24th International Congress on Modelling and Simulation, Sydney, NSW, Australia, 5 to 10 December 2021
mssanz.org.au/modsim2021

INVITED PAPER

Improving sub-seasonal streamflow forecasts across flow regimes

David McInerney[a], Mark Thyer[a], Dmitri Kavetski[a], Richard Laugesen[b], Fitsum Woldemeskel[c], Narendra Tuteja[b] and George Kuczera[d]

[a] School of Civil, Environmental and Mining Engineering, University of Adelaide, SA, Australia
[b] Bureau of Meteorology, Canberra, ACT, Australia
[c] Bureau of Meteorology, Melbourne, Victoria, Australia
[d] School of Engineering, University of Newcastle, Callaghan, NSW, Australia
Email: david.mcinerney@adelaide.edu.au

Abstract: Sub-seasonal streamflow forecasts are important for a range of water resource management applications, with a distinct practical interest in forecasts of high flows (e.g. for managing flood events) and low flows (e.g. for managing environmental flows). Despite this interest, differences in forecast performance for high and low flow events are not routinely investigated. Our study reveals that while forecasts evaluated over the full flow range can appear reliable, stratification into high/low flow ranges highlights significant under/over-estimation of forecast uncertainty, respectively.

This study introduces a flow-dependent (FD) non-parametric component into a post-processing model of hydrological forecasting errors, the Multi-Temporal Hydrological Residual Error (MuTHRE) model, yielding the MuTHRE-FD model. We use a case study with 11 catchments in the Murray Darling Basin, the GR4J rainfall-runoff model and post-processed rainfall forecasts from ACCESS-S, to compare the MuTHRE and MuTHRE-FD models. Through its improved treatment of flow-dependence, the MuTHRE-FD model achieves practically significant improvements over the original MuTHRE model in the reliability of forecasted cumulative volumes for: (i) high flows out to 7 days; (ii) low flows out to 2 days; and (iii) mid flows for majority of lead times. Example cumulative flow time series are provided in Figure 1. The new MUTHRE-FD model provides sub-seasonal forecasts with high quality performance for both high and low flows over a range of lead times. This improvement provides forecast users with increased confidence in using sub-seasonal forecasts across a wide range of applications.

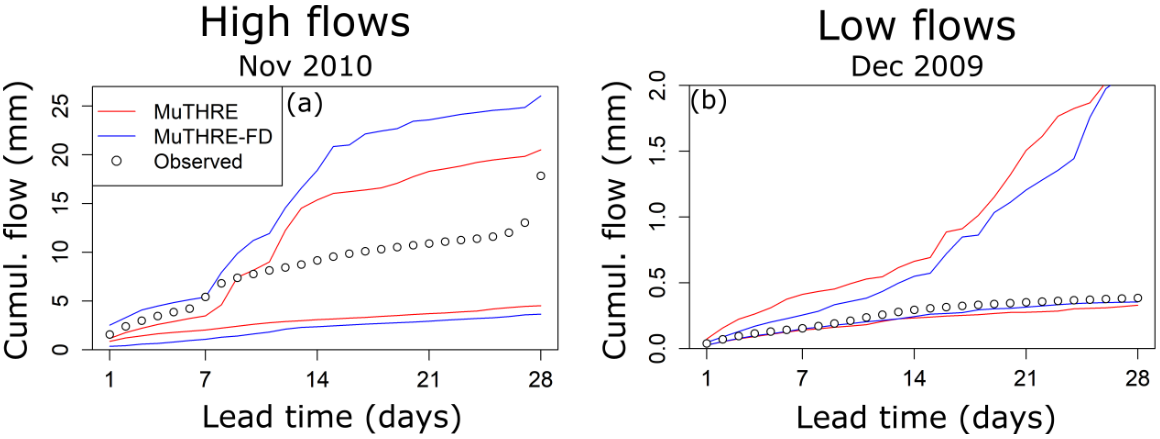

Figure 1. Example time series of predictive limits of cumulative volume forecasts out to 28 days for Hughes Creek (catchment ID 405228). Results are shown for forecasts issued on 1 November 2010, which is a high flow period (left side), and 1 December 2009, which is a low flow period (right side).

