

Better use of prior information in the calibration of river system models

J. Lerat^a, Z. Paydar^a, B. Henderson^b, M. Stenson^c, A. J. M. Van Dijk^a

^aCSIRO Land and Water, Clunies Ross Street, Black Mountain ACT, 2601, Australia

^bCSIRO ICT Centre & CMIS - Canberra Computer Science and Information Technology Building,
North Road, Acton ACT 2601, Australia

^cCSIRO EcoSciences Precinct - Dutton Park, 41 Boggo Road, Dutton Park QLD 4102
Email: julien.lerat@csiro.au

Abstract:

The Australian Water Resources Assessment (AWRA) modelling framework is jointly developed by the Bureau of Meteorology (BOM) and the CSIRO to support the production of the National Water Resources Assessment and the annual National Water Accounts (BOM, 2009). In this context, AWRA models are run over the recent years in hindcast mode to provide estimates of certain terms of the water accounts. This paper is the first step towards a parameterisation strategy for the river system component of AWRA. This component will be used in the accounts to estimate ungauged lateral inflows to the river, transmission losses and storage in the river channel. The model considered in this paper is a simplified river system model coupling a rainfall-runoff, a routing model and a transmission loss model. Management components (e.g. diversions, storage in reservoirs) have not been included in the model based on the assumption that they could be estimated from measured data collected by the BOM.

The calibration of this model against observed data across the Australian continent could be problematic with many areas remaining poorly gauged. The traditional approach based on the minimization of an objective function computed from local data (e.g. the Nash-Sutcliffe efficiency) may be challenged in data scarce environments where only a limited number of data values are available, leading to a loosely constrained optimisation and poorly identifiable parameter sets.

To investigate this question, a simplified daily river system model combining a rainfall-runoff and a routing model was applied on 55 river reaches across Australia. A Bayesian inference scheme was used to derive the posterior distribution of the model parameters. Different priors were used to shift from non-informative to informative priors. The predictive uncertainty obtained with different priors was finally compared using deterministic (root mean squared error, root mean squared error on log transformed flow) and probabilistic scores (continuous rank probability score and reliability).

The chief conclusion from the paper is the importance of the use of a-priori information in the calibration of hydrological models. Calibration data contain error which can introduce bias in the model calibration. An inference scheme constrained with prior knowledge can potentially mitigate this problem. In our study, it led to parameters with better extrapolation capacity for half of the catchments considered. However, the model performance was degraded on the other half. Further work is required to determine the reasons for this counter-performance and better formulate the prior for these catchments.

Keywords: River system model, rainfall-runoff model, calibration, Bayesian inference, Markov chain Monte Carlo, Uncertainty, Australian Water Resources Assessment

1 INTRODUCTION

River system models are frequently used to compare water sharing scenarios in regulated catchments in terms of the impacts for the water users and the environment. These models include a representation of the physical system with modules such as flow routing, rainfall-runoff modelling or groundwater/surface water interactions, as well as management components to take into account infrastructures such as dams, canals or extraction points (Simons *et al.*, 1996). This paper presents a method to calibrate the rainfall-runoff and routing components of a daily river system model in a Bayesian framework combining local information (likelihood) and regional information (informative prior). The work was done as part of the development of the simplified river system component of the Australian Water Resources Assessment system (AWRA), jointly developed by the Bureau of Meteorology and the CSIRO to underpin water accounting and water resources assessment (BOM, 2009).

The model used in this paper combines a lumped rainfall-runoff, a routing model and a linear transmission loss. It is applied to a river reach delimited by several upstream gauging stations and one downstream station as shown in Figure 1. More precisely, the model routes the flow from P upstream catchments ($\{q_u\}_{u=1..P}$) and generates runoff from the intermediate contributing area. The downstream flow is computed as the sum of the components according to the following equation

$$\hat{q}_d = \mathbf{f}(\theta_f, P, E) + \sum_{u=1}^P \mathbf{g}(\theta_g, q_u) \quad (1)$$

where \mathbf{g} is the routing model, θ_g are the routing parameters, \mathbf{f} is the rainfall-runoff model, θ_f are the rainfall-runoff parameters, P and E are the rainfall and potential evapotranspiration inputs on the intermediate contributing area (mm/d) and \hat{q}_d is the streamflow computed at the downstream end of the river reach (m^3/s).

