David McInerney, Mark Thyer, Dmitri Kavetski, Bree Bennett, Julien Lerat, Matthew Gibbs, George Kuczera

A simplified approach to produce probabilistic hydrological model predictions

Environmental Modelling & Software, 2018; 109:306-314

Final publication at http://dx.doi.org/10.1016/j.envsoft.2018.07.001

A simplified approach to produce probabilistic hydrological model predictions

David McInerney[(1)], Mark Thyer[(1)], Dmitri Kavetski[(1)],
Bree Bennett[(1)], Julien Lerat[(2)], Matthew Gibbs[(3)] and George Kuczera[(4)]

(1) School of Civil, Environmental and Mining Engineering, University of Adelaide, SA, Australia
(2) Bureau of Meteorology, Canberra, ACT, Australia
(3) Department of Environment, Water and Natural Resources, Government of South Australia, Australia
(4) School of Engineering, University of Newcastle, Callaghan, NSW, Australia

Abstract

Probabilistic predictions from hydrological models, including rainfall-runoff models, provide valuable information for water and environmental resource risk management. However, traditional "deterministic" usage of rainfall-runoff models remains prevalent in practical applications, in many cases because probabilistic predictions are perceived to be difficult to generate. This paper introduces a simplified approach for hydrological model inference and prediction that bridges the practical gap between "deterministic" and "probabilistic" techniques. This approach combines the Least Squares (LS) technique for calibrating hydrological model parameters with a simple method-of-moments (MoM) estimator of error model parameters (here, the variance and lag-1 autocorrelation of residual errors). A case study using two conceptual hydrological models shows that the LS-MoM approach achieves probabilistic predictions with similar predictive performance to classical maximum-likelihood and Bayesian approaches, but is simpler to implement using common hydrological software and has a lower computational cost. A public web-app to help users implement the simplified approach is available.

Keywords: probabilistic prediction, rainfall-runoff modelling, method of moments, maximum likelihood

Highlights

- New simplified approach for producing probabilistic hydrological predictions
- Similar performance to maximum-likelihood approach, at lower computational cost
- Web-app available to facilitate uptake of probabilistic predictions

Software availability

Product title: Interactive Probabilistic Predictions

Description: Web application for implementing Stage 2 of the LS-MoM approach introduced in this study

Developer: David McInerney, Bree Bennett, Mark Thyer, Dmitri Kavetski

Contact Address: David McInerney, School of Civil, Environmental and Mining Engineering, University of Adelaide, SA, Australia

Contact Email: david.mcinerney@adelaide.edu.au

Software Required: Web browser supported by R Shiny Server (Google Chrome, Mozilla Firefox, Safari)

Available Since: September 2017

Availability: http://www.probabilisticpredictions.org

1. Introduction

Predictions from hydrological models, particularly rainfall-runoff models, provide essential inputs to the planning and operation of water resource systems (Loucks et al., 1981). Probabilistic inference and prediction approaches, where probability models are used to describe data and model uncertainty, are of particular interest to enable uncertainty quantification and risk assessment (Vogel, 2017). Probabilistic techniques are well-known in the hydrological research community and include method-of-moments (MoM), maximum-likelihood (ML) and Bayesian techniques (e.g., Salas, 1993, Martins and Stedinger, 2000), with rainfall-runoff model applications typically employing Bayesian techniques (e.g., Kuczera, 1983, Krzysztofowicz, 2002, Schoups and Vrugt, 2010, Smith et al., 2010, Li et al., 2016, McInerney et al., 2017, Kavetski, 2018). Maximum-likelihood and Bayesian techniques require the specification of a likelihood function, which in rainfall-runoff modelling is typically derived from a residual error model, such as the widely used independent Gaussian error model. In most cases, residual error models include calibrated parameters of their own, such as error variance, lag-1 autocorrelation, and so forth.

In contrast to the research literature, practical hydrological modelling applications tend to rely on "deterministic" approaches, e.g., where rainfall-runoff models are calibrated using goodness-of-fit objective functions and quantification of uncertainty in predictions is typically considered the domain of applied research (Vaze et al., 2012). Least Squares (LS) objective functions (e.g., the sum-of-squared-errors (SSE) and equivalent Nash-Sutcliffe efficiency (NSE)) are widely used in research and practice; they are computed directly or from transformed flows (Chapman, 1970, Chiew et al., 1993, Oudin et al., 2006, Pushpalatha et al., 2012). Many hydrological modelling and calibration platforms implement LS objective functions. For example, the popular calibration package PEST supports weighted SSE (Doherty, 2004), HEC-HMS (Scharffenberg et al., 2006), the Australian "eWater Source" (Welsh et al., 2013) and HBV Light (Seibert, 2005) support log-transformed SSE (often used to better capture low flows), the Hydromad R package (Andrews et al., 2011) allows for objective functions based on Box-Cox transformed flows, and the recent airGR R package (Coron et al., 2017) provides built-in log, square-root and inverse-transformed SSE objective functions. Some of these software packages have capabilities for estimating parameter uncertainty and its impact on predictions. For example, PEST supports linear/nonlinear parameter uncertainty analysis including the null space Monte Carlo method (Tonkin and Doherty, 2009), and Hydromad implements the DREAM MCMC approach of Vrugt et al. (2009) (http://hydromad.catchment.org; see Joseph and Guillaume (2013) for an application).

The statistical modelling needed to derive the likelihood function and estimate the error model parameters creates a perception that probabilistic prediction requires substantial additional effort. For example, in the software packages listed above, it is (relatively) easy to implement new objective functions, but non-trivial to incorporate calibrated error model parameters. This perception can delay the uptake of probabilistic techniques, especially in practical applications. The motivation of this study is to develop a simplified approach that produces high-quality probabilistic rainfall-runoff model predictions at a minor additional effort beyond that required for traditional deterministic predictions.

The specific aims of this study are:

Aim 1. Develop a simplified "LS-MoM" approach to generating probabilistic hydrological predictions, exploiting a combination of Least Squares (LS) and method-of-moments (MoM) approaches;

Aim 2. Empirically compare the LS-MoM, maximum-likelihood and Bayesian approaches in terms of predictive performance and computational cost, in a case study using conceptual hydrological models;

Aim 3. Introduce a public web-app to help practitioners apply the LS-MoM approach.

The paper continues by outlining the likelihood-based framework in Section 2. The LS-MoM approach is developed in Section 3. Section 4 describes the empirical case study methods, with results reported in Section 5. Sections 6-7 discuss and summarize the key findings.

