Dmitri Kavetski, David McInerney, Mark Thyer, Julien Lerat and George Kuczera
Probabilistic streamflow prediction and uncertainty estimation in ephemeral catchments
Proceedings of the Hydrology and Water Resources Symposium (HWRS 2021), 2021, pp.482-496

ISBN number for HWRS 2021 is 978-1-925627-53-4

Probabilistic streamflow prediction and uncertainty estimation in ephemeral catchments

Dmitri Kavetski[1], David McInerney[1], Mark Thyer[1], Julien Lerat[2] and George Kuczera[3]

[1] School of Civil, Environmental and Mining Engineering, University of Adelaide, South Australia

[2] CSIRO, Land and Water, Canberra, ACT

[3] School of Engineering, University of Newcastle, Callaghan, NSW

Corresponding author email: dmitri.kavetski@adelaide.edu.au

ABSTRACT

Probabilistic streamflow predictions at the daily scale are of major practical interest for environmental management and planning, including risk assessment as part of reservoir management operations. Ephemeral catchments, where streamflow is frequently zero or negligible, pose particularly stark challenges in this context, due to asymmetry of the error distribution and the discrete (rather than continuous) nature of zero flows. In this work, our focus is on two practical error modelling approaches where predictive uncertainty is approximated by a (transformed) Gaussian error model. The first approach, termed "pragmatic", does not distinguish between zero and positive flows during calibration, but sets negative flows to zero when making predictions. The second approach, termed "explicit", applies a "censored" Gaussian assumption in both calibration and prediction. We report a comparison of these two approaches over 74 Australian catchments with diverse hydroclimatology, using multiple performance metrics. The performance of the approaches depended on the catchment type as follows: (1) "mid-ephemeral" catchments, where 5-50% of days have zero flows, are best modelled using the "explicit" approach in combination with the Box-Cox streamflow transformation with a power parameter of 0.2; (2) "low-ephemeral" catchments, with fewer than 5% zero flow days, can be modelled using the pragmatic approach with (relatively) little loss of predictive performance; (3) "high-ephemeral" catchments, with more than 50% zero flow days, prove challenging to both approaches, and require more specialised techniques. The findings provide practical guidance towards improving probabilistic streamflow predictions in ephemeral catchments.

INTRODUCTION

Ephemeral catchments, with frequent periods of zero flow, are common and often prevalent in arid and semiarid regions in Australia and worldwide. Almost a third of the Earth's land surface has been described as arid/semiarid (Salem, 1989). Given the limited availability of water resources in these catchments, effective management including allocations for urban and irrigation water use, environmental flows, and so forth) is of particular practical importance. In addition to traditional "point" estimates, a suitable representation of uncertainty in streamflow predictions is required for meaningful risk-based decision making (Vogel, 2017). Probabilistic prediction methods hold particular appeal in these contexts. However, streamflow prediction in ephemeral catchments is challenging for multiple reasons, notably including the representation of uncertainty in zero flows.

Many if not most practical probabilistic modelling systems used in hydrology employ a deterministic rainfall-runoff model coupled with a residual error model to describe streamflow uncertainty. Uncertainty is typically assumed to be Gaussian after flows are transformed using transformations such as the Box-Cox to stabilize the error variance. A literature review suggests that most studies on residual error model development and application are based on catchments with few or no days with zero flows (e.g., Kuczera, 1983; Schoups and Vrugt, 2010; Del Giudice et al., 2013, and many others). The performance of common residual error models in the presence of zero flows has received comparatively less attention. Notably, many residual error models do not recognize that streamflow has a lower bound of zero and that its distribution often has a spike at zero.

The recent work by McInerney et al. (2019) provided a summary of the theoretical foundations of the treatment of zero flows using (relatively) simple Gaussian error models. It introduced the distinction between two "pragmatic" and "explicit" approaches for the treatment of zero flows and considered a range of streamflow transformations. In this paper, we summarize the key findings of direct relevance to practical applications and provide a simplified set of recommendations.

“Pragmatic” approaches are defined by treating zero flows in the same way as other flows during calibration, but dealing with them in a specialised way in prediction. For example, in predictions negative streamflows can be reset to zero (e.g., Pianosi and Raso, 2012; McInerney et al., 2017; Gibbs et al., 2018; Woldemeskel et al., 2018), or, less commonly, resampled until non-negative (Evin et al., 2014). Zero flow data points can also be simply excluded from the calibration (e.g., Morawietz et al., 2011; Westra et al., 2014). These pragmatic approaches are easy to implement, yet their statistical assumptions are inconsistent between calibration and prediction.

“Explicit” approaches, are defined by treating zero flows consistently in calibration and prediction. Several statistical techniques can be used to achieve this. "Censoring” approaches treat zero flows as “censored” data, i.e., as data assumed to lie below a detection threshold, and assign them discrete probability masses (e.g., Tobin, 1958; Wang and Robertson, 2011; Bennett et al., 2016). As such, these approaches represent streamflow uncertainty using a mixed continuous/discrete probability distribution. Other explicit approaches include “zero-inflation”, which also employs a mixed probability distribution but uses an separate probability model (e.g., binomial) to describe the occurrence of zero-valued data (e.g., Min and Agresti, 2002; Smith et al., 2010; Oliveira et al., 2018).

A practical residual error model must balance theoretical rigour with implementation/operational complexity. Pragmatic approaches can use common sum-of-squared-errors objective functions (with optional transformations). In contrast, explicit approaches offer a consistent treatment of zero flows in calibration and prediction – at the practical cost of a more complex error model (likelihood function).

Despite the theoretical appeal of the explicit approach, the lack of a systematic comparison over a wide range of catchments makes it unclear whether explicit approaches offer sufficient performance benefits over pragmatic approaches to justify their extra implementation effort.

The study aims focus on the following questions:

1. Is there a degree of catchment ephemerality (proportion of time steps with zero flow) beyond which explicit approaches are unequivocally preferable?
2. Can explicit approaches handle extremely dry catchments where zero flows dominate the observed record?
3. Does the optimal choice of residual error model – notably its transformation – depend on whether the pragmatic or explicit approaches are implemented?