In the context of the AWRA system, the chief requirement for this model is to be applicable across the Australian continent. Like any other hydrological models, its parameters largely benefit from calibration against observed data (Oudin *et al.*, 2008), which can be problematic when working in a poorly gauged areas. The traditional approach based on the minimisation of a standard objective such as the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) may be challenged in such environments where the objective function is computed from a limited number of data values, leading to a loosely constrained optimisation. In a Bayesian approach, the calibration exercise can be formalised using a probabilistic error model such as

$$y_i = \hat{y}_i(x, \theta) + \epsilon_i \quad (2)$$

where y_i and $\hat{y}_i(x, \theta)$ are the observed and modelled data at time step i respectively, $\theta = \{\theta_f, \theta_g\}$ are the model parameters and $x = \{q_u, P, E\}$ are the model inputs respectively, and ϵ_i is the model residual which is assumed to follow a given probability density (e.g. Gaussian). Finally, the posterior density of the parameters θ given the data (y and x) can be obtained from the Bayes rule leading to

$$P(\theta|x, y) = \mathcal{L}(\theta|x, y)\mathcal{M}(\theta) \quad (3)$$

where \mathcal{L} is the likelihood of the parameters and \mathcal{M} is the prior density. The likelihood function is driven by the data and the assumption made regarding the probability density of the residuals. For example, the sum of squared errors used in the Nash-Sutcliffe efficiency relies on the assumption of independent and Gaussian residuals. The prior density \mathcal{M} introduces a fundamental difference between Bayesian

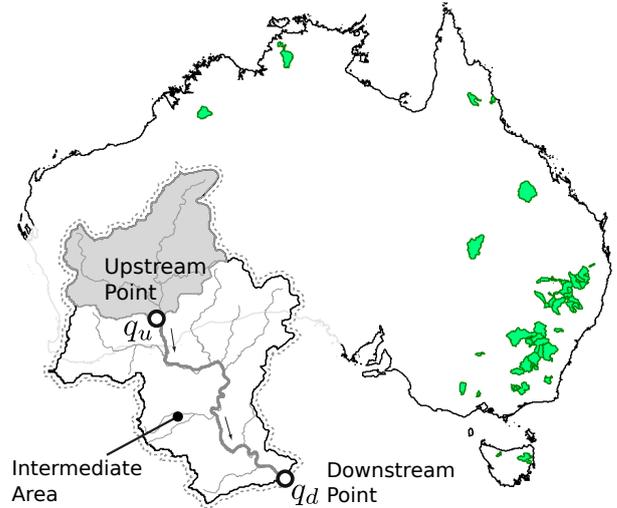


Figure 1. Location of the 55 river reaches and schematic representation of the modelled area with a single upstream point ($P = 1$)

inference and conventional calibration because it contains information that is independent from the data. In the context of AWRA, the use of prior knowledge offers an opportunity to compensate for the lack of data and to increase the identifiability of the parameters. In the hydrological literature, prior knowledge has been defined as either non-informative or subjective based on the authors' best guess. For example, Kavetski *et al.* (2006) and Ajami *et al.* (2007) affect vague priors on the inputs and non-informative priors for the parameters. Wu *et al.* (2010) gave more importance to the priors by using log-normal distributions with a subjective covariance matrix. In this paper, a different approach is proposed by using a parameters library obtained from a previous calibration study (Vaze *et al.*, 2011) where the authors applied six rainfall-runoff models to a large number of catchments in Australia. The calibrated parameter sets constitute a rich knowledge base which can be exploited to define a prior. The main objective of this paper is to compare different options for the definition of an informative prior distribution based this ensemble of parameters.