2. Likelihood-based parameter inference

2.1. Theory

A hydrological (rainfall-runoff) model, *H*, simulates streamflow $\mathbf{Q}^{\boldsymbol{\theta}_H} = \{Q_t^{\boldsymbol{\theta}_H}, t = 1,..,T\}$ over a series of time steps *t*, as a function of forcing data **X**, hydrological model parameters $\boldsymbol{\theta}_H$ and initial conditions $\mathbf{S}_0$,

$$\mathbf{Q}^{\boldsymbol{\theta}_H} = H(\boldsymbol{\theta}_H; \mathbf{X}, \mathbf{S}_0) \tag{1}$$

To estimate $\boldsymbol{\theta}_H$ from observed streamflow data $\tilde{\mathbf{Q}} = \{\tilde{Q}_t, t = 1,..,T\}$ and observed forcing data $\tilde{\mathbf{X}}$ using a maximum-likelihood approach, a likelihood function $\mathcal{L}(\boldsymbol{\theta}_H; \tilde{\mathbf{Q}})$ should be specified and maximized with respect to $\boldsymbol{\theta}_H$. The likelihood function is derived from an assumed probability model of observed data, $\mathcal{L}(\boldsymbol{\theta}_H; \tilde{\mathbf{Q}}) = p(\tilde{\mathbf{Q}} | \boldsymbol{\theta}_H, \tilde{\mathbf{X}})$, e.g., by considering the probability distribution of residual errors assumed to describe the combined contributions of all sources of predictive error (Renard et al., 2011). Residual errors of hydrological model are typically heteroscedastic (larger errors in larger flows) and persistent (similar errors several time steps in a row) (e.g., Sorooshian and Dracup, 1980). In many cases, error heteroscedasticity is represented using streamflow transformations (e.g., logarithmic or Box-Cox), and error persistence is represented using an autoregressive lag-1, AR(1), model (e.g., Sorooshian and Dracup, 1980, Evin et al., 2014). Under these assumptions, and ignoring terms at *t*=1, the (approximate) likelihood is

$$\mathcal{L}_F(\boldsymbol{\theta}_H,\boldsymbol{\theta}_Z,\boldsymbol{\theta}_\varepsilon;\tilde{\mathbf{Q}}) = p(\tilde{\mathbf{Q}}\,|\,\boldsymbol{\theta}_H,\boldsymbol{\theta}_Z,\boldsymbol{\theta}_\varepsilon,\tilde{\mathbf{X}}) = \prod_{t=2}^{T} Z'(\tilde{Q}_t;\boldsymbol{\theta}_Z)\, f_N\left(y_t(\boldsymbol{\theta}_H,\boldsymbol{\theta}_Z,\phi;\tilde{\mathbf{Q}},\tilde{\mathbf{X}});0,\sigma_y^2\right) \quad (2)$$

The terms in equation (2) are as follows:

1) $Z(Q)$ is the streamflow transformation used to describe error heteroscedasticity, and $Z' = dZ/dQ$ is its Jacobian. Here we employ the ubiquitous Box-Cox transformation (Box and Cox, 1964),

$$Z(Q;\lambda,A) = \begin{cases} \dfrac{(Q+A)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(Q+A) & \text{otherwise} \end{cases} \quad (3)$$

where λ and A are transformation parameters, grouped into $\boldsymbol{\theta}_Z$ in equation (2). When $A = 0$, the Box-Cox transformation with $\lambda = 0$, 0.5, and -1 is equivalent to the log, square-root and inverse transformations respectively.

The offset A can be non-dimensionalized by a typical streamflow magnitude, such as the mean observed flow,

$$A^* = A / \text{mean}(\tilde{\mathbf{Q}}) \quad (4)$$

2) The quantity y_t is the "error innovation" at time step t, defined from a zero-mean homoscedastic Gaussian AR(1) model of residuals of transformed streamflows,

$$\eta_t = Z(\tilde{Q}_t;\boldsymbol{\theta}_Z) - Z(Q_t^{\boldsymbol{\theta}_H};\boldsymbol{\theta}_Z) \quad (5)$$

$$y_t = \eta_t - \phi_\eta \eta_{t-1} \quad (6)$$

$$y_t \sim N(0,\sigma_y) \quad (7)$$

where $N(\mu,\sigma)$ denotes the Gaussian distribution with mean μ and variance σ^2, and probability density function (pdf) $f_N(x;\mu,\sigma^2)$. The residual error model in equations (3)-(7) has parameters $\boldsymbol{\theta}_\varepsilon = \{\phi_\eta, \sigma_y\}$, where ϕ_η is the lag-1 autoregressive parameter and σ_y is the standard deviation.

2.2. Two-stage post-processor implementation

A two-stage post-processing (PP) approach for parameter estimation is employed:

Stage 1: Calibrate hydrological and transformation parameters, $\boldsymbol{\theta}_H$ and $\boldsymbol{\theta}_Z$, neglecting error autocorrelation, i.e., maximizing the likelihood in equation (2) while fixing $\phi_\eta = 0$. The parameter σ_y is also calibrated, but then discarded in Stage 2. The transformation parameter λ can be either fixed *a priori* or calibrated (e.g., Wang et al., 2012, McInerney et al., 2017);

Stage 2: Calibrate error model parameters, $\boldsymbol{\theta}_\varepsilon = \{\phi_\eta, \sigma_y\}$, by maximizing the likelihood in equation (2) while keeping $\boldsymbol{\theta}_H$ and $\boldsymbol{\theta}_Z$ fixed at the values estimated in Stage 1. Stage 2 is computationally very fast because it works solely with observed data and optimal streamflow predictions from Stage 1, and hence does not require additional hydrological model runs.

The adopted PP approach is empirically more robust than joint calibration, because it avoids problematic interactions between hydrological and error model parameters (see Evin et al., 2014, and Supplementary Material Section S1). As both stages are implemented using maximum-likelihood, we will refer to this approach as the ML-ML approach.

When parsimonious hydrological models such as GR4J (e.g., Perrin et al., 2003) are calibrated to long observed time series using residual error models such as those in Section 4, the contribution of parametric uncertainty to total predictive uncertainty in streamflow is generally small (Kuczera et al., 2006, Yang et al., 2007, Sun et al., 2017, Kavetski, 2018). For this reason, hydrological prediction and forecasting applications tend to focus on residual errors and often ignore posterior parameter uncertainty (e.g., Engeland and Steinsland, 2014, McInerney et al., 2017). This is the strategy employed in this study, where calibration is undertaken solely through optimization of the likelihood function. The suitability of this approach is illustrated as described in Section 4.1, with limitations discussed in Section 6.4.

3. Simplified approach for parameter inference

The simplified approach has two stages that mimic those of the ML-ML approach:

Stage 1: Estimate hydrological model parameters $\boldsymbol{\theta}_H$ by Least Squares optimization (e.g., by minimizing SSE). Transformation parameters $\boldsymbol{\theta}_Z$ (if any) must be fixed *a priori*;

Stage 2: Estimate error model parameters $\boldsymbol{\theta}_\varepsilon$ from the residuals $\boldsymbol{\eta}$ using the method-of-moments.

We will refer to this approach as LS-MoM; its respective equations are presented next.

3.1. Stage 1

When the transformation parameters $\boldsymbol{\theta}_Z$ are fixed, the Jacobian term in equation (2) no longer depends on any inferred quantity, and represents a proportionality constant. With the additional assumption that $\boldsymbol{\eta}$ is uncorrelated (UC), $\phi_\eta = 0$, equation (2) reduces to

$$\mathcal{L}_{UC}(\boldsymbol{\theta}_H, \sigma_\eta; \tilde{\mathbf{Q}}, \boldsymbol{\theta}_Z) \propto \prod_{t=1}^{T} f_N\left(\eta_t(\boldsymbol{\theta}_H, \boldsymbol{\theta}_Z; \tilde{Q}_t, \tilde{\mathbf{X}}_{1:t}) \,|\, 0, \sigma_\eta^2\right) \tag{8}$$

where σ_η is the standard deviation of $\boldsymbol{\eta}$ (McInerney et al., 2017).

