Seamless streamflow ~~model provides forecasts~~forecasting at ~~all scales from~~ daily to monthly scales: MuTHRE lets you have your cake and ~~matches the performance of non-seamless monthly model~~ eat it too

David McInerney[1], Mark Thyer[1], Dmitri Kavetski[1], Richard Laugesen[2], Fitsum Woldemeskel[3], Narendra Tuteja[2], George Kuczera[4]

[1]School of Civil, Environmental and Mining Engineering, University of Adelaide, SA, Australia
[2]Bureau of Meteorology, ACT, Canberra, Australia
[3]Bureau of Meteorology, VIC, Melbourne, Australia
[4]School of Engineering, University of Newcastle, Callaghan, NSW, Australia

Correspondence to: David McInerney (david.mcinerney@adelaide.edu.au)

Abstract. Subseasonal streamflow forecasts inform a multitude of water management decisions, from early flood warning to reservoir operation. 'Seamless' forecasts, i.e., forecasts that are reliable and sharp over a range of lead times (1-30 days) and ~~when aggregated to multiples~~aggregation time scales (e.g. daily ~~and~~to monthly) are of clear practical interest. However, existing ~~forecasting~~forecast products are often 'non-seamless', i.e., ~~designed~~developed and applied for a single time scale and lead time (e.g. 1 month ahead). If seamless forecasts are to be a viable replacement for existing 'non-seamless' forecasts, it is important that they offer (at least) similar predictive performance at the time scale of the non-seamless forecast.

This study compares forecasts from two probabilistic streamflow post-processing (QPP) models: the recently developed seamless daily Multi-Temporal Hydrological Residual Error (MuTHRE) model ~~to~~and the more traditional (non-seamless) monthly ~~streamflow post-processing (~~QPP~~)~~ model ~~that was~~ used in the Australian Bureau of Meteorology's Dynamic Forecasting System. Streamflow forecasts from both post-processing models are generated for 11 Australian catchments, using the GR4J hydrological model and ~~post~~pre-processed rainfall forecasts from the ACCESS-S ~~climate~~numerical weather prediction model. Evaluating monthly forecasts with key performance metrics (reliability, sharpness, bias and CRPS skill score), we find that the seamless MuTHRE model ~~provides~~achieves essentially the same performance as the non-seamless monthly QPP model for the vast majority of metrics and temporal stratifications (months and years). ~~When this outcome is combined with the numerous practical benefits of~~In other words, seamless forecasts ~~it is clear~~do not require a compromise between capabilities and performance. This finding demonstrates that seamless forecasting technologies, such as the MuTHRE post-processing model, are not only viable, but a preferred choice for future research development and practical adoption ~~of~~in streamflow forecasting.

Style Definition: Heading 4: Font: Italic

Formatted: Normal, Left

1. Introduction

Subseasonal streamflow forecasts (with lead times up to 30 days) can be used to inform a range of water management decisions, from flood warning and reservoir flood management at shorter lead times (e.g. up to a week) to river basin management at time scales up to a month. ~~Since~~The uncertainty in these forecasts is often represented using ensemble and probabilistic methods. Probabilistic streamflow forecasts have traditionally been developed and applied at only a single lead time and time scale (e.g., Souza Filho and Lall, 2003; Pal et al., 2013; Hidalgo-Muñoz et al., 2015; Mendoza et al., 2017; Gibbs et al., 2018). However, since different applications require forecasts over a range of lead times and time scales, recent research has focussed on producing *seamless* forecasts, i.e. forecasts from a single product that are (statistically) reliable and sharp across multiple lead times and aggregation time scales (McInerney et al., 2020a; McInerney et al., 2020b). ~~Current forecasting practice often employs~~ For seamless forecasts to be a viable replacement for more traditional *non-seamless* forecasts~~,~~ (i.e. forecasts ~~that are developed and applicable at only~~for a single lead time and time scale ~~(e.g., Mendoza et al., 2017; Gibbs et al., 2018; Woldemeskel et al., 2018). For seamless forecasts to be a viable replacement for *non-seamless* forecasts,~~) it is important to establish that the performance of seamless forecasts is competitive with their non-seamless counterparts at the native time scale of the latter. ~~This is the focus of our study.~~

Recent research by McInerney et al. (2020a); McInerney et al. (2020b) has shown that seamless subseasonal forecasting is achievable. ~~That study~~McInerney et al. (2020b) developed the Multi-Temporal Hydrological Residual Error (MuTHRE) model~~, which represents seasonality, dynamic biases and non-Gaussian errors.~~ for post-processing daily streamflow forecasts in order to improve reliability across a range of time scales. Using a case study with 11 catchments in the Murray Darling Basin, Australia, it was concluded that subseasonal forecasts generated using the MuTHRE streamflow post-processing model are indeed *seamless:* daily forecasts are consistently reliable (i) for lead times between 1 and 30 days, and (ii) when aggregated to the monthly ~~forecasts~~scale.

Seamless subseasonal forecasts~~, from residual error models such as MuTHRE, produce reliable forecasts~~ are reliable over a wide range of aggregation time scales (e.g. daily to monthly) and lead times (1-30 days). In contrast, non-seamless forecasts are either: (i) only available at a single time scale (e.g. a post-processing model developed directly at the monthly~~), and~~ scale does not generate daily forecasts), or (ii) cannot be reliably aggregated to longer time scales~~.~~ (e.g., from daily to monthly). The practical benefits of ~~this~~seamless forecasts are ~~outlined~~ as follows:

1. **Seamless forecasts can be used to inform decisions at a range of time scales.** Forecast users can utilize seamless subseasonal forecasts to inform a wide range of decisions, including
   - Flood warning, where short-term forecasts (up to 1 week) on individual days are of practical interest (Cloke and Pappenberger, 2009);
   - ~~Hydro-electric reservoir management~~Managing hydropower systems, which can utilize forecasts of inflow between 7 and 15 days to increase production in the electricity grid (Boucher and Ramos, 2019);
   - Managing reservoirs for rural water supply, where forecast volumes over long aggregation scales (e.g. weeks/months), and at long lead times (up to 1 month), are required due to long travel times (Murray-Darling Basin Authority, 2019);

Formatted: Font: Not Italic

Formatted: Normal, Left

- Operation of urban water supply systems, where monthly forecasts are of value (Zhao and Zhao, 2014).

2. **Seamless daily forecasts are easily integrated into river system models used for real-time decision-making.** Perhaps the greatest potential for seamless forecasts is their use as input into real-time decision-making tools used by urban and rural water authorities. These tools include river system models (e.g. eWater Source, Welsh et al., 2013), which run natively at the daily scale and are used to inform resource management decisions over larger time scales. ~~Streamflow~~Non-seamless streamflow forecasts ~~from non-seamless models~~ cannot be used as input into these models, ~~since~~because they do not match the ~~required input~~ time scale of the river system model, and are not reliable when aggregated to longer time scales (e.g. from daily to monthly).

3. **Seamless forecasts simplify forecasting systems, as a single seamless product can serve a range of forecast requirements at different time scales.** As forecasts are often required at multiple time scales (e.g. daily~~-~~ to monthly), non-seamless forecast strategies require developing ~~multiple~~models (e.g. hydrological, statistical or post-processing ~~models at specific~~) for each time ~~scales~~scale of interest (e.g. a daily model and a monthly model). Seamless forecasts offer practical benefits to forecast~~-~~ providers, ~~such as~~e.g. the Australian Bureau of Meteorology, as they reduce the need to develop multiple non-seamless forecasts for different applications. A seamless forecasting system offers a single product that can serve a wide range of forecast requirements.

These practical benefits of seamless forecasts provide a clear motivation for their development and use. However, for seamless forecasts to be a viable replacement for non-seamless forecasts, it is important that they do not come at the cost of a substantial ~~drop in~~loss of performance at the native time scale of the non-seamless forecast. For example, if aggregated forecasts from a seamless daily model were considerably worse than monthly forecasts from an existing non-seamless model, users of the monthly forecasts would prefer to continue using forecasts from the non-seamless model. In general, one might expect forecasts from a non-seamless model, ~~which is~~ developed and calibrated at single time scale, to provide superior performance ~~than~~compared to forecasts from a seamless ~~forecasts~~model calibrated at shorter time scale and then aggregated. While the non-seamless model has only '~~do~~one job to ~~do~~do', which is to provide quality forecasts at single time scale, the seamless model is expected to produce good performance over a range of lead times and aggregation time scales. Herein lies a major challenge of seamless forecasting.

Our interest in comparing the performance of aggregated seamless forecasts with non-seamless forecasts at their native time scale has similarities to previous research in aggregating deterministic streamflow predictions. For example, Wang et al. (2011) found that the WAPABA monthly rainfall-runoff model produced similar/better performance than aggregated predictions from the SIMHYD/AWBM daily rainfall-runoff models, despite only using observed monthly forcing data. Yang et al. (2016) compared daily and sub-daily versions of the SWAT model (with daily and sub-daily observed rainfall inputs) and found large differences in the partitioning of baseflow and direct runoff. However, to the best of the authors' knowledge, no studies have compared aggregated *probabilistic* forecasts from a seamless model against probabilistic forecasts from a non-seamless model. The aim of this study is to *establish whether aggregated forecasts from a (probabilistic) seamless model achieve comparable performance to those from a non-seamless ~~forecasting~~ (probabilistic) model at its native time scale*. This aim is achieved by

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Normal, Left

comparing the monthly forecast performance of the seamless MuTHRE post-processing model (aggregated from daily to monthly) against the non-seamless monthly streamflow post-processing model ~~of Woldemeskel et al. (2018),~~ used in the Australian Bureau of Meteorology's Dynamic Forecasting System~~.~~ (Woldemeskel et al., 2018).

The remainder of the paper is organized as follows. Section ~~2~~2 describes the ~~QPP models and~~forecasting methods, with a focus on the streamflow post-processing models, Section ~~3~~3 introduces the case study methods ~~used to evaluate the QPP models.,~~ Sections ~~4~~4 and 5 present and discuss case study results, ~~while~~and Section 6 provides concluding remarks.

Formatted: Normal, Left

2. Forecasting ~~models~~methods

~~This section describes the seamless MuTHRE daily streamflow post-processing (QPP) model and the non-seamless monthly QPP model.~~

~~2.1. Probability model~~

~~Both QPP models can be represented as a probability model (Q_t) for streamflow q_t at time t ,~~

~~$$q_t \sim Q_t\left(\boldsymbol{\theta}; \mathbf{x}_t, \mathbf{s}_{t-1}, \tilde{\mathbf{q}}_{\omega(t_0)}\right)$$~~

The forecasting methods investigated in this study share a similar general structure but differ in the streamflow post-processing model. To facilitate the presentation, this section is organised as follows. The general structure is outlined in Section 2.1. Common features of the post-processing models are described in Section 2.2. Specific details of the MuTHRE and monthly QPP models are described in Sections 2.3 and 2.4.

2.1. General structure

The forecasting methods in this study employ a deterministic hydrological model forced with an ensemble of rainfall forecasts and combined with a streamflow post-processing (QPP) model. This general structure is illustrated schematically in Figure 1 and detailed next.

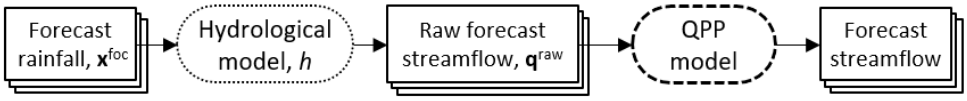


Figure 1: Illustration of general approach used to produce streamflow forecasts. Layers represent ensemble members.

