of autoregressive error models as post-processors for probabilistic streamflow forecasts, Adv. Geosci., 29, 109–118, doi:10.5194/adgeo-29-109-2011, 2011.

HESSD
9, 11199–11225, 2012

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, 1970.
Reggiani, P. and Weerts, A. H.: A Bayesian approach to decision-making under uncertainty: an application to real-time forecasting in the river Rhine, J. Hydrol., 356, 56–69, 2008.
Pagano, T. C., Wang, Q. J., Hapuarachchi, P., and Robertson, D.: A dual-pass error-correction technique for forecasting streamflow, J. Hydrol., 405, 367–381, 2011.
Robertson, D. E., Pokhrel, P., and Wang, Q. J.: Improving statistical forecasts of seasonal streamflows using hydrological model output, Hydrol. Earth Syst. Sci., 2012, submitted.
Seo, D.-J., Herr, H. D., and Schaake, J. C.: A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, Hydrol. Earth Syst. Sci. Discuss., 3, 1987–2035, doi:10.5194/hessd-3-1987-2006, 2006.
Shamseldin, A. Y. and O'Connor, K. M.: A non-linear neural network technique for updating of river flow forecasts, Hydrol. Earth Syst. Sci., 5, 577–598, doi:10.5194/hess-5-577-2001, 2001.
Shi, X., Wood, A. W., and Lettenmaier, D. P.: How essential is hydrologic model calibration to seasonal streamflow forecasting?, J. Hydrometeorol., 9, 1350–1363, 2008.
Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., and Srikanthan, S.: Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: a case study using Bayesian total error analysis, Water Resour. Res., 45, W00B14, doi:10.1029/2008WR006825, 2009.
Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, Int. J. River Basin Manage., 6, 123–137, 2008.
Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and ensemble forecasts, in: Forecast Verification: a Practitioner's Guide in Atmospheric Science, edited by: Jolliffe, I. T. and Stephenson, D. B., Wiley, Chichester, 137–164, 2003.
Wang, Q. J. and Robertson, D. E.: Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, Water Resour. Res., 47, W02546, doi:10.1029/2010wr009333, 2011.
Wang, Q. J., Robertson, D. E., and Chiew, F. H. S.: A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, Water Resour. Res., 45, W05407, doi:10.1029/2008WR007355, 2009.

HESSD
9, 11199–11225, 2012

Wang, Q. J., Pagano, T. C., Zhou, S. L., Hapuarachchi, H. A. P., Zhang, L., and Robertson, D. E.: Monthly versus daily water balance models in simulating monthly runoff, J. Hydrol., 404, 166–175, 2011.

Wang, Q. J., Shrestha, D. L., Robertson, D. E., and Pokhrel, P.: A log-sinh transformation for data normalization and variance stabilization, Water Resour. Res., 48, W05514, doi:10.1029/2011WR010973, 2012.

WMO: Simulated real-time intercomparison of hydrological models, Op. Hydrol. Rep., World Meteorological Organisation, Geneva, 1992.

Xiong, L. H. and O'Connor, K. M.: Comparison of four updating models for real-time river flow forecasting, Hydrol. Sci. J., 47, 621–639, 2002.

Yeo, I. and Johnson, R. A.: A new family of power transformations to improve normality or symmetry, Biometrika, 87, 954–959, 2000.

Zhang, L., Potter, N., Hickel, K., Zhang, Y., and Shao, Q.: Water balance modelling over variable time scales based on the Budyko framework – model development and testing, J. Hydrol., 360, 117–131, 2008.

Zhao, L., Duan, Q., Schaake, J., Ye, A., and Xia, J.: A hydrologic post-processor for ensemble streamflow predictions, Adv. Geosci., 29, 51–59, doi:10.5194/adgeo-29-51-2011, 2011.

HESSD

9, 11199–11225, 2012

Table 1. Brief attributes of the 18 catchments used for the study.

