Advances in subseasonal streamflow forecasting: An overview

Mark Thyer [(a)], **David McInerney** [(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: mark.thyer@adelaide.edu.au

Abstract: Sub-seasonal streamflow forecasts, with lead times up to 30 days, can provide valuable information for water management, including reservoir operation to meet environmental flow, irrigation demands, and managing flood protection storage. A key aim is to produce "seamless" probabilistic forecasts, with high quality performance across the full range of lead times (1-30 days) and time scales (daily to monthly). This paper provides an overview of advances towards subseasonal forecasting, by comparing the recently developed multi-temporal scale hydrological residual error (MuTHRE) model, one of the first approaches that provides seamless subseasonal forecasting, to an existing baseline residual error model and a non-seamless monthly streamflow post-processing (QPP) model. This comparison is in terms of model features and also through forecast evaluation on 11 catchments in the Murray-Darling Basin using multiple performance metrics, across a range of lead times, months and years, and at daily and monthly time scales. Compared to the baseline residual error model, the MuTHRE model is shown to provide improvements, in terms of reliability for short lead times (up to 10 days), in dry months, and dry years. Forecast performance also improved in terms of sharpness (Figure 1). Comparison against the non-seamless monthly QPP model showed MuTHRE provided similar reliability and sharpness for monthly forecasts stratified over months and years. This is a remarkable achievement, given the non-seamless monthly QPP models "sees" the monthly observed streamflow in calibration, whereas the MuTHRE model does not. This study highlights the benefits of modelling multiple temporal characteristics of hydrological errors, and demonstrates the power of the MuTHRE model for producing seamless sub-seasonal streamflow forecasts that have a wide range of practical benefits, as outlined.

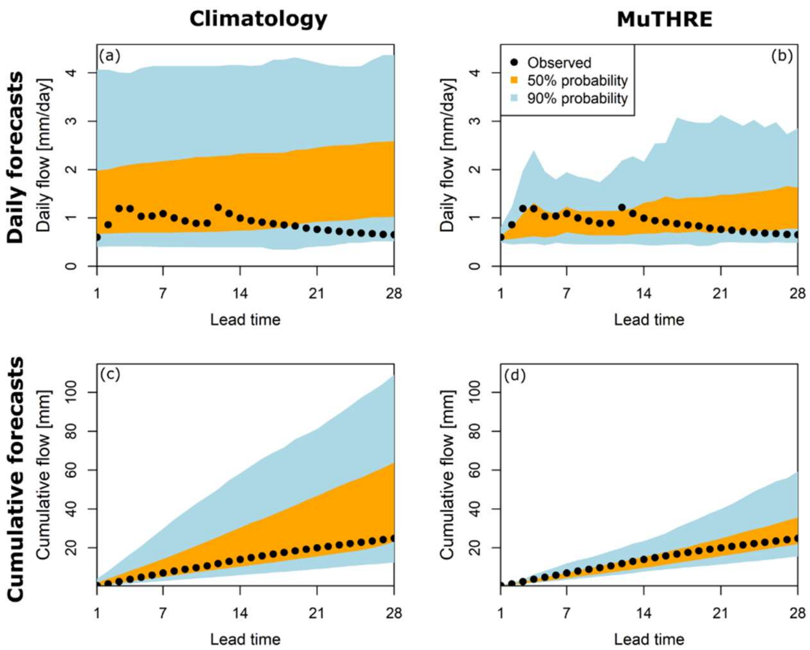

Figure 1. Streamflow forecasts in the Biggara catchment (401012) during August 2002. Climatology (left) is compared with the MuTHRE model (right). Both daily (top) and cumulative forecasts (bottom) are shown.