The deterministic hydrological model, $h(\boldsymbol{\theta}_h; \mathbf{x}_t, \mathbf{s}_{t-1})$, has a (single) set of parameters $\boldsymbol{\theta}_h$, inputs $\mathbf{x}_t$ (including forecast rainfall $\mathbf{x}^{foc}$), and states $\mathbf{s}_{t-1}$ at time $t-1$. In general any rainfall-runoff model can be used for this purpose; in our case study we employ the rainfall-runoff model GR4J (see Section 3.2).

The streamflow forecasts are obtained in two steps. First, an ensemble of N_{foc} rainfall forecasts $\{\mathbf{x}^{foc(f)}; f = 1,...,N_{foc}\}$ generated by a numerical weather prediction model is propagated through the deterministic hydrological model to generate a corresponding ensemble of 'raw' streamflow forecasts, $\{\mathbf{q}^{raw(f)}; f = 1,\ldots,N_{foc}\}$. Second, a probabilistic streamflow post-processing model is applied to the raw forecasts to generate the (post-processed) streamflow forecasts $\{\mathbf{q}^{(f)}; f = 1,\ldots,N_{foc}\}$.

The streamflow post-processing models are constructed using the residual error modelling approach. They comprise a deterministic component and a residual error model. The residual error model employs a streamflow transformation to

Formatted: Space Before: 0 pt

Field Code Changed

Formatted: Normal, Left

represent the heteroscedasticity and skew of the errors, an autoregressive term to represent error persistence, and components to capture other features of errors such as seasonality.

Formatted: Normal, Left

We consider two forecasting methods, which differ in the structure and details of the streamflow post-processing model. A schematic representation of these models is given in Figure 2a.

- **Seamless MuTHRE streamflow post-processing model** (McInerney et al., 2020b). The residual error model is formulated at the *daily* scale and is applied directly to (daily) raw streamflow forecasts. Conceptually, the ensemble of raw streamflow forecasts accounts for forecast rainfall uncertainty and the residual error model accounts for hydrological uncertainty.
- **Non-seamless monthly streamflow post-processing (QPP) model** (Woldemeskel et al., 2018). The residual error model is formulated at the *monthly* scale. It is applied to raw streamflow forecasts aggregated to the monthly scale and collapsed to their medium value. Conceptually, the residual error model accounts for both hydrological and forecast rainfall uncertainty.

The post-processing models also differ in their parameter estimation (calibration) procedure. Figure 2b shows that the MuTHRE model is calibrated using observed daily rainfall and observed daily streamflow, whereas the monthly QPP model is calibrated to forecast daily rainfall and observed monthly streamflow (see Sections 2.3.4 and 2.4.4 for details).

Figure 2c illustrates the key operational distinction between the models. The MuTHRE model produces seamless daily streamflow forecasts that can be used at a range of lead times and aggregation periods (e.g. daily, weekly, fortnightly, monthly). In contrast, the monthly QPP model produces only one-month ahead non-seamless monthly forecasts.

The next section presents common features of the post-processing models, before moving to specific model details.

2.2. Streamflow post-processing model

2.2.1. Deterministic component

The deterministic component q_t^{det} is obtained from the raw streamflow forecasts (Figure 2a). The deterministic component used in the seamless MuTHRE and non-seamless monthly streamflow post-processing approaches are detailed in Sections 2.3.2 and 2.4.2 respectively.

2.2.2. Residual error model

The residual error model describing the relationship between the probabilistic streamflow estimate Q_t and the deterministic component q_t^{det} is formulated as additive in transformed space,

$$z(Q_t;\boldsymbol{\theta}_z) = z(q_t^{\text{det}};\boldsymbol{\theta}_z) + \eta_t \quad (1)$$

~~where $\boldsymbol{\theta}$ are parameters of the hydrological and error models (described below), $\mathbf{x}_t$ are inputs to the hydrological model, including rainfall, $\mathbf{s}_{t-1}$ are states of the hydrological model at time $t-1$, and $\tilde{\mathbf{q}}_{\omega(t_0)}$ are observed streamflow from a preceding time period $\boldsymbol{\omega}(t_0)$ up to time t_0. The units for time t are days for the seamless MuTHRE model, and months for the non-seamless monthly QPP model.~~

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Formatted: Normal, Left

~~The probability model Q_t is a residual error model, where a residual error η_t is added (in transform space) to a deterministic prediction q_t^{det},~~

$$z(Q_t;\boldsymbol{\theta}_z) = z(q_t^{\text{det}};\boldsymbol{\theta}_z) + \eta_t \quad (2)$$

~~The deterministic prediction q_t^{det} is computed from a deterministic hydrological model $h(\boldsymbol{\theta}_h;\mathbf{x}_t,\mathbf{s}_{t-1})$, as detailed in Sections 2.2.2 and 2.3.2 for the seamless and non-seamless approaches respectively. Note that the hydrological model h has parameters $\boldsymbol{\theta}_h$, which for the purposes of describing the QPP model are assumed to be pre-calibrated.~~

~~The model h is used to produce~~

~~i. "Simulated" streamflow $\mathbf{q}^{\text{sim}}$, when observed rainfall $\tilde{\mathbf{x}}$ is used as input forcing in h, and~~

~~ii. An ensemble of "raw" streamflow forecast replicates, $\{\mathbf{q}^{\text{raw}(f)}; f = 1,\ldots,N_{\text{foc}}\}$, when an ensemble of N_{foc} forecast rainfall replicates $\{\mathbf{x}^{\text{foc}(f)}; f = 1,\ldots,N_{\text{foc}}\}$ is used to force h. The raw streamflow forecasts account for forecast rainfall uncertainty, but do not account for hydrological uncertainty associated with errors in hydrological model structure and initial conditions.~~

where η_t is a random residual error term.

Field Code Changed

Formatted: Normal, Left

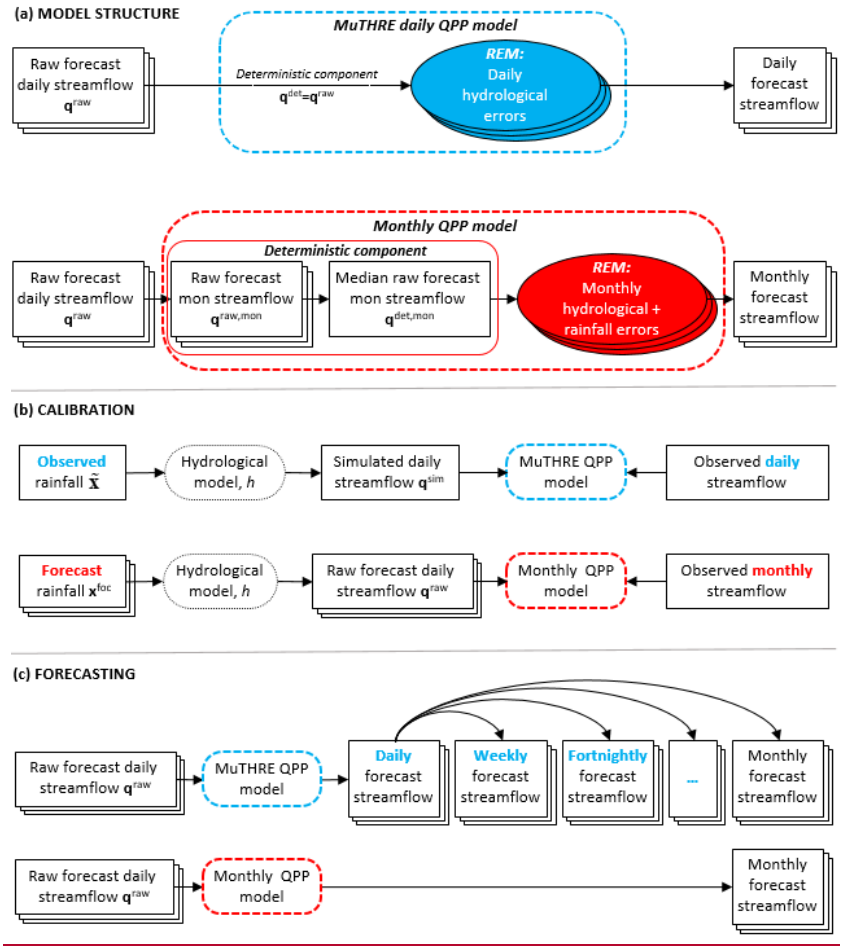


Figure 2 Conceptual diagrams of the seamless MuTHRE model and the non-seamless monthly QPP model. Ensemble components are indicated with multiple 'layers'. Panel (a) shows the post-processing model structure including the deterministic component and the residual error model (REM). Panel (b) shows the calibration approach to estimate the parameters of the streamflow post-processing model. Panel (c) illustrates the key distinction between the forecasting products generated by the models: the MuTHRE model produces seamless daily streamflow forecasts that can be used at a range of lead times and aggregation periods (e.g. daily, weekly, fortnightly, monthly), whereas the monthly QPP model produces only monthly forecasts.

Formatted: Normal, Left

The transformation z, with parameters $\boldsymbol{\theta}_z$, is used to reduce ~~both~~the heteroscedasticity and skewness in residuals. We choose the Box Cox transformation (e.g., Box and Cox, 1964),

$$z(q;\lambda,A)=\begin{cases}\dfrac{(q+A)^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0\\ \log(q+A) & \text{otherwise}\end{cases} \quad (2~~3~~)$$

with parameters $\boldsymbol{\theta}_z=\{\lambda,A\}$. The power parameter λ is set to 0.2 in both ~~QPP~~streamflow post-processing models (McInerney et al., 2017). ~~For~~In the seamless MuTHRE model, the offset parameter A is inferred as part of the hydrological model calibration (McInerney et al., 2020a)~~, while for the non-seamless monthly QPP model it is set to 1% of the mean observed monthly streamflow, i.e.~~(McInerney et al., 2020b), while in the non-seamless monthly QPP model it is set to 1% of the mean observed monthly streamflow, i.e. $A=0.01\times\text{mean}(\tilde{\mathbf{q}}^{\text{mon}})$ (Woldemeskel et al., 2018).

The residual error term η_t is standardized and then modelled as an AR(1) process ~~after standardization~~,

$$\nu_t=(\eta_t-\mu_t)/\alpha_t \quad (3)$$

$$\nu_t=\phi_\eta\nu_{t-1}+y_t \quad (4)$$

~~$\nu_t=(\eta_t-\mu_t)/s_t$ (5)~~

where μ_t and ~~s_t~~ α_t are the (time-varying) mean and scaling factor of η_t, ϕ_η is the lag-1 autoregressive parameter, and y_t is the random component (referred to as the ~~"innovation")~~'innovation') at time t.

When generating forecasts, recent streamflow observations are used to update errors via the AR(1) model, and reduce uncertainty in η_t for short lead times.

~~2.2.~~2.3. Seamless MuTHRE model

~~**2.2.1. Overall approach**~~

2.3.1. Model structure

The seamless MuTHRE post-processing model operates at the daily time scale.

Uncertainty due to ~~both~~ forecast rainfall and hydrological errors is represented using the ensemble dressing approach (Pagano et al., 2013)~~.~~. The ensemble of daily raw streamflow forecasts, $\mathbf{q}^{\text{raw}}$, obtained by propagating an ensemble of rainfall forecasts through the hydrological model ~~h (as described in Section 2.1),~~ h, accounts for forecast rainfall uncertainty. A randomly

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Formatted: Normal, Left

generated replicate of the residual term, $\boldsymbol{\eta}$, is then added to each ~~replicate~~ of the N_{foc} raw streamflow ~~forecasts~~forecast ensemble members to account for hydrological uncertainty. This produces an ensemble of N_{foc} post-processed streamflow forecasts. See schematic in Figure 2a. Note that this approach to capturing forecast rainfall and hydrological uncertainty ~~relies on~~requires the rainfall forecasts ~~being~~to be reliable in order to produce reliable streamflow forecasts (Verkade et al., 2017).