Catchment	Region	Available Record	Catchment area (km^2)	Mean annual rainfall (mm)	Mean annual flow (mm)	Annual runoff coefficient
Burdekin River	Queensland	1967–2008	36 260	567	76 (2765 GL)	0.13
Cape River	Queensland	1967–2008	16 074	456	23 (372 GL)	0.05
Lake Hume	upper Murray	1950–2008	12 184	819	227 (2764 GL)	0.28
Dartmouth Reservoir	upper Murray	1950–2008	3193	1042	279 (890 GL)	0.27
Kiewa River	upper Murray	1965–2008	1748	1099	248 (433 GL)	0.23
Ovens River	upper Murray	1959–2008	7543	963	175 (1320 GL)	0.18
Lake Nillahcootie	central Victoria	1950–2008	422	942	150 (63 GL)	0.16
Lake Eildon	central Victoria	1950–2008	3877	1104	373 (1447 GL)	0.34
Goulburn Weir	central Victoria	1950–2008	7166	769	188 (1349 GL)	0.24
Eppalock Reservoir	central Victoria	1950–2008	1749	630	98 (172 GL)	0.16
Cairn Curran Reservoir	central Victoria	1950–2008	1603	617	72 (115 GL)	0.12
Tullaroop Reservoir	central Victoria	1950–2008	702	633	77 (54 GL)	0.12
Thompson Reservoir	southern Victoria	1950–2008	487	1299	485 (236 GL)	0.37
Upper Yarra Reservoir	southern Victoria	1950–2008	336	1387	443 (149 GL)	0.32
Maroondah Reservoir	southern Victoria	1950–2008	129	1351	577 (74 GL)	0.43
O'Shannassy Reservoir	southern Victoria	1950–2008	127	1404	766 (97 GL)	0.55
Mersey-Forth HES	Tasmania	1950–2008	2698	1900	793 (2141 GL)	0.42
King HES	Tasmania	1950–2008	731	2703	1724 (1260 GL)	0.64

Note: HES stands for Hydro Electric Scheme.

HESSD
9, 11199–11225, 2012

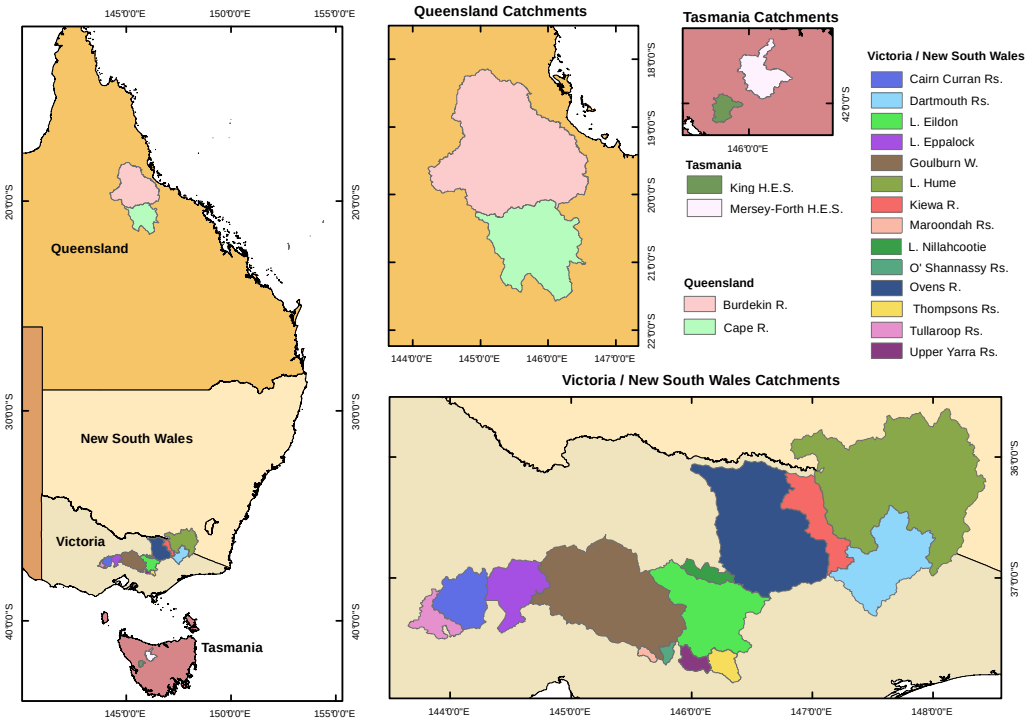

Fig. 1. Location of the 18 catchments used for the study.