The data and models are described in section 2.1 and 2.2 respectively. The inference scheme is presented in section 2.3. Finally, the results are detailed in section 3 and discussed in section 4.

2 METHOD

2.1 Data

The river system model was applied to 55 river reaches located across Australia as shown in Figure 1. Reaches were selected in order to avoid areas with important extractions for consumptive use. For each river reach, daily time-series of spatially averaged rainfall and potential evapotranspiration, as well as upstream and downstream flow were collected for a period of 16 years. The climate data were extracted from the BILO database provided by the Bureau of Meteorology. The streamflow data were obtained from the Government of the Northern Territories, the Queensland Department of Environment and Resource Management, the New-South Wales Office for Water and the Tasmania Department of Primary Industries, Parks and Environment.

2.2 Model

The continuous lumped rainfall-runoff model GR4J was used to estimate the runoff for the lateral contributing area (Perrin *et al.*, 2003). The GR4J model has four calibration parameters: the capacity of the production store S (mm), the intercatchment groundwater exchange coefficient IGF (mm), the capacity of the non-linear routing store R (mm) and the unit hydrograph time base TB (days). The model was selected because its parsimonious structure facilitates the application of Bayesian methods. In the future, alternatives such as the AWRA-L landscape model (Van Dijk, 2010) used by other components of the AWRA system will be considered.

The lag and route propagation model combined with a linear transmission loss model was used to route the upstream flow to the downstream end of the reach. With this model, the model g in Equation 1 is the solution of the linear equation

$$D K \frac{dq}{dt} + q = L q_u (t - D \delta) \quad (4)$$

where q is the routed flow (m^3/s), q_u is the observed upstream flow (m^3/s), D is the length of the river reach (m), δ and K are the two parameters of the routing model that are expressed in s/m. The coefficient L represents the transmission loss. It is dimensionless and bounded between $[0, 1]$. In the case of multiple inflows ($P > 1$ in Equation 1), the same set of parameters δ , K and L is applied to each inflow. The model has seven free parameters: the four GR4J parameters with $\theta_f = \{S, IGF, R, TB\}$ following the notation of Equation 1, the two routing parameters and the loss coefficient with $\theta_g = \{\delta, K, L\}$.

2.3 Model calibration in a Bayesian framework

The parameters of the model were estimated with a Bayesian approach combining a likelihood function and a prior distribution to obtain their posterior distribution according to Equation 3. The error model used to obtain the likelihood function is defined by Equation 2. A transformation combining a square root

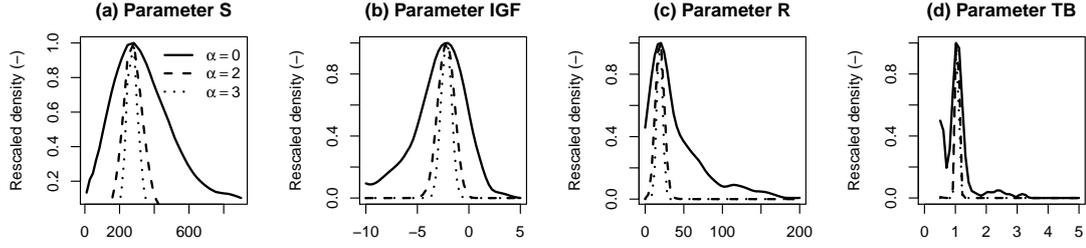


Figure 2. Marginal prior densities $\mathcal{N}(\theta_f)$ obtained with $\alpha = 0, 2, 3$

transform and an auto-regressive process was applied on both the observed and simulated flow data :

$$y_i = \sqrt{q_i} - 0.7\sqrt{q_{i-1}} \quad (5)$$

where q_i is the original flow data and y_i is the transformed data. After applying this transformation, the residuals $\epsilon_i = y_i - \hat{y}_i(x, \theta)$ obtained from Equation 2 were assumed to be independent and to follow a Laplace distribution with mean 0 and scale parameter σ resulting in

$$P(\epsilon_i) = \frac{1}{2\sigma} \exp\left(-\frac{|\epsilon_i|}{\sigma}\right) \quad (6)$$

