A New Method for Postprocessing Numerical Weather Predictions Using Quantile Mapping in the Frequency Domain

Ze Jiang[a] and Fiona Johnson[a]

[a] *School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia*

Corresponding author: Fiona Johnson, f.johnson@unsw.edu.au

File generated with AMS Word template 2.0

Early Online Release: This preliminary version has been accepted for publication in *Monthly Weather Review*, may be fully cited, and has been assigned DOI 10.1175/MWR-D-22-0217.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT

Improving lead time for forecasting floods is important to minimize property damage and ensure the safety of the public and emergency services during flood events. Numerical weather prediction (NWP) models are important components of flood forecasting systems and have been vital in extending forecasting lead-time under complex weather and terrain conditions. However, NWP forecasts still have significant uncertainty associated with the precipitation fields that are the main inputs of the hydrologic models and thus the resulting flood forecasts. An issue often overlooked is the importance of correctly representing variability over a range of different temporal scales. To address this gap, here a new wavelet-based method for post-processing NWP precipitation forecasts is proposed. First, precipitation forecasts are decomposed into the frequency domain using a wavelet transform, providing estimates of the amplitudes and phases of the time series at different frequencies. Quantile mapping is then used to correct bias in the amplitudes of each frequency. Randomized phases are used to generate an ensemble of realizations of the precipitation forecasts. The post-processed precipitation forecasts are reconstructed by taking the inverse of adjusted time-frequency decompositions with the corrected amplitudes and randomized phases. The proposed method was used to post-process NWP precipitation forecasts in the Sydney region, Australia. There is a significant improvement in post-processed precipitation forecasts across multiple time scales in terms of bias and temporal and spatial correlation structures. The post-processed precipitation fields can be used for the modeling of fully distributed hydrologic systems, improving runoff stimulation, flood depth estimation, and flood early warning.

SIGNIFICANCE STATEMENT

A new method accounting for the timing and spatial errors of NWP precipitation forecasts is proposed, and it can improve the skill of forecasts across multiple time scales, especially at short lead times. The proposed method provides a practical and effective way to correct these errors by incorporating spatio-temporal neighborhood information through the frequency domain using sophisticated wavelet transforms. With systematic timing and spatial errors removed, precipitation forecasts will be more skillful, and hydrological modeling using the post-processed forecasts can provide higher accuracy of streamflow estimation.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

1. Introduction

Accurate and timely flood forecasts are vital for minimizing the impacts of these potentially deadly and destructive events. It is the duty of operational flood forecasting centers to assist in the emergency preparation for these extreme flood events, both to mitigate the impacts on human life and the economy. Although major flood events cannot be avoided, early flood warnings can assist in providing adequate time for these emergency responses (He, Wetterhall, Cloke et al., 2009). Flood forecasts driven by observed gauge or radar rainfall data are sufficient for large catchments where the time of concentration is longer than the storm duration. However, a shift to the use of numerical weather prediction (NWP) to forecast precipitation has been vital in extending forecasting lead-time (de Roo, Gouweleeuw, Thielen et al., 2003). NWP models are important tools in the process of generating weather forecasts, including meteorological variables such as wind, pressure, air temperature, precipitation, etc. Although improved computing power and NWP model structure has led to considerable advances in the ability to forecast precipitation, the forecast skill of the NWP models to simulate precipitation is still low, especially for very short lead times and weather systems with small spatial scales (Shahrban, Walker, Wang et al., 2016; Shrestha, Robertson, Wang et al., 2013). The skill of weather forecasts is still affected by both systematic and random errors. These mainly originate from errors in specifying the initial conditions and boundary conditions, and as well as the structure of the NWP models (Vannitsem, Bremnes, Demaeyer et al., 2021). As a result, a range of methods has been developed to post-process precipitation forecasts from NWP models and the hydrological forecasts (e.g., streamflow) that are driven by NWP precipitation.