HESSD
9, 11199–11225, 2012

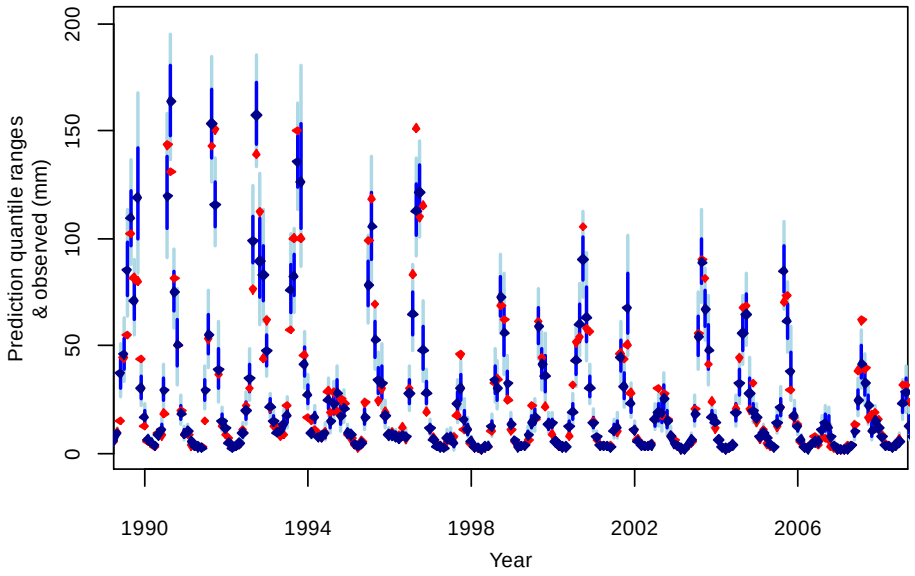

Fig. 2. The time series of the post-processed prediction quantiles against the observed; only a subset of the entire cross validation period is shown. (Light blue lines represent 0.1–0.9 quantiles, dark blue lines represent the 0.25–0.75 quantiles, red dots are the observed streamflow values and the blue dots are the medians of the post-processed predictive distributions.)

HESSD
9, 11199–11225, 2012

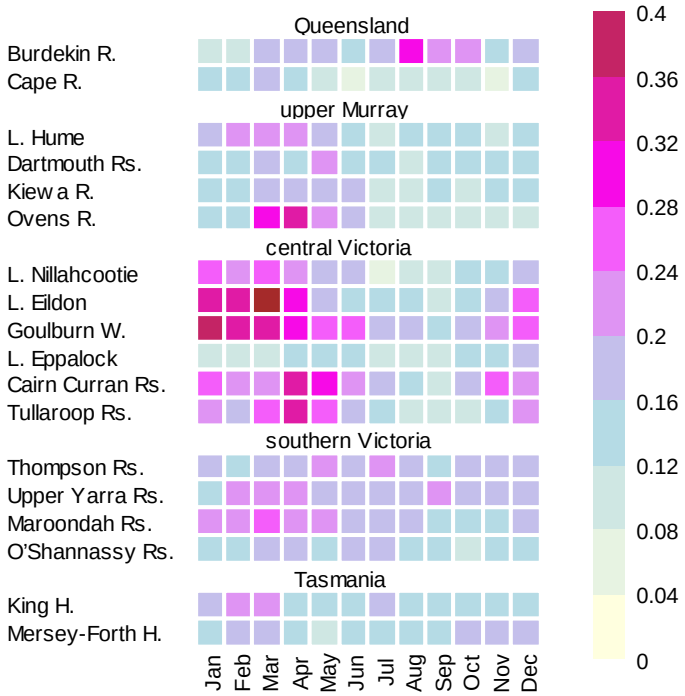

Fig. 3. Performance of WAPABA in all 18 catchments; RMSEP error values calculated for each month.

