Mark Thyer, David McInerney, Dmitri Kavetski, Jason Hunter

Practical approaches to produce high-quality probabilistic predictions and improve risk-based design making

Proceedings of the Hydrology and Water Resources Symposium (HWRS 2021), 2021, pp.612-625

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

Practical approaches to produce high-quality probabilistic predictions and improve risk-based design making

Mark Thyer, David McInerney, Dmitri Kavetski, Jason Hunter

School of Civil, Environmental and Mining Engineering, University of Adelaide

Email: mark.thyer@adelaide.edu.au

ABSTRACT

Probabilistic predictions provide crucial information regarding the uncertainty of hydrological predictions, which are a key input for risk-based decision-making. High-quality probabilistic predictions provide reliable estimates of water resource system risks – avoiding a false sense of security. However, probabilistic predictions are not widely used in hydrological modelling applications because they are perceived to be difficult to construct and interpret. We present a software tool that provides an easy-to-use and simple approach to produce high-quality probabilistic streamflow predictions. The approach integrates the recommendations from multiple research papers over multiple years to provide guidance on selection of robust descriptions of uncertainty (residual error models) for a wide range of hydrological applications. This guidance includes the choice of transformation to handle common features of residual errors (heteroscedasticity, skewness, persistence) and techniques that handles a wide range of common objective functions. A case study illustrating the practical benefits of uncertainty analysis for risk-based decision- making is provided. The case study evaluates fish health in two catchments (Mt. McKenzie and Upper Jacobs) in Barossa Valley, South Australia. The streamflow predictions of environmental flow metrics are combined with a simplified environmental response model to estimate fish health. The outcomes obtained using deterministic streamflow predictions are contrasted to the outcomes obtained from probabilistic predictions. In general, probabilistic predictions provide greater confidence in the predictions of fish health because the uncertainty ranges recognise the differences at the two sites between the quality of hydrological predictions. The uncertainty ranges were generally high, in the range 40-60% (Mt McKenzie) or 4-20% (Upper Jacobs) for predictions of the frequency of years with poor (or worse) fish health. This analysis provides a richer source of information for risk averse decision-makers than the single values provided by deterministic predictions.

INTRODUCTION

Predictions from hydrological models provide essential inputs to the planning and operation of water resource systems (Loucks et al., 1981) – see Figure 1 for examples. 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 with hydrological modelling 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).

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 it not commonly used as it is typically considered the domain of applied research (Vaze et al. 2012). One of the key challenges is the perception that probabilistic predictions require substantial additional effort (McInerney et al. 2018) This perception can lead to reduction in the update of probabilistic techniques. The practical impact of this is that using only deterministic predictions can lead to a "false sense of security" (Vogel 2017), because predictive uncertainty is ignored and system risks are under-estimated (see Figure 1 for illustrative examples)

This paper focuses on addressing this research gap in order to increase the uptake of probabilistic predictions in practical applications. The following aims are pursued:

1. Introduce a simple and easy-to-use software tool that can produce high quality probabilistic predictions
2. Demonstrate the practical benefits of high-quality probabilistic predictions on a case study involving environmental decision making

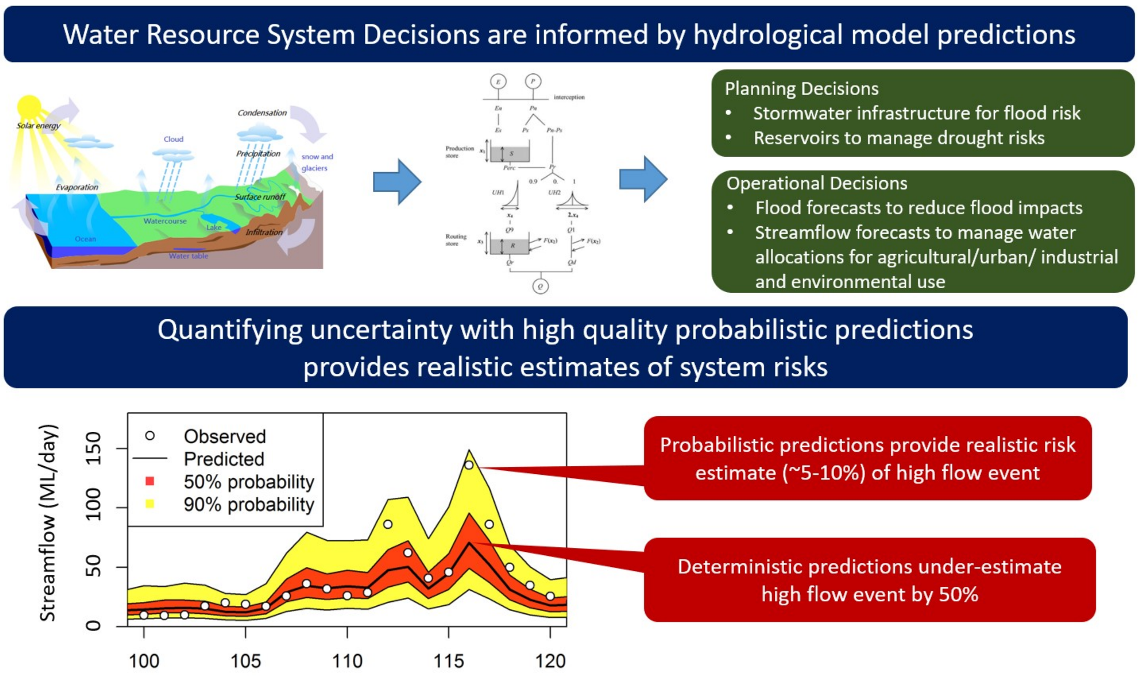


Figure 1: Motivation for use and application of probabilistic predictions.

OVERVIEW OF SOFTWARE TOOL FOR HIGH QUALITY PROBABILISTIC PREDICTIONS

Description

The software tool utilises a range of research advances in residual error model development to provide high quality probabilistic predictions. The tool is easy to use: it is provided as open source software with an easy-to-use web app interface. It is based on the post-processing approach, where probabilistic predictions are constructed based on a pre-existing deterministic prediction using a residual error model. In practical terms, users simply need to upload a time series of their deterministic predictions and available observations. Guidance on selection of residual error model parameters is provided to achieve high quality probabilistic predictions for a range of catchment types (perennial/ephemeral). The aim is to increase the uptake of probabilistic predictions by the researchers and practitioners.

The key features of this software tool are that it

1. Provides guidance on obtaining high quality probabilistic predictions

2. Is simple and easy to use

3. Evaluates predictive performance using a range of commonly used metrics and diagnostics

Each key feature will be explained in more detail below. Note that the current version of the software that is publicly available is a beta version. Users with an interest in utilising probabilistic predictions are encouraged to contact the authors in the first instance to discuss their needs.