The challenges of precipitation forecast postprocessing include but are not limited to (Jankov et al., 2021; Li et al., 2017; Scheuerer & Hamill, 2015): (i) the distributions of both forecast and observed precipitation are mixed discrete/continuous distributions; (ii) extreme events are difficult to capture; (iii) the forecast error is heteroscedastic; and (iv) systematic timing and spatial errors. Methods for postprocessing precipitation forecasts have been developed to address these challenges, including empirical methods, such as quantile mapping and analog method, and other statistical methods, such as conditional distribution-based methods (e.g., Bayesian joint probability), regression-based methods, and ensemble dressing methods (Li, Duan, Miao et al., 2017). Quantile mapping (QM) is one of the most simple yet effective postprocessing methods that adjusts the cumulative distribution function

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

(CDF) of the forecasts according to the CDF of the observations (Cannon, Sobie, & Murdock, 2015), and it has been used extensively for post-processing precipitation forecasts (Gudmundsson, Bremnes, Haugen et al., 2012; Hamill, Engle, Myrick et al., 2017; Hamill & Scheuerer, 2018; Themeßl, Gobiet, & Heinrich, 2012) and can also be used for postprocessing probabilistic precipitation forecasts (Hamill & Scheuerer, 2018). There are also more complex methods, such as Ensemble Model Output Statistics (EMOS) producing a forecast probability distribution using ensemble spread information (Gneiting, Raftery, Westveld et al., 2005), Bayesian Model Averaging (BMA) providing probabilistic forecasts from an ensemble of forecasts (Raftery, Gneiting, Balabdaoui et al., 2005), and Bayesian Joint Probability (BJP) generating a probability model for the forecast (Robertson, Shrestha, & Wang, 2013). For a general review of the literature on statistical post-processing methods for hydro-meteorological forecasting, we recommend (Li, Duan, Miao et al., 2017) which provides a comprehensive review of progress in this area.

Of the four challenges in postprocessing precipitation forecasts outlined above, QM can, to a considerable extent, address the first and second challenges by adjusting the distribution of the precipitation. However, it fails to address the two remaining challenges. Because forecast error is dependent on the magnitude of precipitation amount (i.e., heteroscedasticity in errors), the forecast error varies over different time scales, for example when aggregating fine resolution forecasts to longer time scales. There have been a number of studies to address this issue. For example, Wang, Zhao, Yang et al. (2019) proposed the Seasonally Coherent Calibration method accounting for the seasonality when post-processing or calibrating the daily precipitation forecasts of NWP. Other related methods to post-process the raw simulations from climate models correct bias across multiple time scales using the nested approach for univariate (Johnson & Sharma, 2012) and multivariate (Mehrotra & Sharma, 2015), using frequency-based approaches such as Nguyen, Mehrotra, and Sharma (2016) and (Kusumastuti, Jiang, Mehrotra et al., 2021, 2022). It is noted that these methods generally consider the systematic errors in variables at a large time scale (e.g., monthly, annually, or interannually), and they are yet to be applied for NWP forecasts.

Another key issue that needs more attention during statistical postprocessing is the treatment of systematic timing and spatial errors in raw precipitation forecasts (e.g., loss of temporal and spatial dependency between neighboring locations). Spatial EMOS (Feldmann, Scheuerer, & Thorarinsdottir, 2015; Scheuerer & Büermann, 2014) and spatial BMA

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

(Veronica, Adrian, & Tilmann, 2008), both aim to preserve the observed spatial dependency for precipitation forecasts. Another popular method for preserving the spatio-temporal dependence between neighboring locations is known as the Schaake Shuffle (Clark & Hay, 2004). The method ranks the post-processed ensemble forecast values and randomly selects historical values at the locations of interest, and then reshuffles the ensemble output to match the rank of the historical forecast values at different lead times. The reshuffled ensembles were found to better match the correlation structure of the observations between rain stations, but it has difficulties capturing statistical dependencies between the locations when the forecast is initiated at short lead times (Verkade, Brown, Reggiani et al., 2013). The current study aims to extend these distributional and spatial based approaches into the frequency domain using wavelet transforms to remove the biases in spatio-temporal precipitation forecast across multiple times scales from hourly up to daily.

Previous wavelet-based postprocessing methods have focused on post-processing hydrological variables and showed good performance for correcting error at different time scales, but the issue of spatial error is yet to be explored (Konrad Bogner, Liechti, & Zappa, 2016; K. Bogner & Pappenberger, 2011; Buschow & Friederichs, 2021). These approaches only considered the amplitudes of wavelet decompositions, but ignored the phases component of wavelet decompositions. Addressing this gap could potentially improve the representation of spatial-temporal dependence for generating surrogate data (Brunner & Gilleland, 2020; Chavez & Cazelles, 2019; Lane, 2007; Scovell, 2020). For instance, Brunner and Gilleland (2020) included phases from continuous wavelet transform in their stochastic streamflow simulation, with good skill in preserving spatial dependency across different sites. Here, we develop an approach that decomposes the precipitation forecasts into the frequency domain using wavelet transforms (Torrence & Compo, 1998) where both the amplitudes and phases of the wavelet decomposition are used to correct the systematic errors in quantities, timing, and spatial structure of raw precipitation forecasts.

