

Recent learnings towards achieving high quality probabilistic predictions in practical applications of hydrological models

Dmitri Kavetski, David McInerney, Jason Hunter and Mark Thyer

*School of Civil, Environmental and Mining Engineering, University of Adelaide.
Email: dmitri.kavetski@adelaide.edu.au*

Abstract: Hydrological predictions – and environmental predictions in general – are inevitably affected by substantial uncertainty due to limitations in data quality and model simplifications. Probabilistic predictions, where predictive uncertainty is represented using probability distributions, provide a vehicle for reflecting these uncertainties. Initially limited to research applications, probabilistic predictions are increasingly appealing in practical applications to help guide risk-based decision-making. For example, in water resource modelling, probabilistic predictions can provide reliable estimates of streamflow in the days and months ahead – thus avoiding a false sense of security in applications as diverse as dam operation, drought management, ecological health monitoring, and so forth. Nevertheless, a number of barriers still exist that impede a wider adoption of probabilistic predictions in practical hydrological work. The underlying theory is often perceived as complicated, and the predictions themselves are often perceived to be difficult to construct and interpret.

In this paper we demonstrate that simple practical approaches are increasingly available to achieve high quality probabilistic predictions in a wide range of practical hydrological prediction applications. Our target audience are hydrologists in research and operations who are familiar with traditional hydrological modelling concepts and would like to extend their expertise to probabilistic modelling. A major focus is on providing guidance for robust descriptions of uncertainty using residual error models. This guidance includes the choice of transformation to handle common features of residual errors (heteroscedasticity, skewness, persistence), on estimation techniques for residual error model parameters, guidance on the relative importance of hydrological parameter uncertainty, and finally techniques to achieve high quality predictions from hydrological models already calibrated using a wide range of common objective functions. A software tool that implements these techniques is provided. The approaches are based on key insights from multiple research papers over multiple years, and their robustness has been shown over numerous catchments.

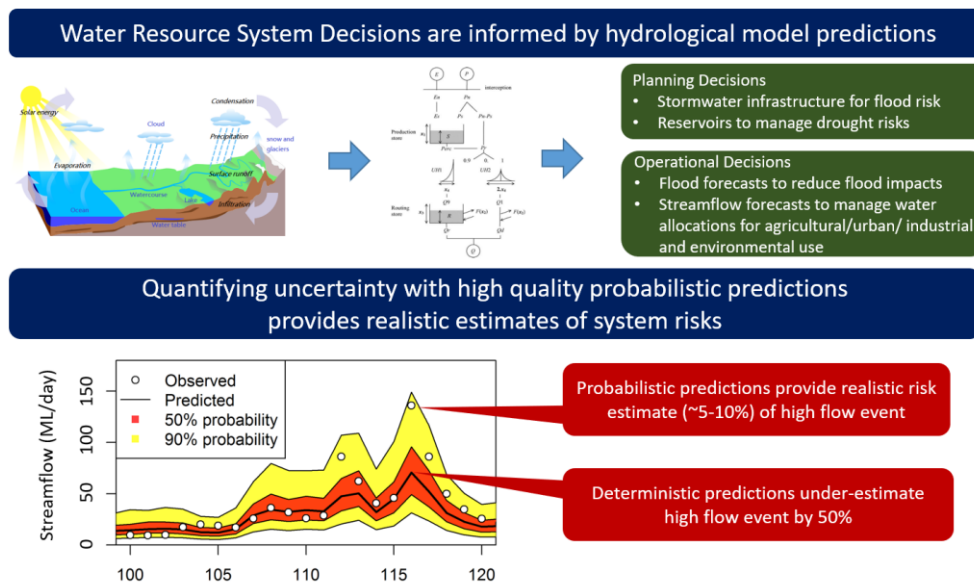


Figure 1. Construction and use of probabilistic predictions in practical hydrological applications