Key Feature 1: Provides high quality probabilistic predictions

A key feature of the software tool is the provision of high quality probabilistic predictions by incorporating research advances to handle common features of residual errors in hydrological models (Figure 2), including:

- Guidance on choice of transformation and parameter selection to handle heteroscedasticity (see Table 1 and (McInerney et al. 2017))
- Incorporation of persistence to provide reliable predictions when aggregating from daily to monthly time scale (Evin et al. 2014)
- Incorporation of flow-dependent mean of errors to achieve reliable predictions for a wide range of objective functions (Hunter et al. 2021)
- Guidance on different catchment types (perennial and ephemeral) (McInerney et al. 2019)

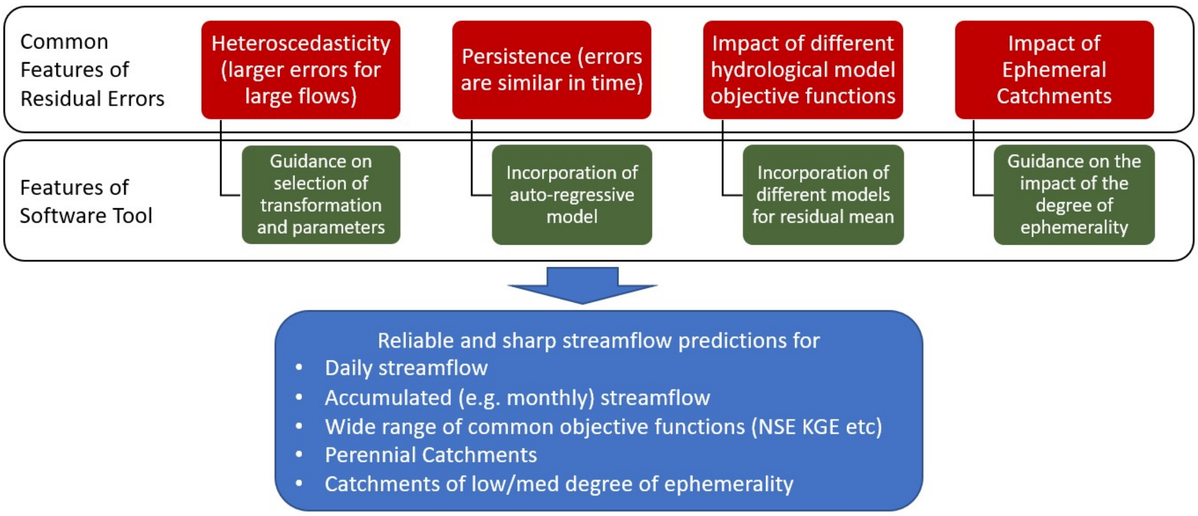


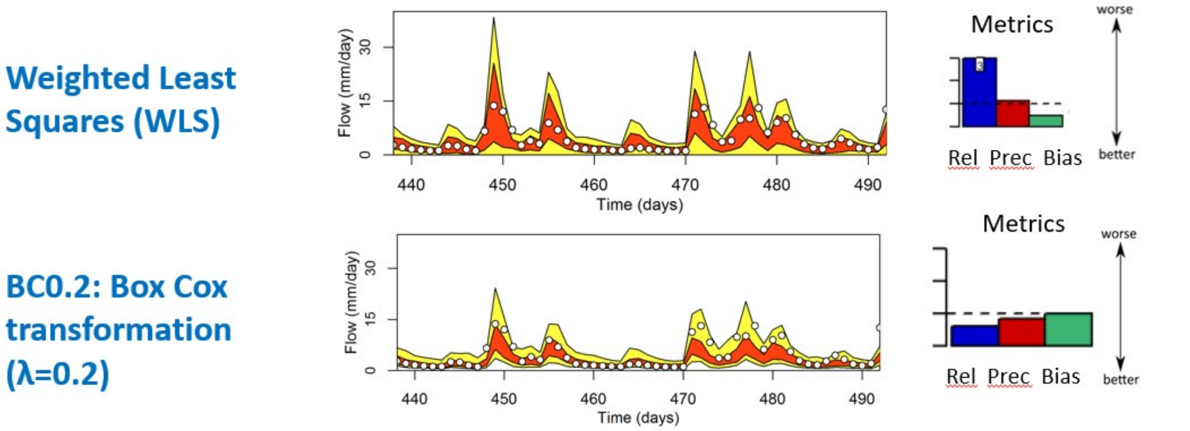
Figure 2: Common features of residual errors and the software tool features that address them and the impact on probabilistic predictions

Table 1: Guidance on Modifying Residual Error Model Parameters

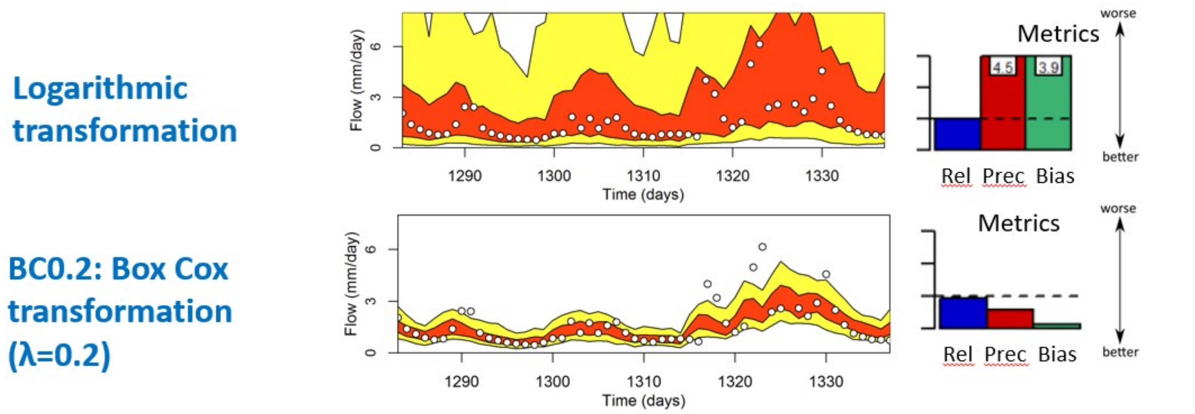
Table: Guidance on Modifying Residual Error Model Parameters			
Catchment Type	**Transformation Parameter (λ)**	**Offset Parameter (A*)**	**Likely performance, but needs evaluation**
Perennial (No zero flows)	Start with λ =0.2 (BC0.2) -Trial λ= 0 (log) to improve reliability -Trial λ= 0.5 (BC0.5) to improve bias	zero	Likely to be high quality (HQ) performance for range of common objective functions
Low ephemeral (0-5% zero flows)	'as above'	Trial non-zero values of A*	Likely to be similar to perennial
Mid ephemeral (5-50% zero flows)	'as above'	Trial non-zero values of A*	Likely to be worse then perennial, may require specialised treatement
High ephemeral (>50% zero flows)	Not recommended		Unlikely to be HQ, requires specialised treatment

'Specialised treatment' refers to residual error model features beyond the current capability of the current software tool (see (McInerney et al. 2019), (Wang et al. 2020)).