To address these questions, we compare a pragmatic approach (treatment of zero flows as censored in prediction but not in calibration) versus an explicit approach (consistent treatment of zero flows as censored in calibration and prediction), in the context of daily catchment-scale modelling. Empirical case studies include 74 Australian catchments with a representative range of hydroclimatology and ephemerality; multiple metrics are employed to quantify predictive performance. Within the error models, two streamflow transformations are considered, namely the logarithmic and Box-Cox transformations (McInerney et al., 2017). As such, the case study is designed to ensure the findings are robust and suitable for selecting methods for operational deployment (e.g., Tuteja et al., 2012; Prudhomme et al., 2017; Woldemeskel et al., 2018).

The remainder of this paper is structured as follows. We begin with an outline of the theory of treating zero flows in residual error models. The case study is described next, followed by results and discussion. The final section summarizes the key conclusions and recommendations.

THEORY

General Model Setup

Let q_t denote streamflow predictions at time step *t* obtained using a deterministic hydrological model *h* with parameters $\boldsymbol{\theta}_H$ and inputs $\mathbf{x}_{1:t}$ (up to step *t*),

$$q_t^{\boldsymbol{\theta}_H} = h(\boldsymbol{\theta}_H; \mathbf{x}_{1:t}) \tag{1}$$

Consider the (common) case where the uncertainty in the deterministic model is represented using Gaussian residual errors in transformed space,

$$z(Q_t; \boldsymbol{\theta}_z) = z(q_t^{\boldsymbol{\theta}_H}; \boldsymbol{\theta}_z) + \eta_t \tag{2}$$

where *z* is a transformation function and η_t is the normalized residual at time *t*,

$$\eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \tag{3}$$

Equation (2) represents a probabilistic model of streamflow Q_t at time *t*, with probability density/mass function (pdf/pmf) $p(q_t \mid \boldsymbol{\theta}, \mathbf{x}_{1:t})$; note the distinction between the random variable Q_t (upper case) and a realization q_t (lower case). Note also that in this work we focus on "total" uncertainty and do not attempt to disaggregate the individual contributions of data and model structural errors.

The transformation z is selected to stabilise the variance of the errors and reduce its dependence on the flow magnitude; in this work we use the Box-Cox transformation and its special case the logarithmic transformation (Box and Cox, 1964).

The probability model in equation (3) can be also expressed directly in terms of a Gaussian distribution centered on the deterministic predictions,

$$z(Q_t; \boldsymbol{\theta}_z) \sim \mathcal{N}\left(z(q_t^{\boldsymbol{\theta}_H}; \boldsymbol{\theta}_z), \sigma_\eta^2\right) \tag{4}$$

This formulation is particularly convenient for specialised handling of zero flows because it enables a direct specification of any constraints on the values of Q_t.

The parameter set of the probabilistic model, $\boldsymbol{\theta}$, includes in addition to $\boldsymbol{\theta}_H$, parameters describing the transformation function and parameters describing the statistical properties of residuals (here, the error variance σ_η^2). Some parameters (e.g., the Box-Cox power parameter) can be fixed a priori based on recommended values (McInerney et al., 2017). Other parameters, including hydrological parameters θ_H and the error variance σ_η^2, are inferred (calibrated) from observed streamflow time series $\mathbf{q}^{\text{obs}}$; here we use an optimization-based maximum likelihood (ML) (McInerney et al., 2018),

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}; \mathbf{q}^{\text{obs}}, \mathbf{x}^{\text{obs}}) \tag{5}$$

The likelihood function $\mathcal{L}$ is defined formally as the pdf/pmf of the probability model evaluated at the observed data and viewed as a function of model parameters, i.e., $\mathcal{L}(\boldsymbol{\theta};\mathbf{q}^{\text{obs}},\mathbf{x}^{\text{obs}}) = p(\mathbf{q}^{\text{obs}} \,|\, \boldsymbol{\theta},\mathbf{x}^{\text{obs}})$. It is known that a poorly specified probability model can lead to poor parameter estimates (e.g., Ang and Tang, 2007).

Following calibration, probabilistic streamflow predictions are generated by sampling replicates (indexed by superscript *r*) from the assumed probability model with optimal parameters $\hat{\boldsymbol{\theta}}$,

$$q_t^{\text{pred},r} \leftarrow Q_t(\hat{\boldsymbol{\theta}},\mathbf{x}_{1:t}) \quad (6)$$

where ← denotes the process of sampling from a probability distribution. The full set of predictive replicates at either a given step *t* or a series of steps will be referred to as the "predictive distribution".

The next sections focus on the treatment of zero flows. Figure 1 illustrates representative assumptions; for simplicity the illustration does not use streamflow transformations.

Pragmatic Approach

The pragmatic approach assumes a Gaussian probability model of (possibly transformed) streamflow, but applies this model in differently in calibration and prediction. In calibration, the likelihood function does not treat observed zero flows any differently from non-zero flows. The common Gaussian likelihood function is used (e.g., equation 20 in McInerney et al., 2017). However, in prediction, in order to avoid negative flows in the predictive distribution, any $q_t < 0$ sampled in equation (4) are set to 0. This approach is similar to standard regression (e.g., Ang and Tang, 2007) but with the "ad hoc" modification to avoid negative flows. It is shown in Figure 1 (panels a and b). It has been used in numerous previous hydrological modelling studies (e.g., Pianosi and Raso, 2012; McInerney et al., 2017; Woldemeskel et al., 2018, and others).

The resetting of negative flows in prediction is a pragmatic way of avoiding the most obvious undesirable consequence of assuming that streamflow uncertainty is Gaussian. However, it does not address a more subtle issue, namely that the (unmodified) Gaussian likelihood function does not recognize the lower streamflow bound of zero and that zero flows form a distinct "spike" in the streamflow distribution. Therefore, the Gaussian likelihood function can be expected to produce potentially poor parameter estimates when the observed data has a frequent occurrence of zero flows. This is the key "theoretical" weakness of the pragmatic approach.

Explicit Approach

The explicit approach employs a more realistic probability model, which is modified specifically to recognise the distinct nature of zero flows. Figure 1cd shows a "censored" Gaussian distribution, is obtained from a "regular" Gaussian distribution by chopping it at a particular value of streamflow and creating a "spike" (probability mass) at $q = 0$. This implementation is based on the censoring method of Tobin (1958); it is well-established in the research literature (e.g., Wang and Robertson, 2011; Li et al., 2013; Pagano et al., 2013) and adopted in operational applications (e.g., Bennett et al., 2016).

A key theoretical advantage of the explicit approach is that the same (censored) probability model is used consistently in calibration and prediction. Notably, the likelihood function in the explicit approach is based on the censored Gaussian distribution. Essentially, when evaluated at a single given time step *t*, the likelihood function value is computed from the bell-shaped section of the curve when $q_t^{\text{obs}} > 0$ and from the spike when $q_t^{\text{obs}} = 0$. As such, this likelihood function is conceptually and computationally more complicated than in the pragmatic approach.

