20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013
www.mssanz.org.au/modsim2013

Challenges for including error updating in real-time hydrological error models

M. Li [a], **Q.J. Wang** [b] **and J.C. Bennett** [b]

[a] *CSIRO Mathematics, Informatics and Statistics, Floreat, WA, Australia*
[b] *CSIRO Land and Water, Highett, Victoria, Australia*
Email: Ming.Li@csiro.au

Abstract: Accurate and reliable quantification of uncertainty in hydrological models promises to make real-time streamflow predictions both more accurate and more useful. One simple but effective approach to quantifying hydrological uncertainty is to apply an error model to streamflow simulations. An error model collates errors from all sources into prediction errors and builds up a statistical model of the error time series. Streamflows are often highly auto-correlated, and updating real-time hydrological error models with information from recent observations is an obvious means to improving the accuracy of the predictions. We attempt to include an error updating component in a two-stage error model. In Stage 1, we use a logarithmic hyperbolic sine transform to normalise both simulated and observed time series and stabilize variance. A bias-correction component is introduced to correct the bias of transformed simulations. In Stage 2, we apply an error updating procedure to the simulations obtained from Stage 1 by using the information from the previous time step. The error updating is based on the auto-correlation of hydrological errors. All model parameters in both stages are assumed to be seasonally dependent, because hydrological models often perform differently for different seasons. The residual term from both stages are assumed to be Gaussian. Stage 1 only uses the information from the present time step and can be applied without Stage 2, while Stage 2 relies on Stage 1 and requires information from the previous time step.

We test the error model on hydrological simulations of four catchments generated with a conceptual daily rainfall-runoff model. In isolation, Stage 1 leads to similar or marginally more accurate predictions than the original hydrological simulations. The Stage 1 uncertainty estimation is generally reliable. Applying the error updating model (Stage 2) markedly improves the accuracy of Stage 1 simulations. However, we show that the uncertainty in the Stage 2 simulations is no longer reliably quantified after the error updating is applied. This is associated with our assumption that the residual term is Gaussian. After applying the error updating, most residuals in the error model are significantly reduced, while a few residuals remain. This causes the distribution of residuals to have a longer tail than a Gaussian distribution.

The two-stage estimation of hydrological prediction uncertainty is shown to be simple but effective. The prediction accuracy is progressively improved after bias correction in the transform domain (Stage 1) and error updating in original domain (Stage 2). However, it remains challenging to offer reliable uncertainty estimation after including error updating. We recommend replacing the Gaussian distribution with more sophisticated distributions in the error model.

Keywords: *Uncertainty estimation, error model, hydrological prediction*

1. INTRODUCTION

Accurate and reliable quantification of uncertainty in hydrological models promises to make real-time streamflow predictions both more accurate and more useful. Various approaches based on full Bayesian modeling have been introduced to estimate the hydrological uncertainty. For example, Generalized Likelihood Uncertainty Estimation (GLUE) (e.g. Beven and Binley, 1992; Freer et al., 1996) aggregates all sources of error to parameter uncertainty and derives the predictive uncertainty based on the posterior distribution of parameters. Full Bayesian modeling requires implementing the Markov chain Monte Carlo (MCMC) algorithm (e.g. Kavetski et al., 2006; Kuczera et al., 2006). A simpler, but still effective approach to quantifying hydrological uncertainty is to apply an error model to streamflow simulations. Unlike GLUE, which transforms prediction uncertainty to parameter uncertainty, an error model lumps all errors into prediction errors and explicitly establishes a statistical model for prediction errors.

We establish error models to describe the prediction errors. There are three key components in the error models we describe here: 1) transforming streamflow data, 2) correlation structure and 3) variance structure. Transformations such as the Box-Cox transform and the logarithmic hyperbolic sine transform (hereafter *log-sinh*) (e.g. Thiemann et al., 2001; Wang et al., 2012) are generally carried out in order to normalize the data and stabilize variance. A time-series model is commonly used to represent the temporal relationship within prediction errors. Variance structure is assumed to be either homoscedastic (constant variance) (e.g. Diskin and Simon, 1977) or heteroscedastic (time-varying variance) (e.g. Schoups and Vrugt, 2010).