Figure 3 illustrates the impact of the streamflow transformation on probabilistic predictive performance (see (McInerney et al. 2017)) for further details). Figure 3(a) illustrates the impact for the perennial catchment, Spring River USA and shows the Box-Cox transformation with fixed lambda=0.2 (BC0.2) outperforms the weighted least squares (WLS) with better reliability and precision (i.e. sharpness). This is because the Box-Cox transformation allows for skew in the empirical residuals, whereas WLS does not. Figure 3(b) illustrates the impact for ephemeral catchment, Rocky River, Australia and shows that the BC0.2 transformation outperforms the log transformation, with better precision and bias, but similar reliability. This is because the log transformation is very sensitive for zero and near-zero flows.



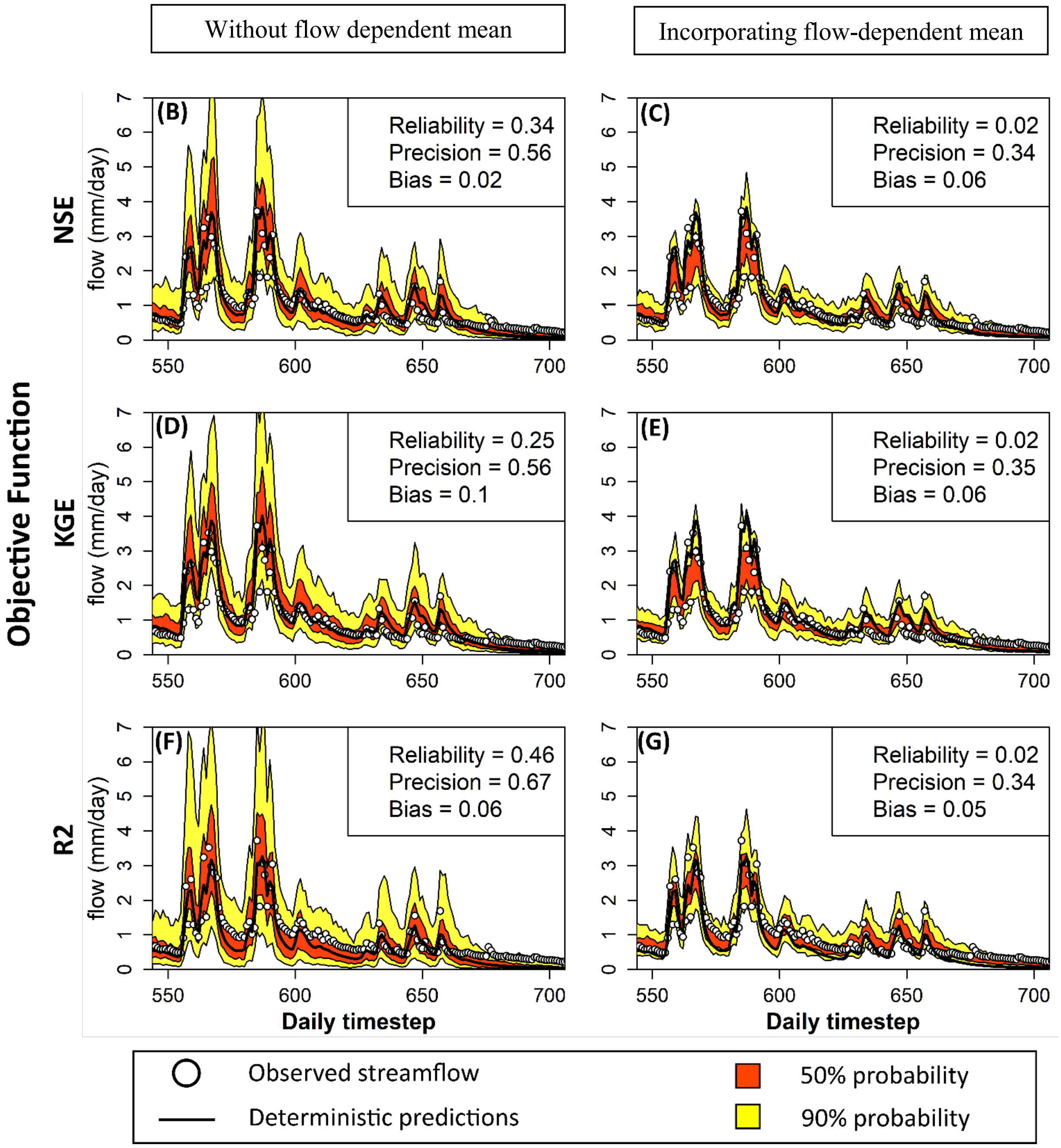
(a) BC0.2 transformation provides better reliability and precision than WLS: Perennial Catchment: Spring River, USA,



(b) BC0.2 transformation has much better precision and low bias than log: Ephemeral Catchment: Rocky River, Australia

Figure 3: Impact of transformation on probabilistic performance (from McInerney et al. 2017)

Figure 4 shows the impact of incorporating a flow-dependent mean for the residuals on the probabilistic predictive performance for range of commonly used objective functions (OF's) - for further details see Hunter et al. (2021). The left column of Figure 4 shows the time series and performance metrics without incorporating residual mean, while the right column shows the impact of incorporating the residual mean. Without incorporating the mean, the reliability metric ranges from 0.25 (KGE) to 0.46 (R2) which dramatically improves to value of 0.02 for all OFs, when the residual mean is included. For the precision, without incorporating the mean, the metric ranges from 0.56 (NSE/KGE) to 0.67 (R2), which reduces to 0.34-0.35 all OFs when the residual mean is included.



Note: for all performance metrics (reliability,precision,bias) lower is better

Figure 4: Illustration of impact of incorporating the flow-dependent residual mean on probabilistic predictive performance for range of commonly used objective functions for Yackandandah catchment (from Hunter et al. 2021).