The predictive distribution is constructed accordingly by sampling from the censored Gaussian distribution. Note that, algorithmically, the censored Gaussian pdf can be conveniently sampled by first drawing from the regular (uncensored) Gaussian, and then imposing the lower bound.

Understanding the Subtle Difference between Explicit vs Pragmatic Approaches

An astute reader will notice that the equations / algorithm used to generate the predictive streamflow distributions in the explicit and pragmatic approaches are the same. Indeed both approaches can be seen to use the censored Gaussian distribution when generating streamflow predictions – as also seen by comparing panels b and d in Figure 1. However, the values of the parameter(s) with which these distributions are applied will be generally different, because different likelihood functions are used to obtain the ML parameter estimates. Hence, due to differences in parameter values, the predictive distributions of streamflow generated using the pragmatic and censored approaches could be very, very different. This behaviour is also shown in Figure 1.

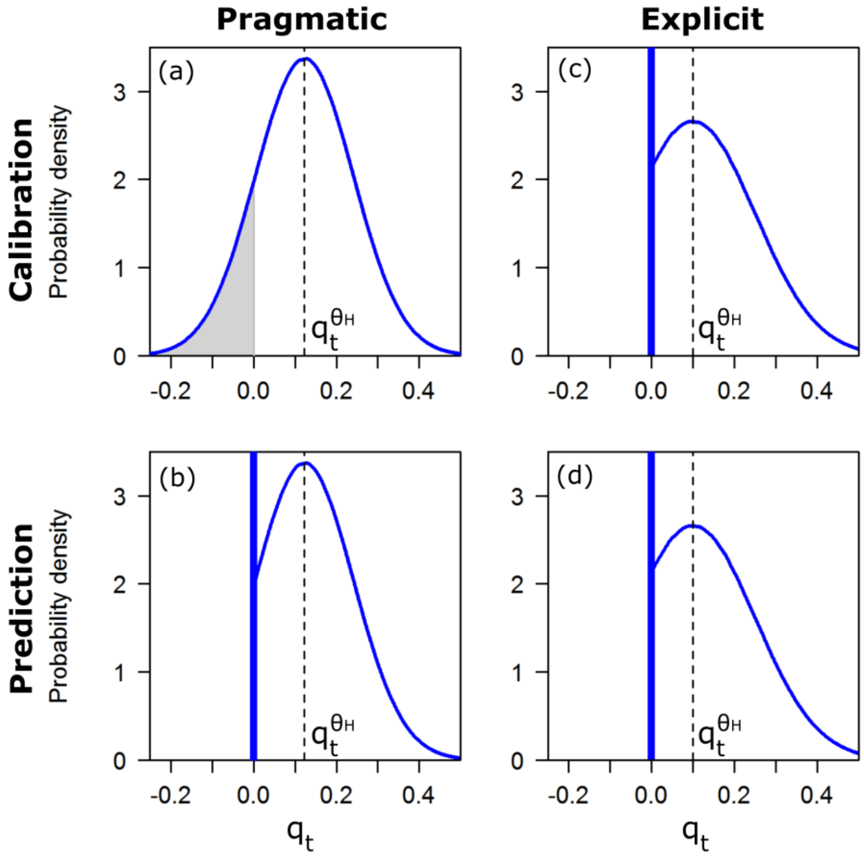

Figure 1. Representation of zero flows in the pragmatic and explicit approaches. The key difference is in calibration, where the likelihood function employed in the pragmatic approach does not recognise the lower bound of zero on streamflow. To keep the illustration simple, no streamflow transformation is used; its impact would be to introduce skew into the distribution. Figure adapted with modifications from McInerney et al. (2019).

Full details of the likelihood function and prediction generation algorithms for the pragmatic and explicit approaches can be found in McInerney et al. (2019). Note that we implement the inference using a post-processor approach in two stages: Stage 1 calibrates the hydrological model parameters θ_H and Stage 2 calibrates the error model parameter σ_η. Stage 1 is the same for both the explicit and pragmatic approach, but Stage 2 is different. Hence, any differences in the probabilistic predictions generated using the explicit and pragmatic approaches are due solely to the differences in the likelihood function used in Stage 2 when estimating the error parameter.

Theoretical Limitation when the Fraction of Zero Flows is High

A notable limitation of the approaches investigated in this work (regardless of censoring and/or transformations) is that they have a theoretical maximum of 50% probability of zero flows, both at a given time step and over the entire predictive distribution. This property is clearest when no transformation is used, in which case the predictive distribution is (symmetric) Gaussian with its mean, mode and median all equal to the deterministic prediction (here, from the GR4J model). Given that hydrological model predictions are never negative (this can be easily ensured in a deterministic model), it follows that at most 50% of the distribution can be censored at any given time step. In turn, this implies that the probability mass of the spike is itself at most 0.5 (50%). The use of streamflow transformations, as long as they are monotonic which is virtually if not always the case, preserves the median of the distribution and thus the 50% limit still holds.

The inability of the predictive distribution to allow for more than a 50% probability of zero flows becomes clearly problematic as the observed proportion of days with zero flows approaches and exceeds 50%, i.e., when dealing with highly ephemeral catchments that are dry over the majority of the year. The practical impact of this limitation will be investigated empirically.

CASE STUDY DESCRIPTION

Case Study Catchments

A total of 74 Australian catchments with a diverse range of hydroclimatic conditions and at least some occurrence of zero flows are considered. Daily rainfall, PET and streamflow time series from a 10-year period (1985-1994) are employed; fewer than 5% days have missing data.

The catchments are classified according to the proportion of zero flows in the observed time series:

• low-ephemeral, defined by fewer than 5% zero flow days: 34 catchments;

• mid-ephemeral, defined by 5-50% zero flow days: 29 catchments;

• high-ephemeral, defined by more than 50% zero flow days: 11 catchments.

These groupings are helpful for identifying common trends in residual error model performance; they are not intended as a general hydrological classification scheme.