The lag-1 autocorrelation and the fit between the proposed and empirical distribution of the residuals was checked to validate the error model. These tests are not presented here due to a lack of space. Further work is required to investigate the impact of the transformation presented in Equation 5 (e.g. allowing the coefficient 0.7 to be variable). Using the error model detailed in Equation 2 and non-informative prior for the scale parameter σ the posterior distribution of the model parameters is given by (Box and Tiao, 1992)

$$P(\theta|y, x) \propto \frac{1}{\sigma^{n+1}} \exp\left(-\sqrt{2}\frac{\sum_i |\epsilon_i|}{\sigma}\right) \mathcal{M}(\theta) \quad (7)$$

with \mathcal{M} the prior density of the parameters and n the number of time steps. In total, height parameters are inferred in this scheme: seven parameters for the hydrological model (see section 2.2) and one parameter for the error model (σ).

Assuming independent prior distributions for the routing and rainfall-runoff parameters, the prior \mathcal{M} was decomposed into two components: $\mathcal{M}(\theta) = \mathcal{N}(\theta_f)\mathcal{O}(\theta_g)$. In all simulations, the prior for the routing parameters was considered as non-informative with $\mathcal{O}(\theta_g) \equiv 1$. Two options are compared for the definition of the prior for the rainfall-runoff parameters. In a first approach, used as a benchmark, \mathcal{N} was considered as non-informative with $\mathcal{N}(\theta_f) \equiv 1$. Alternatively, an informative prior was defined using the parameter library obtained from a calibration exercise performed by Vaze *et al.* (2011) on an independent data set of 240 catchments. This study provided an ensemble of 240 parameter sets for the GR4J model. The density corresponding to this ensemble was estimated with a kernel smoothing algorithm. Finally, the informative prior for the rainfall-runoff parameters was computed as

$$\mathcal{N}(\theta_f) = [\mathcal{KS}(\theta_f)]^{\exp(\alpha)} \quad (8)$$

with \mathcal{KS} the kernel smoothing density and $\alpha \geq 0$ a coefficient that controls the spread of the prior. If $\alpha = 0$, the prior is identical to the kernel smoothing estimate. In this case, the prior is expected to be rather flat because it was estimated from a large ensemble of parameters, resulting in a small impact on the posterior. Conversely, positive values of the α coefficient tend to concentrate the prior around its mode, which increases its impact on the posterior and reduce the impact of the likelihood (and hence of the impact of local data). Figure 2 shows the evolution of the prior \mathcal{N} with $\alpha = 0, 2, 3$ for GR4J parameters. In short, the α coefficient determines the weight of the informative prior in the posterior.

Using the priors defined previously, the simulations were performed by splitting the total simulation period (16 years, see section 2.1) into two equal sub-periods. For each sub-period, the posterior was sampled with a Markov chain Monte Carlo technique using the metropolis algorithm and five chains. Each chain was run 3000 times. The last 100 parameters of each chain were retained, making a total of 500 parameter sets, to generate probabilistic streamflow simulations. An example of simulation is presented in Figure 3.