Li et al. (2013) tested several error models for monthly streamflow predictions. They identified the presence of seasonal dependence in prediction error of monthly streamflow forecasts and recommended the use of seasonally dependent error model parameters. Further, they found that a seasonally variant error model based on autocorrelation in errors produced accurate and reliable monthly error estimates. Monthly flows often vary less and have different autocorrelation than daily flows, and it is not clear that the same model would perform equally well at the daily time step. Here we present a two-stage error model based on Li et al.'s (2013) seasonally variant error model for estimating errors in predictions of daily streamflow.

2. TWO-STAGE UNCERTAINTY ESTIMATION

We denote the actual and simulated streamflow from a hydrological model at a given day t by Q_t and $Q_{S,t}$, respectively. A log-sinh transform (Wang et al., 2012)

$$Z(Q) = b^{-1}\log(\sinh(a+bQ)) \tag{1}$$

is applied to normalize data and stabilize variance. Wang et al. (2012) showed this transform describes hydrological data better than the more commonly used Box-Cox transform. To account for seasonal differences in daily streamflow distributions, we use different transform parameters for different months. We denote the log-sinh transform for month i by Z_i and the inverse log-sinh transform by Z_i^{-1}.

We use a two-stage error model to estimate the prediction uncertainty and attempt to reduce the prediction error progressively. The error model in Stage 1 explicitly accounts for the prediction bias:

$$Z_{i(t)}(Q_t) = \mu_{i(t)} + Z_{i(t)}(Q_{S,t}) + \sigma_{i(t)}\varepsilon_t \tag{2}$$

where $i(t)$ denotes the calendar month of day t and ε_t is assumed to independently follow a standard Gaussian distribution. The error model parameters μ and σ represent the prediction bias in the transform domain and the standard deviation of model residuals, respectively, in Stage 1. For a given month, there are 4 unknown parameters, including two transform parameters and two error model parameters. Forty-eight parameters are estimated in Stage 1.

The error model in Stage 2 is based on the Stage 1 bias-corrected streamflow simulations and uses the information from the previous day for error updating, as follows:

$$Z_{i(t)}(Q_t) = Z_{i(t)}\left(Q_{S,t}^* + \rho_{i(t)}\left(Q_{t-1} - Q_{S,t-1}^*\right)\right) + \tau_{i(t)}\varepsilon_t \tag{3}$$

where $Q_{S,t}^* = \max\left(Z_{i(t)}^{-1}\left(\mu_{i(t)} + Z_{i(t)}(Q_{s,t})\right), 0\right)$ is the prediction mean made from Stage 1 and ε_t is also assumed to independently and identically follow a standard Gaussian distribution. The error model parameters ρ and τ represent the auto-correlation of prediction errors in the original domain and the

standard deviation of model residuals in Stage 2. Stage 2 uses the transformation parameters estimated in Stage 1. Twenty-four parameters are estimated in Stage 2 (2 parameters for a given month).

Stage 1 aims to correct the bias of hydrological simulation and assumes that such bias is constant for a given month. No other information other than the hydrological simulation at the present time is needed. The error model in Stage 1 can be used in isolation and is particularly useful when streamflow observations at the previous time step are not available. Stage 2 is based on the bias-corrected simulation from Stage 1 and further utilizes the information from the previous day in error updating. The error updating is essentially performed in the original domain in order to prevent potential irregular outliers caused by the back-transform.

The error models in both stages are estimated by the maximum likelihood estimation (MLE). MLE provides a point estimation of model parameters that maximizes the likelihood function (conditional probability of Q_t). Explicitly, MLE finds the model parameter that maximizes

$$\text{Stage1: } \prod_{t=1}^{T} P(Q_t \mid Q_{S,t}) , \tag{4}$$

$$\text{Stage2: } \prod_{t=1}^{T} P(Q_t \mid Q_{S,t}^{*}, Q_t, Q_{S,t-1}^{*}) . \tag{5}$$