Figure 2 shows the catchment locations, as well as their ephemerality groupings and Koppen climate characteristics (Stern et al., 2000). The cumulative distribution of the proportion of zero flow days across the catchments is also shown. With some exceptions, low-ephemeral catchments are located primarily in the temperate region on the East Coast of Australia, high-ephemeral catchments are located primarily in the grassland and desert regions (which tend to be semiarid and arid) of Northern and Central Australia, and mid-ephemeral catchments are spread throughout all regions (including in tropical and subtropical regions). Overall, the widespread occurrence of zero flow conditions in Australia is noted, spanning large geographical areas and many climatically diverse regions.

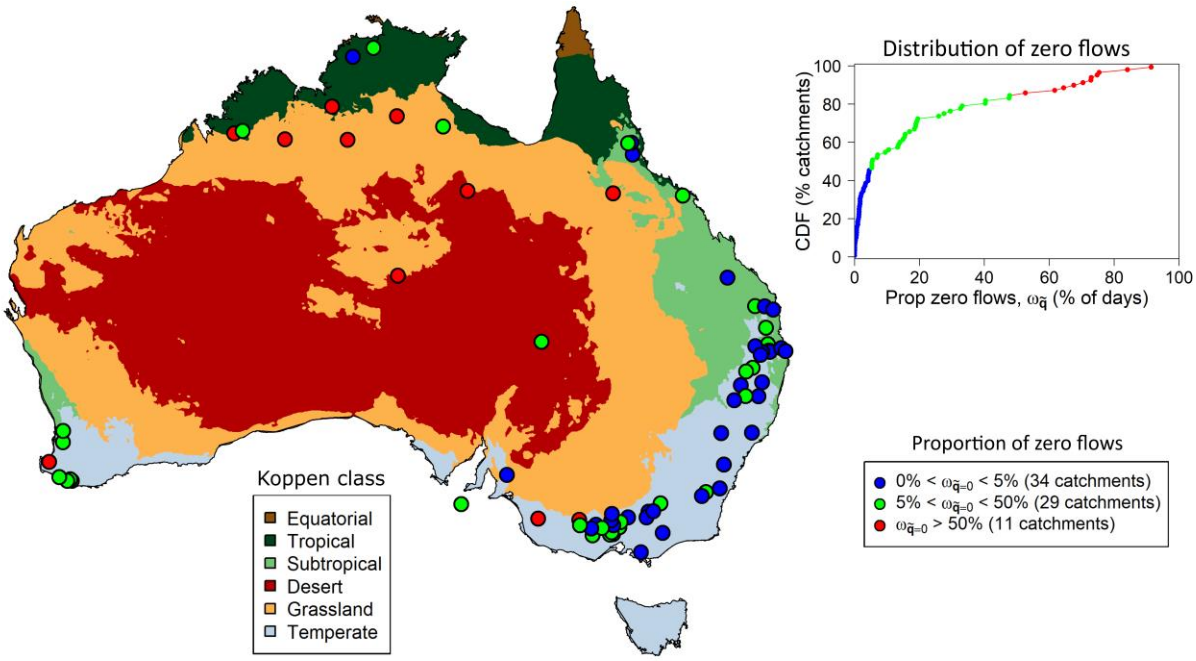

Figure 2. Location of the 74 case study catchments and their classification in terms of the proportion of zero flows. The inset shows the cumulative distribution of the proportion of zero flow days over the 74 catchments. Figure adapted from McInerney et al. (2019).

Hydrological Model

The conceptual rainfall-runoff model GR4J is used in equation (1) to simulate daily catchment-scale streamflow time series from rainfall and PET inputs (Perrin et al., 2003). GR4J has a parsimonious model structure with four parameters to represent interception, infiltration and percolation, and is routinely used in research and operations in Australia and worldwide (e.g., Woldemeskel et al., 2018). Upper and lower parameter bounds are specified as given by Evin et al. (2014).

Residual Error Model And Streamflow Transformations

Two residual error schemes are evaluated, differing in the streamflow transformation z used before the Gaussian error assumptions are applied. We use the term "residual error scheme" (or simply "scheme") to refer to a statistical model of streamflow errors, including the specification of the streamflow transformation function and the probability distribution of the model residuals.

The LOG residual error scheme employs the log transformation $z(q) = \log q$; the BC0.2 scheme employs the Box-Cox transformation $z(q) = (q^{\lambda} - 1) / \lambda$ with the power parameter λ fixed to 0.2. For a broader investigation using the BC0.5 (square root) transformation (McInerney et al., 2017) and the log-sinh transformation (Wang et al., 2012), see McInerney et al. (2019).

Performance Metrics

Four metrics are used to evaluate the quality (performance) of the daily streamflow predictions.

Reliability refers to the degree of statistical consistency between the observed streamflow time series and the predictive distribution. It is assessed visually using the predictive quantile-quantile (PQQ) plot (Thyer et al., 2009) and summarized using the reliability metric of Evin et al. (2014) with an enhancement to accommodate both continuous and discrete values (McInerney et al., 2019). When predictions are perfectly reliable, the PQQ plot follows the 1:1 line. The reliability metric value of 0 represents perfect reliability and a value of 1 represent the worst reliability (all observations lying either above or below the predictive distribution).

Precision (also known as sharpness or resolution; e.g., Franz et al., 2003) refers to the width of the predictive distribution, and is quantified using the average standard deviation of the uncertainty in the daily predicted streamflow (McInerney et al., 2017). Lower values of this metric correspond to better precision, with a value of 0 representing a point prediction.

Volumetric bias refers to the overall water balance error of the predictions, and is quantified as the (scaled) difference in the mean of the predictive distribution and the mean of the observed data. Lower values correspond to better performance, with a value of 0 representing unbiased predictions.

Ability to reproduce the observed proportion of zero flows (PZF), which we quantify as the difference in the proportion of zero flows in the predictions (averaged over both time steps and replicates) and the observed data (averaged over the time steps). Lower values indicate better performance.

Note that all metrics used in this work are dimensionless, which allows a direct comparison of predictive performance across multiple catchments.

The metrics are applied to streamflow time series constructed using a leave-one-year-out cross-validation procedure. A 2-year warmup period and 1-year period are omitted on either side of the evaluation window to reduce storage memory effects (McInerney et al., 2017).

RESULTS

Figure 3 compares the performance of the pragmatic and explicit approaches, displaying distributions of performance metrics for the four residual error schemes over the three catchment groups defined earlier. Figure 4 displays representative streamflow predictions obtained using the pragmatic and explicit approaches for selected residual error schemes and catchments.