~~2.2.2.~~2.3.2. Deterministic ~~model implementation~~component

In the context of equation ~~(2)~~(1), the deterministic ~~term~~component in the MuTHRE model at its *daily* time step *t* is

$$q_t^{\text{det}} = q_t^{\text{raw}(f)} = h(\boldsymbol{\theta}_h; \mathbf{x}_t^{(f)}, \mathbf{s}_{t-1}) \quad (5\text{~~6~~}6)$$

i.e., the residual error model is applied directly to each ~~individual~~ensemble member of the raw ~~forecast.~~forecasts (Figure 2a).

~~2.2.3.~~2.3.3. Residual error model ~~implementation~~

The MuTHRE model assumes that the mean of the residual error – μ_t in equation ~~(5)~~(3) – varies in time due to ~~"seasonality"~~'seasonality' and ~~"~~'dynamic ~~biases"~~biases' (associated with hydrologic non-stationarity),

$$\mu_t = \mu_{d(t)}^{(s)} + \mu_t^{(b)} + \mu^* \quad (6\text{~~7~~}7)$$

The seasonality component $\mu_{d(t)}^{(s)}$ describes the mean value of $\boldsymbol{\mu}$ on the day-of-the-year $d(t)$, the dynamic bias term $\mu_t^{(b)}$ describes the mean value of $\boldsymbol{\mu}$ (after removing seasonality) over the preceding N_b days ($N_b = 30$ is used), and μ^* is a constant to capture the remaining ~~constant~~bias. Full details of these terms are provided in McInerney et al. (2020c).

~~In the MuTHRE model, the~~The scaling factor – ~~s_t~~ α_t in equation ~~(5)~~(3) – is constant (set to 1 for simplicity).

Innovations are modelled using a two-component mixed-Gaussian distribution

$$y_t \sim \mathcal{N}_{\text{mix}}\left(0, \sigma_1^2, 0, \sigma_2^2, w_1\right) \quad (7\text{~~8~~}8)$$

where σ_1 and σ_2 are the standard deviations of the two components, and w_1 is the weight of the first component (with component means set to zero). Compared to a standard Gaussian distribution, the mixed-Gaussian distribution allows for ~~fatter~~fattier tails (i.e., excess kurtosis) in the distribution of innovations, which has been shown to improve reliability of daily forecasts at short lead times (Li et al., 2016; McInerney et al., 2020a). ~~(Li et al., 2016).~~ Note that the mixed-Gaussian distribution does not offer benefits at longer lead times, nor when aggregating forecasts to the monthly scale (McInerney et al., 2020b).

Field Code Changed
Field Code Changed
Formatted: Font: (Default) Times New Roman
Formatted: Space After: 0 pt
Formatted: English (United Kingdom)
Formatted: Font: (Default) Times New Roman
Formatted: Space After: 0 pt
Field Code Changed
Formatted: Font: (Default) Times New Roman
Formatted: Space After: 0 pt
Formatted: Normal, Left

~~2.2.4.~~2.3.4. Calibration ~~of residual error model~~

~~In the seamless MuTHRE model the residual term $\boldsymbol{\eta}$ represents hydrological uncertainty only, i.e. it does not include forecast rainfall uncertainty.~~ The parameters of the residual error model ~~$\{\phi_\eta, \sigma_1^2, \sigma_2^2, w_1\}$~~ $\{\boldsymbol{\mu}^{(s)}, \boldsymbol{\mu}^{(b)}, \mu^*, \phi_\eta, \sigma_1^2, \sigma_2^2, w_1\}$ are ~~calibrated using data at~~estimated from the following *daily* scale~~:~~ data (see Figure 2b):

(i) Daily hydrological model simulations $\mathbf{q}^{sim}$ forced with observed rainfall ~~($\mathbf{q}^{sim}$, described in Section 2.1)~~ $\tilde{\mathbf{x}}$;

(ii) Daily observed streamflow ~~($\tilde{\mathbf{q}}$).~~ $\tilde{\mathbf{q}}$.

~~Full details of the calibration procedure are provided in~~ McInerney et al. (2020a)~~.~~

Seasonality ($\boldsymbol{\mu}^{(s)}$) and dynamic bias ($\boldsymbol{\mu}^{(b)}$) terms are calculated using moving averages, parameters μ^* and ϕ_η are estimated as the sample mean and lag-1 auto-correlation of the de-trended residuals, while the mixed-Gaussian parameters $\{\sigma_1^2, \sigma_2^2, w_1\}$ are estimated using maximum-likelihood. Full details of the calibration procedure are provided in McInerney et al. (2020b).

~~2.3.~~2.4. Non-seamless monthly QPP model

~~2.3.1. Overall approach~~

2.4.1. Model structure

The non-seamless monthly QPP model operates at the monthly time scale. The ~~ensemble of daily~~ raw ~~streamflow~~ forecasts ~~$\{\mathbf{q}^{raw(f)}; f = 1, \ldots, N_{foc}\}$ (obtained using the rainfall ensemble as input to the hydrological model *h*)~~ is~~are~~ aggregated from ~~the~~ daily to monthly ~~time~~ scale ~~$\{\mathbf{q}^{raw,mon(f)}; f = 1, \ldots, N_{foc}\}$~~, and ~~then~~ collapsed to ~~a deterministic forecast by taking the~~their median value ~~at each time step~~ ~~$\mathbf{q}^{det,mon}$~~ ~~(yielding~~ $\mathbf{q}^{det,mon}$, i.e.~~.~~, the uncertainty from the raw streamflow ~~replicates~~ensemble is discarded~~). Combined~~. The combined forecast rainfall uncertainty and hydrological uncertainty ~~is then~~are represented through the residual error term $\boldsymbol{\eta}$~~, with replicates of $\boldsymbol{\eta}$ added to the deterministic forecast $\mathbf{q}^{det,mon}$ to produce the monthly $\boldsymbol{\eta}$~~. Monthly streamflow forecasts are obtained from $\mathbf{q}^{det,mon}$ by adding N_{foc} replicates of $\boldsymbol{\eta}$. See schematic in Figure 2a.

~~2.3.2.~~2.4.2. Deterministic ~~model implementation~~component

~~In the context of equation (2), the deterministic term~~The deterministic component in the non-seamless model at its *monthly* time step *t* is computed as follows,

$$q_t^{raw,mon(f)} = \text{average}\left(q_{t^*}^{raw(f)}; t^* \in T(t)\right) \quad (8)$$

$$q_t^{det} = \text{median}\left(q_t^{raw,mon(f)}; f = 1, \ldots, N_{foc}\right) \quad (9)$$

Field Code Changed
Field Code Changed
Field Code Changed
Formatted: Indent: Left: 0.25 cm, Hanging: 0.75 cm, Space After: 0 pt
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Field Code Changed
Formatted: Normal, Left

~~$$q_t^{\text{det}} = \text{median}\left(q_t^{\text{raw,mon}(f)}; f = 1,\ldots,N_{\text{foc}}\right) \quad (10)$$~~

where $T(t)$ is averaging window (range of days) corresponding to the monthly time step *t*. ~~In other words, the residual error model is applied at the monthly scale and after collapsing the ensemble of raw forecasts to a single time series.~~

~~2.3.3.~~2.4.3. Residual error model ~~implementation~~

~~The~~The residual error model is applied at the monthly scale after collapsing the ensemble of raw forecasts to a single time series.

The monthly residual error model captures seasonality in residuals by varying the mean μ_t and scaling factor ~~s_t by month.~~ α_t in equation (3) by month. Innovations are assumed to be independent and identically distributed Gaussian,

$$y_t \sim \mathcal{N}\left(0, \sigma_y^2\right) \quad (10\text{~~11~~})$$

where σ_y is the standard deviation of the innovations.

~~2.3.4.~~2.4.4. Calibration of residual error model

~~In the non-seamless monthly QPP model the residual term $\boldsymbol{\eta}$ represents combined forecast rainfall and hydrological uncertainty.~~ The parameters of the monthly residual error model ~~$\{\phi_\eta, \sigma_y^2, \{\mu_m, s_m; m = 1,\ldots,12\}\}$~~ $\{\phi_\eta, \sigma_y^2, \{\mu_m, \alpha_m; m = 1,\ldots,12\}\}$ are ~~calibrated using data at~~estimated from the following *monthly* scale~~:~~ data (see Figure 2b):

(i) Monthly deterministic forecasts $\mathbf{q}^{\text{det,mon}}$ obtained using forecast rainfall ~~($\mathbf{q}^{\text{det,mon}}$~~, as described in Section ~~2.3.1)~~2.4.2;

(ii) Monthly observed streamflow ~~($\tilde{\mathbf{q}}^{\text{mon}}$)~~ $\tilde{\mathbf{q}}^{\text{mon}}$.

All parameters are calibrated using the method-of-moments. Full details ~~of the calibration approach~~ are provided in Woldemeskel et al. (2018).

~~2.4. Differences between the MuTHRE and monthly QPP models~~

~~The seamless MuTHRE and non-seamless monthly QPP models differ in their model structure, and hence their approach to calibration and forecasting.~~

~~2.4.1. Differences in model structure~~

~~The residual error models used in the MuTHRE and monthly QPP models represent different sources of uncertainty and have differences in their implementations, as outlined below and shown in Figure 1a:~~

Formatted: Font: (Default) Times New Roman, English (United States)
Field Code Changed
Formatted: Space After: 0 pt
Field Code Changed
Field Code Changed
Formatted: Indent: Left: 0.25 cm, Hanging: 0.75 cm, Space After: 0 pt
Field Code Changed
Formatted: Font: Bold, Font color: Black, English (United Kingdom), Kern at 16 pt
Formatted: Normal, Left

- The seamless MuTHRE model uses a *daily* residual error model to capture *only hydrological uncertainty*. It applies residual errors directly to raw streamflow forecasts
- The non-seamless monthly QPP model uses a *monthly* residual error model to capture *both hydrological and rainfall uncertainty*. Daily raw streamflow forecasts are aggregated to monthly raw streamflow forecasts, then the ensemble is reduced down to its median, and residual errors are applied to this monthly time series.

The residual error model used in the seamless MuTHRE model and non-seamless monthly QPP models differ in their implementation of the following aspects:

a) *Seasonality*. The seamless MuTHRE model only varies the mean with the time of year, while the non-seamless monthly QPP model varies both the mean and scaling factor.

b) *Incorporation of recent streamflow data*. The seamless MuTHRE model uses monthly observed streamflow from the previous month to update recent hydrological biases, as well as daily observed streamflow from the most recent day to update the daily AR(1) model. The non-seamless monthly QPP model only uses monthly observed streamflow from the previous month to update the monthly AR(1) model.