Keywords: *Subseasonal, streamflow forecasting, seamless, water resource management*

1. INTRODUCTION

Water management and operations across large river basins have historically focused on releasing and delivering water for consumptive purposes (e.g. irrigation), under relatively controlled and predictable flow conditions. Prolonged dry conditions and water scarcity in recent decades in many major river basins around the world supporting large populations, have led to the development of integrated water resource management plans that set the amount of water that can be taken from the basin each year, while leaving enough environmental water for the rivers, lakes and wetlands and the plants and animals that depend on them (e.g. Hart, 2016). Environmental water management is complex, requiring the release of large volumes of environmental water from storages to be delivered over long distances at sub-seasonal or longer time scales to achieve a range of environmental targets and outcomes, under both regulated and unregulated conditions. To ensure that future water delivery optimises consumptive as well as environmental outcomes, new forecasting and planning tools, and streamlined processes are necessary especially at sub-seasonal time scale.

Considerable benefit can be obtained by producing probabilistic sub-seasonal (0-30 days) forecasts which are "seamless" in time; i.e. from a single product that is reliable and sharp across a range of lead times and aggregation time scales (White et al., 2017). In large water resource systems, such as the Murray-Darling Basin System, with travel times of days to weeks, forecasts of both *daily flows* and *cumulative volumes* are required both shorter (1-7 days) and longer longer lead times (up to 30 days). Seamless forecasts, i.e., forecasts obtained using a single method that maintain high quality across a range of time scales, are clearly attractive in this practial application, since a single users can rely on single forecast product to achieve multiple benefits at different spatial and temporal scales within the river system .

The study provide an overview of the key advances towards seamless subseasonal forecasting. This is undertaken by comparing the recently developed multi-temporal hydrological residual error (MuTHRE) model, one of the first approaches that provides seamless subseasonal forecasting, to an existing baseline residual error model and a non-seamless monthly QPP model. This comparison is in terms of model features and also through forecast evaluation metrics. The paper ends by providing a summary of the key benefits of the MuTRHE approach to seamless subseasonal forecasting.

2. STREAMFLOW FORECASTING MODELS

2.1. MuTHRE Model for Seamless Subseasonal Forecasting.

Streamflow forecasts are subject to uncertainty in rainfall, associated with predicting future rainfall, and hydrological errors, associated with uncertainty in model structure, initial conditions and parameters. In order to represent both sources of uncertainty in streamflow forecasts, the "ensemble dressing" approach (Pagano et al., 2013) is often implemented, whereby (i) replicates of forecast rainfall are propagated through a rainfall-runoff model, and (ii) a residual error model is used to add hydrological errors to each streamflow replicate.

A key challenge is the development of the residual error model, which must capture relevant features of hydrological errors. It is well-known that hydrological errors are heteroscedastic (larger errors for larger flows) and persistent (similar errors for consecutive times), and these features are typically represented in residual error models (e.g., McInerney et al. (2017)). However, other important features which are less commonly represented include

- *Seasonal variability,* due to hydrological models being unable to appropriately capture seasonal variations in streamflow (e.g., Woldemeskel et al. (2018));
- *Dynamic biases,* i.e., shifts in the mean of hydrological errors over longer time periods (e.g. month to year) due to hydrological non-stationarity (e.g., Westra et al. 2014);
- *Non-Gaussian errors*. The random component (innovation) of residual error models are commonly assumed to follow a Gaussian distribution models (e.g., McInerney et al. 2017). However, recent studies have found that non-Gaussian distributions better capture extreme errors (e.g., Li et al. 2016)

The Multi-Temporal Hydrological Residual Error (MuTHRE) model is the first residual error model (to the best of the authors' knowledge) which represents these three temporal error characteristics (i.e., seasonal variability, dynamic biases and non-Gaussian innovations). See McInerney et al. (2020) for the details of how these components are represented and also for a comprehensive evaluation of how these three features provide the MuTHRE model with ability to provide seamless subseasonal forecasts. model. This paper will provide highlight some key outcomes from McInerney et al. (2020).

2.2. "Baseline" Residual error model

The first model that the MuTHRE is compared against is a "baseline" residual error model. This baseline residual error only includes the heteroscedasticity (using Box-Cox transformation), persistence (using an AR(1) model) and uses Gaussian errors. It does not include seasonality, dynamic biases or mixed Gaussian for extreme errors (Figure 2). In terms of its application in forecasting mode, the baseline residual error model follows the same ensemble dressing approach as the MuTHRE model.