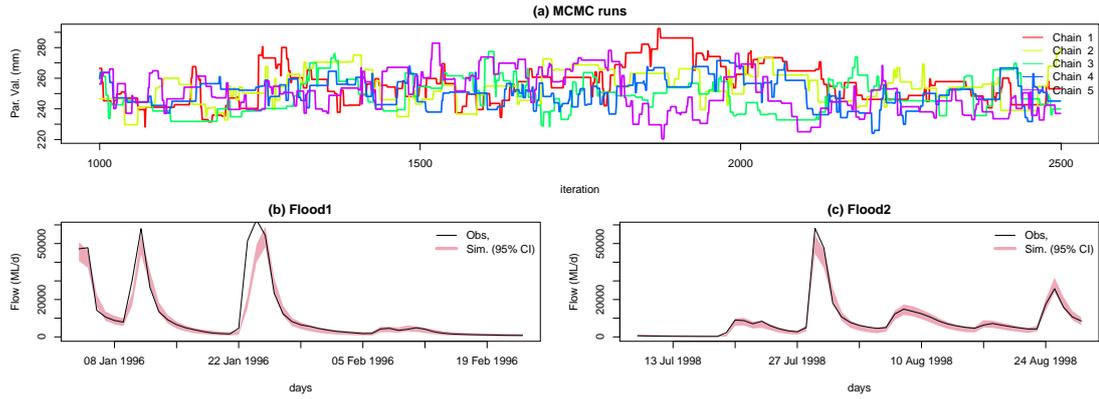


Figure 3. MCMC runs (a) and probabilistic simulations (b,c) in the Border river region (stations 416007)

2.4 Relative performance scores

The comparison between the simulations obtained with the non-informative and informative priors was based on relative performance scores (Nash and Sutcliffe, 1970)

$$R_{\mathcal{D}}(M|B) = 1 - \frac{\mathcal{D}(q, \hat{q}_M)}{\mathcal{D}(q, \hat{q}_B)} \quad (9)$$

where q is the observed streamflow at the downstream end of the reach, \hat{q}_B and \hat{q}_M are two probabilistic streamflow simulations obtained with non-informative prior (benchmark) and informative prior respectively, \mathcal{D} is a metric measuring the discrepancies between simulated and observed streamflow time series. $R_{\mathcal{D}}$ varies between $-\infty$ and 1. If $R_{\mathcal{D}} > 0$, the simulation using the informative prior performs better. Four metrics \mathcal{D} were selected to compute the performance indices. The first two metrics are deterministic and qualify the fit between the median simulation and the observed data on high flows (*rmse*) and low flows (*rmse_{log}*) respectively. The third and fourth metrics are probabilistic and qualify the position and spread of the probabilistic simulation. The third metric is the Continuous Rank Probability Score (CRPS Hersbach, 2000). The fourth metric is the reliability component (named *reliab*) of the CRPS proposed by Hersbach (2000). The first three metrics can be written

$$rmse(q, \hat{q}) = \sqrt{\frac{1}{n} \sum_i [\text{median}(\hat{q}_i) - q_i]^2} \quad (10)$$

$$rmse_{log}(q, \hat{q}) = \sqrt{\frac{1}{n} \sum_i [\log(\text{median}(\hat{q}_i)) - \log(q_i)]^2} \quad (11)$$

$$crps(q, \hat{q}) = \frac{1}{n} \sum_i \int_{-\infty}^{+\infty} [P(\nu \geq \hat{q}_i) - \mathcal{H}(\nu - q_i)]^2 d\nu \quad (12)$$

with n the number of time steps, $\text{median}(\hat{q}_i)$ the median simulation at time step i (m^3/s), ν a dummy variable (m^3/s), $P(\nu \geq \hat{q}_i)$ the cumulative density of the simulated flow at time step i , \mathcal{H} the Heaviside function. The reader is referred to Hersbach (2000) for the expression of the fourth metric.

3 RESULTS

The distributions of the four relative performance scores R_{rmse} , $R_{rmse_{log}}$, R_{crps} and R_{reliab} are presented in Figure 4 and can be commented as follows:

- All the distributions appear symmetric with a median equal to 0. As a result, the informative and non-informative priors lead to equivalent performance over the whole dataset with an equal proportion of positive (informative prior performs best) and negative (non-informative prior performs best) scores.