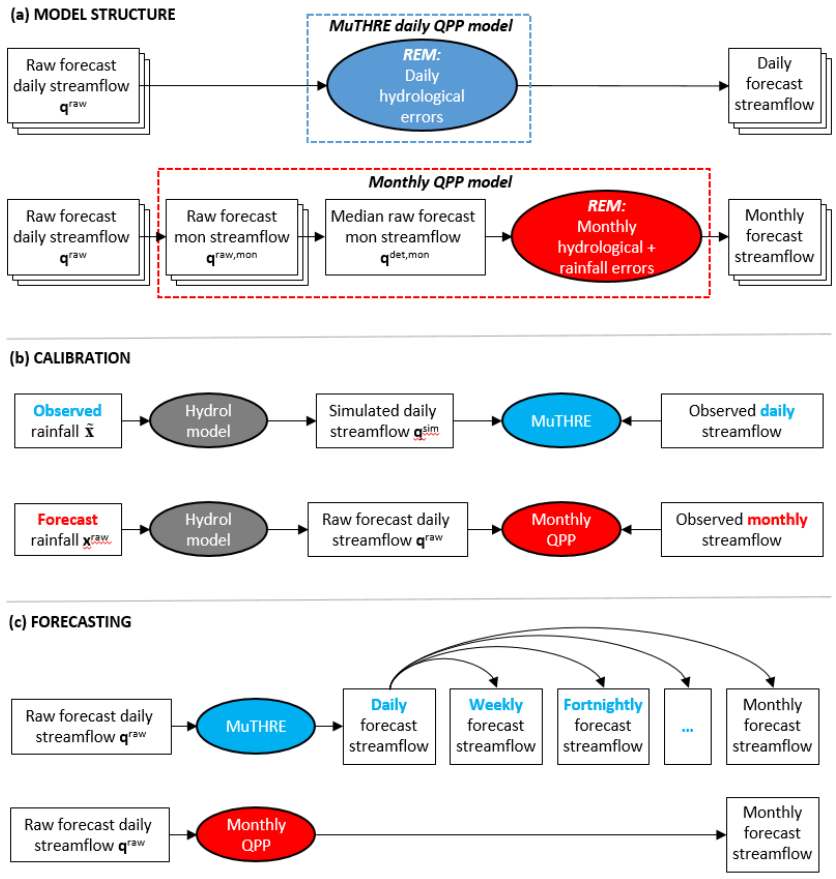
c) *Distribution used to model innovations*. The seamless MuTHRE model uses a mixed-Gaussian distribution for daily innovations, compared with the non-seamless monthly QPP model which uses a standard Gaussian distribution for monthly innovations. The mixed-Gaussian distribution has been shown to improve reliability of daily forecasts at short lead times, but does not impact performance of daily forecasts at longer lead times, or when aggregated to the monthly scale (McInerney et al., 2020a).

2.4.2. Differences in calibration approach

Both approaches use the same deterministic hydrological model calibrated using observed rainfall and streamflow data. However, due to structural differences in their representation of residual errors, the seamless MuTHRE and non-seamless monthly QPP models differ in approach used to calibrate the residual error model parameters. The key differences are illustrated in Figure 1(b) and outlined below:

- The seamless MuTHRE model uses hydrological model simulations forced by *observed* rainfall, while the non-seamless monthly QPP model parameters uses *forecast* rainfall as input to the hydrological model;
- The seamless MuTHRE model is calibrated using *daily* observed streamflow data, while the non-seamless monthly QPP model is calibrated using *monthly* observed streamflow.

Formatted: Normal, Left



~~**Figure 1: Conceptual diagrams illustrating key differences between the seamless MuTHRE model and non-seamless monthly QPP model. The diagram for model structure in panel (a) shows the MuTHRE model applies daily hydrological errors to raw streamflow forecasts to produce an ensemble of daily streamflow forecasts, while the monthly QPP model aggregates raw forecasts to the monthly scale, collapses the ensemble to its median value, and applies a monthly residual error model to accounts for both hydrological and forecast rainfall uncertainty. The diagram for the calibration approach in panel (b) shows key differences in how the models are calibrated (highlighted in colored text) are (1) the MuTHRE model is forced by *observed* rainfall during calibration, while the monthly QPP model is forced with *forecast* rainfall, (2) the MuTHRE model is calibrated using observed *daily* streamflow data, while the monthly QPP model uses observed *monthly* streamflow. The diagram for forecasting in panel (c) shows that the MuTHRE model produces seamless daily streamflow forecasts, which can be used at a *range of lead times and aggregation periods (e.g. daily, weekly, fortnightly, monthly)*, while the monthly QPP model produces only *monthly forecasts*.**~~

Formatted: Normal, Left

~~This difference in calibration provides another practical benefit of the seamless MuTHRE model for forecast providers, in that improvements in rainfall forecasting are easily integrated into the forecasting system. Since the non-seamless monthly QPP model is calibrated using forecast rainfall, it must be recalibrated whenever a new rainfall forecast is to be used. In contrast, the seamless MuTHRE model uses only observed rainfall in calibration and does not require recalibration with different forecast rainfall, allowing for easier use of improved rainfall forecast products in operational settings.~~

~~2.4.3. Differences in forecasting~~

~~The differences in the model structure and calibration approach for the seamless MuTHRE model and non-seamless monthly QPP model results in key differences in terms of the forecasts that each model can produce. Figure 1(c) illustrates these differences and shows that the seamless MuTHRE model produces daily streamflow forecasts that can be used at a *range of lead times and aggregation periods*, while the non-seamless monthly QPP model produces only *one-month ahead monthly forecasts.*~~

3. Case study

3.1. Catchments and Data

~~A~~The case study uses a set of 11 catchments from the Murray Darling Basin in Australia, ~~consisting of~~including four catchments on the Upper Murray River (NSW and Victoria) and seven catchments on the Goulburn River (Victoria~~), is used in the case study.~~). These catchments have winter dominated rainfall which leads to higher streamflow between June and October (see Figure 3~~Figure 2~~), and have fewer than 5% of days with no flow. Catchment properties are summarised in Table 1. This same set of catchments was used to extensively evaluate the MuTHRE model in McInerney et al. (2020a); McInerney et al. (2020b).

Time series of daily observed streamflow over a 22-year period between 1991 and 2012 are obtained from the Hydrologic Reference Stations (HRS) dataset (http://www.bom.gov.au/water/hrs). Observed rainfall and PET data over the same period are obtained from the Australia Bureau of Meteorology's climate data service (www.bom.gov.au/climate), with a climatological average used for PET (McInerney et al., 2021).

Rainfall forecasts are provided by the Australian Community Climate Earth-System Simulator - Seasonal (ACCESS-S) (Hudson et al., 2017). ~~The post-processing~~The ACCESS-S rainfall forecasts are pre-processed using the method of Schepen et al. (2018) ~~is used~~in order to reduce biases and improve the reliability ~~of the ACCESS-S~~in comparison to observed rainfall ~~forecasts. A set~~. An ensemble of 100 ~~replicates of post~~pre-processed rainfall forecasts that begin on the first day of each month and extend out to a maximum lead time of 1 month are used.

3.2. Hydrological model

The conceptual rainfall-runoff model GR4J (Perrin et al., 2003) is used as the deterministic hydrological model h ~~(introduced in Section 2.1)~~ for simulating daily streamflow from rainfall and PET inputs~~.~~ (see Section 2.1). GR4J has been widely used and evaluated over diverse catchment climatologies and physical characteristics (Perrin et al., 2003; Hunter et al., 2021). GR4J represents the processes of interception, infiltration and percolation, and has four calibration parameters: ~~x_1~~ x_1 is the capacity

Field Code Changed

Formatted: Normal, Left

of the production store (mm), ~~x_2~~ x_2 is the water exchange coefficient (mm), ~~x_3~~ x_3 is the capacity of the routing store (mm), and ~~x_4~~ x_4 is the time parameter of the unit hydrograph (days).

360

Field Code Changed

Field Code Changed

Field Code Changed

Formatted: Normal, Left

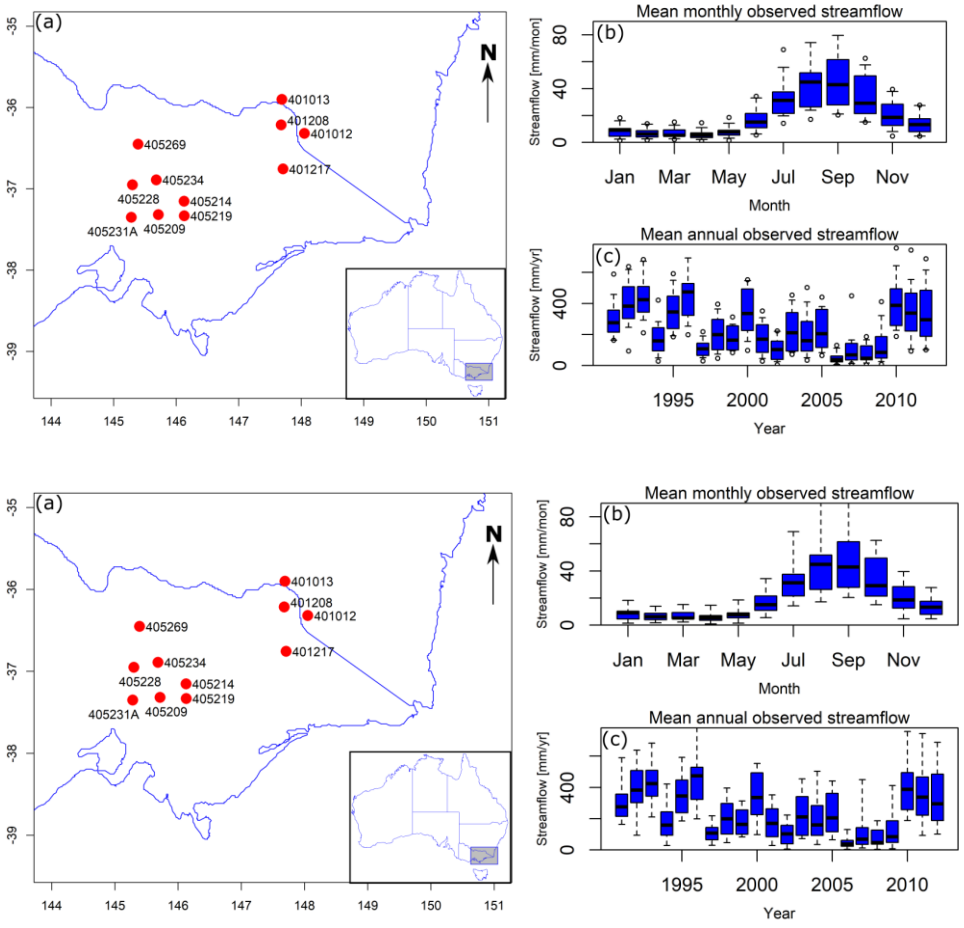


Figure 32: Locations: Location of the 11 case study catchments (panel a), and mean observed streamflow for the 11 case study catchmentseach month (panel b) and each year (panel c). Boxplots in panels (b) and (c) show distributions of mean observed streamflow over the 11 catchments.

365

Formatted: Normal, Left

Table 1: Properties of the 11 case study catchments.