This work develops and applies a wavelet-based quantile mapping post-processing method (WQM) using a case study of the greater Sydney region, in eastern Australia and an Australian developed NWP model. The main contributions of this study are to (i) develop a new approach that could remove the bias in spatio-temporal precipitation forecast at multiple time scales, (ii) demonstrate the method by post-processing short lead-time precipitation

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

forecasts, (iii) evaluate the performance of the approach from temporal, distributional and spatial perspectives.

The paper is organized as follows. Section 2 introduces the study area and ACCESS NWP dataset used in this study, as well as the various metrics for evaluation. Section 3 presents the methodology, including the basis of wavelet transforms, the process of amplitude correction in the frequency domain as well as the phase randomization and reshuffling for generating ensemble forecasts. Section 4 shows the characterization of bias in precipitation forecasts from the Australia NWP model, as well as the evaluation of the proposed method from three different perspectives. Conclusions are provided in Section 5.

2. Study area and datasets

a. Study area

The new post-processing method was applied to precipitation stations over the Sydney region, New South Wales, Australia that covers a wide range of hydrological characteristics. Fig. 1 shows the study area of the ACCESS NWP model covering the Sydney domain as well as locations of the precipitation stations within the state of New South Wales (NSW), Australia. Two stations (red triangle) in the two selected catchments (green and purple polygons) are used to demonstrate key elements of the method and results in Sections 3 and 4. Station No. 210018 is in the Hunter River catchment while Station No. 41000269 is in the Murrumbidgee River catchment.

b. Rainfall station data

Observed precipitation station data were obtained from the Bureau of Meteorology (BoM), Australia. The precipitation station data collected and maintained by BoM were under quality control. For the 173 rain gauges in NSW state, the shortest time interval of raw data was 15 minutes, and the longest period of record was available from November 1908 till now. There are 158 stations (locations are shown in Fig. 1) within the study domain of ACCESS NWP Sydney with a complete record from March 2018 to September 2021. The observed precipitation was aggregated to hourly time step for calibrating and evaluating the ACCESS NWP hourly precipitation forecasts.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

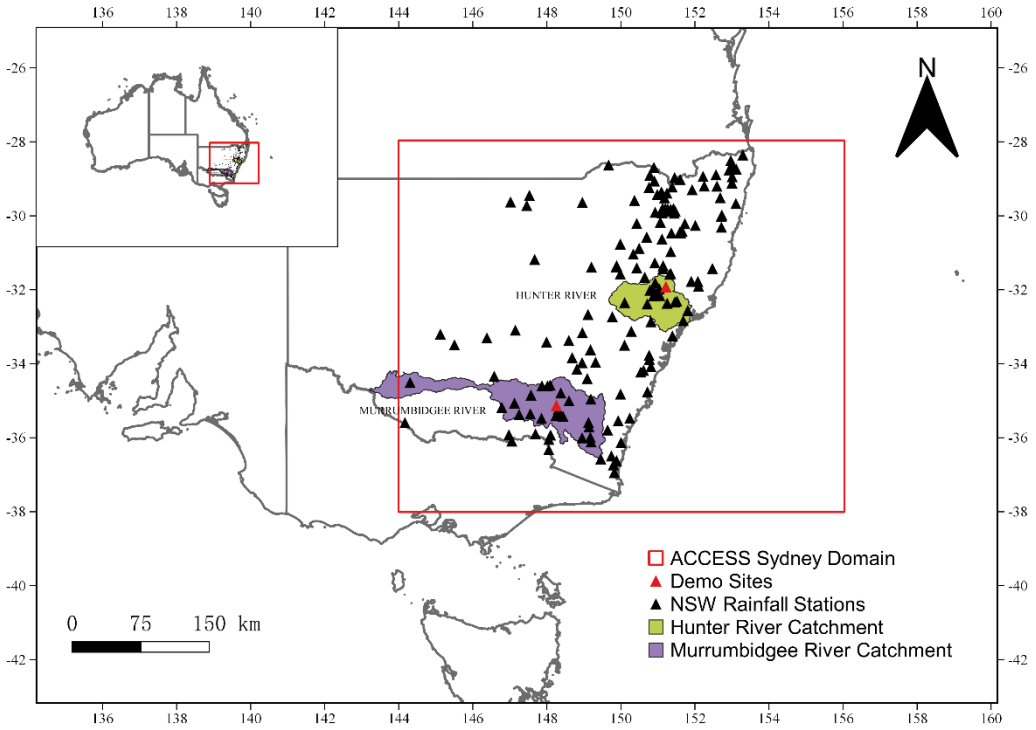

Fig. 1. Study area of ACCESS NWP model covering domain around Sydney, Australia. Two highlighted catchments (green and purple polygons) and two stations (red triangle) in the two catchments are used for demonstration in this study.