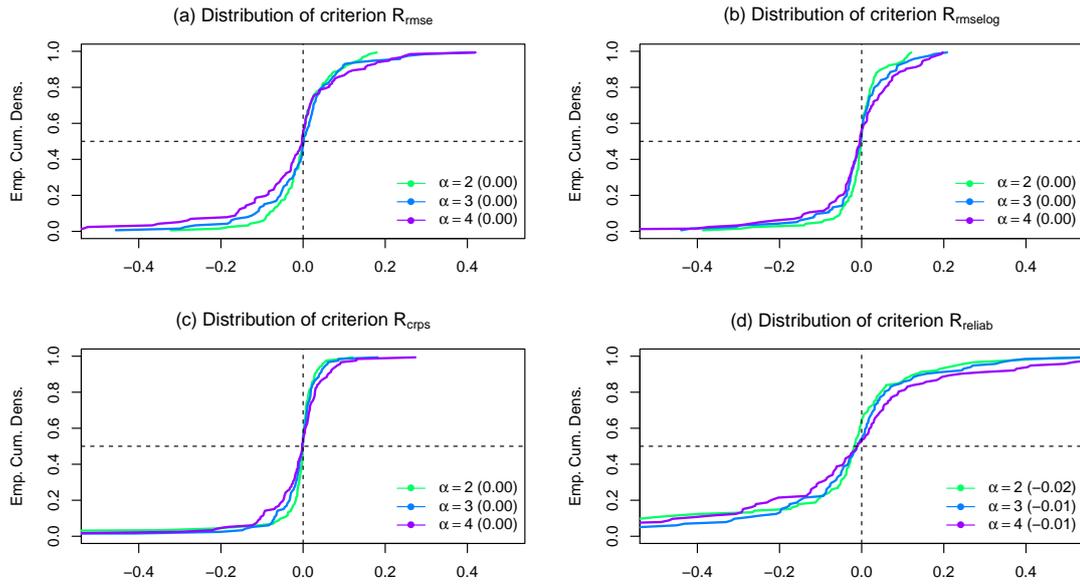


Figure 4. Empirical cumulative distribution (Y axis) of the four relative performance scores R_{rmse} , $R_{rmselog}$, R_{crps} and R_{reliab} (X axis, see Equation 9 to 12) in validation mode over 8-year-sub-periods and 55 river reaches. The scores compare the performance of a model calibrated with a non-informative prior and an informative prior \mathcal{N} (see Equation 8) with $\alpha = 2$ (green), 3 (blue) and 4 (purple). The bracketed numbers are the median values of the distributions.

- Smaller values of α result in performance scores that remain close to 0, with distributions concentrated around the vertical line $R = 0$. This behaviour was expected because low values of α keep the prior close to the kernel smoothing estimate, which is in turn close to non-informativeness. The prior starts to have a strong impact on the performance for a value of α between 2 and 3.
- A value of $\alpha = 4$ (purple curve) tends to degrade significantly the performance on some catchments with highly negative scores. As a result, it is recommended to avoid α values significantly greater than 4.
- Similar trends are observed for the deterministic (R_{rmse} , R_{rmse}) and probabilistic scores (R_{crps} , R_{reliab}).

4 DISCUSSION

The previous results demonstrate that prior knowledge can have a strong impact on the calibration process. An improvement of performance compared to non-informative prior was obtained for half of the sample and performance was degraded on the other half, which does not justify the use of the informative prior defined in section 2.3 for operational applications. However, the results are encouraging considering the extensive validation scheme (four criteria, validation mode and large sample of catchments) and the simple nature of the prior, which was identical for all the catchments. The definition of catchment specific priors, e.g. based on catchment descriptors or nearest neighbours, constitutes the main direction for further research.