Expanding $f_N(x; \mu, \sigma^2)$ and taking logarithms, equation (8) can be re-written as

$$\log \mathcal{L}_{UC}(\boldsymbol{\theta}_H, \sigma_\eta; \tilde{\mathbf{Q}}, \boldsymbol{\theta}_Z) = -\Psi_1(\sigma_\eta) - \Psi_2(\sigma_\eta)\, \Phi_{SSE}(\boldsymbol{\theta}_H; \tilde{\mathbf{Q}}_t, \tilde{\mathbf{X}}_{1:t}, \boldsymbol{\theta}_Z) + \text{const} \tag{9}$$

where $\Psi_1 = T \ln(\sigma_\eta^2)/2$ and $\Psi_2(\sigma_\eta) = 1/2\sigma_\eta^2$ are functions solely of σ_η, and

$$\Phi_{SSE}(\boldsymbol{\theta}_H; \bullet) = \sum_{t=1}^{T} \eta_t(\boldsymbol{\theta}_H; \bullet)^2 \tag{10}$$

is the sum of squared errors (SSE) of transformed flows, viewed solely as a function of $\boldsymbol{\theta}_H$.

Noting that $\Psi_2 > 0$, the hydrological parameter values $\boldsymbol{\theta}_H$ that maximise $\log \mathcal{L}_{UC}$ (and hence $\mathcal{L}_{UC}$) are the same ones that minimize Φ_{SSE}. This equivalence is verified algebraically in Supplementary Material Section S2, and is well-known in the statistical literature (e.g., Charnes et al., 1976).

In other words, under the assumptions of uncorrelated Gaussian residuals and provided the transformation parameters are fixed, Stage 1 of the ML-ML approach can proceed through Least Squares optimization of transformed flows. Table 1 provides the correspondence between common objective functions and the SSE applied to Box-Cox transformed flows.

Note that, given the reduction in the number of optimized quantities in Stage 1 – which is the only stage that requires running the hydrological model – it can be expected that LS-MoM is computationally cheaper than ML-ML for a given optimization algorithm.

Table 1. Correspondence between common objective functions used in the hydrological literature and Box-Cox transformation parameters applied to SSE.

Objective function	Transformation parameters	References
Sum of squared errors (SSE) of untransformed flows Root mean squared error (RMSE) Nash Sutcliffe Efficiency (NSE)	$\lambda = 1$ and $A^* = 0$	Servat and Dezetter (1991), Gan et al. (1997), Oudin et al. (2006), Kumar et al. (2010)
NSE of square root transformed flows	$\lambda = 0.5$ and $A^* = 0$	Chapman (1970), Chiew et al. (1993), Ye et al. (1998), Perrin et al. (2003), Oudin et al. (2006), Pushpalatha et al. (2012)
NSE of log transformed flows	$\lambda = 0$ and $A^* = 0$	Dawdy and Lichty (1968), Chapman (1970), Oudin et al. (2006), Kumar et al. (2010)
Likelihood function based on log transformed flow with non-zero offset	$\lambda = 0$ and $A^* \neq 0$	Bates and Campbell (2001), Smith et al. (2010)

3.2. Stage 2

Given estimated values $\hat{\boldsymbol{\theta}}_H$ from Stage 1 and fixed values of $\boldsymbol{\theta}_Z$, the estimated residuals $\hat{\boldsymbol{\eta}}$ in equation (5) are themselves fixed. The method-of-moments can then be used to estimate the error model parameters $\hat{\boldsymbol{\theta}}_\varepsilon$ from sample statistics of $\hat{\boldsymbol{\eta}}$.

The lag-1 autoregressive parameter $\hat{\phi}_\eta$ is estimated as the sample lag-1 autocorrelation coefficient

$$\hat{\phi}_\eta = \mathrm{acorr}_{\ell=1}[\hat{\boldsymbol{\eta}}] = \frac{1}{(T-1)\, s_{\hat{\eta}}^2} \sum_{t=2}^{T} \left(\hat{\eta}_t - m_{\hat{\eta}}\right)\left(\hat{\eta}_{t-1} - m_{\hat{\eta}}\right) \tag{11}$$

where $m_{\hat{\eta}} = \mathrm{mean}[\hat{\boldsymbol{\eta}}]$ and $s_{\hat{\eta}}^2 = \mathrm{var}[\hat{\boldsymbol{\eta}}]$ denote, respectively, the sample mean and variance of $\hat{\boldsymbol{\eta}}$.

The innovation variance $\hat{\sigma}_y^2$ is estimated from the well-known relationship between conditional and marginal variances of an AR(1) process (Box and Jenkins, 1970),

$$\hat{\sigma}_\eta^2 = s_{\hat{\eta}}^2 = \mathrm{var}[\hat{\boldsymbol{\eta}}] = \frac{1}{T-1} \sum_{t=1}^{T} \left(\hat{\eta}_t - m_{\hat{\eta}}\right)^2 \tag{12}$$

$$\hat{\sigma}_y^2 = \hat{\sigma}_\eta^2 \left(1 - \hat{\phi}_\eta^2\right) \tag{13}$$

Once again, no additional hydrological model runs are required in Stage 2.

4. Case study methods

4.1. Experiments and residual error schemes

The objective of the case study is to establish if the simple LS-MoM approach (described in Section 3) is competitive with the more complex ML-ML approach (described in Section 2.2) in hydrological modelling applications. This comparison is carried out for the Box-Cox error models recommended by McInerney et al. (2017), as follows:

1) Benchmarking of the LS-MoM approach against a "well-performing" ML-ML approach:
   a) For the residual error schemes recommended by McInerney et al. (2017), namely the Log ($\lambda=0$), BC0.2 ($\lambda=0.2$) and BC0.5 ($\lambda=0.5$) schemes in perennial catchments, and the BC0.2 and BC0.5 schemes in ephemeral/low-flow catchments, we compare LS-MoM with fixed $A^*=0$ against the ML-ML approach with inferred A^* (Section 2.2). The value $A^*=0$ is of particular interest in the LS-MoM approach because it provides the closest correspondence to common objective functions (Table 1);
   b) When applying the Log scheme in ephemeral/low-flow catchments, ML-ML with inferred A^* performs poorly (McInerney et al., 2017), and LS-MoM with $A^*=0$ is not applicable. Hence, in these scenarios, we set $A^*=10^{-1}$ in both the ML-ML and LS-MoM approaches;
2) Analysis of the LS-MoM approach with $A^*=0$, 10^{-4} and 10^{-1} (with $A^*=0$ excluded when using the Log scheme in ephemeral/low-flow catchments). This experiment establishes the impact of the offset, which must be specified *a priori* in the LS-MoM approach and could potentially impact on calibration and prediction.

Given that the LS-MoM and ML-ML approaches compared in this work are set to ignore parameter uncertainty, the contribution of parameter uncertainty to total predictive uncertainty in streamflow is evaluated by comparing LS-MoM and ML-ML to two Bayesian setups, namely to a full Bayesian approach where $\boldsymbol{\theta}_H$ and $\boldsymbol{\theta}_Z$ are inferred jointly, and to a Bayesian implementation of Stage 1 from Section 2.2. The details of this comparison are reported in Supplementary Material Section S1.

4.2. Hydrological data and models

The case study setup from McInerney et al. (2017) is used, with 11 Australian catchments (http://www.bom.gov.au/water/hrs) and 12 US catchments (Duan et al., 2006). For modelling purposes, catchments are classified into two types: 11 catchments where the minimum observed flow is below 2% of the mean observed flow are referred to as "ephemeral/low-flow"; the remaining 12 catchments are termed "perennial" (see Supplementary Material Table S1 for details of the catchments). This classification was found to correlate better with probabilistic model performance than an earlier classification based on the proportion of zero flow days (McInerney et al., 2017), and is not intended as a classification from a hydrological process perspective.