Low-Ephemeral Catchments

In low-ephemeral catchments (PZF < 5%), the explicit approach has a generally small influence on predictive performance (Figure 3, left column). The choice of residual error scheme has greater impact. For example, for both the pragmatic and explicit approaches, the BC0.2 scheme has markedly better reliability (median metric of 0.16 versus 0.03 for the LOG schemes) and proportion of zero flows (median error of 10% versus 1% for the LOG scheme). In contrast, the Log schemes has notably worse precision (median metric of 0.6 and 0.8 respectively versus 0.45 for BC0.2).

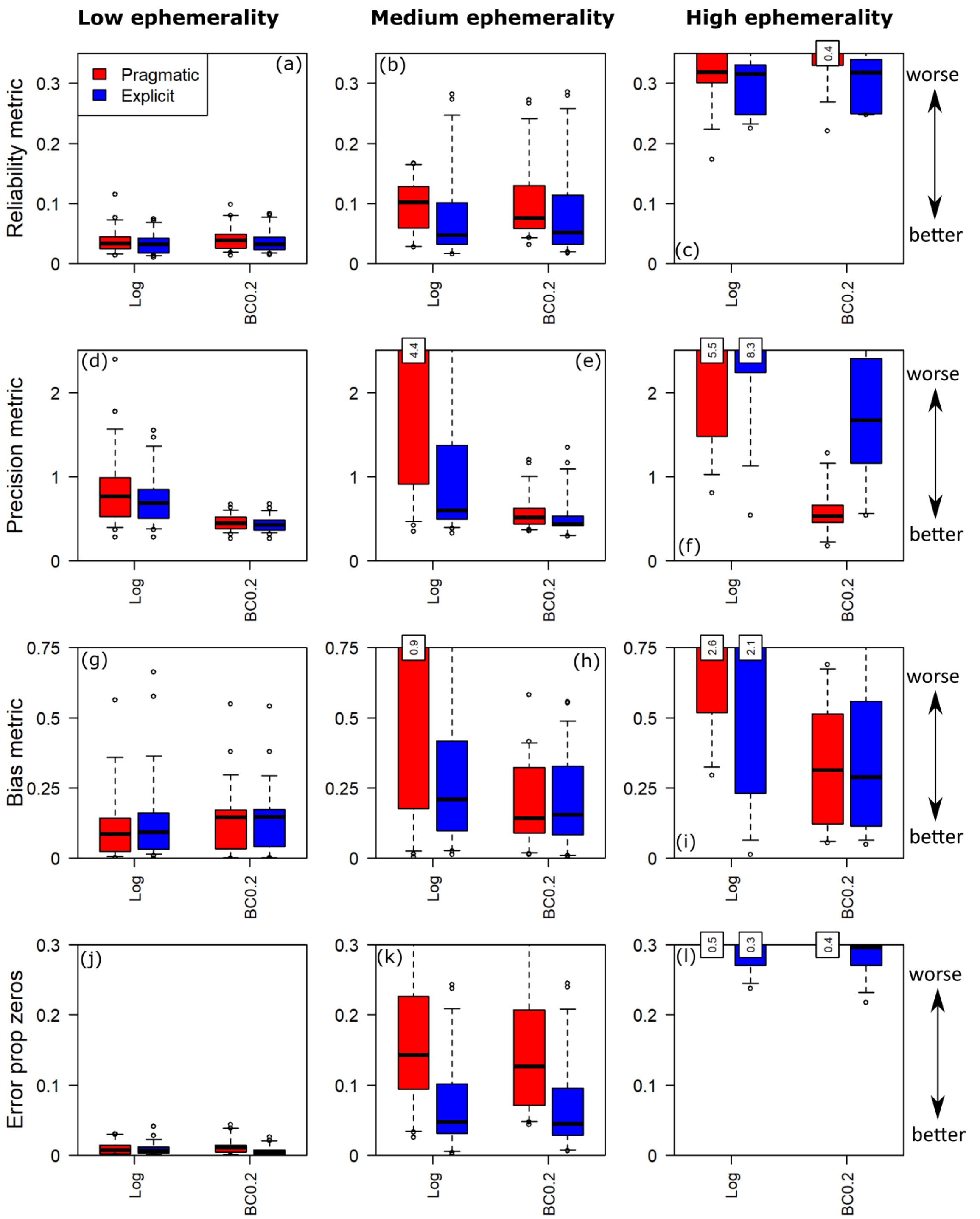

Figure 3. Predictive performance of the pragmatic and explicit approaches for handling zero flows. Results shown for two streamflow transformations, namely BC0.2 and LOG. Textboxes are used to state the median value of the metric if the latter is beyond the y-axis limits. Figure adapted with modification from McInerney et al. (2019).

That said, the treatment of zero flows still makes a notable impact in some low-ephemeral catchments. For example, in some catchments the explicit approach improves considerably the reliability metric for both the LOG and BC0.2 schemes. In addition, in some instances, the explicit approach improves the precision metric by a factor of ≈1.5 (corresponding to an average reduction in the spread of the predictions by a third) for the LOG schemes. A more detailed inspection indicates that improvements in reliability achieved by the explicit approach tend to become more pronounced as the proportion of zero flows increases from 0% to 5%, though this trend is noisy (McInerney et al., 2019).

Figure 4a,b illustrates, using the BC0.2 scheme as a representative example, how little is gained by the explicit approach in the low-ephemeral catchment 422334A (which has 3% zero flow days).