Keywords: *Subseasonal streamflow forecasting, high and low flows, non-parametric*

1. INTRODUCTION

Sub-seasonal streamflow forecasts offer valuable information for a range of real-time water resource applications. For example, forecasts of high flows can be used for flood storage reservoir operations, while forecasts of high and low flows are useful for environmental flow management. In large water resource systems, with travel times of days to weeks, forecasts of daily flows and cumulative volumes are required at lead times up to the maximum travel time within the system. In this context, seamless forecasts, i.e., forecasts obtained using a single method that maintains high quality across a range of time scales, are attractive in both research and practical applications.

Streamflow forecasts are uncertain due to rainfall forecast uncertainty (associated with predicting future rainfall) and hydrological uncertainty (associated with model structure errors, initial conditions, etc). Hydrological uncertainty is often represented using residual error models. These models should reflect the complex statistical characteristics of hydrological errors, including heteroscedasticity (larger errors for larger flows), persistence (similar errors for consecutive times), non-Gaussianity (skewness and kurtosis), and other temporal variations (e.g. due to seasonality and changing catchment conditions).

A key question is whether common residual error models are able to provide high quality characterization of forecast uncertainty in both high and low flows. Despite most water resource decisions being made based on high and low flows, forecast performance for high and low flow events are not routinely investigated.

The study has the following aims:

1) Investigate the performance of sub-seasonal streamflow forecasts using a high/low flow stratification;

2) Improve the reliability of high/low flow forecasts by representing the flow-dependence of innovations.

The analysis is carried out using the Multi-Temporal Hydrological Residual Error (MuTHRE) model (McInerney et al. 2020), which was shown to achieve seamless forecasts with good reliability across a range of lead times (1-30 days), stratifications (months and years), and time scales (daily and monthly).

2. PROBABILISTIC MODEL FOR STREAMFLOW FORECASTING

The methods in this study employ the ensemble dressing approach to probabilistic streamflow forecasting (Pagano et al. 2013). Hydrological uncertainty is characterized using the MuTHRE model (Section 2.1), with two different approaches for modelling innovations (Section 2.2). Rainfall uncertainty is represented using multiple rainfall forecast replicates (Section 2.3).

2.1. Summary of MuTHRE model for representing hydrological uncertainty

The MuTHRE model represents hydrological uncertainty in streamflow q_t (at time step t) through a probability model Q_t (i.e. $q_t \sim Q_t$) which combines a deterministic hydrological prediction q_t^{det} and a residual error term η_t in transformed space,

$$z(Q_t\,;\boldsymbol{\theta}_z) = z(q_t^{\mathrm{det}};\boldsymbol{\theta}_z) + \eta_t \tag{1}$$

Here z represents the Box-Cox transformation (Box and Cox 1964), with power parameter $\lambda = 0.2$ (McInerney et al. 2017). The deterministic term is

$$q_t^{\mathrm{det}} = h(\boldsymbol{\theta}_h;\mathbf{x}_t,\mathbf{s}_{t-1}) \tag{2}$$

where h is a rainfall-runoff model with parameters $\boldsymbol{\theta}_h$, inputs $\mathbf{x}_t$ and initial conditions $\mathbf{s}_{t-1}$. The residual error term follows as AR(1) model

$$\eta_t = \phi_\eta(\eta_{t-1} - \mu_{t-1}) + \mu_t + y_t \tag{3}$$

where ϕ_η is the lag-1 autoregressive parameter, μ_t is the time-varying mean of η_t which accounts for seasonality and dynamic biases (McInerney et al. 2020), and y_t is the innovation (i.e. random component) at time t. We consider two models for y_t in Section 2.2.

A conceptual diagram of the MuTHRE model is shown in Figure 2.

The deterministic model in equation (2) can be used to generate the following streamflow estimates:

i. "Simulated" streamflow $\mathbf{q}^{\text{sim}}$, when *h* is forced with observed rainfall $\tilde{\mathbf{x}}$, which is used for calibrating the set of parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_h, \boldsymbol{\theta}_z, \boldsymbol{\theta}_\eta\}$ using the observed streamflow time series $\tilde{\mathbf{q}}$. Given estimated parameters $\hat{\boldsymbol{\theta}}$ we can compute simulated streamflow $\mathbf{q}^{\text{sim}}$ and compare against observed streamflow $\tilde{\mathbf{q}}$ to calculate "empirical innovations" $\tilde{\mathbf{y}}$ using equations (1)-(3).
ii. An ensemble of "raw" streamflow forecasts, $\{\mathbf{q}^{\text{raw}(f)}; f = 1,\ldots,N_{\text{foc}}\}$, when *h* is forced using an ensemble of N_{foc} forecast rainfall replicates $\{\mathbf{x}^{\text{foc}(f)}; f = 1,\ldots,N_{\text{foc}}\}$, which is used in forecasting.