Keywords: *Hydrological modelling, streamflow, predictive uncertainty, probabilistic prediction, residual errors, practical model applications*

1. INTRODUCTION

Streamflow predictions from hydrological models are essential inputs to the planning and operation of water resource systems; see Figure 1. Amongst the key requirements is the ability to represent the considerable predictive uncertainty arising from the combined effect of data and model uncertainties. Probabilistic inference and prediction, where uncertainty is described using probability distributions, is of particular interest in this regard. Probabilistic techniques are well-known in the hydrological research community, ranging from methods-of-moments to frequentist methods to full Bayesian techniques (e.g., Kuczera, 1983; Krzysztofowicz, 2002; Schoups and Vrugt, 2010; Smith et al., 2010; Kavetski, 2019; Li et al., 2016; McInerney et al., 2017).

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 are used to generate “point” predictions. Quantification of uncertainty in predictions is often considered the domain of applied research (Vaze et al., 2011). One of the key challenges is the perception that probabilistic predictions require substantial additional effort (McInerney et al., 2018). However, limiting decision-making to deterministic predictions can lead to a “false sense of security” (Vogel, 2017): predictive uncertainty is ignored and system risks can be substantially under-estimated.

This paper aims to increase the uptake of probabilistic predictions in practical applications, by providing guidance on the techniques found to achieve high quality predictions in a wide range of scenarios.

2. PROBABILISTIC MODELLING AT A GLANCE

2.1. How to construct a probabilistic model from a deterministic model?

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

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

The simplest approach to describe predictive uncertainty in such a model is to add a “residual error” term η_t ,

$$z(Q_t) = z(q_t^\theta) + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \tag{2}$$

In equations (2)-(3) the error is assumed to be Gaussian with zero mean and constant variance *after* a streamflow transformation z is applied. The choice of transformation is described below.

Following model calibration, the simplest approach to generate probabilistic streamflow predictions is by sampling replicates from the assumed probability model with optimal parameters $\hat{\theta}$,

$$q_t^{\text{pred},r} \leftarrow Q_t(\hat{\theta}, \mathbf{x}_{1:t}) \tag{3}$$

The full set of predictive replicates, $r = 1, \dots, N_{\text{rep}}$, either at a given step t or a series of steps, will be referred to as the “predictive distribution”. Here, the predictive distribution is a (transformed) Gaussian distribution with a time-varying median set to the deterministic predictions. Figure 2 provides a simple schematic illustration.

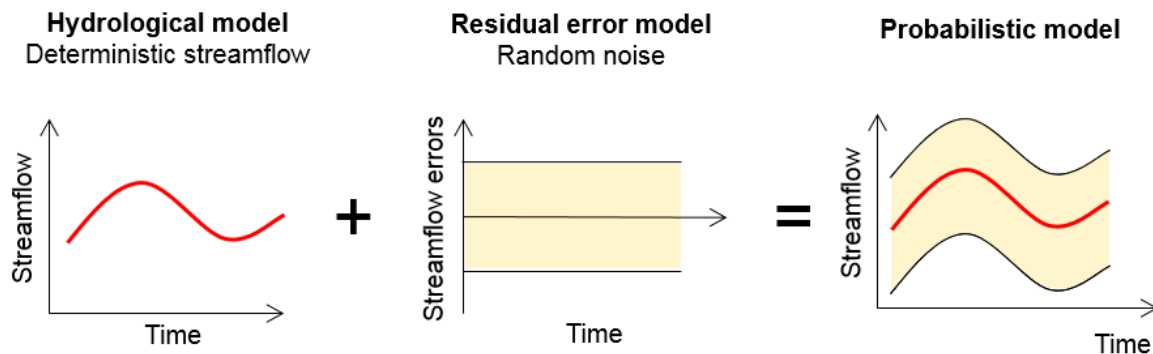


Figure 2. Constructing a probabilistic model from a deterministic model by adding a residual error model. For simplicity, the schematic shows a constant variance error model.