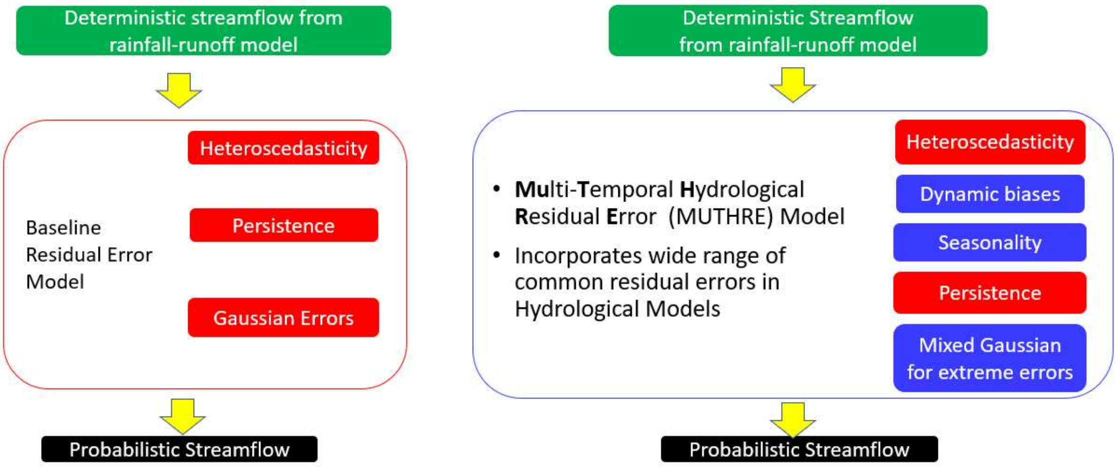

Figure 2. Key differences between MuTHRE and Baseline Residual Error Model

2.3. Non-seamless Monthly QPP Model

The second model used for comparison is the non-seamless Monthly QPP model of Woldemeskel et al. (2018). This model is referred to as non-seamless because it only provides streamflow forecasts at the monthly time scale, in comparison to the MuTHRE model which provides daily streamflow forecasts which can be easily aggregated to the monthly time scale (see Figure 2). Another key difference is that during model calibration, the non-seamless monthly QPP is calibrated against monthly observed streamflow, hence, its parameters "see" the monthly observed streamflow. In contrast the MuTHRE model is calibrated to daily observed streamflow, so its parameters do not "see" the monthly observed streamflow - when its MuTHRE monthly forecasts are evaluated it is essential undertaking temporal extrapolation. Hence, comparison of these two models' ability to match monthly observed streamflow will be a strong test of the MuTHRE model's seamless streamflow forecasting capabilities.

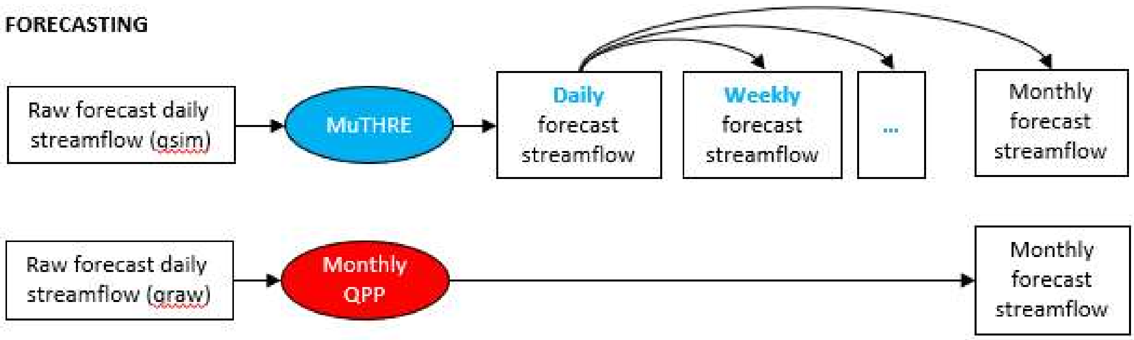

Figure 3. Key differences between seamless daily forecasts of MuTHRE and non-seamless monthly QPP in forecasting mode