HESSD
9, 11199–11225, 2012

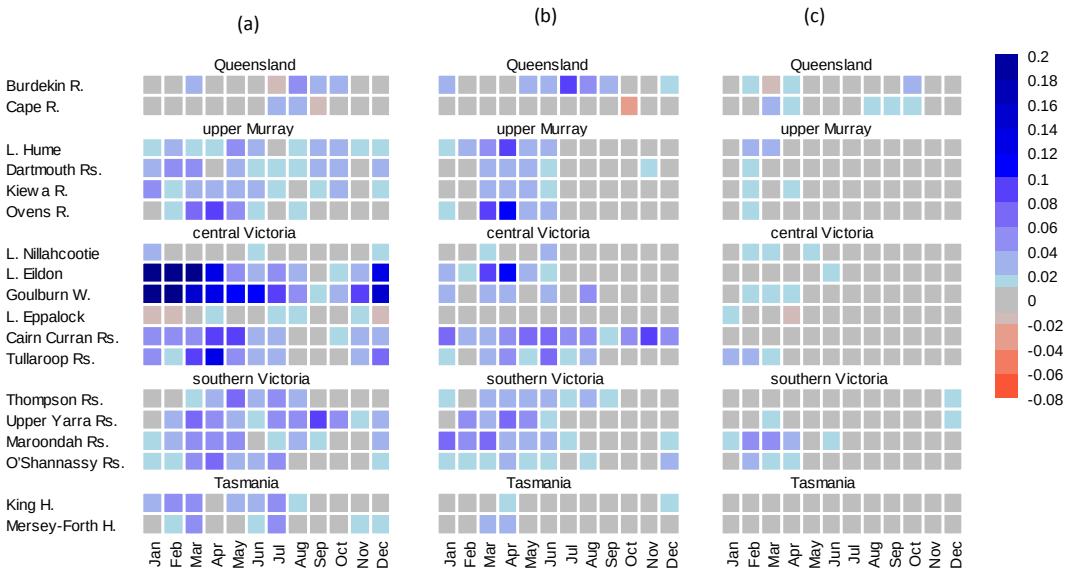

Fig. 4. (a) Difference in RMSEP values between WAPABA predictions and method A (WAPABA – method A); **(b)** difference in RMSEP values between method A and B (method A – method B); **(c)** difference in RMSEP values between method B and C (method B – method C).

HESSD
9, 11199–11225, 2012

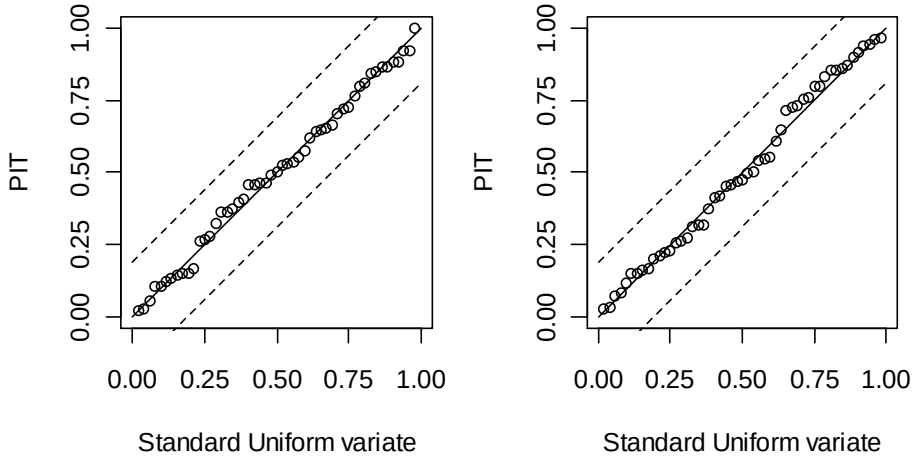

Fig. 5. PIT uniform probability plots of post-processed streamflow predictions for two months (February and July) in Lake Eildon; (1 : 1 solid line, theoretical uniform distribution; broken lines, Kolmogorov 5 % significance band; circle, PIT value of observed streamflow).