Multiple technical questions arise when implementing the probabilistic model and interpreting its predictions:

1. How to select the streamflow transformation and represent the error persistence?
2. How to estimate the hydrological and residual error model parameters?
3. How important is parameter uncertainty versus total predictive uncertainty?
4. How to evaluate the goodness of fit of a probabilistic prediction
5. How to produce probabilistic predictions from already calibrated hydrological models?

The following sections attempt to "de-mystify" these questions for a general hydrologist.

3. IMPLEMENTING A PROBABILISTIC MODEL IN PRACTICE

3.1.1. How to select the streamflow transformation?

A key feature of residual errors of hydrological models is their overall dependence on the flow magnitude: predictions of high flows are typically more uncertain than those of low flows; see Figure 3. The streamflow transformation is intended to stabilise the error variance – in technical terms, reduce error "heteroscedasticity". A common choice is the Box-Cox transformation $z(q) = (q^\lambda - 1) / \lambda$, which reduces to the log transformation $z(q) = \log q$ when $\lambda = 0$ (Box and Cox, 1964). An offset parameter can be added to avoid numerical problems when $q \approx 0$. The log transformation corresponds to the assumption that the errors are multiplicative rather than additive, which intuitively provides a better match to the actual errors. An additional benefit is the representation of skew and fatter tails in the distribution of errors: the probabilistic prediction is log-normal with its (time-varying) median equal to the deterministic prediction. For these reasons, the log transformation tends to outperform Weighted Least Squares approaches that make the error variance flow-dependent but retain the symmetric Gaussian error model (Figure 4a; McInerney et al., 2017).

However, the log transformation is sensitive for zero and near-zero flows, degrading predictions in arid and semi-arid ephemeral catchments. A "milder" transformation, given by $\lambda = 0.2$, tends to be preferable. Another choice, $\lambda = 0.5$, which corresponds to the square-root transformation, can be advantageous if emphasis is on capturing high flows (Figure 4b). Tradeoffs in the choice of λ are discussed in McInerney et al. (2017).

3.1.2. How to represent the persistence of hydrological model errors?

Another notable feature of residual error is their persistence. Figure 3 shows streamflow errors at consecutive time steps often form long stretches with the same sign and similar magnitude. This behavior arises because of the storage memory effects in hydrological models – and indeed in the underlying physical catchment processes. The presence of persistence has two implications on predictions: (i) it enables more precise short-term predictions, because the error at the next time can be related to the error at the current (observed) time step; and (ii) it impacts on aggregated predictions, in particular causing an under-estimation of uncertainty if the error model ignores this persistence and instead assumes independent errors (Evin et al., 2014).

Error persistence can be represented using an auto-regressive model, in the simplest case an AR(1) model. Some care is needed when combining the treatment of heteroscedasticity and persistence: as shown by Evin et al. (2013), heteroscedasticity should be treated first followed by the application of the AR(1) assumptions.