3. CASE STUDY

3.1. Hydrological data and model

We compare forecasts from the MuTHRE, baseline, and non-seamless monthly QPP 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 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 forecast performance in terms of reliability and sharpness. The reliability of forecasts is defined such that a forecast is deemed reliable if the forecast probabilities are statistically consistent with observations, (i.e. the 90% forecast probability limits capture 90% of the observations). This is quantified using the commonly used reliability metric (see McInerney et al, 2017 for details). Sharpness (i.e., uncertainty/width/spread 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. Lower metric values indicate better performance. See McInerney et al. (2020) for further details of metrics.

Forecasts are evaluated over (i) multiple time scales from daily to aggregated monthly forecasts, and (ii) multiple stratification types, including by lead time, month and year. Practical significance tests are used to determine whether the MuTHRE model has better or worse performance metrics than the baseline residual error model or the non-seamless monthly QPP over the range of catchments, and whether these differences are of practical relevance (defined as a difference by more than 10% of the median metric value for the baseline or non-seamless monthly QPP, depending on which model is used for comparison.

4. RESULTS

4.1. Forecast Time series

Figure 1 provides an illustration of daily and cumulative streamflow forecasts from the MuTHRE model in the Biggara catchment (401012) during August 2002. The comparison is against 'climatology' which uses the historical range of streamflow from the observed record, and is commonly used in industry for streamflow forecasting. For daily forecasts, the observations lie within the 90% predictive limits for both climatology (Figure 1a) and the MuTHRE model (Figure 1b). However, forecasts from the MuTHRE model are much sharper than climatology. These forecasts are sharpest for short lead times, but are still considerably sharper than climatology for longer lead times. Similarly, 90% limits for cumulative forecasts from the MuTHRE model (Figure 1d) capture the observed values, and are much sharper than climatology.

4.2. Comparison against Baseline Daily Residual Error Model

Daily forecasts

Figure 4 compares performance of *daily* forecasts from the MuTHRE and baseline models for different lead times. The MuTHRE model provides consistent good reliability over all lead times (Figure 4a), and practically significant improvements over the baseline model for short lead times (10 out of the first 11 days). MuTHRE forecasts are much sharper than climatology, especially for short lead times (e.g. median value of 0.2 for lead time of 1 day corresponds to 80% reduction in uncertainty). Compared to the baseline model, sharpness improves for all lead times (Figure 4b), although these are not classified as practically significant.

Monthly forecasts

Figure 5 compares the reliability of *monthly* forecasts when stratified by month and year. The MuTHRE model provides consistent reliability over all months (Figure 5a), with practically significant improvements over the baseline model in 6/7 dry months (November-May). The MuTHRE model also provides improvements in reliability when stratified by year – these are largest in dry years (2006-2009).