Two conceptual rainfall-runoff models, GR4J (Perrin et al., 2003) and HBV (Bergström, 1995) are used. A cross-validation framework is implemented over a 10-year period (McInerney et al., 2017, Table 5) and used to produce a concatenated 10-year series of daily streamflow predictions. Predictive distributions are computed as described in Appendix A. Parameter optima are obtained from 100 quasi-Newton optimizations (Kavetski and Clark, 2010). The offset A^* is given a lower bound of 10^{-7} to avoid the Jacobian in equation (2) becoming undefined for $\tilde{Q}_t = 0$; all other bounds are taken from McInerney et al. (2017). "Typical" parameter values are obtained from a single calibration over the entire 10-year period.

4.3. Evaluation criteria

Predictive performance is assessed in terms of reliability, precision and bias, using the metrics from McInerney et al. (2017) (see Supplementary Material Section S4 for details). Reliability describes the degree of statistical consistency of predictive distributions and observations; precision refers to the width of predictive distributions; bias measures overall water balance errors. In all metrics, lower values indicate better performance.

Estimates of parameters A^*, ϕ_η and σ_y from the LS-MoM and ML-ML approaches are compared, including a check of how often the inferred A^* lies within machine precision of its lower bound.

Computational cost is quantified by the number of objective function evaluations (equivalent to the number of hydrological model calls), averaged over 100 independent optimizations, required for parameter optimization in Stage 1. This stage dominates the total cost in all schemes, because each objective function call in Stage 1 requires running the hydrological model; the cost of a *single* objective function evaluation in Stage 1 is essentially the same in all schemes.

5. Results

Figure 1 shows the predictive performance of all approaches. To facilitate comparison, Figure 2 shows the distribution of differences in metric values against a baseline approach. The baseline is taken as LS-MoM with $A^* = 0$ in all scenarios except for the use of the Log scheme in ephemeral/low-flow catchments, where $A^* = 0$ is not applicable and the baseline is hence LS-MoM with $A^* = 10^{-1}$.

5.1. Comparison of LS-MoM and ML-ML approaches

Figure 1 shows that the LS-MoM approach with $A^* = 0$ (red) has similar performance to the ML-ML approach (dark blue), for most residual error schemes, catchments and metrics (excepting the Log scheme applied in ephemeral/low-flow catchments). In most cases, performance metrics vary by about ±0.01 (Figure 2). Even the largest difference between LS-MoM and ML-ML approaches, in the precision of the BC0.2 scheme in ephemeral/low-flow catchments (Figure 2d, LS-MoM approach is better by a median value ≈0.02), is much smaller than the differences between BC0.2 vs BC0.5 schemes (median differences of ≈0.11 and ≈0.08 for LS-MoM and ML-ML approaches respectively).

The values of the offset and error model parameters in the two approaches are also similar. In the ML-ML approach, the inferred value of A^* is at its lower bound of 10^{-7} in 114 of 116 scenarios (excluding Log in ephemeral/low-flow catchments), effectively matching the value $A^* = 0$ used in the LS-MoM approach. The values of error model parameters ϕ_η and σ_y estimated using the two approaches differ by less than 1%.

In ephemeral/low-flow catchments, the Log scheme with inferred offset A^* yields poor precision and large biases (Figure 1d,f; see also McInerney et al. (2017)). Figure 1 shows that fixing the offset A^* to a larger value of 10^{-1} is highly beneficial, making the Log scheme competitive with the BC0.2 and BC0.5 schemes; importantly ML-ML and LS-MoM approaches once again perform very similarly.

5.2. Sensitivity of LS-MoM approach to the offset parameter

The impact of the offset A^* on the LS-MoM approach is shown in Figures 1 and 2 for fixed values of $A^* = 0$ (red), $A^* = 10^{-4}$ (green) and $A^* = 10^{-1}$ (cyan).

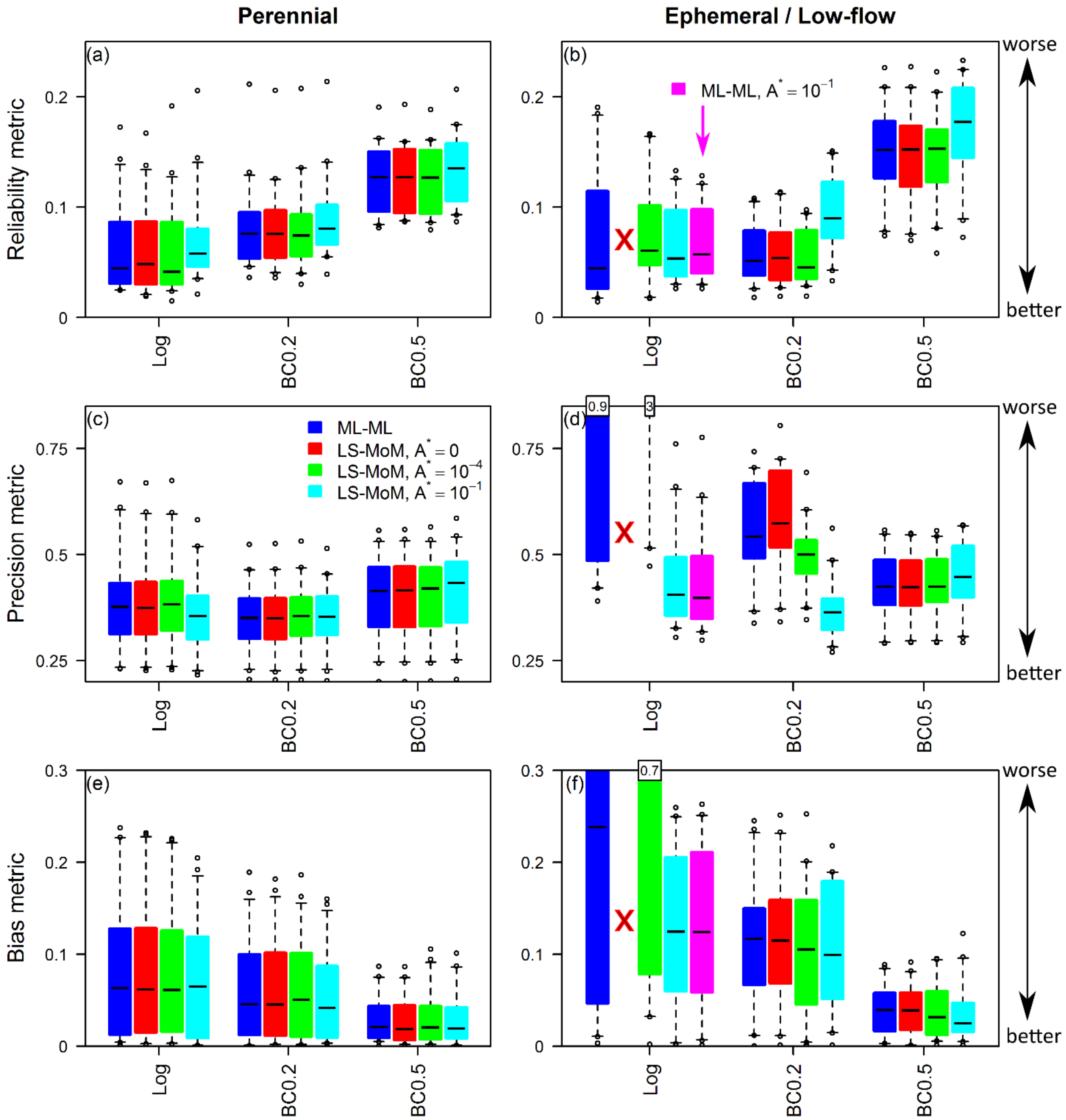

Figure 1: Predictive performance metrics of the LS-MoM approach with fixed $A^* \in \{0, 10^{-4}, 10^{-1}\}$ and the ML-ML approach with inferred A^*. The whiskers represent 90% probability limits computed over the 23 case study catchments. Results of applying the Log scheme in ephemeral/low-flow catchments are presented with modifications: (i) LS-MoM with $A^* = 0$ is not applicable (marked by red X), (ii) ML-ML with $A^* = 10^{-1}$ is included because ML-ML with fitted A^* performs very poorly.