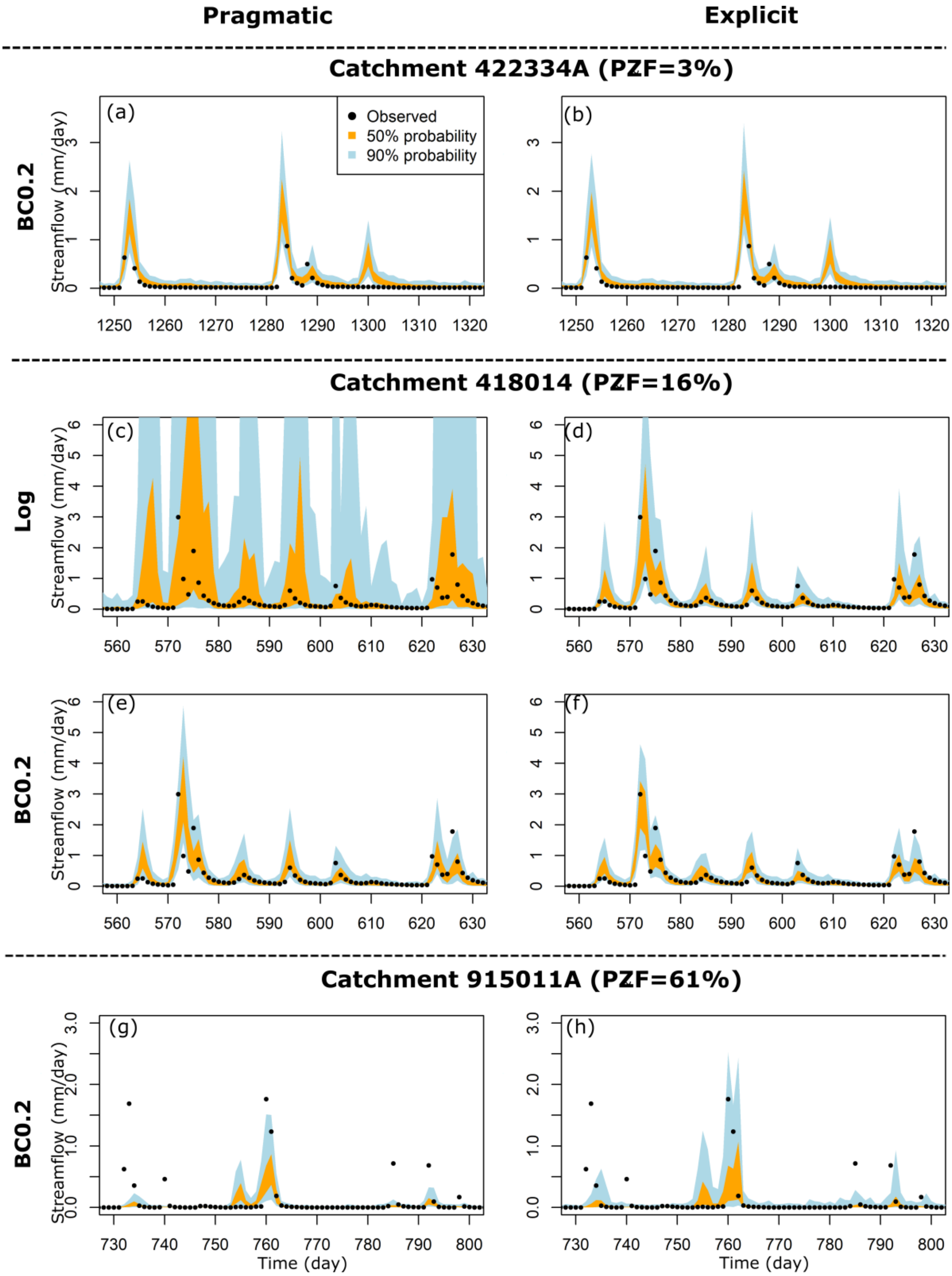

Figure 4. Representative probabilistic predictions of streamflow obtained using the pragmatic and explicit approaches in low-ephemeral, mid-ephemeral, and high-ephemeral catchments. Figure adapted with modification from McInerney et al. (2019).

Mid-Ephemeral Catchments

In mid-ephemeral catchments (PZF of 5-50%), the explicit approach achieves a clear improvement in predictive performance (Figure 3, middle column). The following results are noted:

• Reliability (Figure 3b): the explicit approach improves performance of all residual error schemes. For example, the median reliability metric is improved by around 50% (from 0.1 to 0.05) for the LOG schemes, and by around 30% (from 0.076 to 0.053) for the BC0.2 scheme.

• Precision (Figure 3e): the explicit approach is substantially superior to the pragmatic approach for the LOG scheme, with the median precision metric improving from 4.3 to 0.6. Indeed a detailed inspection (not shown) indicates that the explicit approach improves precision in all catchments of this group. For the BC0.2 scheme, the explicit approach improves precision but by a much smaller amount, and when PZF > 20% the results are mixed.

• Volumetric bias (Figure 3h): the explicit approach clearly improves the performance of the LOG scheme (e.g., median bias metric improves by a factor of 4), but is less useful for the BC0.2 scheme.

• Ability to reproduce the proportion of zero flows (Figure 3k): the explicit approach improves the performance of both the Log and BC0.2 schemes, e.g., the median value of the error in PZF reduces from 0.14 to 0.05.

Figure 4c-f shows how the treatment of zero flows impacts probabilistic predictions in the mid-ephemeral catchment 418014 (PZF = 16%). For the LOG scheme, the explicit approach is clearly beneficial, tightening the predictive distribution by a huge factor of nearly 20 (precision metric improves from 19 to 0.9). For the BC0.2 scheme, the predictions generated using the two approaches are more similar, with the explicit approach still achieving slightly tighter predictions (precision metric improves from 0.66 to 0.53).

High-Ephemeral Catchments

In high-ephemeral catchments (PZF > 50%), predictive performance metrics are poor regardless of the approach used (Figure 3, right column). For example, in terms of reliability, the explicit approach slightly outperforms the pragmatic approach (e.g., median reliability metric improves in all residual error schemes in Figure 3c), yet is still very poor compared to the performance achieved in less ephemeral catchments. Other performance metrics of the pragmatic and explicit approaches are similarly much worse than in the other ephemerality groups.

Figure 4g,h shows representative probabilistic predictions in the high-ephemeral catchment 915011A (PZF = 61%). The explicit approach generates predictive limits that capture observed data better than the pragmatic approach, yet neither approach is reliable. Most notably, the predictions substantially under-estimate most observed high flows. For example, over the time period shown in the plot, the 95% limits for the pragmatic approach lie below 88% of observed flows larger than 0.25 mm/day; for the explicit approach this fraction is 63%.

DISCUSSION

Is Explicit treatment of Zero Flows Beneficial and Why?

The improved performance of the explicit approach can be attributed to the theoretical arguments presented in the earlier sections. The deficiency of the likelihood function in the pragmatic approach, particularly its lack of treatment of zero flows, results in poor parameter estimates and hence poor streamflow predictions – even if the predictive distribution is censored to avoid negative flows.

In low-ephemeral catchments, there is little/minor impact on overall predictive performance (Figure 3, left column), especially for catchments where PZF < 3% (McInerney et al. 2019). This finding is unsurprising: when the proportion of zero flow observations is small, the inconsistency in the pragmatic approach can only manifest in a small number of days and can make at most a small impact on the numerical values of the likelihood function.