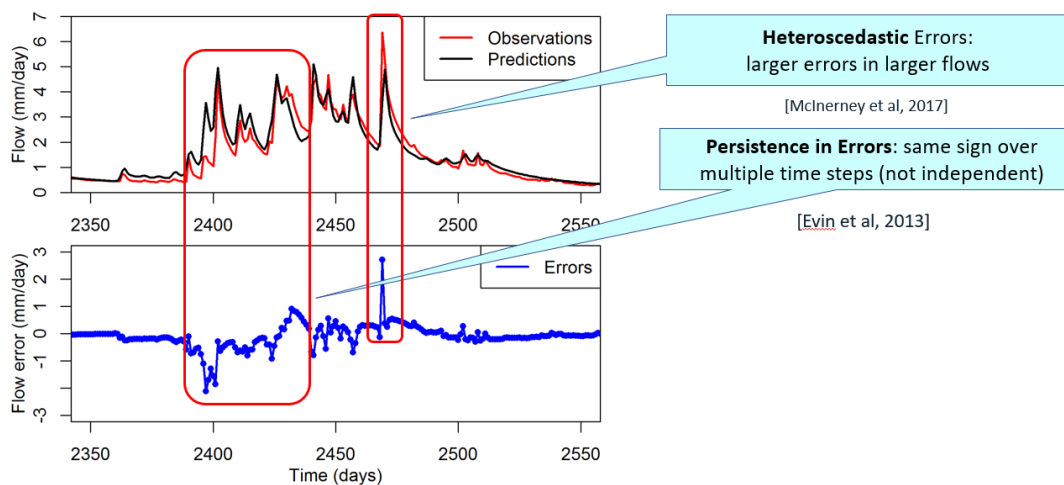
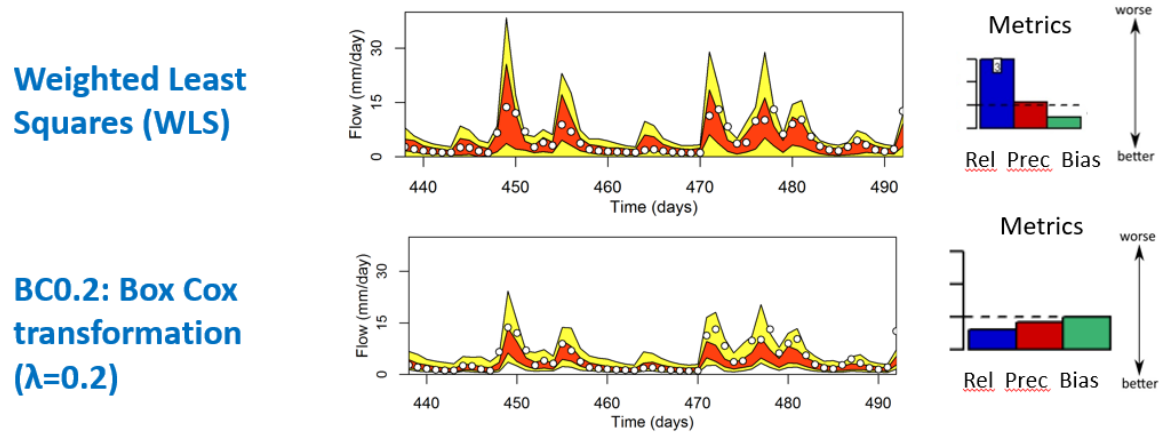
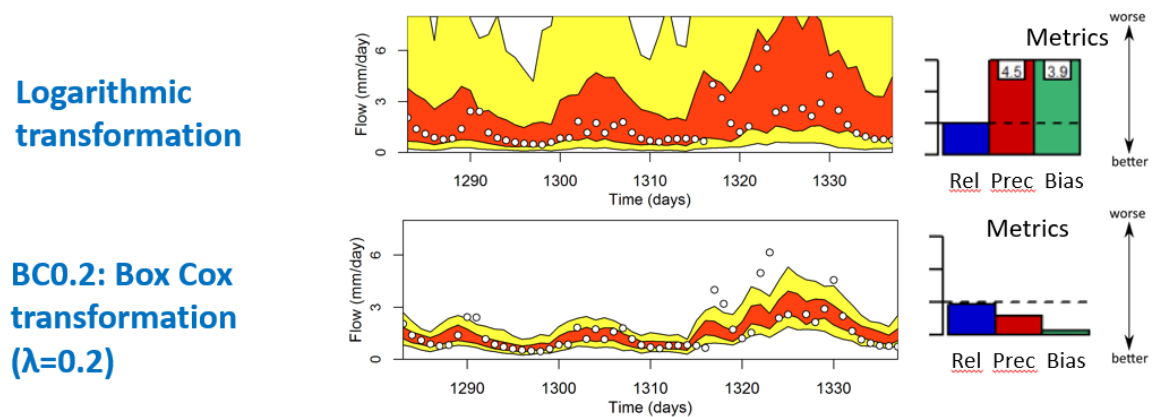


Figure 3. Residual errors of hydrological models and their key characteristics.