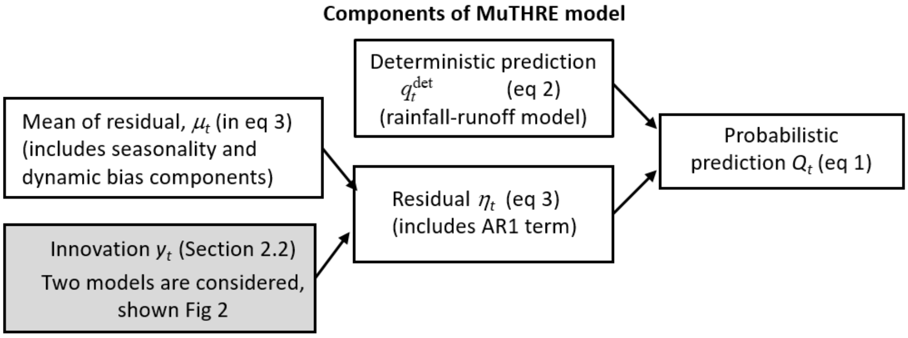
Components of MuTHRE model
Deterministic prediction q_t^{det} (eq 2) (rainfall-runoff model)
Mean of residual, μ_t (in eq 3) (includes seasonality and dynamic bias components)
Residual η_t (eq 3) (includes AR1 term)
Probabilistic prediction Q_t (eq 1)
Innovation y_t (Section 2.2) Two models are considered, shown Fig 2

Figure 2. Components of MuTHRE model

2.2. Models of innovations

Mixed-Gaussian model

The original MuTHRE model represents innovations using a two-component mixed-Gaussian distribution, which was found to improve the reliability of forecasts at short lead times by allowing for excess kurtosis in the innovations (Li et al. 2016; McInerney et al. 2020). This innovation model assumes the distribution of innovations does not depend on the flow magnitude (see Figure 3a).

Flow-dependent model

The flow-dependent (FD) innovation model allows for a dependence (conditioning) of the distribution of innovations y_t on the flow magnitude. For a given q_t^{det}, the (conditional) innovation $y_t \mid q_t^{\text{det}}$ is sampled non-parametrically as follows.

1. Construct the set of time steps $\mathbf{T}$ in the calibration period with predicted flows "similar" to q_t^{det}, chosen here as times where CDF values of $q_{\mathbf{T}}^{\text{det}}$ are within 0.05 of the CDF value of q_t^{det}.
2. Extract the flow-dependent empirical distribution $\tilde{\mathbf{y}}_{\text{T}}$
3. $y_t \mid q_t^{\text{det}}$ is selected randomly with replacement from $\tilde{\mathbf{y}}_{\text{T}}$.

The sampling procedure is illustrated in Figure 3b using two examples of $\mathbf{T}$ for low and high flows. For a deterministic prediction $q_t^{\text{det}} = P_{\text{cal}}^{-1}(0.15)$ (i.e., equal to the bottom 15th percentile of flows from the calibration period), the set $\mathbf{T}$ (and corresponding innovations) are indicated using blue points. The empirical innovations from these time steps make up the flow-dependent empirical distribution for that value of q_t^{det}, as given by the blue histogram. Similarly, for $q_t^{\text{det}} = P_{\text{cal}}^{-1}(0.9)$, the set $\mathbf{T}$ (and corresponding innovations) are indicated using red points, and the corresponding flow-dependent empirical distribution is given by the red histogram.

See McInerney et al. (2021) for full equations and justification and discussion on the potential implications of the modelling choices listed above. The new residual error model with flow-dependent (FD) innovations is referred to as MuTHRE-FD.

2.3. Generation of streamflow forecasts accounting for rainfall forecast uncertainty

Hydrological uncertainty (represented by the MuTHRE/MuTHRE-FD models) is then added to the raw streamflow forecasts, ($\mathbf{q}^{\text{raw}}$ in Section 2.1), to produce post-processed streamflow forecasts. See McInerney et al. (2020) for a detailed description of this procedure. Forecasts of cumulative flow volumes at lead time ℓ are obtained by aggregating the daily forecasts between t_0+1 and ℓ.