c. ACCESS Numerical Weather Predictions

In this study, we post-process hindcasts from the Australian Community Climate and Earth-System Simulator (ACCESS) NWP model made available by the Bureau of Meteorology since August 2010 (Bureau of Meteorology, 2010). Several variants of ACCESS are run operationally, extending from a coarse resolution global model (ACCESS-G) down to the high-resolution city-based models (ACCESS-C). Collectively, these ACCESS variants form the Australian Parallel Suite (APS), including APS0, APS1, APS2, and APS3. The upgrade of the ACCESS suite of NWP models has resulted in increases in the skill of precipitation forecasts (Bureau of Meteorology, 2012). ACCESS-G of ACCESS NWP models extends over the global domain and is provided at a nominal 12 km grid spacing while ACCESS-C consists of convection-allowing kilometer-scale regional models covering capital cities and major population centers, and is provided at a nominal 1.5 km grid spacing. So far, the ACCESS-C models only have one single deterministic forecast publicly available. An ensemble prediction system for ACCESS-C (ACCESS-CE) is under development for

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

three domain over Australia (Roff, Bermous, Dietachmayer et al., 2022). Here precipitation forecasts from APS2 and APS3 of ACCESS NWP across the Sydney region covering the period from March 2018 to September 2021 were used to assess the proposed WQM method.

The evaluation consists of three parts: (i) Evaluation of temporal characteristics, including correlation coefficient (*r*) and root mean square error (RMSE); (ii) Evaluation of distributional characteristics using probability density functions (PDF) skill scores (S_{score}); and (iii) Evaluation of the spatial correlation (correlation between rainfall sites against their distance). Note that the PDF skill score is a measure of how similar two PDFs are (Perkins, Pitman, Holbrook et al., 2007). It quantifies the overlap between the two PDFs by measuring the common area between them. The definitions of all evaluation metrics are shown in Table 1. All evaluations were based on two-fold cross-validation. The first half of the study period was used as the training set while the other half was used as the testing set, and then they were swapped.

Table 1 Definition of evaluation metrics[a]

Metric	Definition	Range	Optimal
Root mean square error, RMSE	$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}}$	$[0, \infty]$	0
Correlation coefficient, *r*	$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$	[-1, 1]	1
Probability density functions skill scores, S_{score}	$S_{\text{score}} = \sum_{i=1}^{k} minmum(Z_x, Z_y)$	[0, 1]	1

[a] x is simulated forecasts from NWP models and y is observations. N is the number of data pairs. k is the number of bins used to estimate the PDF, Z_x is the frequency of values in a given bin from the NWP simulations while Z_y is the frequency of values from the observed data.

In addition to assessement using the metircs above, we adopted the Hering-Genton (HG) test (Gilleland, Hering, Fowler et al., 2018) to calculate the forecast skill, which can account for both temporal dependence and contemporaneous correlation. As the proposed new method creates an ensemble of forecasts from the single deterministc ACCESS-C forecast,

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

the metrics, including RMSE, *r*, and S_{score} from the post-processed forecasts presented in the Section 4 are derived from the ensemble mean of the forecasts consisting of all realizations. Rank histograms (Hamill, 2001), generated by repeatedly tallying the rank of the observations relative to values from an ensemble sorted from lowest to highest, are used to measures how well does the ensemble spread of the forecast represent the true variability (uncertainty) of the observations. The results of rank histograms are given in Fig. S4 of the Supplemental Material.