We derive the likelihood at each stage as follows:

$$\text{Stage1: } P(Q_t \mid Q_{S,t}) = J_{z\to Q}\phi\left(\varepsilon_{i(t)}^{-1}\left(Z_{i(t)}\left(Q_t\right) - \mu_{i(t)} - Z_{i(t)}\left(Q_{S,t}\right)\right)\right), \tag{6}$$

$$\text{Stage2: } P(Q_t \mid Q_{S,t}^{*}, Q_t, Q_{S,t-1}^{*}) = J_{z\to Q}\phi\left(\tau_{i(t)}^{-1}\left(Z_{i(t)}\left(Q_t\right) - Z_{i(t)}\left(Q_{S,t}^{*} + \rho_{i(t)}\left(Q_{t-1} - Q_{S,t-1}^{*}\right)\right)\right)\right), \tag{7}$$

where $J_{z\to Q} = \left(\tanh(a_{i(t)} + b_{i(t)}Q_t)\right)^{-1}$ is the Jacobian determinant of the log-sinh transform and ϕ is the probability density function of a standard normal distribution. We apply the zero flow treatment introduced by Wang and Robertson (2011) and consider zero flows as censored data with unknown value less than or equal to zero. The likelihood function is revised below when zero flows occur:

$$\text{Stage1: } P(Q_t = 0 \mid Q_{S,t}) = \Phi\left(\varepsilon_{i(t)}^{-1}\left(Z_{i(t)}\left(Q_t\right) - \mu_{i(t)} - Z_{i(t)}\left(Q_{S,t}\right)\right) \mid Q_t = 0\right), \tag{8}$$

$$\text{Stage2: } P(Q_t = 0 \mid Q_{S,t}^{*}, Q_t, Q_{S,t-1}^{*}) = \Phi\left(\tau_{i(t)}^{-1}\left(Z_{i(t)}\left(Q_t\right) - Z_{i(t)}\left(Q_{S,t}^{*} + \rho_{i(t)}\left(Q_{t-1} - Q_{S,t-1}^{*}\right)\right)\right) \mid Q_t = 0\right). \tag{9}$$

Two evaluation statistics are used to quantify the accuracy of the prediction at each stage. The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) is used to assess the accuracy of the prediction mean:

$$NSE = 1 - \sum_{t=1}^{T}\left(Q_{m,t} - Q_t\right)^2 \Big/ \sum_{t=1}^{T}\left(\bar{Q} - Q_t\right)^2 , \tag{10}$$

where $Q_{m,t}$ is the prediction mean and $\bar{Q}$ is the average streamflow. A greater value of NSE indicates a better prediction. The continuous ranked probability score (CRPS) (Gneiting et al., 2005) is used to evaluate the accuracy of prediction probability distributions:

$$CRPS = T^{-1}\sum_{t=1}^{T}\int\left(F_t(x) - I(x \geq Q_t)\right)^2 dx ,$$

where F_t is the cumulative probability distribution of streamflow prediction at t. A highly useful characteristic of the CRPS is that it collapses to the mean average error for deterministic forecasts, allowing comparison between deterministic and ensemble forecasts. Smaller values of CRPS indicate a more accurate prediction. The raw hydrological simulation (which is a single value) does not provide any information about uncertainty. Therefore, we do not apply the CRPS to the raw hydrological simulation. We use a quantile-quantile (Q-Q) plot to compare the distribution of the simulated prediction residual with a standard normal distribution. If most points in a Q-Q plot align with a straight line, the test prediction is reliable.

3. CASE STUDY

We simulate daily streamflow by the GR4J model (Perrin et al., 2003) and apply the two-stage uncertainty estimation to four catchments in Southeast Australian (Figure 1), from a range of climatic and hydrological conditions (Table 1). GR4J is a 4-parameter unit hydrograph model that simulates daily streamflows from daily rainfall and potential evaporation (PE). Daily rainfall and PE are aggregated to catchment areas from the Australian water availability project (AWAP) gridded dataset (Jones et al., 2009). We use data from 1980 to 2004 for all catchments. Missing streamflow data occur in all catchments (Table 1). As a small fraction of data is missing, we simply omit the missing data from calculating the likelihood function.

We calibrate GR4J with the Shuffled Complex Evolution (SCE) algorithm (Duan et al., 1994), using NSE as the objective function. The data from the first year (i.e. 1980) is used to warm up the GR4J model and the remaining data are used for calibration.