Catchment	Site ID	Area (km^2)	Mean rainfall (mm/yr)	Mean runoff (mm/yr)	Runoff ratio	Zero flow days (%)	Aridity index
Murray River at Biggara	401012	1257	1117	370	0.33	0	0.99
Jingellic Creek at Jingellic	401013	390	876	112	0.13	1.1	0.68
Cudgewa Creek at Berringama	401208	351	1127	209	0.19	0	0.90
Gibbo River at Gibbo Park	401217	390	1138	273	0.24	0	1.01
Acheron River at Taggerty	405209	629	1234	443	0.36	0	1.2
Delatite River at TongaBridge	405214	368	959	248	0.26	0	0.85
Goulburn River at Dohertys	405219	700	1156	424	0.37	0	1.0
Hughes Creek at Tarcombe Rd	405228	475	760	116	0.15	1.3	0.65
King Parrot Creek at Flowerdale	405231A	181	999	187	0.19	0	0.95
Seven at D/S Polly McQuinns Weir	405234	148	852	226	0.27	0	0.71
Seven Creeks River at Kialla West	405269	1513	655	93	0.14	3.0	0.53

Formatted: Normal, Left

3.3. Calibration/evaluation procedure

Calibration of model parameters and evaluation of forecasts is performed using a leave-one-year-out cross validation procedure (McInerney et al., 2020a). For each(McInerney et al., 2020b). For each calendar year j, hydrological and residual error model parameters are calibrated using observed streamflow data from the entire evaluation period, except for year j and the subsequent years $j+1$ to ~~$j+4$ (which are excluded to reduce the influence of system memory on model evaluation).~~ $j+4$ (which are excluded to reduce the influence of system memory on model evaluation, as described in Pokhrel et al., 2013).

Hydrological model parameters are estimated using likelihood maximisation based on the BC0.2 error model (McInerney et al., 2020a; McInerney et al., 2020b), implemented using a quasi-Newton optimization algorithm run with 100 independent multistarts (Kavetski and Clark, 2010). ~~Methods for estimating residual error model parameters are described in~~ McInerney et al. (2020a) ~~and Woldemeskel et al. (2018).~~Methods for estimating residual error model parameters are described in Sections 2.3.4 and 2.4.4.

~~Calibrated~~Note that in this work we do not consider parametric uncertainty (in the hydrological and residual error models), which is expected to be a (relatively) minor contributor to total forecast uncertainty given the long data period used in the estimation; this simplification is common in contemporary forecasting implementations (e.g., Engeland and Steinsland, 2014; Verkade et al., 2017).

For each year j, calibrated hydrological and error models are ~~then~~ used to generate ~~probabilistic~~an ensemble of 100 streamflow forecasts ~~for year j~~. Daily forecasts from the MuTHRE model begin on the first day of each month, and extend out to a maximum lead time of 1 month (which is the same as the rainfall forecasts).

This calibration/forecasting process is repeated for all 22 years, resulting in 22 sets of one-year forecasts, which are subsequently merged into a single 22-year forecast to facilitate evaluation against streamflow observations ~~(as described in Section 3.4).~~

Field Code Changed

Field Code Changed

Formatted: Font: Bold

Formatted: Normal, Left

3.4. Forecast evaluation

3.4.1. Performance metrics

Streamflow forecasts are evaluated using numerical metrics for the following attributes:

Reliability~~, which~~ refers to the degree of statistical consistency between the forecast distribution and ~~observations,~~the observed data. It is evaluated using the reliability metric of Evin et al. (2014) ~~(which is based on the predictive quantile-quantile plot).~~ Lower metric values are better, with 0 ~~indicating~~representing perfect reliability, and 1 ~~being~~ representing the worst reliability.

Sharpness refers to the spread of the forecast distribution, with sharper forecasts those with lower ~~uncertainty~~spread. We use the sharpness metric of McInerney et al. (2020a); McInerney et al. (2020b), which is based on the ratio of the average 90% inter-quantile range (IQR) of the forecasts and a climatological distribution (described below). Lower values are better, with 0 representing a deterministic forecast (with no ~~uncertainty~~spread) and 1 representing the same sharpness as climatology. In contrast to the other attributes considered here, sharpness is a property of the forecast only and does not depend on the observed data.

Volumetric bias refers to the long-term water balance error. It is quantified using the metric of McInerney et al. (2017) as the relative absolute difference between total observed streamflow and the total forecast streamflow (averaged over the forecast ~~replicates~~ensemble). Lower values are better, with 0 representing unbiased forecasts.

Combined performance is quantified using the continuous ranked probability score (CRPS). The CRPS is defined as the sum of squared differences between forecast cumulative distribution function (CDF) and the empirical CDF of the observation. Note that the CRPS can be decomposed into terms representing individual performance aspects, namely reliability, and uncertainty/resolution (related to sharpness) (Hersbach, 2000). We express this metric as a skill score (CRPSS) relative to the climatological distribution. Higher CRPSS values are better, with a value of 1 ~~indicating~~representing a perfectly accurate deterministic forecast, and 0 ~~indicating~~representing the same skill as the climatological distribution.

The *climatological distribution* represents the distribution of daily streamflow for a given time of the year based solely on previously observed streamflow at that time of the year. The climatological distribution is constructed using a 29 day moving-window approach, described in detail in McInerney et al. (2020a); McInerney et al. (2020b).

3.4.2. Aggregation and stratification

The ~~main aim of this paper is to compare~~study focuses on the performance of the ~~seamless MuTHRE model and the non-seamless monthly QPP model~~streamflow post-processing models at the monthly scale. The monthly MuTHRE forecasts are obtained by aggregating daily forecasts to the monthly scale. The monthly QPP model generates monthly forecasts directly.

Overall evaluation of monthly forecasts is performed using data from the entire evaluation period, i.e. all months and years, with more detailed *stratified* performance evaluation performed for individual months and years.

We also demonstrate the ability of the MuTHRE model to produce seamless forecasts, which are reliable over a range of lead times and aggregation scales. This is achieved by evaluating both (i) daily forecasts stratified by lead times from 1-28 days, and (ii) cumulative flow forecasts for periods 1-28 days. The forecast is considered 'seamless' if reliability metrics are similar across all lead times and aggregation scales. The evaluation of cumulative flow forecasts expands on the analysis of McInerney

Formatted: Normal, Left

et al. (2020b)~~, which~~McInerney et al. (2020b), who evaluated only daily and monthly forecasts, and provides and important demonstration of seamless forecasting over the entire range of time scales ~~between~~from 1 ~~and~~to 28 days. ~~–~~We note that cumulative flow forecasts over 1 month correspond to monthly forecasts.

3.4.3. Evaluation of ~~performance~~practical significance of differences between ~~QPP~~streamflow post-processing models

Forecast performance ~~for~~of the two streamflow post-processing models is compared across multiple catchments using *practical significance tests*, as described next. For each combination of performance metric (e.g., reliability~~, CRPSS~~) and stratification (e.g., month~~, year~~), a statistical test is used to determine whether differences in metric values over the range of catchments ~~are of~~exceed a pre-defined margin representing practical significance (relevance~~. Statistical~~).

The statistical tests are performed using the paired Wilcoxon signed rank test (Bauer, 1972), with controls applied to reduce the false discovery rate to 5%, corresponding to a confidence level of 95% (Benjamini and Hochberg, 1995; Wilks, 2006). ~~Practically relevant differences are~~The practical significance margin is taken as 20% of the median metric value for the non-seamless monthly QPP model (following McInerney et al., 2020b)~~.~~ (following McInerney et al., 2020b).

Formatted: Normal, Left

4. Results

4.1. Demonstration of seamless forecasting capabilities of the MuTHRE model

Daily forecasts

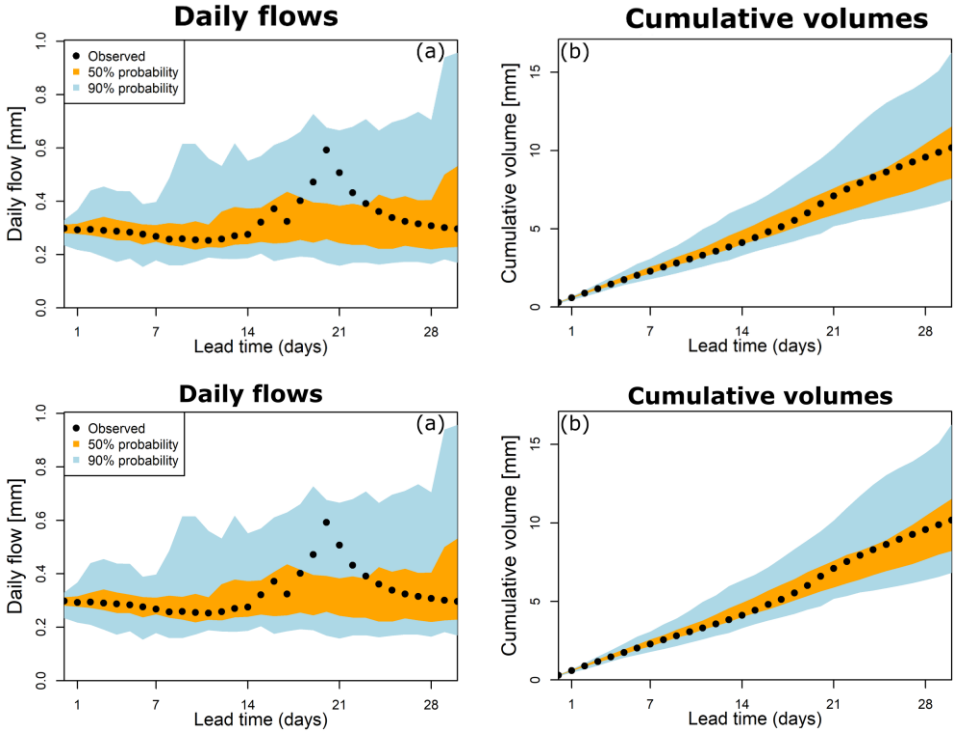


Figure ~~4~~3: Time series of daily and cumulative probabilistic forecasts from the seamless MuTHRE model for Murray River at Biggara (401012, see Figure 3~~Figure 2~~) for May 2002. The non-seamless monthly QPP model does not have the capability to produce these forecasts.

Figure 4~~Figure 3 provides illustrations of probabilistic~~ illustrates the streamflow forecast time series in the Biggara catchment (Catchment ID 401012, see Figure 3~~Figure 2~~). Daily forecasts from the seamless MuTHRE model~~,~~ for a representative time period beginning on 1st May 2002~~,~~ are shown in Figure 4~~Figure 3~~a. The observed daily streamflow lies within the 90% probability limits of the MuTHRE forecasts for each lead time. As expected, the probability limits are tight for short lead times (when forecast rainfall uncertainty and hydrological uncertainty are small), and widen for longer lead times.

Formatted: Normal, Left

Figure 5~~Figure 4~~ (left column) shows the performance of ~~the~~ daily forecasts from the MuTHRE model for lead times of 1-28 days, evaluated over ~~the full range of~~all case study catchments. The key finding from this analysis is that reliability is relatively constant over all lead times, with median metric values lying in the tight range of 0.04-0.06 (Figure 5~~Figure 4~~a). We also note that forecasts are sharper and have better CRPSS at short lead times, and that bias is relatively constant.