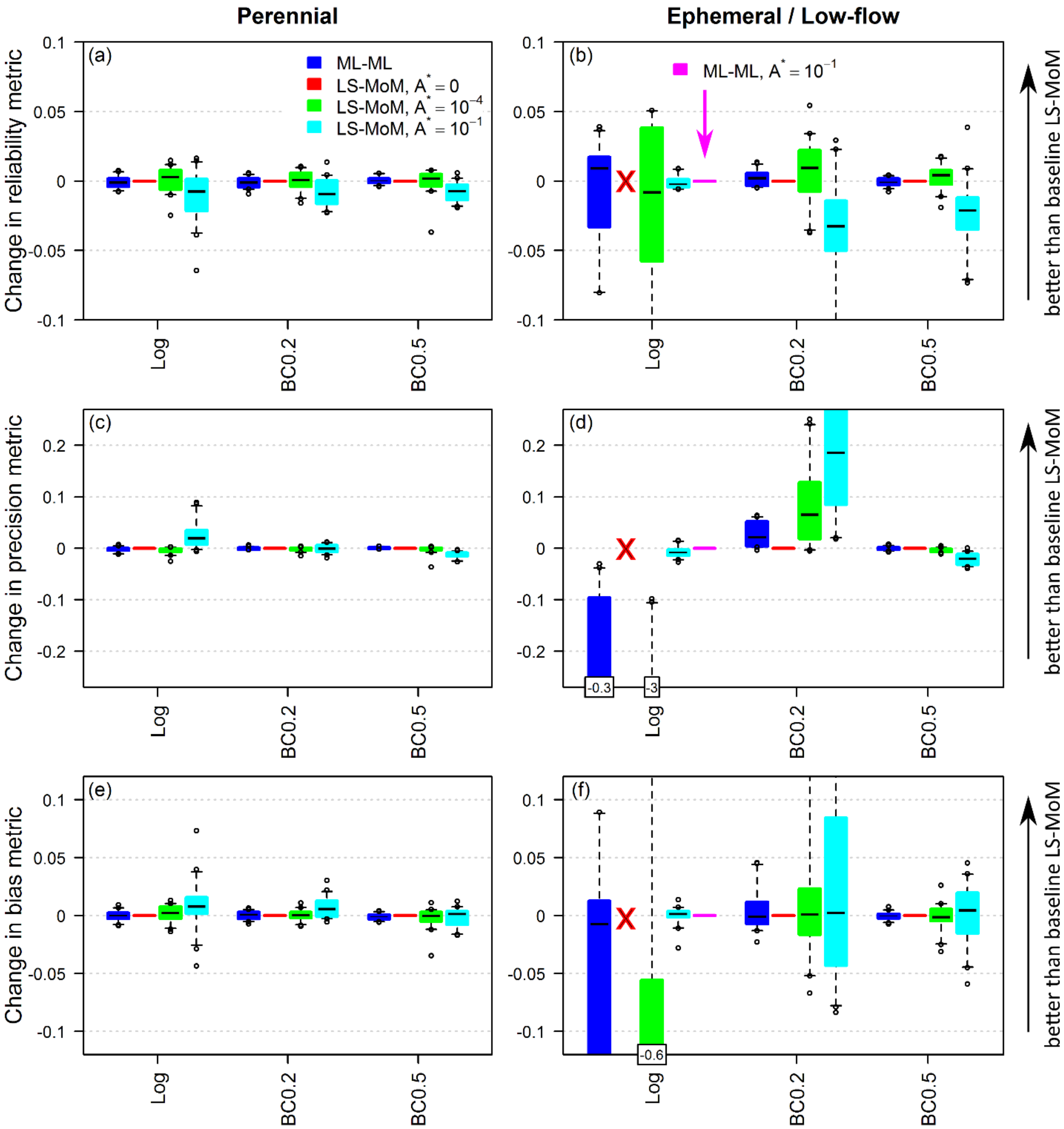

Figure 2: Difference in predictive performance metrics of the LS-MoM and ML-ML approaches in Figure 1. The baseline is given by the LS-MoM approach with $A^* = 0$, except for applications of the Log scheme in ephemeral/low-flow catchments, where the baseline is LS-MoM with $A^* = 10^{-1}$. Positive differences indicate schemes with better performance than the baseline LS-MoM approach.

273

In perennial catchments, all three values of A^* lead to similar predictive performance, with the largest changes occurring when $A^* = 10^{-1}$. For example, increasing A^* from 0 to 10^{-1} worsens reliability in all schemes (Figure 2a, median change ≈0.01). However, this difference is much smaller than differences between the Log and BC0.5 schemes (Log scheme is better by a median value ≈0.06).

In ephemeral/low-flow catchments, the offset parameter plays a bigger role. The impact is most evident in the Log scheme, where $A^* = 0$ is not applicable and increasing A^* from 10^{-4} to 10^{-1} substantially improves predictive performance (median precision tightens from ≈3 to ≈0.4, and median bias reduces from ≈0.7 to ≈0.12). For the BC0.2 scheme, increasing A^* from 0 to 10^{-4} improves reliability (Figure 2b, median change ≈0.01) and precision (Figure 2d, median change ≈0.07). Increasing A^* to 10^{-1} worsens reliability (median increase ≈0.03), but further improves precision (median change ≈0.18). The offset value is less important in the BC0.5 scheme; the most noticeable difference is the worsening of reliability when $A^* = 10^{-1}$ (median change ≈0.02).

5.3. Effect of ignoring posterior parametric uncertainty

Supplementary Material Section S1 reports the results of comparing LS-MoM and ML-ML against Bayesian implementations of the same residual error schemes. As shown in Supplementary Material Figure S1, the contribution of posterior parameter uncertainty to total predictive uncertainty in streamflow is small to negligible, and predictive performance metrics of LS-MoM are comparable to or better than the Bayesian approaches over the majority of catchments. These results are in line with theoretical expectations and previous empirical investigations (Kuczera et al., 2006, Yang et al., 2007, Sun et al., 2017, Kavetski, 2018, and others).

5.4. Comparison of computational cost

Figure 3 compares the number of objective function evaluations required for calibrating GR4J and HBV using the ML-ML approach with fitted A^* versus the LS-MoM approach with $A^* = 0$ (excluding the Log scheme in ephemeral/low-flow catchments). When using GR4J, the LS-MoM approach more than halves the computational cost (based on the median value over all scenarios). When using HBV, the savings are slightly smaller, around 40%.