(a) Perennial catchments: BC0.2 transformation provides better reliability and precision than the WLS scheme with error variance proportional to streamflow: Representative catchment: Spring River, USA,



(b) Ephemeral catchments: BC0.2 transformation achieves much better precision and low bias than the log transformation. Representative catchment: Rocky River, Australia

Figure 4. Streamflow transformation and probabilistic performance. From McInerney et al. (2017).

3.1.3. How to estimate the hydrological and residual error model parameters?

A probabilistic model, in addition to the parameters of the (original) hydrological model, θ , also has error model parameters. In particular, the error variance σ_η^2 is seldom known a priori – but can be estimated in the calibration period. Traditional statistical estimation focuses on "joint" (simultaneous) estimation of hydrological and error model parameters using a likelihood function (e.g., Kuczera, 1983). However, in practice we recommend a "postprocessor" strategy, where Stage 1 estimates hydrological parameters using an objective function and Stage 2 estimates error model parameters keeping the hydrological parameters fixed.

The postprocessor approach offers two important pragmatic benefits:

1. Stage 1 can be implemented using common objective functions such as NSE or RMSE, for which correspondence to common Gaussian likelihood functions can be shown (e.g., McInerney et al., 2018). Moreover, in cases where a calibrated hydrological model is already available – the postprocessor approach spares the modeler the considerable effort of re-calibrating it (Section 5.1.1). A key simplification is that error persistence is ignored in Stage 1 (see point 3 below). Note that the objective functions can themselves employ the Box-Cox transformation for consistency with the error model;
2. Stage 2 can employ conceptually simple statistical methods, e.g., using the Method of Moments to estimate the error variance and autoregressive coefficient from the (fixed) time series of residuals
3. More robust numerical behavior, avoiding problematic interactions between mass balance parameters and the autoregressive error parameter reported by Evin et al. (2014) in joint inference.

As such, the postprocessor approach is an appealing pragmatic choice for practical work (Hunter et al., 2021).

4. HOW "GOOD" IS A PROBABILISTIC PREDICTION?

Probabilistic performance can be viewed in terms of three key attributes: reliability, precision and bias.

Reliability refers to the degree of statistical similarity between the predictive distribution and the distribution of observed data. This perspective connects to traditional statistical tools such as quantile-quantile plots (and metrics such as the Kolmogorov-Smirnoff statistic) – but is applied in a way that recognizes that the predictive distribution is time-varying. Visual tools include the Predictive Quantile-Quantile (PQQ) plot (Thyer et al., 2009), which – similar to regular quantile-quantile plots – follow a 1:1 line if the predictions are reliable. Numerical reliability metrics can be devised to quantify this discrepancy (Evin et al., 2014).

Precision (also known as sharpness or resolution) refers to the width of the predictive distribution, once again generalised to accommodate its time-varying nature. E.g., the standard deviation of the predictive distribution, averaged over the time steps, can be used (McInerney et al., 2017). Note the distinction between precision versus the width of the marginal distribution of streamflow (flow duration curve). The former is a measure of probabilistic performance (tighter is generally better), whereas the latter is a characteristic of streamflow itself.

Volumetric bias (overall water balance error of the predictions) is of key importance in many hydrological applications, especially those concerned with long-term yield estimation. It can be quantified as the difference between the mean of the predictive distribution (across time) and the mean of the observed data.