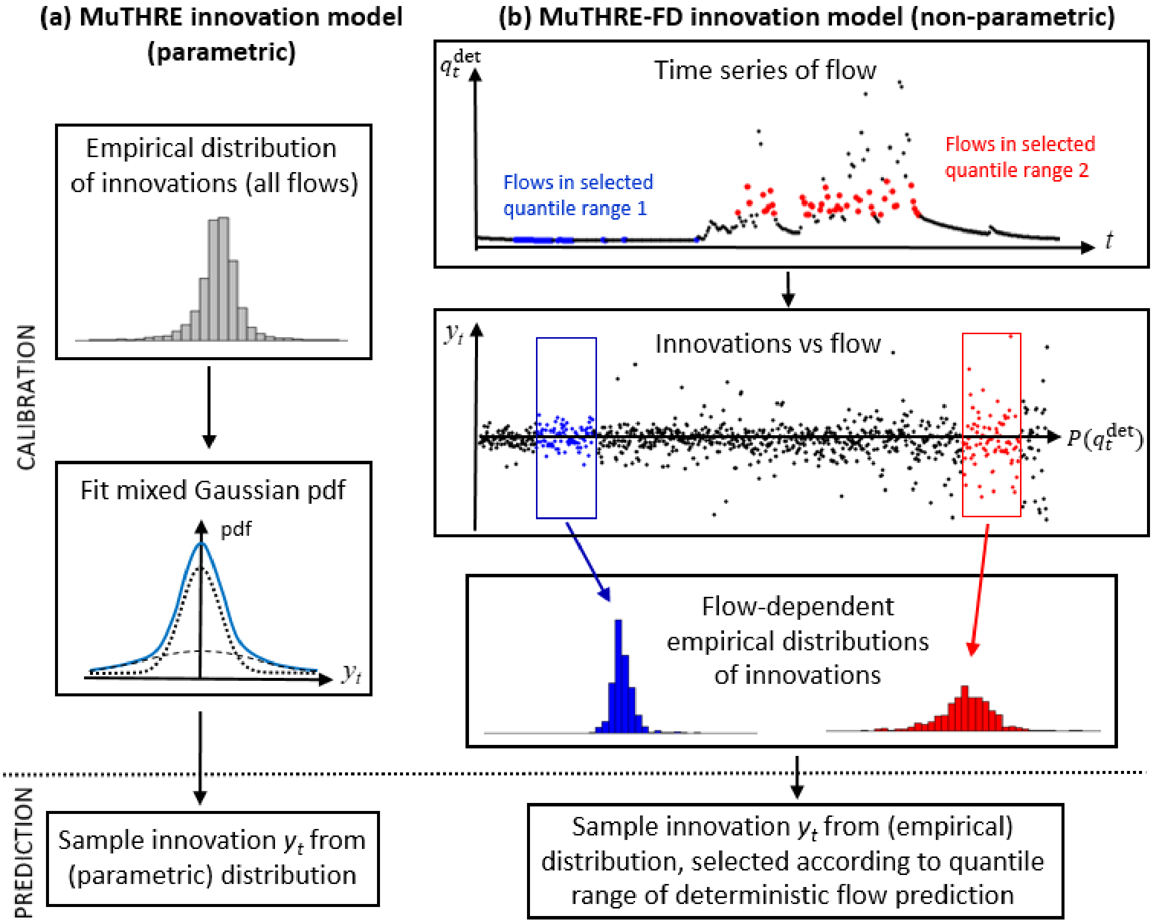
(a) MuTHRE innovation model (parametric)
(b) MuTHRE-FD innovation model (non-parametric)
CALIBRATION
PREDICTION
Empirical distribution of innovations (all flows)
Fit mixed Gaussian pdf
pdf
y_t
Sample innovation y_t from (parametric) distribution
q_t^{det}
Time series of flow
Flows in selected quantile range 1
Flows in selected quantile range 2
t
y_t
Innovations vs flow
$P(q_t^{det})$
Flow-dependent empirical distributions of innovations
Sample innovation y_t from (empirical) distribution, selected according to quantile range of deterministic flow prediction

Figure 3. (a) mixed-Gaussian innovation model used in MuTHRE, which uses a single distribution to sample innovations when producing streamflow forecast replicates and (b) flow-dependent innovation model used in MuTHRE-FD, which samples non-parametrically from different subsets of the empirical distribution of innovations, according to the magnitude of predicted flow.