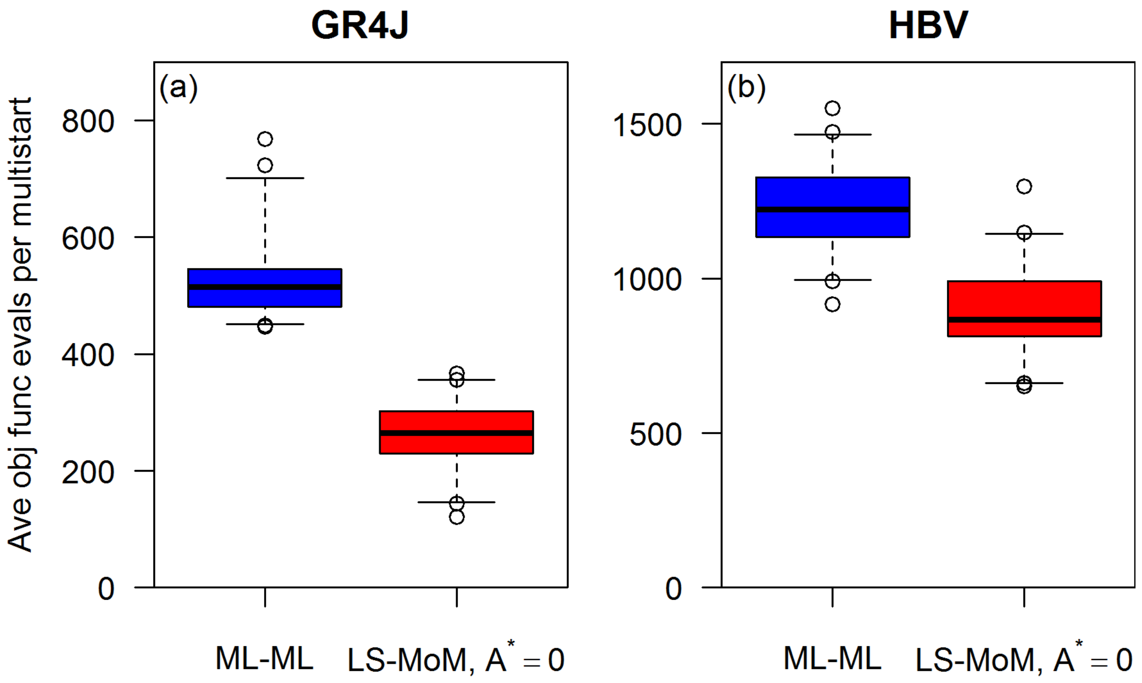

Figure 3: Computational cost of parameter optimization in Stage 1 of the ML-ML vs LS-MoM approaches. The number of objective function calls per invocation of a quasi-Newton optimizer is shown (averaged over 100 multistarts). Boxplots indicate results over all catchments and residual error schemes in Figures 1 and 2 (except for the Log scheme in ephemeral/low-flow catchments).

6. Discussion

6.1. Bona fides of the LS-MoM approach

The similar predictive performance of the LS-MoM and ML-ML approaches is explained by the calibrated parameters having similar values. In particular, with the exception of the Log scheme applied in ephemeral/low-flow catchments, the inferred A^* is generally close to 0, and hence to the fixed values used in the LS-MoM approach with $A^*=0$. Given similar values of A^*, the similarity of the two approaches is expected from theory: the equivalence of Stage 1 hydrological parameter optima is shown in Section 3.1, and the similarity of Stage 2 error parameter estimates reflects the general consistency of maximum-likelihood and method-of-moments estimators of AR(1) process parameters.

The computational savings of the LS-MoM approach can be attributed to fewer estimated parameters, as Stage 1 no longer calibrates A^* and σ_y. For example, in the case of GR4J, the dimension of the search space is reduced by 33%. The cost savings might vary depending on the particular optimization algorithm used, and further savings are likely if optimization algorithms adapted to LS-type objective functions, such as the Levenberg-Marquardt method (e.g., Doherty, 2004), are exploited. Cost savings are expected to be even larger in comparison to a Bayesian approach using MCMC.

6.2. Selection of offset value

In the experiments reported here, as A^* increases, precision generally improves but reliability worsens. This trade-off is most evident in ephemeral/low-flow catchments, especially when the BC0.2 scheme is used, and is reminiscent of the trade-offs seen when changing the Box-Cox parameter λ (McInerney et al., 2017). Given that the LS-MoM approach requires all transformation parameters, including A^*, to be fixed *a priori*, we recommend starting with a value of $A^* = 0$ and increasing it while monitoring relevant aspects of predictive performance. The exception is when the Log scheme is used in ephemeral/low-flow catchments, in which case a larger offset of $A^* = 10^{-1}$ can improve the precision and reduce bias.

6.3. Web-app implementing the LS-MoM approach

A public-access web-app is provided at *www.algorithmik.org.au/apps/probabilisticPredictions* to implement Stage 2 of the LS-MoM approach. The web-app assumes the user has already calibrated their hydrological model (Stage 1), using their preferred software and objective function (e.g., Table 1). The user uploads the observed and calibrated streamflow time series, and specifies $\boldsymbol{\theta}_Z = \{\lambda, A^*\}$ used in Stage 1. The web-app then estimates the error model parameters $\boldsymbol{\theta}_\varepsilon = \{\phi_\eta, \sigma_y\}$ (Stage 2) and generates probabilistic predictions. The web-app includes interactive display of probabilistic predictions and observed data time series, performance metrics and residual diagnostic plots. Figure 4 demonstrates the application of the web-app to the Gingera catchment on Cotter River (Australia), using GR4J pre-calibrated to the log-flow NSE ($\lambda = 0$ and $A^* = 0$).