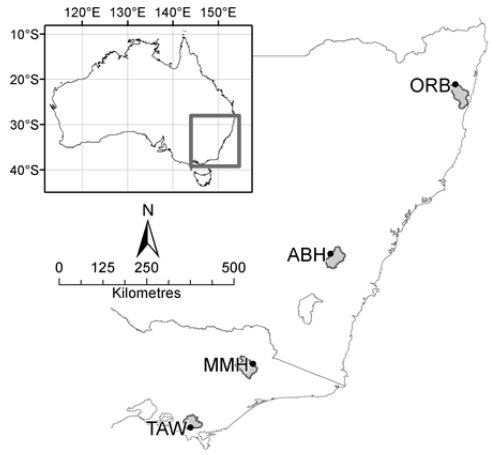

Figure 1: Catchment location

Table 1: Catchment characteristics

Id	Gauge Site	Area (km^2)	Rainfall (mm/yr)	Streamflow (mm/yr)	Runoff coefficient	Zero flows	Missing values
ABH	Abercrombie River at Hadley no. 2	1447	804	96	0.12	10.1%	1.1%
MMH	Mitta Mitta River at Hinnomunjie	1527	1307	262	0.20	0	2.5%
ORB	Orara River at Bawden Bridge	1868	1310	352	0.27	0.4%	0.7%
TAM	Tarwin River at Meeniyan	1066	1070	210	0.20	0	3.8%

The error model parameters estimated for each catchment are shown in Figure 2. All parameters are seasonally dependent. Most estimated values of ρ are greater than 0.5 (except for ABH) and are generally higher in dry months than in wet months. This confirms that the daily streamflows in these catchments are highly autocorrelated, especially in dry months. The parameters of σ and τ represent the standard deviation of the residual at Stage 1 and Stage 2, respectively. The estimate of τ is less than that of σ at each individual month. This suggests that Stage 2 can explain the prediction error that Stage 1 cannot. The model residual is generally reduced when the component of error updating is introduced at Stage 2.

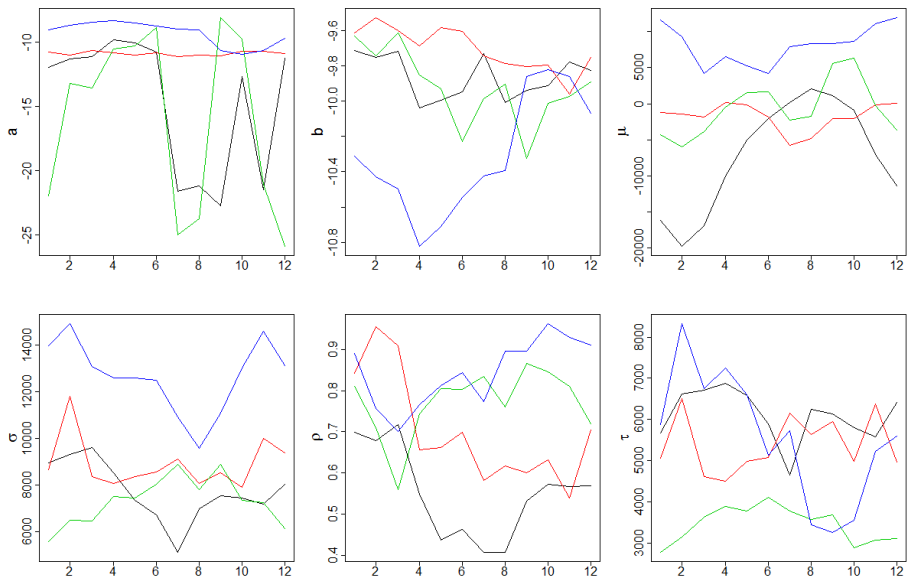

Figure 2: Error model parameters for each catchment: ABH (black), MMH (red), ORB (green) and TAM (blue).

Figure 3 compares the evaluation statistics calculated from each stage. The NSE of the raw GR4J simulation is at least 0.80 for the ABH, ORB and TAM catchments, while it is 0.72 for the MMH catchment. This implies that GR4J simulates daily streamflow reasonably well for all four catchments. When the error model at Stage 1 is applied, the NSE is slightly increased for the MMH catchment from 0.71 to 0.72 but is slightly decreased for the other catchments. The simple bias-correction in transform domain results in little or no improvement in the model accuracy (NSE) at least in part because the NSE is used as an objective function in calibration. The NSE from Stage 2 is consistently greater than that from Stage 1 and is higher than NSE for the raw GR4J simulation at MMH and TAM catchments. With the information from the previous time step, Stage 2 can improve the accuracy of model prediction particularly for the catchments where the raw simulation from the hydrological model is not good. There is little room to make such an improvement if the hydrological model already works very well. The improvement of model accuracy is clearer in the comparison of CRPS. Instead of only considering the prediction mean in NSE, CRPS compares the model accuracy in terms of probabilistic prediction. In addition, CRPS does not square the error term, as happens in NSE. This means that NSE can emphasise a few, large errors, which typically occur at high flows. CRPS gives a broader indication of model performance at a range of flows, both small and large. The error updating at Stage 2 consistently improves the accuracy of probabilistic prediction made from Stage 1. More improvement happens in wet catchments, such as the ORB catchment.