3. Methodology

The proposed method for postprocessing NWP precipitation forecasts consists of three main steps. First, precipitation forecasts were decomposed into the frequency domain using the continuous wavelet transform, resulting in estimates of the amplitudes and phases of the time series across the frequency spectrum. Quantile mapping was then used to correct bias in the amplitudes of each frequency. Random time series were used to generate phases (hereinafter referred to as randomized phases) for reconstructing the post-processed forecasts. In the third step, post-processed precipitation forecasts were reconstructed by taking the inverse wavelet transform of the corrected amplitudes and randomized phases to create an ensemble of precipitation forecasts. We provide a detailed breakdown of the procedures for each step, with a summary of the overall procedures presented at the end. To aid in understanding, a general flowchart illustrating the methodology is provided in Fig. 2.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

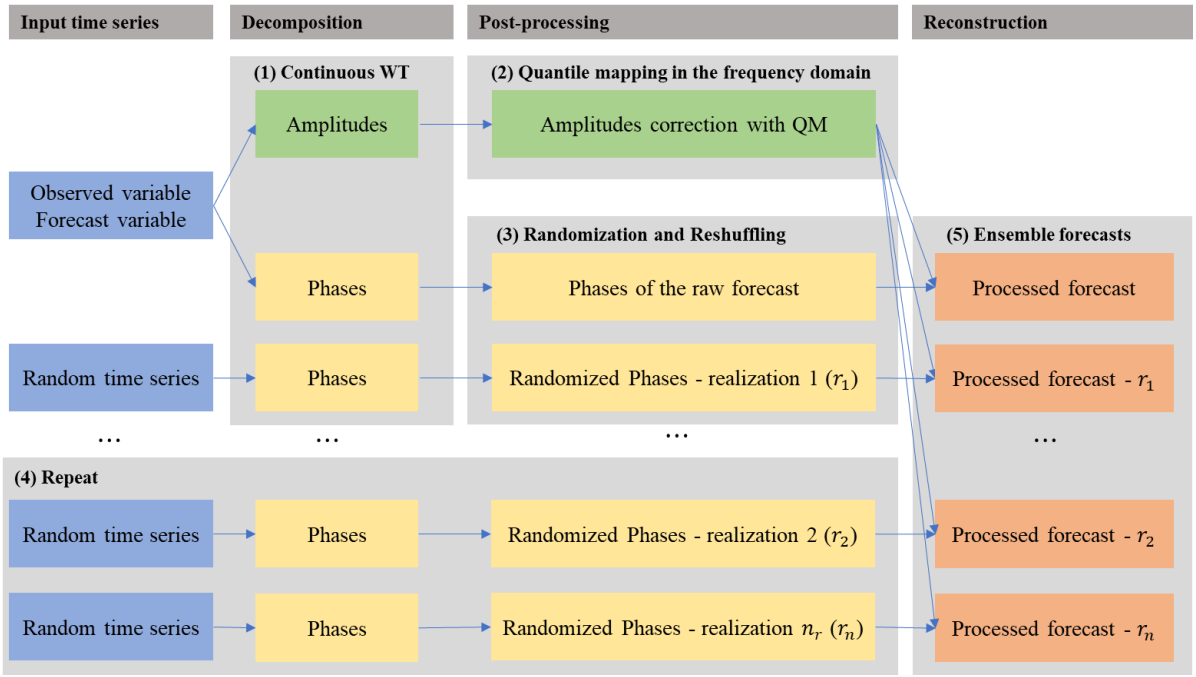

Fig. 2. Flowchart of the methodology which includes five main steps: (1) Decompostion of observed, raw forecasts as well as random times series using continuous wavelet transform (CWT); (2) Corrections of amplitudes from raw forecasts using quantile mapping (QM); (3) Randomization and reshuffling of phases from random time series; (4) Repeation of step (3) to generate multiple realizations of randomized phased; and (5) Inverse wavelet transform to obtain reconstructed ensemble forecasts.

a. Decomposition in the frequency domain using continuous wavelet transform

Wavelet transforms can be used to analyze time series or signals that contain information at many different frequencies (Daubechies, 1992; Torrence & Compo, 1998). The continuous wavelet transform decomposes the original time series into a set of coefficients, consisting of amplitudes and phases, across the time-frequency domain. The wavelet function used for the transform (e.g., orthogonal or nonorthogonal wavelets) should fit for the purpose, such as denoising or image processing (Sang, 2013; Torrence & Compo, 1998). A generalized wavelet function (or a daughter wavelet at scale s with shift n), $\psi(\eta)$, that depends on a nondimensional "time" parameter η, is shown as follows (Torrence & Compo, 1998):

$$\psi(\eta) = \psi\left[\frac{(n'-n)\delta t}{s}\right] = \left(\frac{1}{s}\right)^{1/2} \psi_0\left[\frac{(n'-n)\delta t}{s}\right] \quad (1)$$