Key Feature 2: Simple and Easy to Use

The software tool is designed to be simple and easy-to-use to increase the uptake and use of probabilistic predictions in research and industry. The following features improve usability

- Interactive webapp for single site analysis (Figure 6)
- Users simply upload a time series of (deterministic) predictions and observations (Figure 6a)
- Method-of-moments approach to parameter estimation enables fast computation (McInerney et al. 2018)
- Users can download time series of probabilistic predictions/probability limits and summary metrics.
- Command-line functionality in R package provides opportunity for automated analysis of a large number of sites

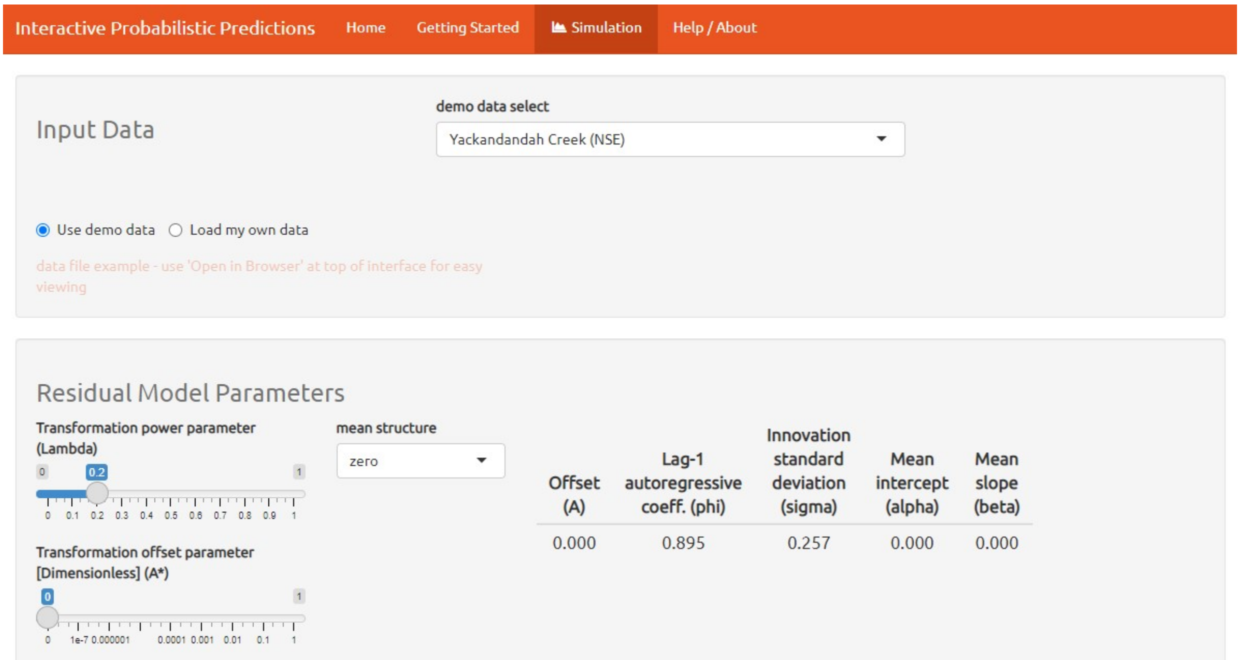
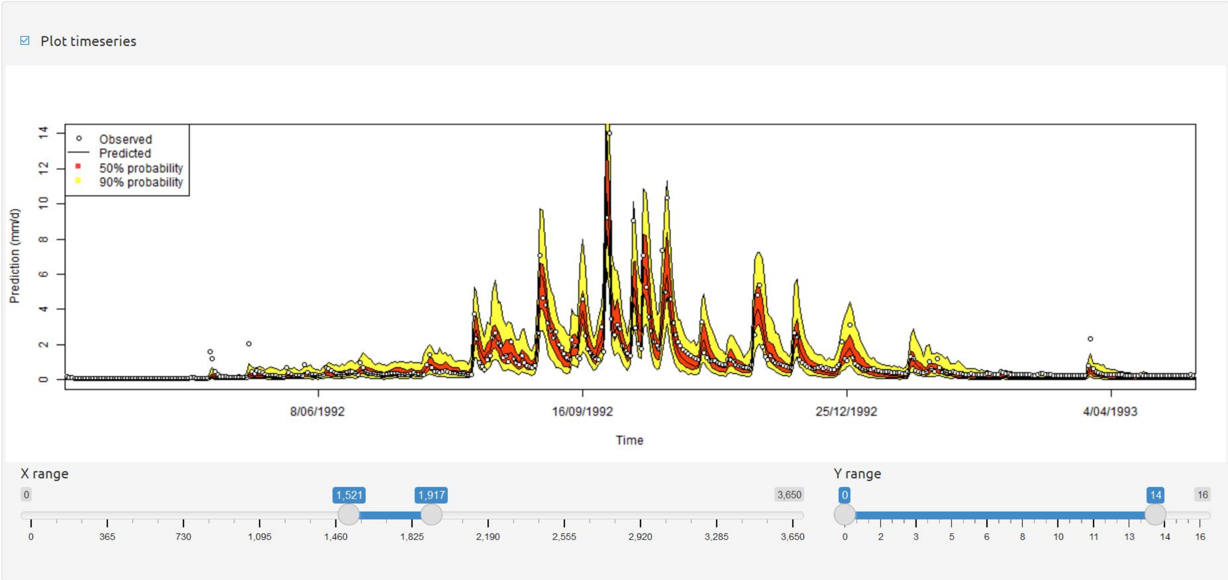


Figure 5: Webapp Input Interface (Data Input and Parameters)

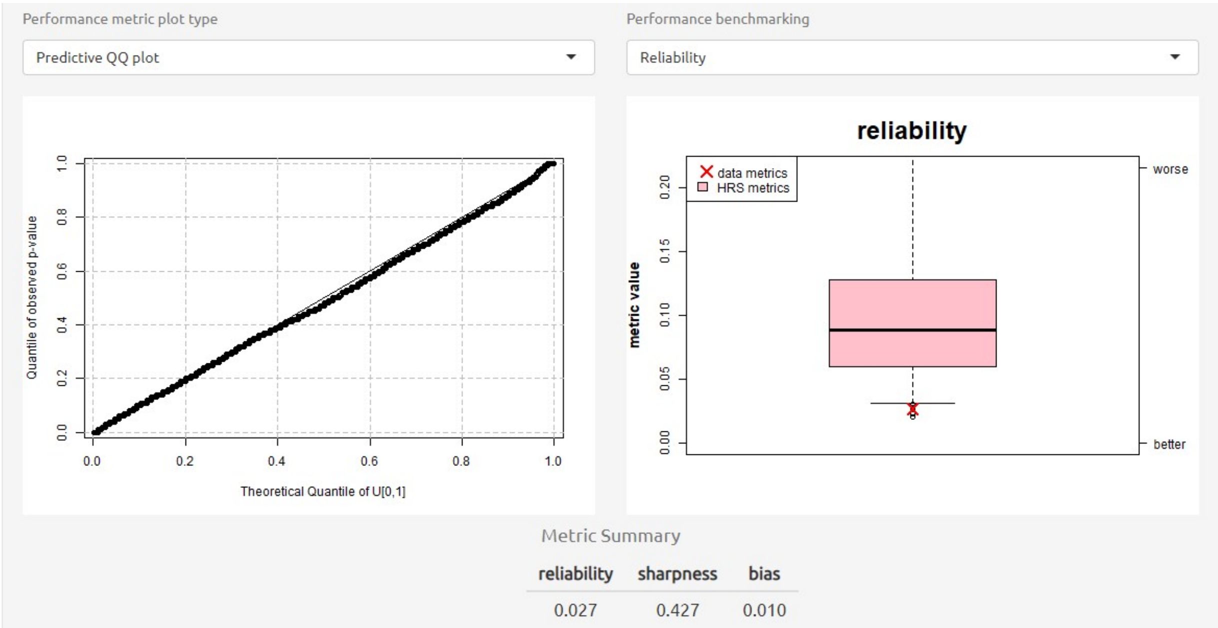
Key Feature 3: Evaluates predictive performance using a range of commonly used metrics and diagnostics