Formatted: Normal, Left

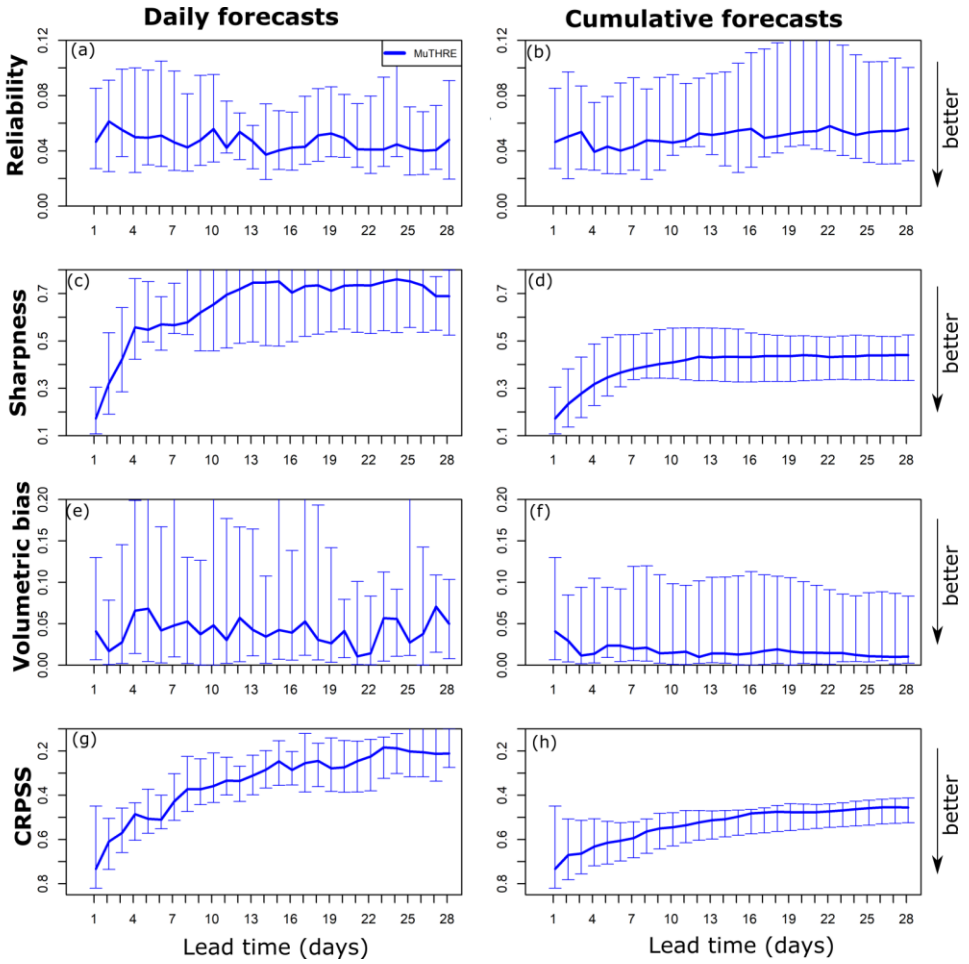


Figure 54: Performance of MuTHRE forecasts in terms of daily streamflow (left) and cumulative flow (right). Metrics shown ~~are~~ for reliability (top row), sharpness (2nd row), volumetric bias (3rd row) and CRPSS (bottom row). ~~Thick lines show median~~ The bars indicate the full range of metric values across the 11 case study catchments and ~~vertical bars represent 80% probability limits.~~ the line indicates the median metric values. Note the inverted *y*-axis for CRPSS~~.~~, for visual consistency with the other metrics.

Formatted: Normal, Left

Cumulative flow forecasts

Figure 4~~Figure 3~~b shows cumulative flow forecasts out to 28 days in the Biggara catchment for the representative time period. The cumulative flows based on observed streamflow lie well within the 90% probability limits of the MuTHRE forecasts for all lead times.

Figure 5~~Figure 4~~ (right column) shows the performance of ~~the~~ cumulative flow forecasts from the MuTHRE model for lead times of 1-28 days over all catchments. Again, we see that reliability is relatively constant over all lead times, with median metric values between 0.04 and 0.06 (Figure 5~~Figure 4~~b). We also note that sharpness, volumetric bias and CRPSS metrics are typically better for cumulative forecasts than for daily forecasts (compare left and right columns in Figure 5~~Figure 4~~).

In summary, the forecasts from the MuTHRE model are seamless, ~~since~~because they are reliable over (a) the range of lead times, and (b) multiple aggregation scales, from the shortest scale of 1 day, to the longest scale of 1 month, and everything in between. This result confirms and extends previous findings in McInerney et al. (2020a).~~.~~McInerney et al. (2020b) who focused on daily and monthly scales only. In contrast to the seamless MuTHRE model, the non-seamless monthly QPP model does not have the capability to produce forecasts of daily streamflow and cumulative flows for time periods ~~less than~~below one month.

Formatted: Normal, Left

4.2. Comparison ~~between~~of monthly forecasts

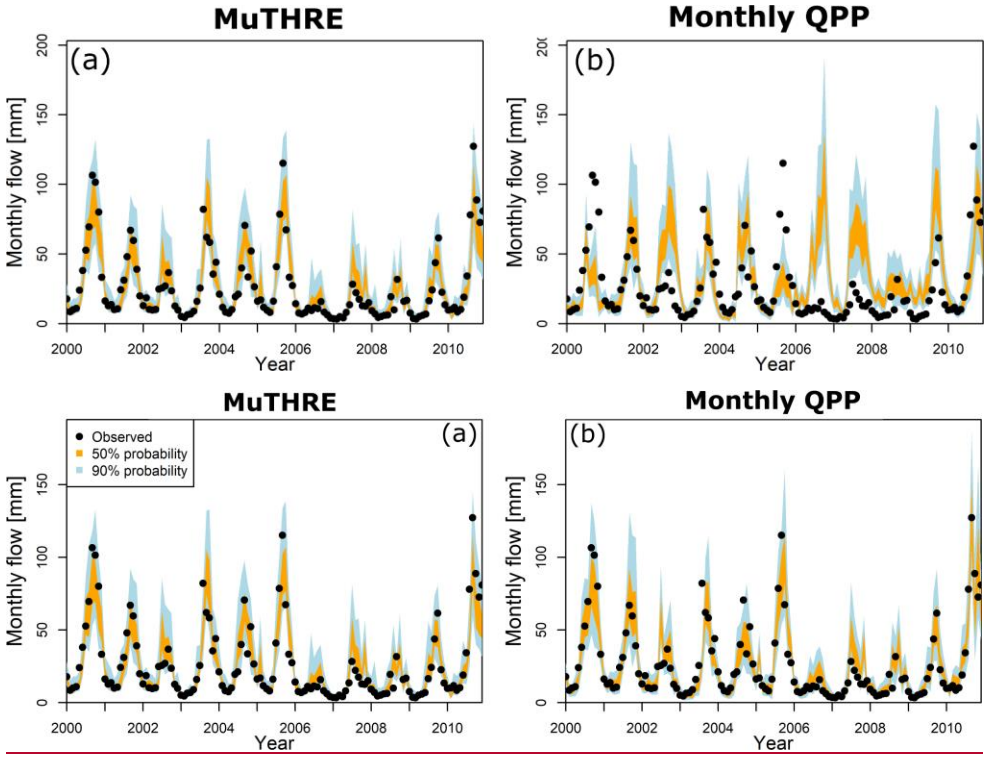


Figure ~~6~~5: Time series of monthly probabilistic forecasts for Murray River at Biggara (401012, see Figure 3~~Figure 2~~) for the seamless MuTHRE model and non-seamless monthly QPP model. Results are shown between the years 2000 and 2011.

Figure 6~~Figure 5~~ compares monthly forecasts from the seamless MuTHRE model and non-seamless monthly QPP model for the Biggara catchment. ~~For high flow periods the 90% prediction limits from the seamless MuTHRE model appear to be tighter (i.e. have less uncertainty) than those from the non-seamless monthly QPP model, while still capturing the observed streamflow~~While there are some minor differences between the two forecasts (e.g. the monthly QPP model produces larger spread than the MuTHRE model during 2010), the two forecasts are clearly very similar.

Formatted: Normal, Left

~~The~~Figure 7 compares monthly forecasts from the MuTHRE and monthly QPP models ~~are compared in Figure 6~~ in terms of overall performance (left column), and when stratified by month (middle column), and year (right column). The key findings are as follows.

Reliability. Figure 7~~Figure 6~~a shows ~~that the~~similar overall reliability of monthly forecasts from the MuTHRE and monthly QPP models ~~is similar~~. While the median metric value ~~is~~of 0.06 for the ~~seamless~~ MuTHRE model is ~~larger~~worse than the median value of 0.04 for the ~~non-seamless~~ monthly QPP model, these differences are not practically significant~~.~~ (based on the test described in Section 3.4.3). Figure 7~~Figure 6~~b shows that when performance is stratified by month, the two models have similar reliability (i.e. not practically significant) for all 12 months. When stratified by year, the MuTHRE model ~~offers~~achieves similar reliability to the monthly QPP model for 20 out of the 22 years, ~~with~~while the ~~non-seamless~~ monthly QPP model ~~offering~~achieves practically significant improvements in 2 of the 22 years (Figure 7~~Figure 6~~c).

Sharpness. Figure 7~~Figure 6~~d shows that the overall sharpness of monthly forecasts from the ~~seamless~~ MuTHRE model is slightly better than the ~~non-seamless~~ monthly QPP model (median metric values of 0.44 c.f. 0.49), although differences are not practically significant. Figure 7~~Figure 6~~e 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 7~~Figure 6~~f shows sharpness stratified by year is similar for both models for all years.

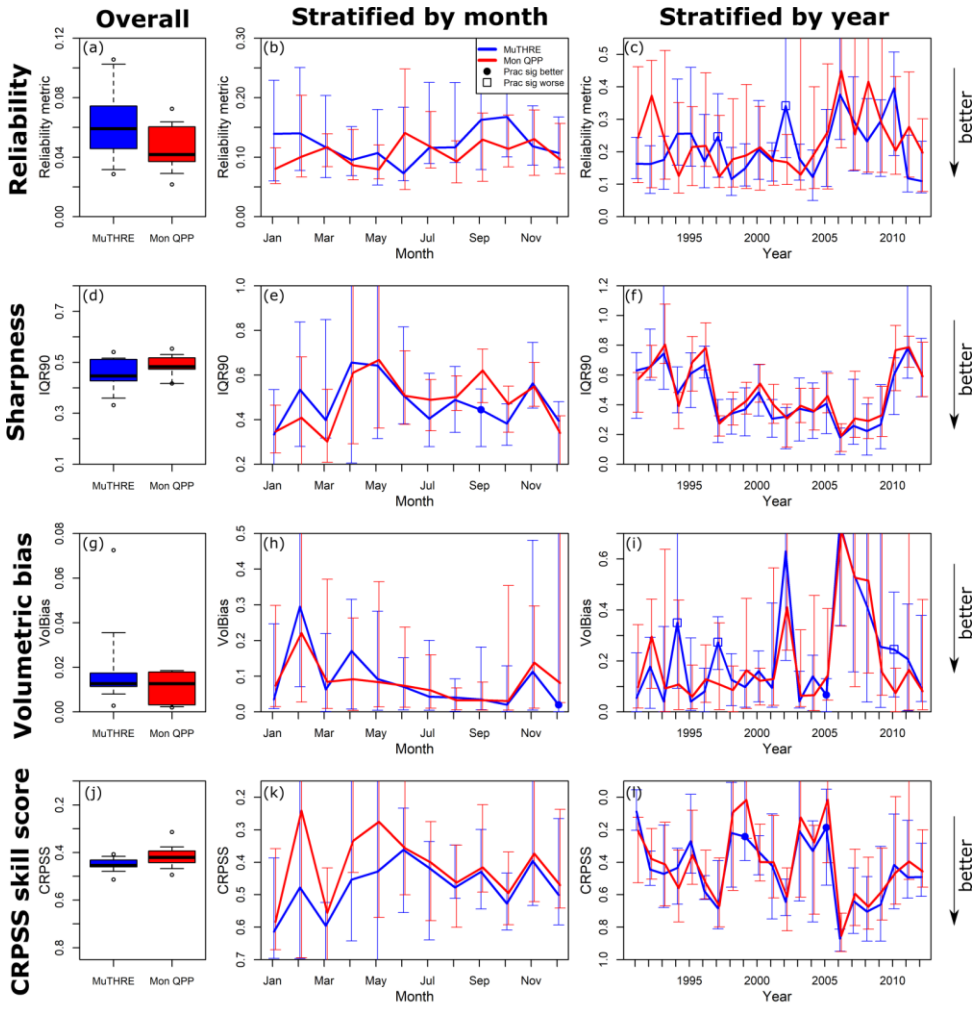
Volumetric bias. Figure 7~~Figure 6~~g shows that the overall volumetric bias from both models is similar (median of 0.01). Figure 7~~Figure 6~~h shows that when stratified by month, the MuTHRE model produces ~~similar/better performance in all months, with~~ practically significant improvements in December and similar performance in the remaining 11 months. Figure 7~~Figure 6~~i shows that when stratified by year, the MuTHRE model produces ~~similar/better performance in 19 out of 22 years, with practical~~practically significant improvements in ~~one~~1 year~~, while~~ (2005), the monthly QPP model provides practically significant improvements in 3 years~~.~~, with similar performance in the remaining 18 years.