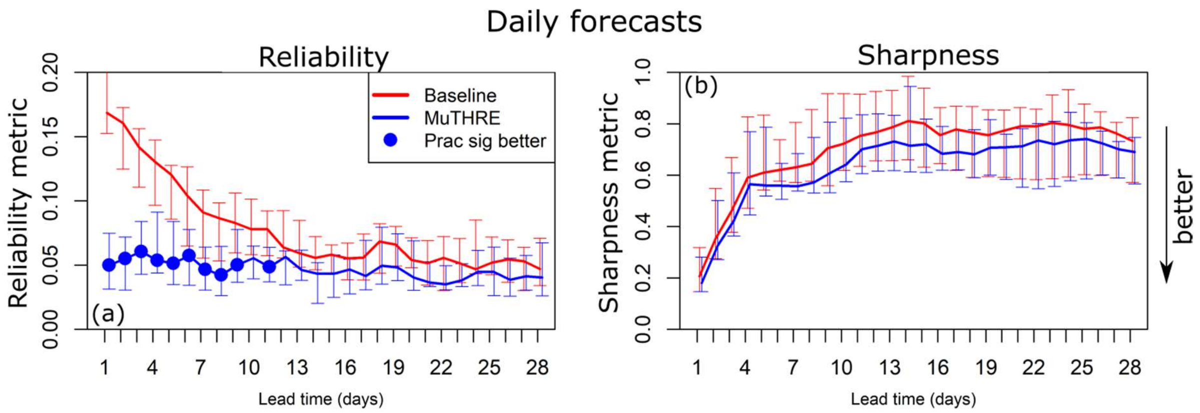

Figure 4. (a) Reliability and (b) sharpness metrics when streamflow forecasts from the MuTHRE and baseline residual error model are stratified by lead time. Lines represent median metric values calculated over the 11 catchments, whiskers represent 90% limits, and circles indicates lead times for which the MuTHRE model produces practically significant better performance than the baseline model.

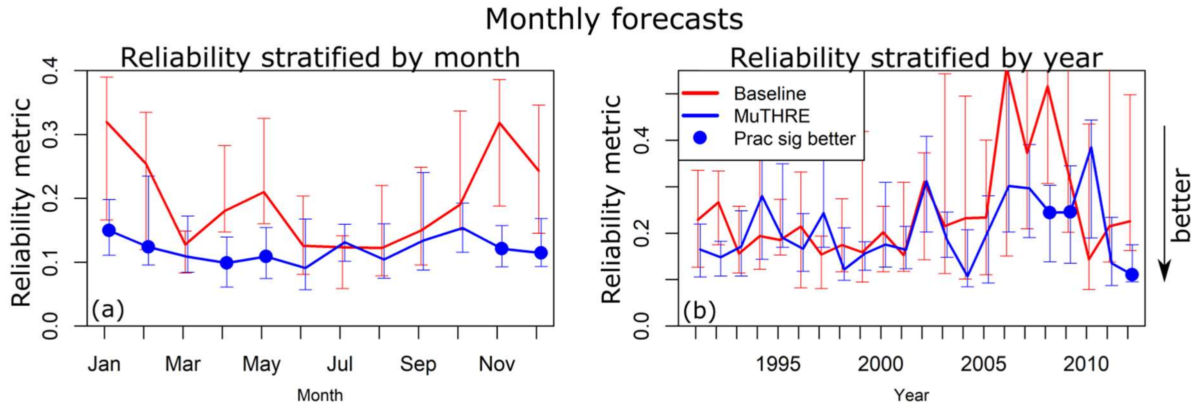

Figure 5. Reliability metrics when monthly forecasts from the MuTHRE and baseline residual error model are stratified by (a) month, and (b) year.

4.3. Comparison against non-seamless Monthly QPP model

The performance of the monthly forecasts from the seamless MuTHRE and the non-seamless monthly QPP is shown in Figure 6, when stratified by month (left column) and year (right column).

In terms of reliability, Figure 6a shows that when performance is stratified by month, the two models have similar reliability for all 12 months. When stratified by year Figure 6b, the MuTHRE model offers similar reliability to the monthly QPP model for 20 out of the 22 years, with the non-seamless monthly QPP model offering practically significant improvements in 2 of the 22 years (Figure 5b). In terms of sharpness, Figure 6c shows that when sharpness is stratified by month, the seamless MuTHRE model provides practically significant improvement in 1 month (September) and similar performance in the other 11 months. Figure 6d shows sharpness stratified by year is similar for both models for all years.

This similarity in performance between the seamless MuTHRE model and non-seamless monthly QPP model is a remarkable achievement that illustrates the seamless forecasting capability of the MuTHRE model. As outlined in section 2.3, when evaluating the ability of monthly forecasts to capture the monthly observed streamflow, the MuTHRE model's ability to undertake temporal extrapolation is being evaluated – it does not "see" the monthly observed streamflow during its model calibration. In contrast, the non-seamless monthly QPP model is being evaluated against its ability to capture monthly observed streamflow that it "sees" during model calibration. Given this key difference the result that they give similar performance is remarkable.