The software tool uses a wide range of commonly used metrics and diagnostics. This feature is important because each catchment can have its own nuances. Hence it is important to undertake evaluation of performance for each individual catchment. The list of evaluation techniques supported by the software tool is as follows:

- Comparison of probabilistic time series against observations (Figure 6b)
- Reliability using Predictive QQ Plots (see Figure 6b)
- Commonly used metrics to evaluate probabilistic predictions (Reliability, Sharpness and Bias) - see
- Comparison against reference set of metrics based on the Bureau of Meteorology Hydrological Reference Stations (Figure 6c)
- Wide range of residual diagnostics to improve predictive performance



(a) Comparison of probabilistic time series against observations



(b) Probabilistic Performance Diagnostics and Metrics

Figure 6: Probabilistic Performance Evaluation of Software Tool

Future Development of Software Tools

The software tool is enhanced regularly to incorporate ongoing research advances in residual error model development and probabilistic diagnostics. Features currently being incorporated (work in progress) include aspects such as multi-time scale features of the residual errors, such as seasonality ((Li, Wang, and Bennett 2013), (McInerney et al. 2020)) and impacts of hydrological non-stationarity, e.g. as included in the MUTHRE model of (McInerney et al. 2020))

CASE STUDY USING ENVIRONMENTAL DECISION-MAKING

Overview

The case study illustrates the use of probabilistic predictions to provide a more reliable evaluation of the impact of management actions on the ecological health of fish species (fish health) by simplifying the analysis undertaken by Department of Environment, Water and Natural Resources in the Barossa catchment (Green et al. 2014). Fish health is dependent on the flow regime within the catchment. This flow regime and the management actions were modelling using hydrological models at a range of sites in the Barossa Catchment. Our illustration considers two sites in the Barossa Catchment (Figure 7), Mt McKenzie and Upper Jacobs catchment. These two sites were chosen due to the differences in quality of the fit of the hydrological model (GR4J) to the streamflow data; Mt McKenzie, was classified as good, with Nash-Sutcliffe Efficiency (NSE) =0.67, while Upper Jacobs was classified as poor, with NSE=0.44 (Figure 7).

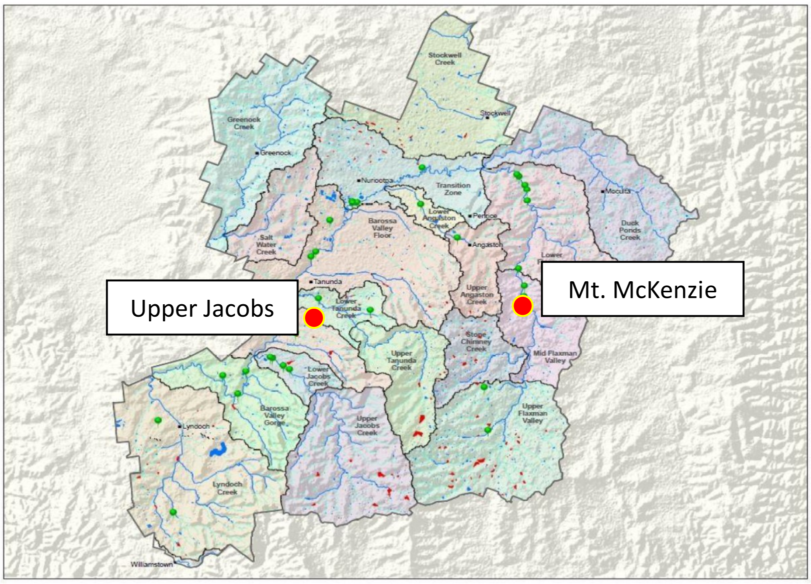


Figure 7: Barossa Valley Region and Case Study Catchments (Mt McKenzie and Upper Jacobs)

Methodology: Incorporation of Probabilistic Predictions

In probabilistic decision-making frameworks, risk is defined as the likelihood of occurrence multiplied by the consequence of occurrence (Figure 8). This study uses uncertainty analysis on hydrological predictions to inform the likelihood of occurrence of environmental flow metrics. Ecological response models are used to inform the consequence of these environmental flow metrics on risks to fish health.

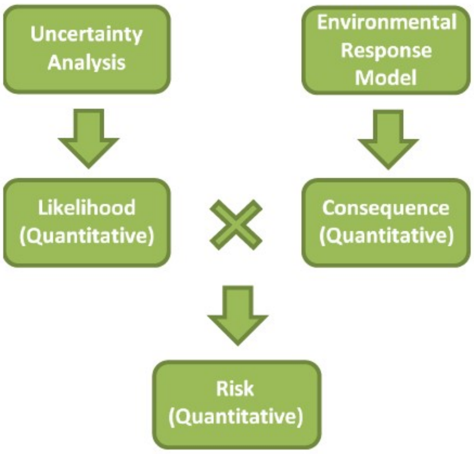


Figure 8: Framework for Quantitative Risk Assessment

The overall process is outlined in Figure 9. A series of management options (Do Nothing, Add Low Flow Bypass, Removed Dams) were simulated using a hydrological catchment model. The uncertainty in the hydrological predictions was estimated using the software tool. These uncertain hydrological predictions were then converted into distributions of environmental flow metrics (riffle flow days, days between May and November, where flow>12cm.), in turn used as input into an environmental response model to determine the ecological health assessment of the fish species.