In high-ephemeral catchments, with PZF > 50%, both pragmatic and explicit approaches suffer poor performance (Figure 3), with residual error assumptions breaking down in both approaches (McInerney et al. 2019). This pronounced loss of performance, in particular, of reliability, is likely a consequence of the theoretical maximum of 50% probability of zero flows in the predictive distributions of these approaches, as elaborated in an earlier section.

Selecting Streamflow Transformations within the Residual Error Schemes

In probabilistic modelling, reliability is generally seen as paramount (Gneiting and Katzfuss, 2014), as it essentially corresponds to predictive uncertainty being characterized adequately (in a statistical sense). For example, in applications such daily streamflow forecasting and environmental flow modelling, where predictive uncertainty is typically large and hence of major interest, reliability is a particularly important performance attribute. Both the BC0.2 and LOG schemes achieve (relatively) high reliability and are hence suitable for probabilistic prediction. However, the BC0.2 scheme has better precision, i.e., it makes tighter predictions. The combination of reliability and precision is particular desirable, as a reliable but highly imprecise (vague) prediction may not provide a suitable basis for decision-making.

Note that in some practical applications hydrologists may choose to sacrifice some reliability to meet other critical performance attributes. For example, in catchment yield estimation (e.g., Nathan and McMahon, 2017), the primary focus is on reducing biases in estimated long-term streamflow volumes, the uncertainty in which tends to be small due to averaging/aggregation. In this case, different streamflow transformations may be of interest. For example, McInerney et al. (2019) find that the BC0.5 scheme (with Box-Cox power parameter of 0.5., i.e., the square-root transformation) has smaller bias than the BC0.2 scheme, albeit at the expense of poorer reliability.

Practical Considerations

Practical applications generally favour ease of implementation. Explicit approaches hold some disadvantages in this respect. For example, existing software or calibration approaches, such the Least Squares / Method of Moments (LS-MoM) approach of McInerney et al. (2018) and the enhanced linear-mean error model of Hunter et al. (2021) may be difficult to augment to censor zero flows. If a modeller elects to use the pragmatic approach and does not treat zero flows explicitly, the case studies presented here provide an indication of potential losses in predictive performance, which are reduced (though not eliminated) for the BC0.2 scheme. Generally speaking, residual error analysis can be used to help guide the selection of pragmatic versus explicit approaches in operational applications. Techniques for implementing residual error diagnostics suitable for regular and censored Gaussian error models are provided by McInerney et al. (2019).

A notable finding with practical implications is that the performance of the BC0.2 scheme is less affected by the choice of zero flow handling approach (pragmatic or explicit) than the performance of the LOG scheme. Therefore, if a modeller elects to not treat zeros flows explicitly, e.g., in order to take advantage of the simplicity of the pragmatic approach, the BC0.2 scheme remains a good practical choice.

Limitations and Future Work

A key limitation of the explicit approach used in this study is its theoretical maximum of 50% probability of zero flows, as explained in an earlier section. This limitation could be addressed by allowing the median of the predictive distribution to go below zero using additional shift/skew parameters (e.g., Schoups and Vrugt, 2010), or by explicitly modelling the probability of zero flows, e.g., by using zero-inflation approaches (e.g., Smith et al., 2010; Oliveira et al., 2018), or by specifying both simulated and observed streamflow as being samples from censored distributions (Wang et al., 2020).

Another important limitation is lack of treatment of persistence (autocorrelation) in the hydrological model residuals. Streamflow error persistence is important when aggregating predictions to longer time scales, e.g., when undertaking one-step-ahead prediction (Evin et al., 2014), constructing "seamless" forecasts (McInerney et al., 2020), and when predicting the duration of zero-flow periods relevant to ecological functions (e.g., Leigh and Sheldon, 2008).

These research directions will be pursued in future studies.

CONCLUSIONS AND RECOMMENDATIONS

Streamflow prediction in ephemeral catchments is challenged by the (often frequent) occurrence of days with zero observed streamflow. This study compared two post-processor approaches for treating zero flow: an "explicit" approach that employs censored Gaussian distributions in both calibration and prediction vs a simpler "pragmatic" approach that uses regular (uncensored) Gaussian error models in calibration and resets negative flows to zero in prediction.

Empirical case studies were conducted using a grand total of 74 Australian catchments with ephemerality in the range from virtually no zero flows to the majority of observed flows being zero. The daily catchment-scale rainfall runoff model GR4J was used. Within the residual error schemes, several common streamflow transformations were contemplated.

The following conclusions were reached:

1) Mid-ephemeral catchments, where 5-50% of daily flows are zero, are best modelled using the explicit approach. Performance improvements are especially pronounced in terms of reliability and ability to reproduce the proportion of zero flows. These findings can be attributed to the explicit approach making more realistic residual error assumptions in the likelihood function used to calibrate the hydrological and residual error model parameters.

In addition, the BC0.2 residual error scheme, where streamflow is transformed using the Box-Cox transformation with power parameter set to 0.2 was found to be preferable to LOG residual error scheme, which used the logarithmic transformation. For general probabilistic predictions, where reliability is of paramount importance, the BC0.2 scheme is hence recommended. The Log transformation yields worse precision and does not offer practical improvement with respect to other performance metrics. In addition, it is noted that the BC0.2 scheme reduces (though does not eliminate) the loss of predictive performance when the pragmatic approach is used.

3) Low-ephemeral catchments, with zero flows occurring on fewer than 0-5% of days, can be modelled satisfactorily using the pragmatic approach: the benefits of explicit treatment of zero flows are minor. Moreover, little difference is found between the BC0.2 and LOG schemes.

4) High-ephemeral catchments, where the (observed) proportion of zero flows exceeds 50%, pose considerable problems for both the pragmatic and explicit approaches and for both transformations. More complex error models, particularly those allowing for more than 50% probability of zero flows at any particular day, are therefore required.

The empirical findings agree with theoretical insights on the importance of implementing realistic probability models (here, censored Gaussian residual error models) consistently in calibration and prediction. In addition, the empirical results shed light on the balance between complexity and benefit in residual error models. Future work is recommended on streamflow prediction in high-ephemeral catchments, where limitations of the censored Gaussian probability model result in notably worse predictive performance than in low and mid-ephemeral catchments.

ACKNOWLEDGMENTS

This work was partially funded by Australian Research Council Linkage Grant LP140100978 in partnership with the Australian Bureau of Meteorology and South East Queensland Water. Computations were performed on the Phoenix cluster at the University of Adelaide. We are grateful to the Australian Bureau of Meteorology for providing the data underlying this research. Observed rainfall was obtained from `www.bom.gov.au/climate` and observed streamflow was obtained from `www.bom.gov.au/waterdata`. The fourth author (JL) was employed at the Bureau of Meteorology at the time the project was carried out. We thank Christopher Pickett-Heaps, Richard Laugesen and Narendra Tuteja for insightful comments on earlier versions of this work.