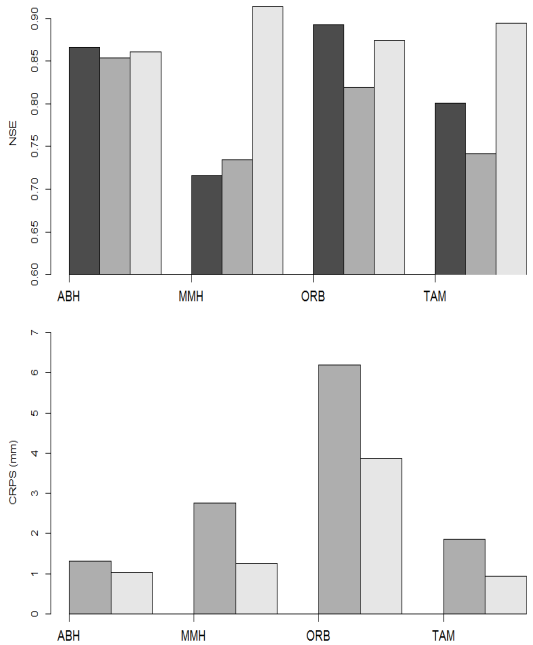

Figure 3: Comparison of NSE (top panel) and CRPS (bottom panel) for raw GR4J simulation (dark), Stage 1 (dark grey) and Stage 2 (light grey). Higher values of NSE indicate better performance, while lower CRPS values indicate better performance.

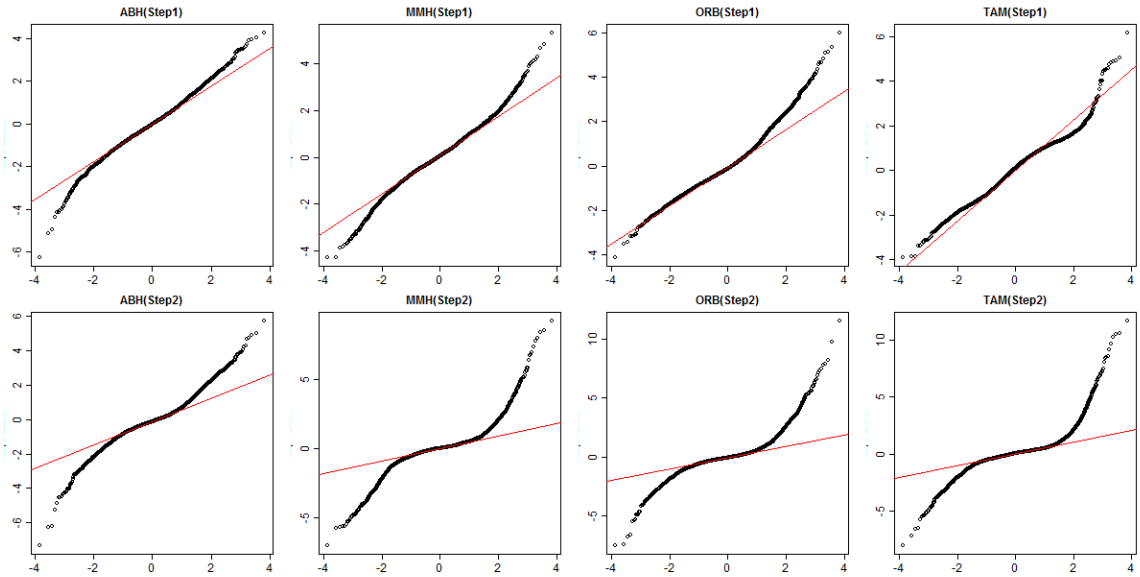

Figure 4: The quantile-quantile (Q-Q) plots of estimated model residual at Stage 1 and Stage 2.