In practice, there are tradeoffs between performance attributes. For example, tight (precise) predictions often come at the expense of lower reliability (e.g., McInerney et al. 2017). In general, reliability is seen as paramount in probabilistic modelling, because an (overconfident) prediction that is tight but unreliable is ultimately misleading to the user. On the other hand extremely vague predictions are themselves of little value. It is also possible to devise composite metrics, such as the Continuous Rank Probability Score (CRPS; Hersbach, 2000) that combine all three attributes – though we find it more informative to use more targeted metrics.

5. MAKING PROBABILISTIC PREDICTIONS FROM ALREADY CALIBRATED MODELS

5.1.1. The linear-mean error model to avoid biases with "inconsistent" objective functions

In practice, a hydrological model may already have been calibrated using some pre-selected objective function. A natural question is then whether such a model can be used directly as the basis for probabilistic predictions, i.e., without recalibration (or repeating Stage 1). For example, in the postprocessor approach of Section 3.1.3, could we treat the already calibrated model as the outcome of Stage 1 and proceed directly to Stage 2?

A naïve implementation of this approach using the error model in equation (3) runs into problems. For example, if the hydrological model was originally calibrated using the (Stage 1) NSE objective function without streamflow transformation (i.e., $\lambda = 1$), using an (Stage 2) error model that employs a Box-Cox transformation with $\lambda = 0-0.5$ will tend to introduce strong biases in the predictions (see Hunter et al. 2021 for details). The root cause is the "inconsistency" between the stages, here due to differences in their respective transformations.

A remedy for the bias problem is to allow the mean of the errors to depend on the simulated (deterministic) streamflow. The following linear model is suggested,

$$\mu_t = \alpha + \beta \cdot z(q_t^0) \quad (4)$$

The parameters α and β of this linear model can be estimated from the residuals along with the error variance.

As seen in Figure 5, the "linear-mean" enhancement restores the quality of the predictions to similar levels as when a "consistent" objective function is used (same value of λ in both stages). Remarkably, the improvements hold for a wide range of objective functions, from Nash-Sutcliffe and Kling-Gupta efficiencies with Box-Cox transformation parameter values of $\lambda = 0-1$, to composite objective functions combining the NSE with mass balance penalty terms, to the Pearson correlation coefficient, across 54 study catchments (Hunter et al. 2021).

The ability to obtain high quality probabilistic predictions with a wide range of objective functions substantially broadens the range of practical modelling problems for which probabilistic predictions can be employed.

5.1.2. How important is parameter uncertainty versus total predictive uncertainty?

It is commonly assumed that, given the considerable simplifications made in environmental models, parameter uncertainty must be necessarily high (e.g., Beven et al., 2006). However, if predictive uncertainty is represented using a residual error model, its magnitude is controlled by the error variance. When a parsimonious rainfall-

runoff model is calibrated to a sufficiently long time series (e.g., multiple years of daily streamflow), neglecting parameter uncertainty often has a very minor impact on probabilistic predictions (e.g., see Figure 9 in Kavetski et al., 2019; Sun et al. 2017, and others). When applicable, this pragmatic choice can produce substantial computational savings, e.g., by avoiding costly Monte Carlo parameter sampling, etc.