REFERENCES

Ang, A. H.-S. & Tang, W. H. 2007. *Probability concepts in engineering: emphasis on applications in civil & environmental engineering*, Wiley New York.

Bennett, J. C., Wang, Q. J., Li, M., Robertson, D. E. & Schepen, A. 2016. Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model. *Water Resources Research*, 52, 8238-8259.

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

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P. & Rieckermann, J. 2013. Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. *Hydrol. Earth Syst. Sci*., 17, 4209-4225.

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.

Franz, K. J., Hartmann, H. C., Sorooshian, S. & Bales, R. 2003. Verification of National Weather Service Ensemble Streamflow Predictions for Water Supply Forecasting in the Colorado River Basin. *Journal of Hydrometeorology*, 4, 1105-1118.

Gibbs, M, McInerney, D., Humphrey, G., Thyer, M. A., Maier, H., Dandy, G. & Kavetski, D. 2018. State updating and calibration period selection to improve dynamic monthly streamflow forecasts for an environmental flow management application. *Hydrol. Earth Syst. Sci*., 22, 871-887.

Gneiting, T. & Katzfuss, M. 2014. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1, 125-151.

Hunter J., Thyer M., McInerney D., Kavetski D. (2021) Achieving high-quality probabilistic predictions from hydrological models calibrated with a wide range of objective functions, *Journal of Hydrology*, in press, doi:10.1016/j.jhydrol.2021.126578.

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

Leigh, C. & Sheldon, F. 2008. Hydrological changes and ecological impacts associated with water resource development in large floodplain rivers in the Australian tropics. *River Research and Applications*, 24, 1251-1270.

Li, M., Wang, Q. J. & Bennett, J. 2013. Accounting for seasonal dependence in hydrological model errors and prediction uncertainty. *Water Resources Research*, 49, 5913-5929.

McInerney, D., Thyer, M., Kavetski, D., Bennett, B., Lerat, J., Gibbs, M. & Kuczera, G. 2018. A simplified approach to produce probabilistic hydrological model predictions. *Environmental Modelling & Software*, 109, 306-314.

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*, 53.

McInerney, D., Kavetski, D., Thyer, M., Lerat, J. and Kuczera, G. (2019) Benefits of explicit treatment of zero flows in probabilistic hydrological modeling of ephemeral catchments, *Water Resources Research*, 55, 11035-11060, doi:10.1029/2017WR024148.

McInerney D., Thyer M., Kavetski D., Laugesen R., Tuteja N., Kuczera G. (2020) Multi-temporal hydrological residual error modelling for seamless sub-seasonal streamflow forecasting. *Water Resources Research*, 56, doi:10.1029/2019WR026979.

Min, Y. & Agresti, A. 2002. Modeling nonnegative data with clumping at zero: a survey. *Journal of the Iranian Statistical Society*, 1, 7-33.

Morawietz, M., Xu, C.-Y., Gottschalk, L. & Tallaksen, L. M. 2011. Systematic evaluation of autoregressive error models as post-processors for a probabilistic streamflow forecast system. *Journal of Hydrology*, 407, 58-72, doi:10.1016/j.jhydrol.2011.07.007.

Nathan, R. J. & McMahon, T. A. 2017. Recommended practice for hydrologic investigations and reporting. *Australasian Journal of Water Resources*, 21, 3-19.

Oliveira, D., Chaffe, P. & J., S. 2018. Extending the Applicability of the Generalized Likelihood Function for Zero-Inflated Data Series. *Water Resources Research*, 54, 2494-2506.

Pagano, T. C., Shrestha, D. L., Wang, Q. J., Robertson, D. & Hapuarachchi, P. 2013. Ensemble dressing for hydrological applications. *Hydrological Processes*, 27, 106-116.

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.

Pianosi, F. & Raso, L. 2012. Dynamic modeling of predictive uncertainty by regression on absolute errors. *Water Resources Research*, 48, W03516, doi:10.1029/2011WR010603.

Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., et al. 2017. Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales. *Hydrological Sciences Journal*, 62, 2753-2768.

Salem, B. 1989. Arid zone forestry: a guide for field technicians, *Food and Agriculture Organization*.

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.

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.

Stern, H., De Hoedt, G. & Ernst, J. 2000. Objective classification of Australian climates. *Australian Meteorological Magazine*, 49, 87-96.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. & Srikanthan, S. 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modelling: a case study using bayesian total error analysis. *Water Resources Research*, 45, W00B14, doi:10.1029/2008WR006825.

Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26, 24-36.

Tuteja, N. K., Shin, D., Laugesen, R., Khan, U., Shao, Q., Wang, E., Li, M., Zheng, H., Kuczera, G. & Kavetski, D. 2012. *Experimental evaluation of the dynamic seasonal streamflow forecasting approach*. Bureau of Meteorology, Melbourne.

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

Wang, Q. J. & Robertson, D. E. 2011. Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, 47, W02546.

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, W05514.

Wang, Q. J., Bennett, J. C., Robertson, D. E., & Li, M. (2020). A data censoring approach for predictive error modeling of flow in ephemeral rivers. *Water Resources Research*, 56, e2019WR026128. Doi: 10.1029/2019WR026128.

Westra, S., Thyer, M., Leonard, M., Kavetski, D. & Lambert, M. 2014. A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resources Research*, 50, 5090-5113.

Woldemeskel, F., McInerney, D., Lerat, J., Thyer, M., Kavetski, D., Shin, D., Tuteja, N. & Kuczera, G. 2018. Evaluating residual error approaches for post-processing monthly and seasonal streamflow forecasts. *Hydrol. Earth Syst. Sci. Discuss*., 2018, 1-40.

BIOGRAPHY

Dmitri Kavetski is a Professor of Civil and Environmental Engineering at the University of Adelaide. His research work focuses on hydrological modelling, where he has contributed a range of new techniques for model development, calibration and prediction. A particular emphasis has been on understanding, quantifying, and where possible reducing, the sources of uncertainty in streamflow predictions at the catchment scale. Dmitri's broader research interests include probabilistic methods, numerical modelling and optimization. Dmitri also delivers regular short courses and workshops on hydrological modelling to Australian and international audiences.