3. CASE STUDY

3.1. Hydrological data and model

We compare forecasts from the MuTHRE and MuTHRE-FD models using a case study with 11 catchments in the Murray Darling Basin (McInerney et al. 2020). Daily time series of observed rainfall, PET and streamflow over a 22-year period from 1991-2012 are obtained from the Bureau of Meteorology's Hydrologic Reference Stations (HRS) dataset. Rainfall forecasts are taken from the Australian Community Climate Earth-System Simulator - Seasonal (ACCESS-S), and post-processed to reduce biases and improve reliability (Schepen et al. 2018). The GR4J rainfall-runoff model (Perrin, Michel, and Andreassian 2003) is used as the deterministic model. A moving-window leave-one-year-out cross-validation procedure is used for calibration and evaluation. Forecasts are produced from the first day of each month, and extended out to lead times of one month.

3.2. Forecast evaluation

We evaluate performance of cumulative volume forecasts under leave-one-year-out cross-validation in terms of reliability and sharpness. Reliability (i.e. the degree of statistical consistency between observations and the forecast distribution) is quantified using the metric of Evin et al. (2014). Sharpness (i.e., uncertainty in the forecast distribution) is quantified as a skill score by the average ratio of the 90% limits of the forecast distribution and the 90% limits of the climatology (Woldemeskel et al. 2018). Lower metric values indicate better performance. Performance metrics for cumulative volumes are computed for each lead time between 1 and 28 days, and for specific flow volume ranges (high/mid/low/all, based on the median ensemble forecast). Volumes in the top 5% are referred to as "high" flows, those in the bottom 50% are "low" flows, and the rest are "mid" flows. This stratification is based on the shape of the predicted flow duration curve for the case study catchments, which (broadly speaking) is very steep for the top ~5% of flows, and flat for the bottom ~50% of flows. Practical significance tests are used to determine whether the MuTHRE-FD model has better or worse performance metrics than the MuTHRE model over the range of catchments, and whether these differences are of practical relevance (defined as a difference by more than 10% of the median value for MuTHRE model).

4. RESULTS

Figure 1 and Figure 4 illustrate how the innovation model impacts on the reliability and prediction limits of forecasts in the Hughes Creek catchment. Figure 4a shows PQQ (reliability) plots for high flows at a lead time of 1 day. The shape of the PQQ plots indicates that the original MuTHRE model underestimates uncertainty in high flows, resulting in poor reliability (large departure from 1:1 line). The new MuTHRE-FD model largely resolves this problem and achieves reliable forecasts, with the PQQ plot lying within the 90% uncertainty limits. Figure 1a shows cumulative volume forecasts for a specific high flow period, beginning on the 1-Nov-2010. For short lead times, the MuTHRE model is over-confident, with observed flows outside the 90% prediction limits for the first 8 days. In contrast, the wider prediction limits of the MuTHRE-FD model encompass all observed values.

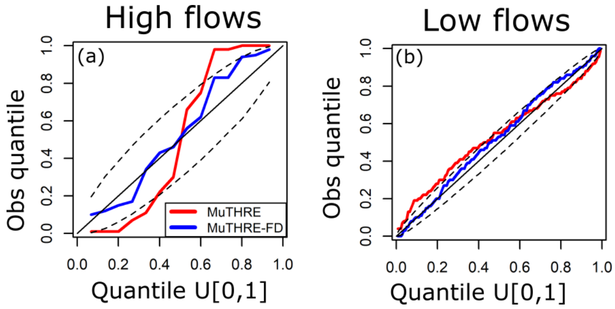

Figure 4. PQQ plots for day-ahead forecasts for Hughes Creek (catchment ID 405228). Results are shown for high flows (left side) and low flows (right side).

For low flow periods, the PQQ plots in Figure 4b show that the MuTHRE model overestimates uncertainty for day-ahead forecasts, while the MuTHRE-FD model provides more reliable forecasts. Figure 1b shows forecasts for a specific low flow period, beginning on 1-Dec-2009. For both models, the observations lie within the 90% prediction limits, but the MuTHRE-FD forecasts are sharper, especially for short lead times.