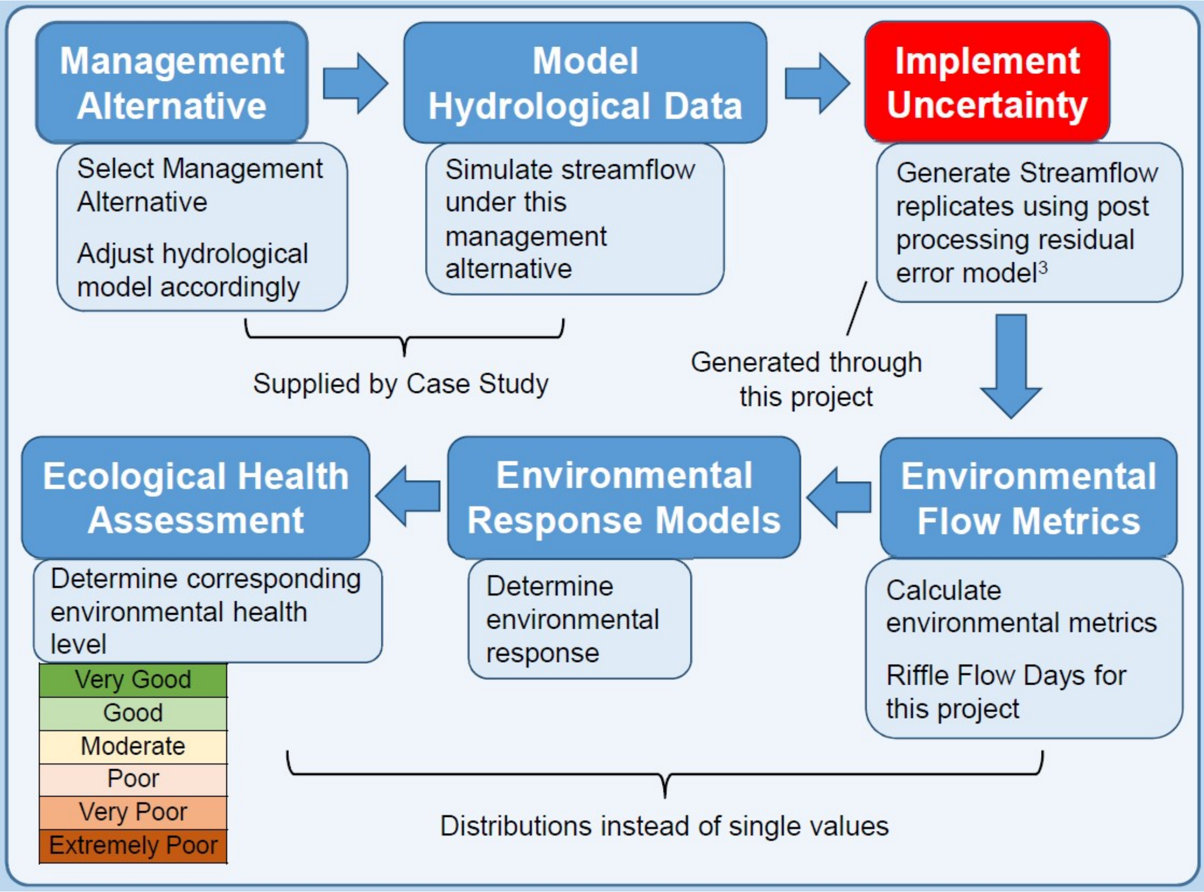


Figure 9: Overview of process for quantitative risk assessment incorporating predictive uncertainty

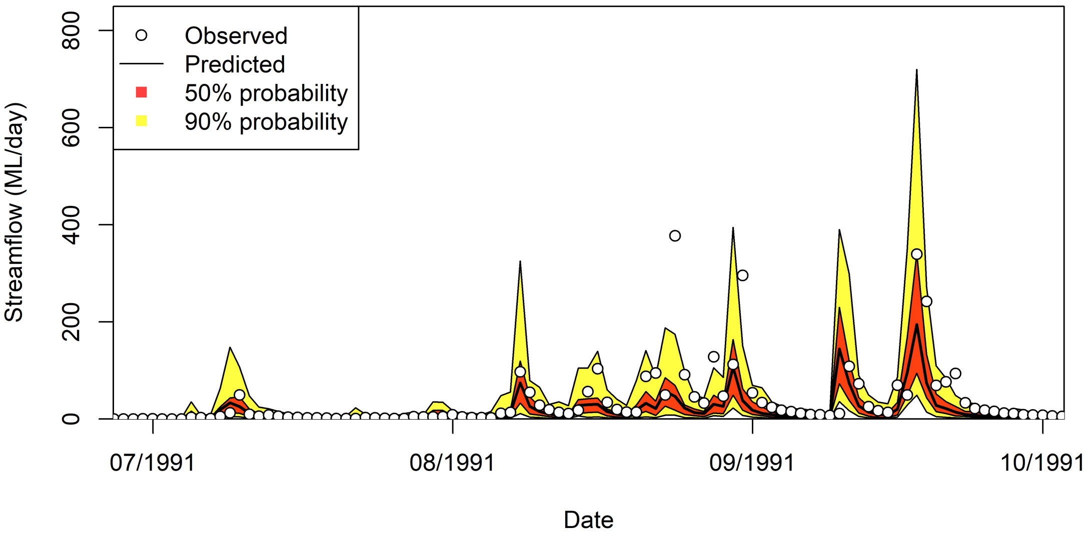
Key Steps of Case Study

1. Estimate uncertainty in streamflow predictions using software tool outlined in section 1.
2. Simulate deterministic streamflow predictions for selected management alternative (e.g. Current/Add low flow bypasses on Dams) using hydrological model
3. Generate probabilistic streamflow by implementing the uncertainty analysis using the outcomes from step 1 to the deterministic predictions from step 2.
4. Generate probabilistic predictions of environmental flow metrics (no. of riffle flow days in May to November for each year, "Annual Riffle Flow Days") based on the probabilistic streamflow predictions
5. Generate ecological health assessment by using a simplified environmental response model to convert environmental flow metrics into fish health level (Annual riffle flow days -> Probability of Successful Recruitment -> Fish Health), see (Green et al. 2014)
6. Convert the annual time series of fish health level (Extremely Poor, Very Poor etc) into the main outcome with is the frequency of years with different levels of fish health.

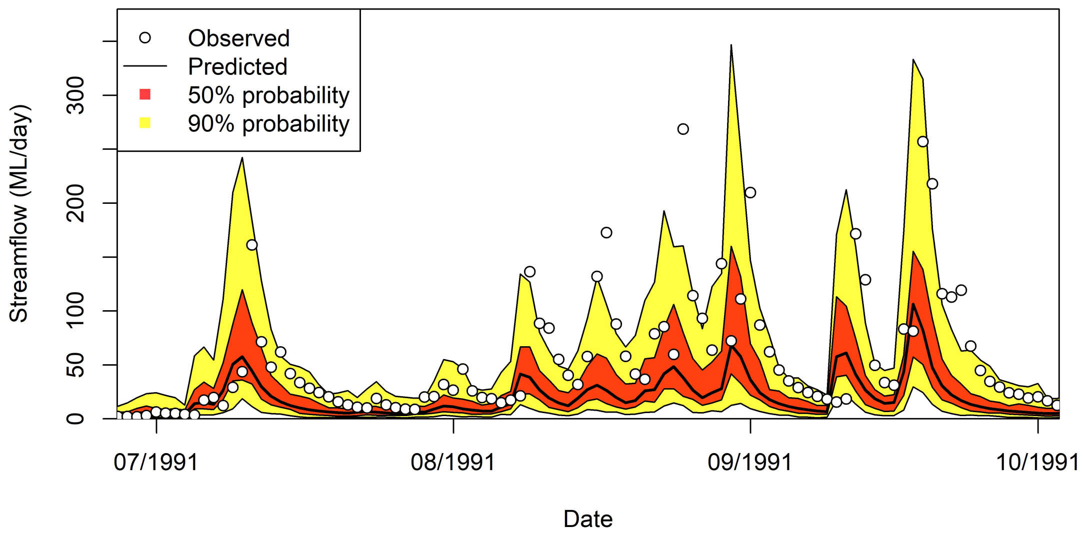
Note the ecological response model was greatly simplified for the purposes of this case study, where a single environmental flow metric (riffle flow days) was related to the probability of a successful recruitment event and used an indicator of fish health. In a more comprehensive study, multiple environmental flow metrics and multiple ecological response models would be used to estimate fish health (Green et al. 2014). Also note that as the publicly available software tool outlined is a beta version some of the features needed to conduct this study are still to be fully implemented. Users interested in applying this tool are encouraged to contact the authors in the first instance to discuss their needs.