CRPSS. In terms of overall CRPSS, Figure 7~~Figure 6~~j shows that the ~~seamless~~ MuTHRE model (median metric value of 0.45) provides slight improvement over the ~~non-seamless~~ monthly QPP model (median metric value of 0.42), although these differences are not practically significant. Figure 7~~Figure 6~~k shows that when stratified by month, the ~~seamless~~ MuTHRE model ~~actually~~ provides similar performance in all 12 months. Figure 7~~Figure 6~~i shows that when performance is stratified by year, the ~~seamless~~ MuTHRE model ~~actually~~ provides practically significant *improvements* in CRPSS in 2 out of 22, and similar performance in the remaining 20 years.

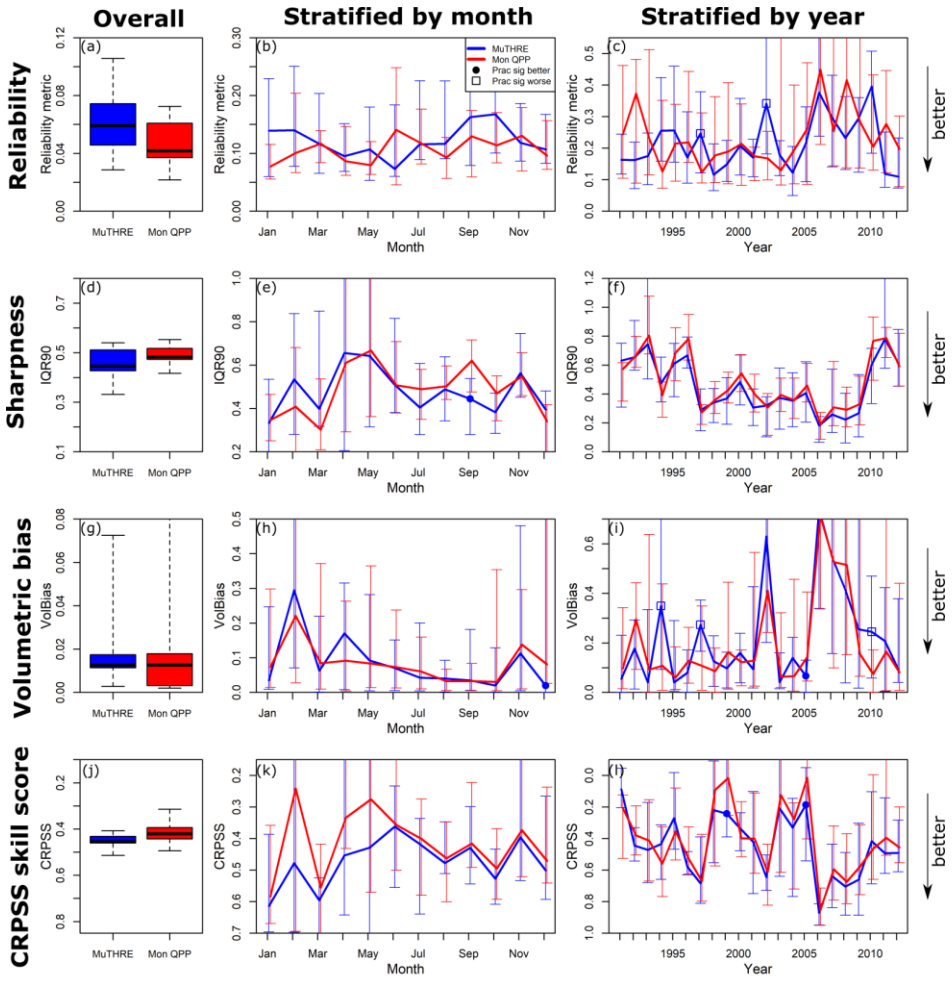
In summary, aggregated forecasts from the seamless MuTHRE model offer similar (not practically significant), and in some cases superior performance, to forecasts from the non-seamless monthly QPP model, for the vast majority of performance metrics and stratifications considered in this study.

Formatted: Font: Italic

Formatted: Normal, Left



Formatted: Normal, Left



520 **Figure 76: ~~Monthly~~Overall performance ~~of the seamless MuTHRE and non-seamless monthly QPP forecasts for (~~all ~~data (~~months and years, left~~), column), performance** stratified by month (middle column) and performance stratified by year (right~~), column), of monthly forecasts from the seamless MuTHRE and non-seamless monthly QPP models.~~ Results are shown for reliability (top row), sharpness (2nd row), volumetric bias (3rd row) and CRPSS (bottom row). ~~Circles~~Boxplots in the left column show the distribution of metric values over the 11 catchments. In the other columns, vertical bars indicate the full range of metric values across the

Formatted: Normal, Left

catchments, the line indicates the median metric values, and circles/squares indicate that the MuTHRE model performs practically significantly better/worse than the monthly QPP model.

5. Discussion

5.1. Interpretation of key findings

The empirical results show that the seamless MuTHRE model achieves essentially the same performance as the non-seamless monthly QPP model at the monthly time scale, and even provides improvement in some aspects. At first glance, this outcome may seem surprising for the following reasons:

- The seamless MuTHRE model is required to produce reliable forecasts over a range of lead times and aggregations scales, whereas the non-seamless monthly QPP model is only required to produce reliable monthly streamflow forecasts.;
- The seamless MuTHRE model is calibrated at the daily scale, using only observed daily streamflow during calibration, while the non-seamless monthly QPP model is calibrated to match the observed monthly streamflow.;
- The seamless MuTHRE model does not ~~see~~'see' the forecast rainfall during calibration, whereas the non-seamless monthly QPP model does.

The subsections below describe how the seamless MuTHRE model is able to achieve comparable/better performance than the non-seamless monthly QPP model despite these apparent challenges.

5.1.1. Time scale of forecasting/calibration

The seamless MuTHRE model produces daily forecasts that can be aggregated from time scales of one day to one month, whereas the non-seamless monthly QPP model produces forecasts only at the monthly scale. One might expect the enhanced capability obtained from the seamless MuTHRE model to come at some cost in performance at the monthly scale. Encouragingly, this is not the case.

The ability to reliably aggregate daily forecasts to the monthly scale demonstrates that the seamless MuTHRE model is adequately capturing temporal persistence in daily forecasts. The MuTHRE model represents temporal persistence in *hydrological errors* using the daily AR(1) model, and the (30-day) dynamic bias component. This is important because neglecting temporal persistence in hydrological errors can result in an underestimation of hydrological uncertainty for aggregated predictions/forecasts (Evin et al., 2014). The reliability of aggregated forecasts also suggests that the (~~post~~pre-processed) *rainfall forecasts* are capturing the day-to-day temporal persistence of observed rainfall required to produce reliable monthly rainfall forecasts (see Section 5.1.2).

The seamless MuTHRE model is not calibrated to optimize performance at the monthly scale, ~~since~~as it uses only observed daily streamflow during calibration. On the other hand, the non-seamless monthly QPP model is calibrated to match the observed monthly streamflow, which could lead to improved performance at the monthly scale compared ~~with~~to the seamless MuTHRE model. ~~The~~As such, the comparable performance of the MuTHRE model at the monthly scale is ~~particular impressive since~~particularly encouraging given that monthly data is not used in its calibration.

Formatted: Font: Not Italic

Formatted: Normal, Left

5.1.2. Use of observed vs forecast rainfall used in calibration

Both approaches use the same deterministic hydrological model calibrated using observed rainfall and streamflow data. However, due to structural differences in their representation of residual errors, the seamless MuTHRE and non-seamless monthly QPP models differ in the approach used to calibrate the residual error model parameters. The residual error model in the non-seamless monthly QPP model represents combined rainfall and hydrological uncertainty. It uses forecast rainfall during calibration, and can (in theory) correct for biases and under/over-dispersion in rainfall forecasts. In contrast, the seamless MuTHRE model represents only hydrological uncertainty, and is calibrated using observed rainfall. Uncertainty due to forecast rainfall is represented by propagating rainfall forecasts through the hydrological model. Since the MuTHRE model does not correct for forecast rainfall errors, this approach ~~relies on~~requires rainfall forecasts being reliable in order to produce reliable streamflow forecasts (Verkade et al., 2017).

In this study we have used rainfall forecasts from the ACCESS-S ~~climate simulator~~numerical weather prediction model (Hudson et al., 2017), which were ~~post~~pre-processed at the catchment scale with the aim of reducing biases and over/under-dispersion at the daily scale, and capturing the temporal persistence in rainfall (Schepen et al., 2018). As a result, these ~~post~~pre-processed rainfall forecasts are reliable and sharp at both the daily and monthly scale, and do not have a detrimental impact on the performance of the seamless MuTHRE model.

Since the seamless MuTHRE model uses only observed rainfall in calibration, it does not require re-calibration if an improved rainfall forecast product becomes available, which is a useful advantage in operational settings. In contrast, the monthly QPP model is calibrated using forecast rainfall, and must be re-calibrated whenever a new rainfall forecast is to be used. Note that this is a benefit of the ensemble dressing approach used in the MuTHRE model, rather than of the forecasts being seamless.

5.1.3. Use of daily vs monthly streamflow observations to update forecasts

The MuTHRE and monthly QPP models differ in their use of recently observed streamflow data. The MuTHRE model uses daily streamflow observations to update both (i) the dynamic bias component of the error model, to account for monthly errors, and (ii) the daily AR1 model to account for recent daily errors and improve sharpness of forecasts for short lead times. In contrast, the monthly model only uses monthly aggregated streamflow observations to update the monthly AR1 model. The ability to utilize the most recent time series of daily streamflow observations provides the MuTHRE model with a potential advantage over the monthly QPP model (which see only monthly totals) and may be another reason why the MuTHRE model performs so well compared with the monthly QPP model.

~~5.1.4~~5.2. Summary of practical benefits of seamless MuTHRE forecasts

The key practical benefits ~~of these differences~~the seamless MuTHRE model provides over the non-seamless monthly QPP model are summarized below:

1. Seamless forecasts can be used to inform decisions at a range of time scales~~.~~;
2. Seamless daily forecasts are easily integrated into river system models used for real-time decision-making;
3. ~~Simplifies~~Simplified forecasting system as a single seamless product can serve range of forecast requirements at different time scales~~.~~;

Formatted: Heading 2

Formatted: Normal, Left

4. Improvements in rainfall forecasting are easily integrated into the forecasting system~~.~~ (as described in Section 5.1.2).

The competitive performance of the seamless MuTHRE model even at the native scale of the non-seamless monthly QPP forecasts is clearly encouraging – it indicates that seamless forecasts do not require a compromise between capabilities (range of available forecast time scales) and~~, in combination with the inherent benefits listed above,~~ quality of performance. In other words, a user of seamless forecasts can have their cake and eat it too! This finding provides further motivation to adopt seamless forecasts in research and practical work.