Figure 5 (top row) show reliability of cumulative volume forecasts from the MuTHRE and MuTHRE-FD models out to 28 days over all case study catchments. When all flows are lumped together, forecast reliability is similar for both models (Figure 5a). When high flows are considered separately (Figure 5b), the MuTHRE-FD model provides practically significant improvements in reliability of cumulative volumes for the first 8 lead times. Largest improvements are for the lead time of 1 day: MuTHRE suffers from poor reliability with median metric value of 0.32, while MuTHRE-FD achieves a (median) reliability of 0.17. Importantly, unlike the MuTHRE model, the reliability of high-flow forecasts from the MuTHRE-FD model is relatively stable across all lead times.

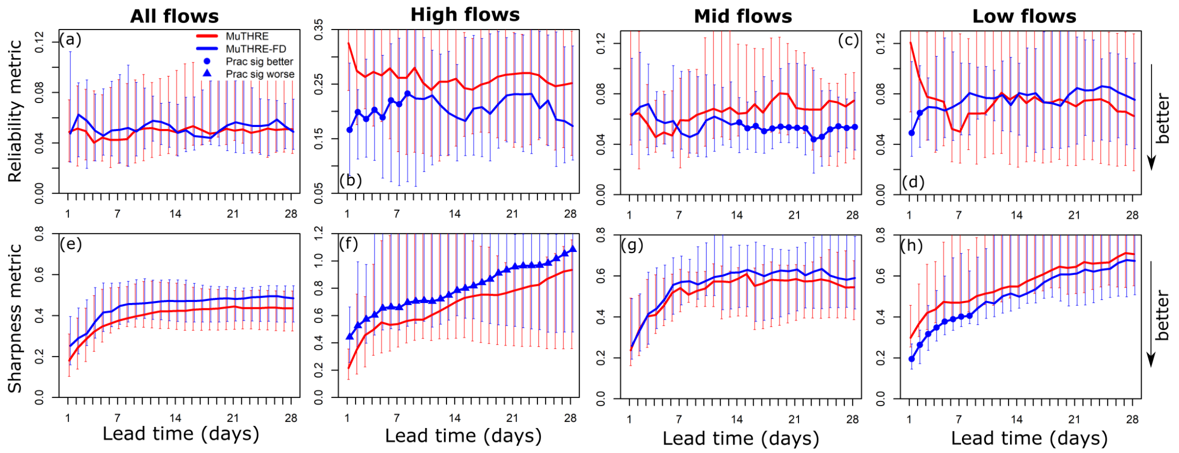

Figure 5. Performance metrics for cumulative volume forecasts from the two models, computed using all flows (Column 1) and stratified by flow magnitude (Columns 2-4). For a given model, the bars indicate the full range of metric values across the catchments, the line indicates the median metric values, and the symbols indicate whether MuTHRE-FD forecasts are practically better/worse than the MuTHRE forecasts.

For mid-flows, the MuTHRE-FD model provides practically significant improvements in reliability for all lead times greater than 13 days (Figure 5c). For low flows, the MuTHRE-FD model provides practically significant improvements for the first 2 lead times (Figure 5). Similar to high-flows, the largest improvements are for the lead time of 1 day, where the median reliability metric for low flows improves from 0.12 (MuTHRE) to 0.05 (MuTHRE-FD).

Figure 5 (bottom row) shows sharpness of forecasts. MuTHRE-FD offers improvements in sharpness of low flows (which make up 50% of days), which are practically significant for the first 8 lead times (Figure 5h). This comes at the cost of practically significant worsening of sharpness for high flows (Figure 5f). The loss of sharpness in high-flow forecasts of the MuTHRE-FD model occurs due to forecasts no longer being over-confident, which was a problem for the MuTHRE model where uncertainty in high flows was under-estimated.

We note that MuTHRE-FD model also produces improvements in terms of volumetric bias and continuous ranked probability score (CRPS) metrics for high and low flows. See McInerney et al. (2021) for results.