The concentrated nature of the likelihood remains the main challenge to define an informative prior. In a typical hydrological calibration, the argument of the exponential in Equation 7 can vary by order of magnitude of 100 which results in a likelihood being close to a Dirac δ function. As a result, an informative prior has to be concentrated in similar way to balance the likelihood and influence the posterior. This is the reason why an exponent defined as $\exp(\alpha)$ with $\alpha = 2, 3, 4$ was needed in Equation 8 (see Figure 2). This solution is not optimal because it dampens the secondary modes of the prior as can be seen on Figure 2.d. Alternative formulations are currently investigated.

Finally, the calibration/validation exercise was conducted here with two sub-periods of equal length (8 years). With a reduced calibration period, the likelihood function may not be able to extract sufficient information from the data. The use of an informative prior could become more beneficial in this case.

5 CONCLUSION

This paper presented a method to calibrate a simplified daily river system model combining a rainfall-runoff and a routing model. The calibration was performed in a Bayesian inference framework combining local data and a priori knowledge. The a priori knowledge was derived from the empirical distribution of the parameters obtained from a previous calibration study. The performance of the calibrated models was assessed by computing deterministic and probabilistic scores in a split-sample test approach. Finally, the whole methodology was applied on 55 river reaches located on the Eastern part of the Australian continent using data covering a 16 years period.

The chief conclusion from the paper is the importance of the use of a-prior information in the calibration of hydrological models. Local data are far from error-free and can introduce large uncertainty in the model calibration. An inference scheme that is constrained with prior knowledge lead to parameters with better extrapolation capacity for half of the catchments considered. However, the model performance was degraded on the other half. Further work is required to determine the reasons for this counter-performance and better formulate the prior for these catchments.

ACKNOWLEDGEMENTS

This work was funded by the WIRADA research alliance between the Bureau of Meteorology and the CSIRO. The authors acknowledge Cuan Petheram, Mick Hartcher and David Post for their help during the compilation of the streamflow data. The comments of two anonymous reviewers are also acknowledged.

REFERENCES

- Ajami, N. K., Q. Y. Duan, and S. Sorooshian (2007). An integrated hydrologic bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour. Res.* 43(1), W01403.
- BOM (2009). Regional water balance: Initial methods review 2009. Technical report, Commonwealth of Australia, Melbourne.
- Box, G. and G. Tiao (1992). *Bayesian inference in statistical analysis*. Wiley classics library. Wiley.
- Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* 15(5), 559–570.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. theory. *Water Resour. Res.* 42, W03 407.
- Nash, J. E. and J. V. Sutcliffe (1970). River flow forecasting through conceptual models. part 1: a discussion of principles. *J. Hydrol.* 10(3), 282–290.
- Oudin, L., V. Andréassian, J. Lerat, and C. Michel (2008). Has land cover a significant impact on mean annual streamflow? An international assessment using 1508 catchments. *J. Hydrol.* 357(3-4), 303–316.
- Perrin, C., C. Michel, and V. Andréassian (2003). Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* 279(1-4), 275–289.
- Simons, M., G. Podger, and R. Cooke (1996). IQQM—A hydrologic modelling tool for water resource and salinity management. *Environmental Software* 11(1-3), 185 – 192.
- Van Dijk, A. (2010). The Australian Water Resources Assessment System, Technical Report 3. Landscape Model (version 0.5) Technical Description. Technical report, CSIRO: Water for a Healthy Country National Research Flagship.
- Vaze, J., F. H. S. Chiew, J. Perraud, N. Viney, D. A. Post, J. Teng, B. Wang, J. Lerat, and M. Goswami (2011). Rainfall-runoff modelling across southeast Australia: datasets, models and results. *Australian Journal of Water Resources* 14 (2), 101–116.
- Wu, W., J. Clark, and J. Vose (2010). Assimilating multi-source uncertainties of a parsimonious conceptual hydrological model using hierarchical bayesian modeling. *J. Hydrol.* 394(3-4), 436–446.