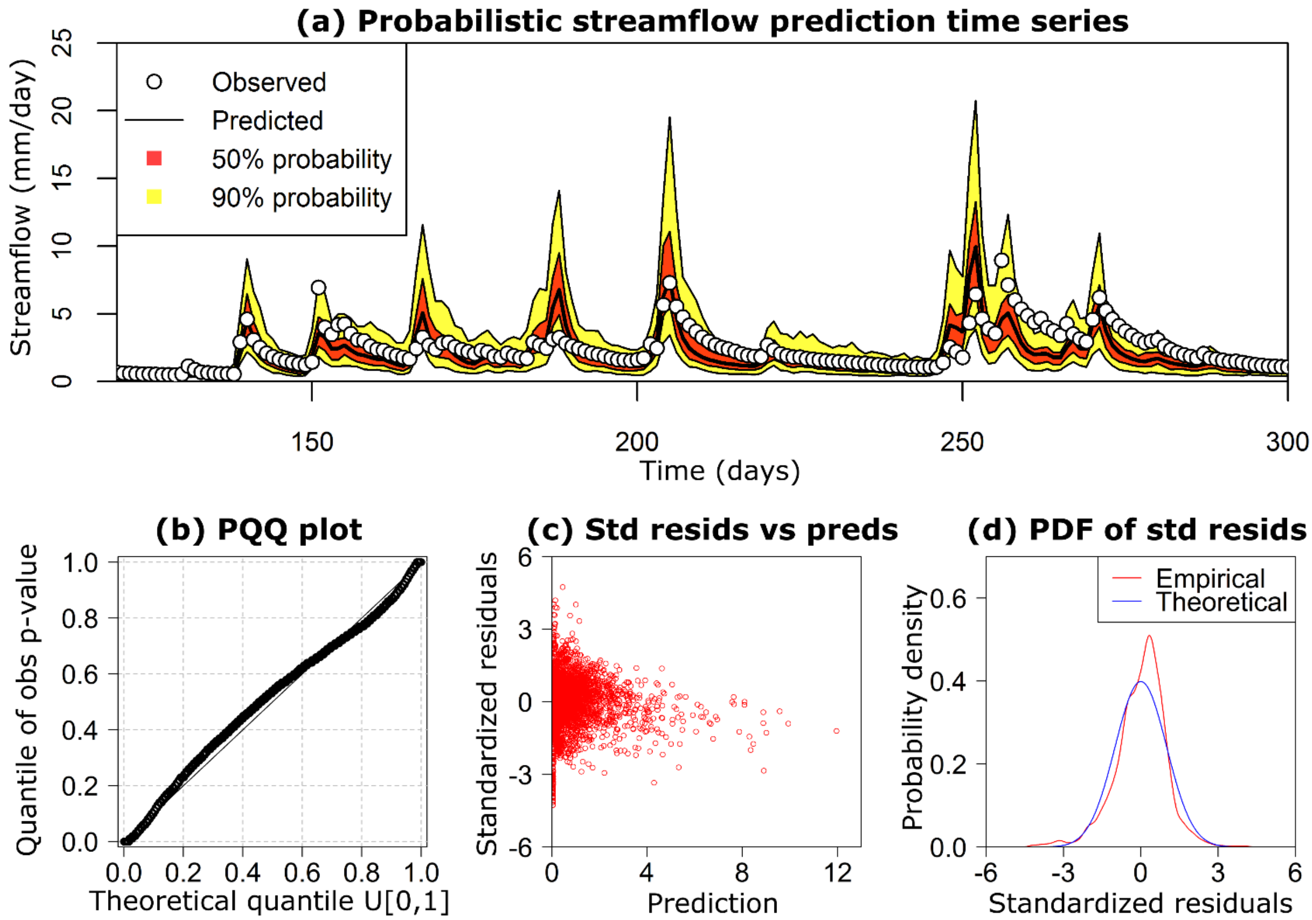

Figure 4: A selection of figures constructed from results obtained using LS-MoM approach web-app. Predictions for the Gingera catchment over the period May-October 1978 are shown, based on the GR4J model pre-calibrated to the NSE of log transformed flows. Shown are (a) 50% and 90% probability limits of the streamflow time series; (b) predictive quantile-quantile (PQQ) plot to assess the reliability of predictions; (c) residual error diagnostic plot of the dependence of standardized residuals η on the predicted streamflow; and (d) probability density of standardized residuals compared to the assumed error model. See Evin et al. (2013) for additional details on these diagnostics.

6.4. Limitations and future work

Several limitations of the LS-MoM approach warrant further investigation:

1. The *a priori* fixed transformation parameters can affect performance. Guidance is available for selecting the Box-Cox parameters λ (McInerney et al., 2017) and A^* (Section 6.2). For other transformations, such as the log-sinh (Wang et al., 2012), less guidance might be available;
2. The LS-MoM approach may be difficult to apply to more complex non-Gaussian residual error models, including those that treat zero flows (Smith et al., 2010, Wang and Robertson, 2011), use mixture-based distributions (Schaefli et al., 2007), include skewness/kurtosis (Schoups and Vrugt, 2010), etc.;

3. The assumption that posterior parametric uncertainty is small, while often appropriate when parsimonious hydrological models are calibrated to long time series using simple residual error models (Supplementary Material Section S1), might break down for more heavily parameterized models, and/or when working with short data sets (e.g., Thyer et al., 2002). Under these scenarios, especially if independent information is available, Bayesian approaches will be preferable. Further analysis is recommended to clarify the range of hydrological model complexity and data length for which posterior parametric uncertainty is sufficiently small to be ignored in practical streamflow prediction contexts.

The LS-MoM approach and the web-app can be used for environmental modelling applications beyond hydrology, whenever the residual error assumptions hold and parametric uncertainty is relatively small. In addition, LS-MoM and the web-app can be used with environmental models calibrated using methods other than (transformed) Least Squares objective functions, taking particular care to monitor predictive performance metrics and residual error diagnostics because inconsistencies between the objective function and the error model can lead to poor probabilistic predictions. These model setups are of practical interest (Li et al., 2016) and warrant further investigation.

7. Conclusions

This study introduces a simplified approach for generating probabilistic predictions. The LS-MoM approach uses Least Squares (LS) optimization to estimate hydrological model parameters and simple method-of-moments (MoM) estimators of error model parameters to describe uncertainty in predictions. It can be used in combination with many existing hydrological modelling packages, and achieves similar predictive performance to more complicated maximum-likelihood and Bayesian approaches while reducing computational costs by factors of two or more. A public web-app is made available to help users apply the LS-MoM approach, and bridge the gap between deterministic and probabilistic prediction techniques in practical hydrological applications.

Appendix A. Generation of probabilistic predictions

In both the LS-MoM and ML-ML approaches, probabilistic predictions are represented using replicates $\mathbf{Q}^{(r)} = \{Q_t^{(r)}, t = 1,..,T\}$ for $r = 1,..,R$. Given parameter values $\{\boldsymbol{\theta}_H, \boldsymbol{\theta}_Z, \boldsymbol{\theta}_\varepsilon\}$, the *r*th replicate is generated as follows:

1. At time step t, sample innovation $y_t^{(r)} \leftarrow N(0,\sigma_y^2)$ and calculate residual $\eta_t^{(r)} = \phi_\eta \eta_{t-1}^{(r)} + y_t^{(r)}$, as per equations (6)-(7). Note that for $t=1$, we directly sample $\eta_1^{(r)} \leftarrow N(0,\sigma_\eta^2)$;
2. Calculate replicate $Q_t^{(r)}$ by rearranging equation (5),

$$Q_t^{(r)} = Z^{-1}(Z(Q_t^{\boldsymbol{\theta}_H}) + \eta_t^{(r)}) \quad (14)$$

3. Repeat for $t+1$, etc.

For practical purposes, $Q_t^{(r)}$ is truncated if it falls outside $Q_{\min} = 0$ and $Q_{\max} = 10 \times \max(\tilde{\mathbf{Q}})$.

Acknowledgments

This work was funded by Australian Research Council grant LP140100978 with the Australian Bureau of Meteorology and South East Queensland Water. We thank the Associate Editor and four anonymous reviewers for their insightful comments and suggestions. Simulations were performed on the Phoenix cluster at the University of Adelaide. We thank Michael Leonard for hosting the web-app on his server. Data for the Australian catchments was obtained from the Hydrologic Reference Stations database provided by the Australian Bureau of Meteorology (http://www.bom.gov.au/water/hrs). The MOPEX dataset for the USA catchments is available on request from Qingyun Duan.

References

ANDREWS, F. T., CROKE, B. F. W. & JAKEMAN, A. J. 2011. An open software environment for hydrological model assessment and development. *Environmental Modelling & Software,* 26**,** 1171-1185.

BATES, B. C. & CAMPBELL, E. P. 2001. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research,* 37**,** 937-947, doi:10.1029/2000WR900363.

BERGSTRÖM, S. 1995. The HBV model. *In:* SINGH, V. P. (ed.) *Computer Models of Watershed Hydrology.* Water Resources Publications, Highlands Ranch, CO.

BOX, G. E. P. & COX, D. R. 1964. An analysis of transformations. *Journal ofthe Royal Statistical Society, Series B,* 26**,** 211-252.

BOX, G. E. P. & JENKINS, G. M. 1970. *Time series analysis: forecasting and control,* San Francisco, CA, Holden-Day.

CHAPMAN, T. G. Optimization of a rainfall-runoff model for an arid zone catchment. I.A.S.H.-UNESCO Symposium on the Results of Research on Representative and Experimental Basins, 1970 Wellington, New Zealand. IASH-AISH Publ., 126-144.

CHARNES, A., FROME, E. L. & YU, P. L. 1976. The Equivalence of Generalized Least Squares and Maximum Likelihood Estimates in the Exponential Family. *Journal of the American Statistical Association,* 71**,** 169-171.

CHIEW, F. H. S., STEWARDSON, M. J. & MCMAHON, T. A. 1993. Comparison of six rainfall-runoff modelling approaches. *Journal of Hydrology,* 147**,** 1-36.

CORON, L., THIREL, G., DELAIGUE, O., PERRIN, C. & ANDRÉASSIAN, V. 2017. The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software,* 94, 166-171.