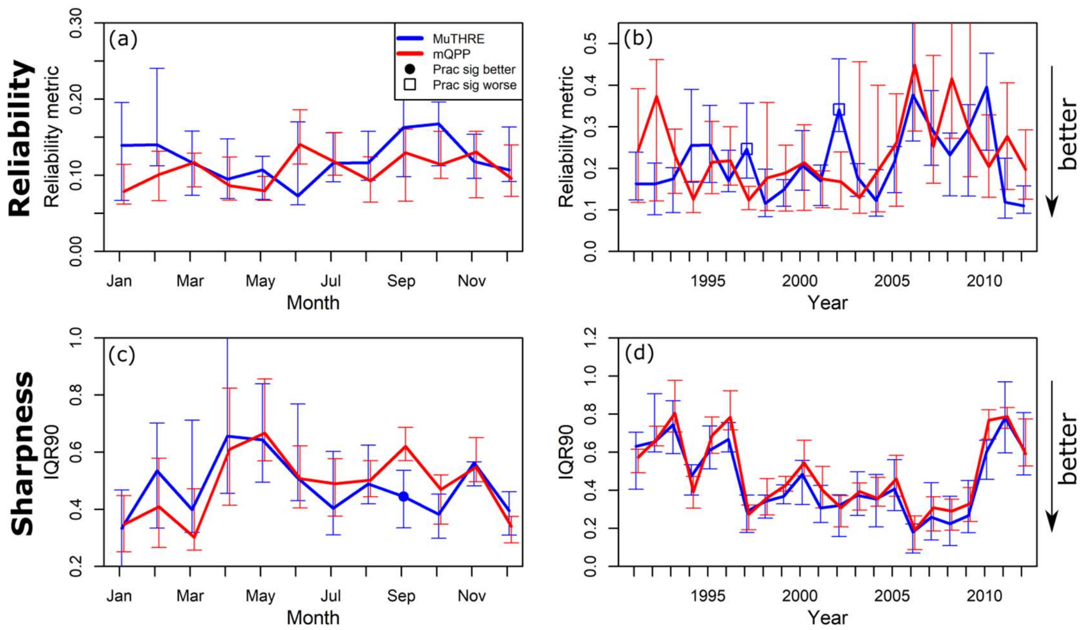

Figure 6. Monthly performance of the seamless MuTHRE and non-seamless monthly QPP forecasts in terms of reliability (top row), sharpness (2nd row), Stratification is performed by month of the year (left column) and by year (right column). Circles/squares indicate that the MuTHRE model performs practically significant better/worse than the monthly QPP model.

5. PRACTICAL BENEFITS OF SEAMLESS SUB-SEASONAL STREAMFLOW FORECASTS

The results presented showed the MuTHRE model's ability to produce seamless sub-seasonal streamflow forecasts. Seamless subseasonal streamflow forecasts with consistent reliability enables water resource managers to confidently utilize sub-seasonal forecasts for decision support in a wide range of applications, including:

- Easily integrate daily streamflow forecasts into exisiting daily river system models, such as eWater Source. River system models are commonly run with historical streamflow inputs (i.e. climatology), so utilizing reliable and sharp sub-seasonal streamflow forecasts would enable improved decision making through better quantifying uncertainty.
- Utilising forecasts for lead times up to 1 month can improved reservoir management of rural water supplies with irrigations demands and environmental flows (Murray-Darling Basin Authority, 2019). For example, if a high streamflow event is forecast with high degree of reliability, the manager could delay/avoid releasing water for environmental flows, and prevent wasting water.
- Forecast informed flood control. Sub-seasonal forecasts can inform the management of multi-purpose reservoirs that serve as both water supply and downstream flood protection services. For example, large volumes of streamflow are forecast, reservoir operators can release water in advance to provide additional flood storage and reduce risks of flooding. This is dependent on the evaluation of forecasts for high flow events, relevent for flood applications - see McInerney et al (2021) for further details.
- Operation of urban water supply systems, which benefit from aggregated monthly forecasts (Zhao and Zhao, 2014). Reliable forecasts can inform managers about whether urban demand can be met from river flows, or whether water needs to be transferred between multiple reservoirs or sourced from desalination.