HESSD
9, 11199–11225, 2012

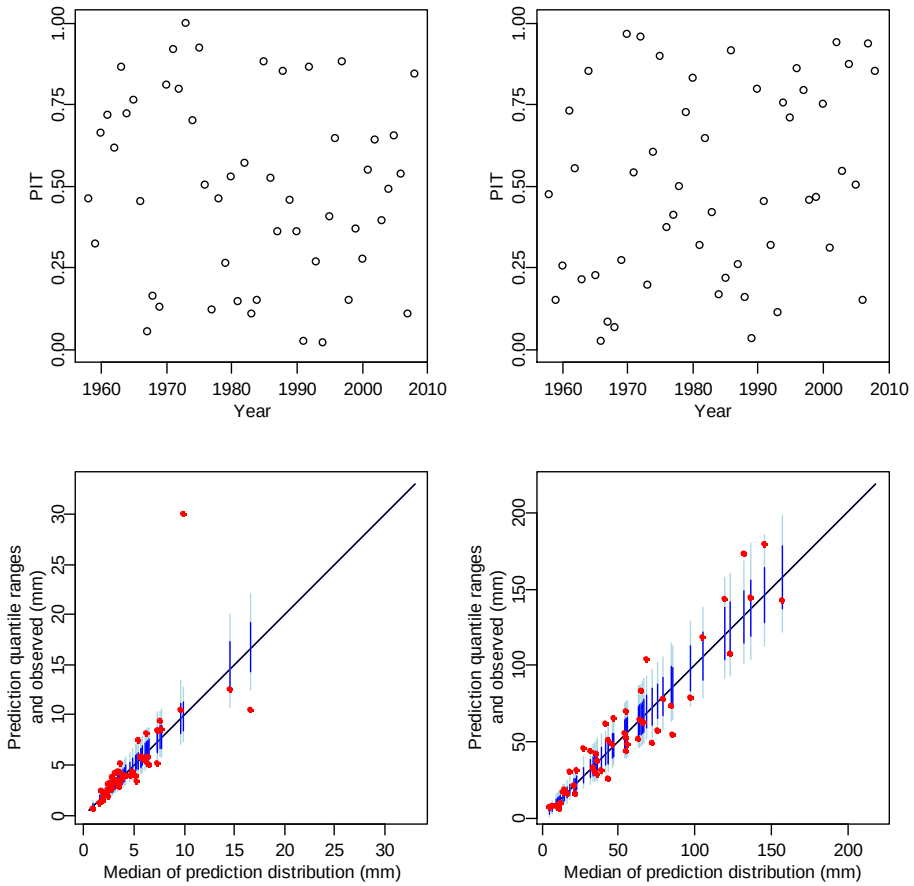

Fig. 6. Top row – Probability Integral Transform (PIT) values plotted against time for months of February and July; Bottom row – post-processed quantiles plotted against their median values for months of February and July. The red dots represent the observed streamflows, light blue vertical lines are the 0.1–0.9 quantiles and the dark blue lines are the 0.25–0.75 quantiles.

HESSD
9, 11199–11225, 2012

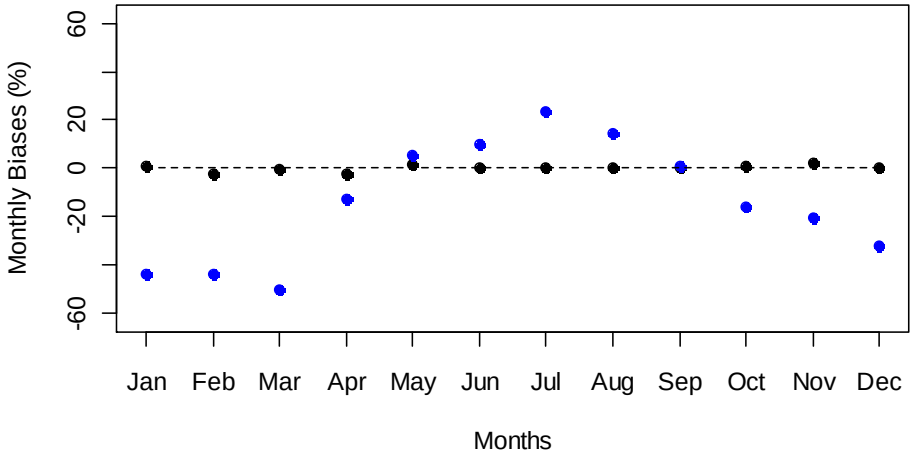

Fig. 7. Monthly biases % in Lake Eildon, raw WAPABA predictions (blue) and method A (black). (The bias for method A = (Mean monthly streamflow observations – Mean probabilistic predictions averaged each month)/(Average observed monthly streamflow); the bias for WAPABA = (mean monthly streamflow observations – mean monthly WAPABA predictions)/(Average observed monthly streamflow)).

HESSD
9, 11199–11225, 2012

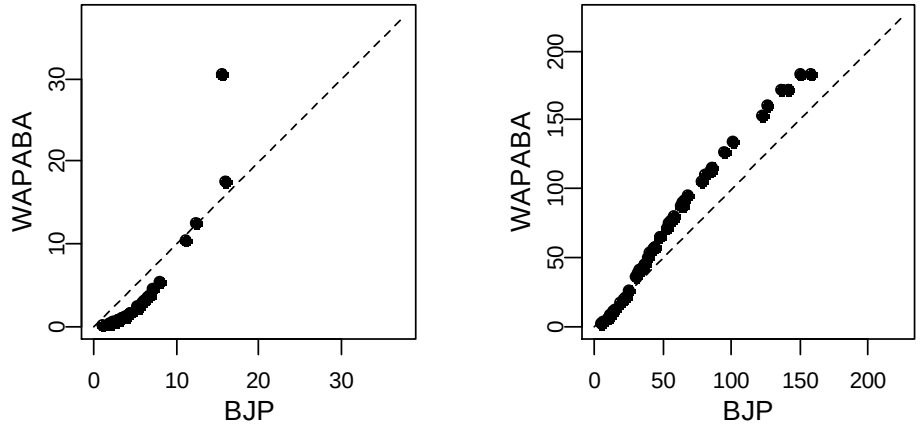

Fig. 8. WAPABA predictions vs. the medians of post-processed predictions produced by method A in Lake Eildon for months of February (left) and July (right), showing the example of non-linear error corrections by the BJP post-processor.

HESSD
9, 11199–11225, 2012