Results

Figure 10 shows the deterministic and probabilistic time series for each site, Mt McKenzie and Upper Jacobs. For both sites, the probability limits of the probabilistic time series better captures the observed data than the deterministic predictions, this is especially so for the Upper Jacobs site (Figure 10b). Comparing the two sites, Upper Jacobs has wider probability limits than Mt McKenzie, which shows how the worse model fit at Upper Jacobs (NSE=0.43 c.f. NSE=0.67 for Mt McKenzie) is represented by the probabilistic predictions. Without the probabilistic predictions, it would be not possible to represent this difference.



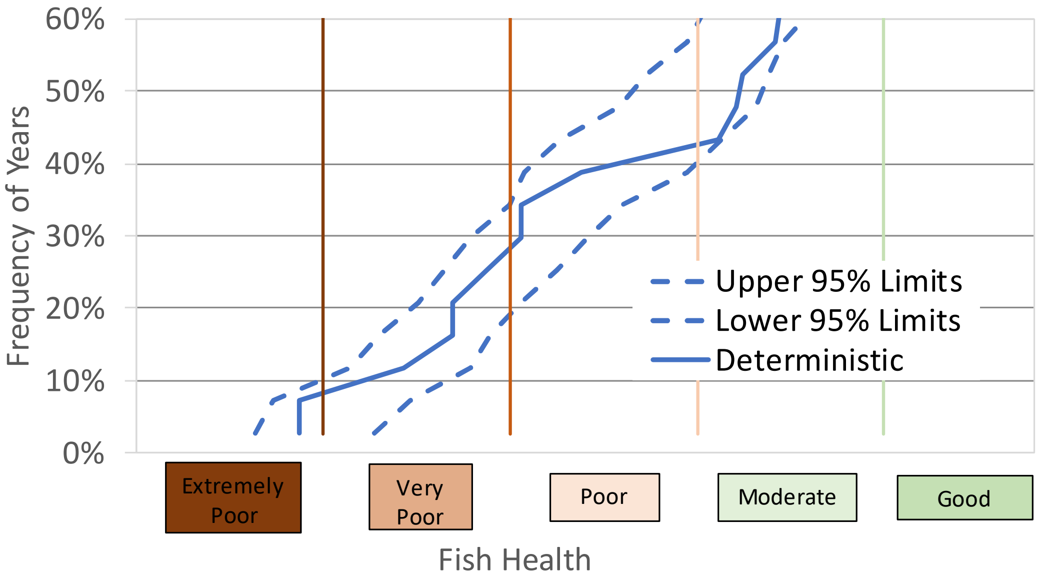
(a) Mt McKenzie



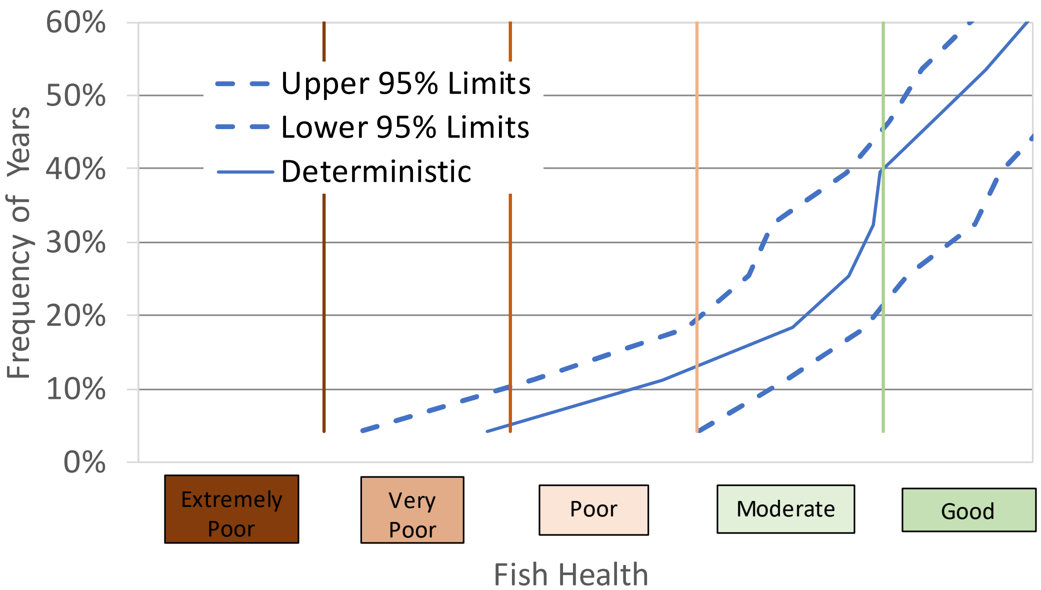
(b) Upper Jacobs

Figure 10: Comparison of Deterministic ("Predicted") and Probabilistic Time Series for Mt McKenzie (NSE=0.67) and Upper Jacobs (NSE=0.44)

Figure 11 compares the deterministic and probabilistic predictions of the frequency of years with different levels of fish health (Extremely Poor, Very Poor, Poor etc.) for both sites. For both sites the results highlight that the deterministic predictions, do not represent the large uncertainty that is represented by the probabilistic limits generated by the probabilistic predictions. Figure 11a shows that for Mt McKenzie, the frequency of years with poor (or worse) Fish Health is between 40% to 60%, a range of a factor of 1.5 In contrast the deterministic prediction provide a single value of 43%. For the Upper Jacobs sites (Figure 11b) the probability limits are much wider than Mt McKenzie, reflecting the poorer fit of the site to the calibration data. The frequency of years with poor (or worse) fish health is estimated to be between 4% and 20%, while the deterministic predictions provide only a single value of ~13%.