6. CONCLUSIONS

This paper provided an overview of the advances in subseasonal forecasting, provided by the development of the Multi-Temporal Hydrological Residual Error (MuTHRE) model. The key differences between the MuTHRE and baseline residual error models are that the MuTHRE model accounts for wide range of charactersitics of hydrological errors at multiple time scales, including seasonality, dynamic biases, and extreme errors. Comparison agaisnt an existing non-seamless monthly QPP showed the key difference was that the MuTHRE model was able to produce seamless subseasonal forecasts at daily, weekly and monthly time scales, while the non-seamless monthly QPP model could only produce forecasts at monthly time scales.

This comparison included forecast evaluation on 11 case study catchments in the Murray Darling Basin. The results showed that the MuTHRE model outperforms an existing baseline residual error model to produce seamless sub-seasonal forecasts, with consistent reliability over lead times (1-30 days), timescales (daily to monthly), and all months and years. In particular the MuTHRE model provides large improvements in relaibility for short lead times, dry months and dry years, as well as improvements in sharpness.

Comparison against an existing non-seamless monthly QPP model showed the MuTHRE model provided similar reliability and sharpness for monthly forecasts stratified over months and years. This is a remarkable achievement, given that non-seamless monthly QPP model "sees" the monthly observed streamflow in calibration, whereas the MuTHRE model does not, and instead is undertaking temporal extrapolation from daily to monthly.

The practical benefits of sub-seasonal streamflow forecasts for a wide range of water management applications were outlined. The consistent high quality performance of the MuTHRE model over multiple lead times (1-30 days) and time scales (daily to monthly) provides confidence in the suitability of forecasts for multiple practical applications, including their use in river system models to optimize water delivery for irrigation and enorinmental outcomes.

Further information on the MuTHRE model, including comprehensive analysis of forecast performance and detailed algorithms, can be found in McInerney et al. (2020).

ACKNOWLEDGMENTS

This work was funded by the Australian Bureau of Meteorology. Computations were performed on the Phoenix cluster at the University of Adelaide. The data underlying this research can be accessed as follows: observed rainfall data from www.bom.gov.au/climate, observed streamflow data from www.bom.gov.au/waterdata and post-processed ACCESS-S forecast rainfall from https://doi.org/10.25909/14604180. .

REFERENCES

Hart, B. T. (2016), The Australian Murray–Darling basin plan: challenges in its implementation (part 1), International Journal of Water Resources Development, 32 (6),819-834

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 (9),3561-3579

McInerney et al (2021). Improving the reliability of sub-seasonal forecasts of high and low flows by using a flow-dependent non-parametric model, Water Resources Research https://doi.org/10.1029/2020WR029317

McInerney, D., et al. (2020), Multi-temporal Hydrological Residual Error Modeling for Seamless Subseasonal Streamflow Forecasting, Water Resources Research, 56 (11),2019WR026979

McInerney, D., M. Thyer, D. Kavetski, J. Lerat, and G. Kuczera (2017), Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors, Water Resources Research, 53 (3),2199-2239

Murray-Darling Basin Authority (2019). Joint management of the River Murray. https://www.mdba.gov.au/river-murray-system/joint-management-river-murray, Accessed on 11/12/2019

Pagano, T. C., D. L. Shrestha, Q. J. Wang, D. Robertson, and P. Hapuarachchi (2013), Ensemble dressing for hydrological applications, Hydrological Processes, 27 (1),106-116

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 (2),1615-1628

Westra, S., M. Thyer, M. Leonard, D. Kavetski, and M. Lambert (2014), A strategy for diagnosing and interpreting hydrological model nonstationarity, Water Resources Research, 50 (6),5090-5113

White, C. J., et al. (2017), Potential applications of subseasonal-to-seasonal (S2S) predictions, Meteorological Applications, 24 (3),315-325

Woldemeskel, F., et al. (2018), Evaluating post-processing approaches for monthly and seasonal streamflow forecasts, Hydrology and Earth System Sciences, 22 (12),6257-6278

Zhao, T., and J. Zhao (2014), Joint and respective effects of long- and short-term forecast uncertainties on reservoir operations, Journal of Hydrology, 517,83-94