We present the Q-Q plots of estimated model residuals from each stage in Figure 3 to investigate the reliability of uncertainty estimation. The middle part of Q-Q plots (e.g. from -2 to 2) from both stages approximately lie on a straight line. Stage 1 provides reliable uncertainty estimation for low flow, while Stage 2 tends to underestimate the occurrence of low flow. Both stages underestimate the occurrence of high

flows, and this tendency is particularly strong in the Stage 2 model. The distribution of estimated model residuals from Stage 1 is close to a standard normal distribution except for high flow. The tails (both upper and lower tails) of the residual distribution from Stage 2 are longer than that of a Gaussian distribution. This suggests that the Gaussian distribution does not adequately describe the residuals, and an alternative distribution must be sought. We will identify and test more appropriate distributions in our future research.

4. CONCLUSION AND DISCUSSION

In this study we introduce a two-stage uncertainty estimation to quantify the prediction uncertainty from a hydrological model. Stage 1 corrects the prediction bias in the log-sinh transform domain and Stage 2 updates the error in the original domain. The residuals from both stages are assumed to be Gaussian. Stage 1 only uses rainfall information from the present time step and can be applied without Stage 2, while Stage 2 requires more streamflow data from the previous time step. A case study based on four catchments in Southeast Australia is performed to assess the model performance. The overall prediction accuracy is progressively improved from Stage 1 to Stage 2.

The reliability of the uncertainty estimation from Stage 1 is generally acceptable. However, we show that uncertainty after including error updating is not reliably estimated. The use of seasonally dependent parameters and a log-sinh transform is not enough to solve this issue. After applying the error model of Stage 2, many prediction residuals have been largely reduced but a number of large prediction residuals still remain. The estimated model residuals from low and high flows are not well approximated by a Gaussian distribution and the occurrence of low and high flows tends to be underestimated. One possible way to improve reliability is to adopt a more sophisticated distribution to represent the prediction residual. This will be the focus of our future research.

ACKNOWLEDGMENTS

This work is fully funded by a strategic project of Water for a Healthy Country Flagship.

REFERENCES

Beven, K.J. and Binley, A.M. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, 6, 279-298, doi: 10.1002/hyp.3360060305.

Diskin, M.H. and Simon, E. (1977), A procedure for the selection of objective functions for hydrologic simulation models, *Journal of Hydrology*, 34(1/2), 129-149, doi: 10.1016/0022-1694(77)90066-X.

Duan, Q., Sorooshian, S. and Gupta, V. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology* 158, 265-284, doi:10.1016/0022-1694(94)90057-4.

Freer, J., Beven, K., and Ambroise, B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resources Research*, 32(7), 2161-2173, doi: 10.1029/95WR03723.

Gneiting, T., Raftery, A.E., Westveld, A.H. and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 33, 1098-1118, doi: 10.1175/MWR2904.1.

Jones, D.A., Wang, W. and Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, *58*, 233-248.

Kavetski, D., Kuczera, G. and Franks, S. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resources Research*, 42, W03407, doi:10.1029/2005WR004368.

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *Journal of Hydrology*, 331(1-2), 161-177, doi: 10.1016/j.jhydrol.2006.05.010.

Li, M. Wang, Q.J. and Bennett J.C. (2013). Accounting for seasonal dependency in hydrological model errors and prediction uncertainty. *Water Resources Research,* 49, doi: 10.1002/wrcr.20445.

Nash, J.E. and Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, 10 (3), 282-290, doi: 10.1016/0022-1694(70)90255-6.

Perrin, C., Michel, C. and Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4), 275-289, doi:10.1016/S0022-1694(03)00225-7.

Schoups, G. and Vrugt, J.A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors, *Water Resources Research*, 46, W10531, doi:10.1029/2009WR008933.

Thiemann, M., Trosset, M., Gupta, H. and Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrologic models, *Water Resources Research* 37(10), 2521-2535, doi:10.1029/2000WR900405.

Wang, Q.J. and Robertson, D.E. (2011). Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, 47, W02546, doi:10.1029/2010WR009333.

Wang, Q.J., Shrestha, D.L., Robertson, D.E. and Pokhrel, P. (2012). A log-sinh transformation for data normalisation and variance stabilisation. *Water Resources Research*, 48, W05514, doi:10.1029/2011WR010973.