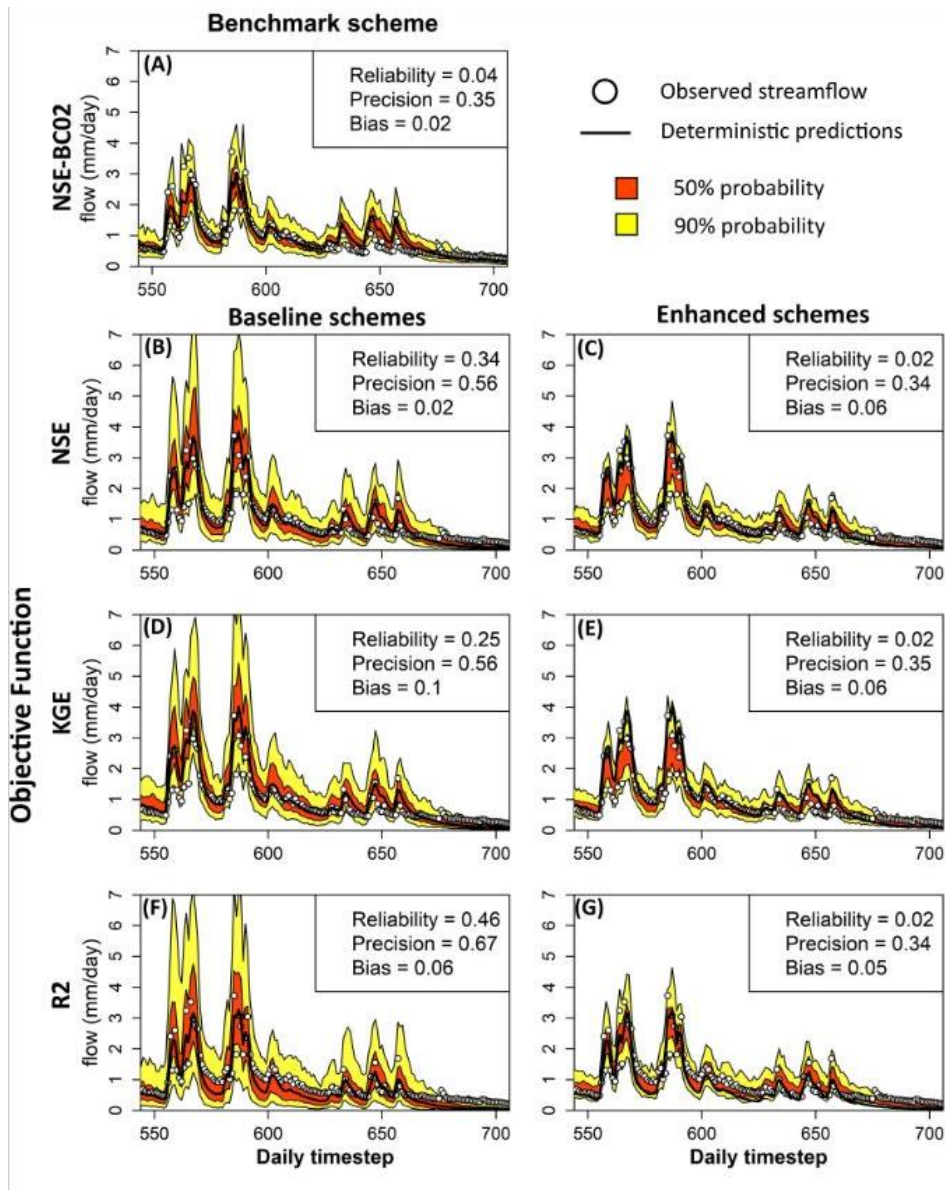


Figure 5. Achieving high quality predictions from hydrological models calibrated using a wide range of objective functions. Results shown for Yackandadah catchment. Adapted from *Hunter et al. (2021)*.

5.1.3. Software tool

The residual error modelling methods summarized in this work are implemented in the R package "probabilisticPredictions", available at https://github.com/Jasenter/Probabilistic_App. The software can be interfaced as either an R-function for batch computation or an R-shiny web application for interactive analysis.

6. SUMMARY AND CONCLUSIONS

Probabilistic methods offer the promise of reliable description of uncertainty in practical hydrological and environmental modelling applications – but suffer from a perception that these methods are difficult to apply.

This paper reviews a number of advances in probabilistic streamflow predictions, with an emphasis on practicality. The following key learnings are reported:

1. Probabilistic predictions can be obtained by enhancing a deterministic hydrological model with a residual error model. Their parameters can be estimated in two respective stages. This strategy enables enhancing rather than replacing existing modelling tools.
2. The use of Box-Cox streamflow transformation with power parameter $\lambda = 0-0.5$ is important to capture error heteroscedasticity, with value $\lambda = 0.2$ offering a good practical tradeoff in most cases. Error persistence can be represented using an AR(1) model, which can be applied after the streamflow transformation.
3. Allowing the error mean to depend on predicted streamflow enables high-quality predictions to be achievable with hydrological models already calibrated using common objective functions.