~~5.2.~~5.3. Future work

Future work ~~in this area will focus~~is recommended on the following aspects:

- Further testing and development of the MuTHRE model on a wide range of catchments. The monthly QPP model has been comprehensively evaluated on 300 catchments around Australia (Woldemeskel et al., 2018), whereas the MuTHRE model has currently been evaluated on 11 catchments in the Murray Darling Basin. Evaluation of the MuTHRE model over a wide range of hydro-climatic conditions is required to ensure the findings of this study are robust. Potential ~~modifications to~~enhancements of the MuTHRE model, including ~~the~~specialised treatment of zero flows in ephemeral catchments (McInerney et al., 2019; Wang et al., 2020), may be required to ensure the MuTHRE model remains competitive with the monthly QPP model over a wider range of ~~catchment types.~~ flow regimes.
- Deeper understanding of the ~~specific~~ reasons for the MuTHRE model matching the monthly QPP model at the monthly scale. ~~Further work could look at~~ For example, systematic testing of different combinations of MuTHRE~~/~~ and monthly QPP model components ~~to systematically~~could help diagnose the specific reasons why the MuTHRE model performs so well. ~~This was not feasible in this study.~~
- ~~Evaluating~~Evaluation of how the ~~use~~quality of ~~different~~ rainfall forecasts~~, with differing quality,~~ impacts on the performance of the seamless MuTHRE model and its ability to match/improve on the performance of the non-seamless monthly QPP model at the monthly scale.

6. Conclusions

Subseasonal streamflow forecasts at time scales ranging from daily to monthly are of major interest in water management. This study ~~has explored~~compares two streamflow post-processing (QPP) models, namely the ~~question of whether aggregated seamless forecasts from the~~'seamless' daily Multi-Temporal Hydrological Residual Error (MuTHRE) model ~~can produce similar performance at the monthly time scale to forecasts from the~~ and the more traditional 'non-~~seamless~~seamless' monthly ~~streamflow post-processing (~~QPP~~)~~ model used in the Australian Bureau of Meteorology's Dynamic Forecasting System. The MuTHRE model is designed at the daily scale and can be aggregated up to the monthly scale, whereas the monthly QPP model is designed directly at the monthly scale and does not produce forecasts at the daily scale. A case study with 11 catchments in south-east Australia, the GR4J conceptual rainfall-runoff model, and ~~post~~pre-processed ACCESS-S rainfall forecasts, is reported.

The key finding is that the seamless MuTHRE model ~~is able to achieve~~achieves essentially the same monthly-scale performance as the non-seamless monthly QPP model for the majority of metrics (reliability, sharpness, bias and CRPSS) and

Formatted: Normal, Left

stratifications (monthly and yearly). Remarkably, the seamless post-processing model achieves high quality forecasts (based on the metrics considered in this study) at its native daily scale *and* matches the performance of the non-seamless monthly model at the monthly scale ~~(~~, despite not being calibrated at that time scale~~).~~.

Seamless subseasonal forecasts, which are reliable over a wide range of lead times (1-30 days) and time scales (daily-monthly), offer numerous practical benefits over non-seamless forecasts. For users, seamless subseasonal forecasts can inform a wide range of management decisions from flood warning to water supply operation, while for service providers, seamless forecasts will reduce the number of forecast products that require development and operation. As such it represents a single modelling tool with great versatility. The ~~clearly~~ encouraging results from this study help ~~pave the way for seamless forecasts to replace non~~ motivate broader adoption of seamless forecasts, as they offer additional capability without loss in performance ~~at the time scale of non-seamless forecasts~~.

Acknowledgments

The ~~recent~~research presented in this paper was funded by the Australian Bureau of Meteorology. Support with supercomputing resources was provided by the Phoenix HPC service at the University of Adelaide. Data used in the case studies is available as follows: observed rainfall data from ~~www.bom.gov.au/climate, observed streamflow data from www.bom.gov.au/waterdata, and post~~www.bom.gov.au/climate, observed streamflow data from www.bom.gov.au/waterdata, and pre-processed ACCESS-S forecast rainfall from https://doi.org/10.25909/14604180. We ~~gratefully acknowledge insightful feedback from~~thank Surendra Rauniyar and Christopher Pickett-Heaps for their insightful feedback during the Bureau review of this manuscript, and two anonymous HESS reviewers for their constructive comments, all of which helped improve the quality of this manuscript.

Formatted: Font: Italic

Formatted: Indent: Left: 0 cm, Hanging: 0.63 cm

Formatted: Normal, Left

References

Bauer, D. F. 1972. Constructing Confidence Sets Using Rank Statistics. *Journal of the American Statistical Association*, 67**,** 687-690.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, 57**,** 289-300.

Boucher, M.-A. & Ramos, M.-H. 2019. Ensemble Streamflow Forecasts for Hydropower Systems. *In:* DUAN, Q., PAPPENBERGER, F., WOOD, A., CLOKE, H. L. & SCHAAKE, J. C. (eds.) *Handbook of Hydrometeorological Ensemble Forecasting*. Berlin, Heidelberg: Springer Berlin Heidelberg.

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

Cloke, H. L. & Pappenberger, F. 2009. Ensemble flood forecasting: A review. *Journal of Hydrology*, 375**,** 613-626.

Engeland, K. & Steinsland, I. 2014. Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times. *Water Resources Research*, 50**,** 182-197.

Evin, G., Thyer, M., Kavetski, D., McInerney, D. & Kuczera, G. 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, 50**,** 2350-2375.

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

Hersbach, H. 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, 15**,** 559-570.

Hidalgo-Muñoz, J. M., Gámiz-Fortis, S. R., Castro-Díez, Y., Argüeso, D. & Esteban-Parra, M. J. 2015. Long-range seasonal streamflow forecasting over the Iberian Peninsula using large-scale atmospheric and oceanic information. *Water Resources Research*, 51**,** 3543-3567.

Hudson, D., Alves, O., Hendon, H. H., Lim, E., Liu, G., Luo, J. J., MacLachlan, C., Marshall, A. G., Shi, L., Wang, G., Wedd, R., Young, G., Zhao, M. & Zhou, X. 2017. ACCESS-S1 The new Bureau of Meteorology multi-week to seasonal prediction system. *Journal of Southern Hemisphere Earth System Sciences*, 67**,** 132-159.

Hunter, J., Thyer, M., McInerney, D. & Kavetski, D. 2021. Achieving high-quality probabilistic predictions from hydrological models calibrated with a wide range of objective functions. *Journal of Hydrology*, 603**,** 126578.

Kavetski, D. & Clark, M. P. 2010. Ancient numerical daemons of conceptual hydrological modeling. Part 2: Impact of time stepping scheme on model analysis and prediction. *Water Resources Research*, 46**,** W10511, doi:10.1029/2009WR008896.

Li, M., Wang, Q. J., Bennett, J. C. & Robertson, D. E. 2016. Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting. *Hydrol. Earth Syst. Sci.*, 20**,** 3561-3579.

McInerney, D., Kavetski, D., Thyer, M., Lerat, J. & Kuczera, G. 2019. Benefits of explicit treatment of zero flows in probabilistic hydrological modelling of ephemeral catchments. *Water Resources Research*, 55**,** 11035– 11060.

McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Tuteja, N. & Kuczera, G. 2020a. Multi-temporal hydrological residual error modelling for seamless sub-seasonal streamflow forecasting. *Water Resources Research*, n/a**,** e2019WR026979.

McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Tuteja, N. & Kuczera, G. 2020b. Multi-temporal hydrological residual error modelling for seamless sub-seasonal streamflow forecasting. *Water Resources Research*, 56**,** e2019WR026979.

McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Tuteja, N. & Kuczera, G. 2020c. Multi-temporal hydrological residual error modelling for seamless sub-seasonal streamflow forecasting. *Water Resources Research*, n/a56**,** e2019WR026979.

McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Woldemeskel, F., Tuteja, N. & Kuczera, G. 2021. Improving the Reliability of Sub-Seasonal Forecasts of High and Low Flows by Using a Flow-Dependent Nonparametric Model. *Water Resources Research*, 57**,** e2020WR029317.

McInerney, D., Thyer, M., Kavetski, D., Lerat, J. & Kuczera, G. 2017. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, 53.

Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., Brekke, L. D. & Arnold, J. R. 2017. An intercomparison of approaches for improving operational seasonal streamflow forecasts. *Hydrol. Earth Syst. Sci.*, 21**,** 3915-3935.

Murray-Darling Basin Authority. 2019. *Joint management of the River Murray* [Online]. Available: https://www.mdba.gov.au/river-murray-system/joint-management-river-murray [Accessed 11/12/2019 2019].

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

Pal, I., Lall, U., Robertson, A. W., Cane, M. A. & Bansal, R. 2013. Predictability of Western Himalayan river flow: melt seasonal inflow into Bhakra Reservoir in northern India. *Hydrol. Earth Syst. Sci.*, 17**,** 2131-2146.

Perrin, C., Michel, C. & Andreassian, V. 2003. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279**,** 275-289, doi:10.1016/S0022-1694(03)00225-7.

Pokhrel, P., Wang, Q. J. & Robertson, D. E. 2013. The value of model averaging and dynamical climate model predictions for improving statistical seasonal streamflow forecasts over Australia. *Water Resources Research*, 49**,** 6671-6687.

Formatted: Normal, Left

Schepen, A., Zhao, T., Wang, Q. J. & Robertson, D. E. 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**,** 1615-1628.

Souza Filho, F. A. & Lall, U. 2003. Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. *Water Resources Research*, 39.

Verkade, J. S., Brown, J. D., Davids, F., Reggiani, P. & Weerts, A. H. 2017. Estimating predictive hydrological uncertainty by dressing deterministic and ensemble forecasts; a comparison, with application to Meuse and Rhine. *Journal of Hydrology*, 555**,** 257-277.

Wang, Q. J., Bennett, J. C., Robertson, D. E. & Li, M. 2020. A Data Censoring Approach for Predictive Error Modeling of Flow in Ephemeral Rivers. *Water Resources Research*, 56**,** e2019WR026128.

Wang, Q. J., Pagano, T. C., Zhou, S. L., Hapuarachchi, H. A. P., Zhang, L. & Robertson, D. E. 2011. Monthly versus daily water balance models in simulating monthly runoff. *Journal of Hydrology*, 404**,** 166-175.

Welsh, W. D., Vaze, J., Dutta, D., Rassam, D., Rahman, J. M., Jolly, I. D., Wallbrink, P., Podger, G. M., Bethune, M., Hardy, M. J., Teng, J. & Lerat, J. 2013. An integrated modelling framework for regulated river systems. *Environmental Modelling & Software*, 39**,** 81-102.

Wilks, D. S. 2006. On "field significance" and the false discovery rate. *Journal of Applied Meteorology and Climatology*, 45**,** 1181-1189.

Woldemeskel, F., McInerney, D., Lerat, J., Thyer, M., Kavetski, D., Shin, D., Tuteja, N. & Kuczera, G. 2018. Evaluating post-processing approaches for monthly and seasonal streamflow forecasts. *Hydrol. Earth Syst. Sci.*, 22**,** 6257-6278.

Yang, X., Liu, Q., He, Y., Luo, X. & Zhang, X. 2016. Comparison of daily and sub-daily SWAT models for daily streamflow simulation in the Upper Huai River Basin of China. *Stochastic Environmental Research and Risk Assessment*, 30**,** 959-972.

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

Formatted: Normal, Left