(a) Mt McKenzie (NSE=0.67)



(b) Upper Jacobs (NSE=0.44)

Figure 11: Comparison of deterministic and probabilistic predictions of frequency of years with different Fish Health for Mt McKenzie and Upper Jacobs

Table 2 summarises the practical benefits of uncertainty analysis at the two sites and shows that probabilistic predictions provide more information than deterministic predictions which greatly under-estimate the uncertainty in estimate of fish health and also are capture the impacts of the poorer hydrological model fit on key decision variables. This information will be highly valuable for decision-makers who are risk averse.

Table 2. Summary of Practical Benefits of Uncertainty Analysis

Site	Benefits of Uncertainty Analysis
Mt. McKenzie (NSE=0.67)	• Probabilistic predictions better capture observations • Shows frequency of years with poor (or worse) fish health ranges between 40% to 60% (c.f. single value of 32% with deterministic predictions)
Upper Jacobs (NSE=0.44)	• Probabilistic predictions better capture observations • Poorer fit of the hydrological model is captured in the wider probabilistic predictions and is propagated through to the wider limits for the key decision metric, the fish health • Shows frequency of years with poor (or worse) fish health ranges between 4% to 20% (c.f. single value of 13% with deterministic predictions)

FUTURE PRACTICAL CASE STUDIES

As hydrological models are used in wide range of water resource contexts, there is potential for future case studies to evaluate the impact of predictive uncertainty on risk estimates for highly uncertain events such as flood and droughts. For flood risk events, it would be required to extend the analysis undertaken here to event-based hydrological models or predictions from hourly models. For drought risk analysis, the analysis of long-time scale errors in hydrological predictions would be required. In general, rare events are challenging to predict using hydrological models and it is likely that uncertainty analyses will help improve the risk estimation of floods and droughts with long recurrence periods. Users interested in using the software tool in a case study are encouraged to contact the authors to discuss their needs in the first instance.

CONCLUSIONS

This paper presented a software tool that provides produce high-quality probabilistic streamflow predictions. Its key features are (1) Provides high quality probabilistic predictions by incorporating research advances and guidance on residual error model development and application by multiple research papers over several years (2) Simple and easy to use to facilitate greater uptake of probabilistic predictions in research and industry (3) Evaluates probabilistic predictive performance on specific catchment of interest using a wide range of commonly used diagnostics and metrics, and includes benchmarking of performance against Bureau of Meteorology Hydrological References Stations.

A case study illustrating the practical benefits of uncertainty analysis for risk-based decision- making is provided. The case study at two sites in the Barossa Valley, South Australia evaluates predictions of fish health, by combining streamflow predictions of environmental flow metrics with a simplified environmental response model. The outcomes using probabilistic streamflow predictions are contrasted to the results from deterministic probabilistic predictions. In general, the probabilistic predictions provided more insights about the confidence in the predictions of fish health as the uncertainty ranges recognised the differences between the quality of the hydrological predictions at the two different case study sites. The uncertainty ranges were found to be high, with the frequency of years with poor (or worse) fish health being estimated in the range of 40-60% (Mt McKenzie site) or 4-20% (Upper Jacobs site). This provides richer information on the uncertainty of key variables used in decision-making, which would be especially useful for risk averse decision-makers.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge the South Australian Department of Environment, Water and Natural Resources who provided partial funding for the development of the software tool and the information related to the Barossa Valley case study - in particular, Dr Matt Gibbs, principal hydrologist and Dr Douglas Green, senior aquatic ecologist who provided advice on range of aspects of the case study. The honours students, Joshua Anfiteatro, Tomaso Gaetan, Declan Mead, Yikai Mao who undertook the initial analysis for the case study. Australian Bureau of Meteorology is gratefully knowledge for supporting the research work that lead to the development of this software tool.

REFERENCES

Evin, Guillaume, Mark Thyer, Dmitri Kavetski, David McInerney, and George Kuczera. 2014. 'Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity', *Water Resources Research*, 50: 2350-75.

Green, D., S. Maxwell, J. VanLaarhoven, and D Deane. 2014. "Barossa Valley Prescribed Water Resources Area hydro-ecological risk assessment,." In.: Government of South Australia, through the Department of Environment, Water and Natural Resources, Adelaide.

Hunter, Jason, Mark Thyer, David McInerney, and Dmitri Kavetski. 2021. 'Achieving high-quality probabilistic predictions from hydrological models calibrated with a wide range of objective functions', *Journal of Hydrology*: 126578.

Kavetski, D. 2018. 'Parameter Estimation and Predictive Uncertainty Quantification in Hydrological Modelling.' in Q. Duan, F. Pappenberger, J. Thielen, A. Wood, H.L. Cloke and J.C. Schaake (eds.), *Handbook of Hydrometeorological Ensemble Forecasting* (Springer Berlin Heidelberg).

Krzysztofowicz, Roman. 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: 1151-62, doi:10.029/WR019i005p01151.

Li, M., Q. J. Wang, J. C. Bennett, and D. E. Robertson. 2016. 'Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting', *Hydrology and Earth System Sciences*, 20: 3561-79.

Li, Ming, Q. J. Wang, and James Bennett. 2013. 'Accounting for seasonal dependence in hydrological model errors and prediction uncertainty', *Water Resources Research*, 49: 5913-29.

McInerney, David, Dmitri Kavetski, Mark Thyer, Julien Lerat, and George Kuczera. 2019. 'Benefits of Explicit Treatment of Zero Flows in Probabilistic Hydrological Modeling of Ephemeral Catchments', *Water Resources Research*, 55: 11035-60.

McInerney, David, Mark Thyer, Dmitri Kavetski, Bree Bennett, Julien Lerat, Matthew Gibbs, and George Kuczera. 2018. 'A simplified approach to produce probabilistic hydrological model predictions', *Environmental Modelling & Software*, 109: 306-14.

McInerney, David, Mark Thyer, Dmitri Kavetski, Richard Laugesen, Narendra Tuteja, and George Kuczera. 2020. 'Multi-temporal Hydrological Residual Error Modeling for Seamless Subseasonal Streamflow Forecasting', *Water Resources Research*, 56: 2019WR026979.

McInerney, David, Mark Thyer, Dmitri Kavetski, Julien Lerat, and George Kuczera. 2017. 'Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors', *Water Resources Research*.

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: W10531, doi:10.1029/2009WR008933.

Smith, T., A. Sharma, L. Marshall, R. Mehrotra, and S. Sisson. 2010. 'Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments', *Water Resources Research*, 46: W12551, doi:10.1029/2010WR009514.

Vaze, J, P Jordan, R Beecham, A Frost, and G Summerell. 2012. 'Guidelines for rainfall-runoff modelling'.

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

Wang, Quan J., James C. Bennett, David E. Robertson, and Ming Li. 2020. 'A Data Censoring Approach for Predictive Error Modeling of Flow in Ephemeral Rivers', *Water Resources Research*, 56: e2019WR026128.