Collectively, these recommendations enable high-quality probabilistic predictions to be obtained in a wide range of practical applications – paving the way for enhanced uptake of these methods.

REFERENCES

- Beven K. (2006) A manifesto for the equifinality thesis, *Journal of Hydrology*, 320 (1-2), 18-36.
- Box G. E. P. and D. R. Cox (1964) An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, 26, 211-252
- Evin G., Thyer M., Kavetski D., McInerney D. and G. Kuczera (2014) Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity, *Water Resources Research*, 50, 2350-2375.
- Evin G., Kavetski D., Thyer M. and G. Kuczera (2013) Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration, *Water Resources Research*, 49(7), 4518-4524, doi:10.1002/wrcr.20284.
- Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast*, 15(5), 559–570.
- Hunter J., Thyer M., McInerney D. and Kavetski D. (2021) Achieving high-quality probabilistic predictions from hydrological models calibrated with a wide range of objective functions, *Journal of Hydrology*, 603A, doi:10.1016/j.jhydrol.2021.126578.
- Kavetski D. (2019) Parameter estimation and predictive uncertainty quantification in hydrological modelling, in *Handbook of Hydrometeorological Ensemble Forecasting*, Duan Q., Pappenberger F. Wood A., Cloke H. and Schaake J. (eds), Springer Berlin Heidelberg. doi: 10.1007/978-3-642-39925-1_25.
- Krzysztofowicz R. (2002) Bayesian system for probabilistic river stage forecasting, *Journal of Hydrology*, 268(1-4), 16-40.
- Kuczera G. (1983) Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Water Resources Research*, 19(5), 1151-1162, doi: 10.1029/WR019i005p01151.
- Li M., Wang Q. J., Bennett J. C. and D. E. Robertson (2016) Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting, *Hydrol. Earth Syst. Sci.*, 20(9), 3561-3579, doi: 10.5194/hess-20-3561-2016.
- McInerney D., Thyer M., Kavetski D., Lerat J., and G. Kuczera (2017) Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors, *Water Resour Res*, 53(3), 2199-2239, doi: 10.1002/2016WR019168.
- McInerney D., Thyer M., Kavetski D., Bennett B., Lerat J., Gibbs M., and G. Kuczera (2018) A simplified approach to produce probabilistic hydrological model predictions, *Environmental Modelling & Software*, 109, 306-314, doi: https://doi.org/10.1016/j.envsoft.2018.07.001.
- Schoups G., and J. A. Vrugt (2010) A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors, *Water Resources Research*, 46(10), W10531, doi:10.1029/2009WR008933.
- Smith T., Sharma A., Marshall L., Mehrotra R., and S. Sisson (2010) Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, *Water Resources Research*, 46(12), W12551, doi:10.1029/2010WR009514.
- Sun, R., Huiling, Y. and Liu X. (2017) Effect of heteroscedasticity treatment in residual error models on model calibration and prediction uncertainty estimation, *Journal of Hydrology*, 554, 680-692.
- Thyer M., Renard B., Kavetski D., Kuczera G., Franks S. W. and Srikanthan S. (2009) Critical evaluation of consistency and predictive uncertainty in hydrological modelling: A case study using Bayesian total error analysis, *Water Resources Research*, 45, W00B14.
- Vaze, J., P. Jordan, R. Beecham, A. Frost, and G. Summerell (2011) Guidelines for rainfall-runoff modelling.
- Vogel, R. M. (2017) Stochastic watershed models for hydrologic risk management, *Water Security*, 1, 28-35, doi: http://dx.doi.org/10.1016/j.wasec.2017.06.001.