5. DISCUSSION

The key findings from Section 4 can be interpreted as follows:

Reliability. Forecasts of high and low cumulative volumes are more reliable for the MuTHRE-FD model, especially for shorter lead times (1-8 days). These improvements can be attributed to the improved representation of innovations (random component of errors), which controls the hydrological uncertainty at short lead times, and which in turn dominates the total forecast uncertainty at short lead times. At longer lead times (e.g. after 2 weeks), the impact of the error model decreases as streamflow forecast uncertainty becomes dominated by rainfall forecast uncertainty.

Sharpness. The MuTHRE-FD model produces sharper forecasts of low flows, and less sharp forecasts of high flows; this holds for both daily and cumulative values. Sharper forecasts at low flows are due to the *lower* variability of innovations at low flows, which is captured by the flow-dependent innovation model in MuTHRE-FD. Reduced sharpness of high flows is due to the *higher* variability of innovations at high flows, which is also reflected by the flow dependent model, and which represents the reason for improved reliability, i.e., MuTHRE-FD is not over-confident.

For high flows the MuTHRE-FD model offers practically significant improvements in reliability at the expense of practically significant worsening in sharpness. The MuTHRE model under-estimates the uncertainty in high-flow forecasts, and will therefore under-estimate the risk of high-flow events. The MuTHRE-FD model overcomes these problems. A common paradigm in forecasting is that reliability takes precedence over sharpness, because a prediction that is sharp but unreliable represents overconfidence (e.g., Gneiting and Katzfuss 2014). It follows that the large improvements in reliability of high flows obtained by the MuTHRE-FD model are worth the sacrifice in sharpness.

6. CONCLUSIONS

This work examines flow dependencies in sub-seasonal forecast performance using the MuTHRE model. A case study based on 11 catchments in the Murray Darling Basin is employed. The following findings and developments are contributed:

1. Flow stratified performance evaluation indicates that the MuTHRE model under/over-estimates the uncertainty for high/low flows, despite the unstratified evaluation across the full flow range suggesting forecasts are reliable;

2. A new non-parametric model is introduced to improve the representation of flow dependencies (FD) in the random component (innovations) of the residual error model;

3. The MuTHRE-FD model improves cumulative volume forecasts, with:

i. Practically significant improvements in the reliability of high flows out to lead times of 7 days, low flows for the first 2 days, and mid flows for 15 out of 28 days;

ii. Practically significant improvements in sharpness for low flows over all lead times, but practically significant widening of forecasts of high flows (reflecting their higher uncertainty);

More generally, this work demonstrates the benefits of capturing flow dependencies in the residual error structure of hydrological models, and the insights achievable from stratified performance assessment when evaluating forecast quality. Further information on the MuTHRE-FD model, including full equations and evaluation using additional metrics, can be found in McInerney et al. (2021).

ACKNOWLEDGMENTS

This work was funded by the Australian Bureau of Meteorology. Computations were performed on the Phoenix cluster at the University of Adelaide.

REFERENCES

Box, George EP, and David R Cox. 1964. 'An analysis of transformations', *Journal of the Royal Statistical Society. Series B (Methodological)*: 211-52.

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

Gneiting, Tilmann, and Matthias Katzfuss. 2014. 'Probabilistic forecasting', *Annual Review of Statistics and Its Application*, 1: 125-51.

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', *Hydrol. Earth Syst. Sci.*, 20: 3561-79.

McInerney, D, M Thyer, D Kavetski, R Laugesen, F. Woldemeskel, N. Tuteja, and G Kuczera. 2021. 'Improving the reliability of sub-seasonal forecasts of high and low flows by using a flow-dependent non-parametric model', *Water Resources Research*, In press.

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: e2019WR026979.

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*, 53.

Pagano, T. C., Durga Lal Shrestha, Q. J. Wang, David Robertson, and Prasantha Hapuarachchi. 2013. 'Ensemble dressing for hydrological applications', *Hydrological Processes*, 27: 106-16.

Perrin, C., C. Michel, and V. Andreassian. 2003. 'Improvement of a parsimonious model for streamflow simulation', *Journal of Hydrology*, 279: 275-89, doi:10.1016/S0022-1694(03)00225-7.

Schepen, A., T. Zhao, Q. J. Wang, and D. E. Robertson. 2018. 'A Bayesian modelling method for post-processing daily sub-seasonal to seasonal rainfall forecasts from global climate models and evaluation for 12 Australian catchments', *Hydrol. Earth Syst. Sci.*, 22: 1615-28.

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