DAWDY, D. R. & LICHTY, R. 1968. Methodology of hydrologic model building. *Proceedings, use of analog and digital computers in hydrology,* 2, 347-355.

DOHERTY, J. 2004. PEST—Model independent parameter estimation 5th edition. Watermark Numerical Computing. Australia.

DUAN, Q., SCHAAKE, J., ANDREASSIAN, V., FRANKS, S. W., GOTETI, G., GUPTA, H. V., GUSEV, Y. M., HABETS, F., HALL, A., HAY, L., HOGUE, T., HUANG, M., LEAVESLEY, G., LIANG, X., NASONOVA, O. N., NOILHAN, J., OUDIN, L., SOROOSHIAN, S., WAGENER, T. & WOOD, E. F. 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology,* 320, 3-17.

ENGELAND, K. & STEINSLAND, I. 2014. Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times. *Water Resources Research,* 50, 182-197.

EVIN, G., KAVETSKI, D., THYER, M. & KUCZERA, G. 2013. Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resources Research,* In Press, W012972, doi:10.1002/wrcr.20284.

EVIN, G., THYER, M., KAVETSKI, D., MCINERNEY, D. & KUCZERA, G. 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research,* 50, 2350-2375.

GAN, T. Y., DLAMINI, E. M. & BIFTY, G. F. 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology,* 192, 81-103.

JOSEPH, J. F. & GUILLAUME, J. H. A. 2013. Using a parallelized MCMC algorithm in R to identify appropriate likelihood functions for SWAT. *Environmental Modelling & Software,* 46, 292-298.

KAVETSKI, D. 2018. Parameter Estimation and Predictive Uncertainty Quantification in Hydrological Modelling. *In:* DUAN, Q., PAPPENBERGER, F., THIELEN, J., WOOD, A., CLOKE, H. L. & SCHAAKE, J. C. (eds.) *Handbook of Hydrometeorological Ensemble Forecasting.* Springer Berlin Heidelberg.

KAVETSKI, D. & CLARK, M. P. 2010. Ancient numerical daemons of conceptual hydrological modeling. Part 2: Impact of time stepping scheme on model analysis and prediction. *Water Resources Research,* 46, W10511, doi:10.1029/2009WR008896.

KRZYSZTOFOWICZ, R. 2002. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology,* 268, 16-40.

KUCZERA, G. 1983. Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty. *Water Resources Research,* 19, 1151-1162, doi:10.1029/WR019i005p01151.

KUCZERA, G., KAVETSKI, D., FRANKS, S. & 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, 161-177, doi:10.1016/j.jhydrol.2006.05.010.

KUMAR, R., SAMANIEGO, L. & ATTINGER, S. 2010. The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics. *Journal of Hydrology,* 392, 54-69.

LI, M., WANG, Q. J., BENNETT, J. C. & ROBERTSON, D. E. 2016. Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting. *Hydrol. Earth Syst. Sci.,* 20, 3561-3579.

LOUCKS, D. P., STEDINGER, J. R. & HAITH, D. A. 1981. *Water resource systems planning and analysis*, Prentice-Hall.

MARTINS, E. S. & STEDINGER, J. R. 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research,* 36, 737-744.

MCINERNEY, D., THYER, M., KAVETSKI, D., LERAT, J. & KUCZERA, G. 2017. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*.

OUDIN, L., ANDREASSIAN, V., MATHEVET, T., PERRIN, C. & MICHEL, C. 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research,* 42.

PERRIN, C., MICHEL, C. & ANDREASSIAN, V. 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology,* 279**,** 275-289, doi:10.1016/S0022-1694(03)00225-7.

PUSHPALATHA, R., PERRIN, C., MOINE, N. L. & ANDRÉASSIAN, V. 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology,* 420**,** 171-182.

RENARD, B., KAVETSKI, D., LEBLOIS, E., THYER, M., KUCZERA, G. & FRANKS, S. W. 2011. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research,* 47**,** W11516, doi:10.1029/2011WR010643.

SALAS, J. D. 1993. Analysis and Modeling of Hydrologic Time Series. *In:* MAIDMENT, D. R. (ed.) *The McGraw Hill Handbook of Hydrology.* New York, USA: McGraw-Hill.

SCHAEFLI, B., TALAMBA, D. B. & MUSY, A. 2007. Quantifying hydrological modeling errors through a mixture of normal distributions. *Journal of Hydrology,* 332**,** 303-315, doi:10.1016/j.jhydrol.2006.07.005.

SCHARFFENBERG, W. A., FLEMING, M. J. & CENTER, H. E. 2006. Hydrologic Modeling System HEC-HMS: User's Manual.: US Army Corps of Engineers, Hydrologic Engineering Center.

SCHOUPS, G. & 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.

SEIBERT, J. 2005. HBV light version 2, user's manual. Stockholm University.

SERVAT, E. & DEZETTER, A. 1991. Selection of calibration objective fonctions in the context of rainfall-ronoff modelling in a Sudanese savannah area. *Hydrological Sciences Journal,* 36**,** 307-330.

SMITH, T., SHARMA, A., MARSHALL, L., MEHROTRA, R. & SISSON, S. 2010. Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments. *Water Resources Research,* 46**,** W12551, doi:10.1029/2010WR009514.

SOROOSHIAN, S. & DRACUP, J. A. 1980. Stochastic parameter estimation procedures for hydrological rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resources Research,* 16**,** 430-442, doi:10.1029/WR016i002p00430.

SUN, R., YUAN, H. & LIU, X. 2017. Effect of heteroscedasticity treatment in residual error models on model calibration and prediction uncertainty estimation. *Journal of Hydrology*.

THYER, M., KUCZERA, G. & WANG, Q. J. 2002. Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation. *Journal of Hydrology,* 265**,** 246-257, doi:10.1016/S0022-1694(02)00113-0.

TONKIN, M. & DOHERTY, J. 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. *Water Resources Research,* 45.

VAZE, J., JORDAN, P., BEECHAM, R., FROST, A. & SUMMERELL, G. 2012. Guidelines for rainfall-runoff modelling.

VOGEL, R. M. 2017. Stochastic watershed models for hydrologic risk management. *Water Security,* 1**,** 28-35.

VRUGT, J. A., TER BRAAK, C. J. F., DIKS, C. G. H., ROBINSON, B. A., HYMAN, J. M. & HIGDON, D. 2009. Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. *International Journal of Nonlinear Sciences and Numerical Simulation,* 10**,** 273-290.

WANG, Q. J. & ROBERTSON, D. E. 2011. Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research,* 47.

WANG, Q. J., SHRESTHA, D. L., ROBERTSON, D. E. & POKHREL, P. 2012. A log-sinh transformation for data normalization and variance stabilization. *Water Resources Research,* 48.

WELSH, W. D., VAZE, J., DUTTA, D., RASSAM, D., RAHMAN, J. M., JOLLY, I. D., WALLBRINK, P., PODGER, G. M., BETHUNE, M., HARDY, M. J., TENG, J. & LERAT, J. 2013. An integrated modelling framework for regulated river systems. *Environmental Modelling & Software,* 39**,** 81-102.

YANG, J., REICHERT, P., ABBASPOUR, K. C. & YANG, H. 2007. Hydrological modelling of the chaohe basin in china: Statistical model formulation and Bayesian inference. *Journal of Hydrology,* 340**,** 167-182.

YE, W., JAKEMAN, A. J. & YOUNG, P. C. 1998. Identification of improved rainfall-runoff models for an ephemeral low-yielding Australian catchment. *Environmental Modelling & Software,* 13**,** 59-74.