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

where $\psi_0(\eta)$ is normalized $\psi(\eta)$ to have unit energy. *n* is the localized time index and *s* is the wavelet scale (proportional to the wavelength of the frequency associated with a Fourier transformation). Due to its smooth features and wide applications in hydro-climatological applications (Brunner & Gilleland, 2020; Fossa, Dieppois, Massei et al., 2021; Hu & Si, 2021; Schaefli, Maraun, & Holschneider, 2007), the Morlet wavelet was adopted in this study because it provides a balanced trade-off between the detection of scales and localization of the oscillations in time (Torrence & Compo, 1998). The continuous wavelet transform of a discrete-time series, $\boldsymbol{x_n}$, is defined as the convolution of $\boldsymbol{x_n}$ with a scaled and translated version of $\psi_0(\eta)$:

$$\boldsymbol{W_n(s)} = \sum_{n\prime=0}^{N-1} \boldsymbol{x_{n\prime}} \psi^*[\frac{(n'-n)\delta t}{s}] \tag{2}$$

where star ($*$) denotes complex conjugate functions of $\psi(\eta)$. Since the wavelet function $\psi(\eta)$ is in general complex, the wavelet decompositions $W_n(s)$ is also complex, including the real part, $\Re\{\boldsymbol{W_n}(s_j)\}$, and imaginary part, $\Im\{\boldsymbol{W_n}(s_j)\}$. The complex values can be transferred to the polar coordinate system, consisting of amplitudes and phases across the spectrum (Torrence and Compo 1998). So, in a polar coordinate system, the complex values of wavelet decompositions can be divided into amplitudes ($\boldsymbol{A} = |\boldsymbol{W_n(s)}|$), and phases ($\boldsymbol{\phi} = \tan^{-1}[\Im\{\boldsymbol{W_n(s)}\}/\Re\{\boldsymbol{W_n(s)}\}]$). The wavelet decompositions measure the similarity of *x*(*t*) to the wavelet ψ at the current scale *s* (Percival & Walden, 2000).

Finally, with the basis wavelet function and wavelet decompositions, the original time series can be reconstructed by the summation of the real part of the wavelet decompositions across all scales using a simple delta (δ) function (Farge, 1992). This inverse transform yields (Torrence & Compo, 1998):

$$\boldsymbol{x_n} = \frac{\delta j \delta t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=0}^{J} \frac{\Re\{\boldsymbol{W_n}(\boldsymbol{s_j})\}}{s_j^{1/2}} \tag{3}$$

where C_δ is a constant that depends on the specific wavelet $\psi(\eta)$ used, and *J* is the largest scale investigated. The choice of scales s_j and *J* can be written as a fractional power of two:

$$s_j = s_0 2^{j\delta j}, \quad j = 0,1,\cdots,J \tag{4}$$

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

$$J = \delta j^{-1} \log_2 \left(\frac{N \delta t}{s_0}\right) \quad (5)$$

where s_0 is the smallest resolvable scale, and *J* determines the largest number of scales. δt is the time step of the time series while δj depends on the width of the wavelet function used and smaller values give finer frequency resolution. In our study, we have carefully selected the value of δj to ensure that the smallest possible number of wavelet scales, *J*, is used while still characterizing the time series spectrum given a length of the dataset. Readers are referred to Torrence and Compo (1998) for additional details on the background of the continuous wavelet transform. In addition, the same number of scales, J, should be used when we decompose the training and target precipitation forecasts. However, in real applications if the target precipitation forecasts are for a short lead time only, then to ensure that there are sufficient data points for a decomposition with J scales to be valid, the target or future forecasts should be concatenated with the training forecasts prior to the wavelet decomposition.

To demonstrate the wavelet transform, we applied the method to a time series of observed precipitation at a selected station (Station No. 210018 in Hunter River catchment shown in Fig. 1). The variability of precipitation differs in the localized time-frequency domain (Fig. 3) with differences in the amplitudes and phases at different time scales. A small number of wavelet scales were used here (a total number of 9 wavelet scales with $J = 8 \; and \; \delta j = 0.125$). It is clear that the variances at the small wavelet scales are much larger than the longer wavelet scales. Precipitation is a temporally dynamic meteorological variable (i.e., high-frequency signal), and thus the amplitudes at short time scales have large variability. Similarly, the phases vary quickly between $-\pi$ and π at small time scales and this determines the timing of the precipitation occurrence in the wavelet reconstruction. Due to the fact that the majority of precipitation information is at smaller wavelet scales, the influence of boundary effects on the precipitation decomposition at smaller time scales is limited (Torrence & Compo, 1998). However, when rainfall events occur at the edge of the time series, it is recommended to pad the beginning and end of the time series with sufficient zeros and then remove them afterward.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

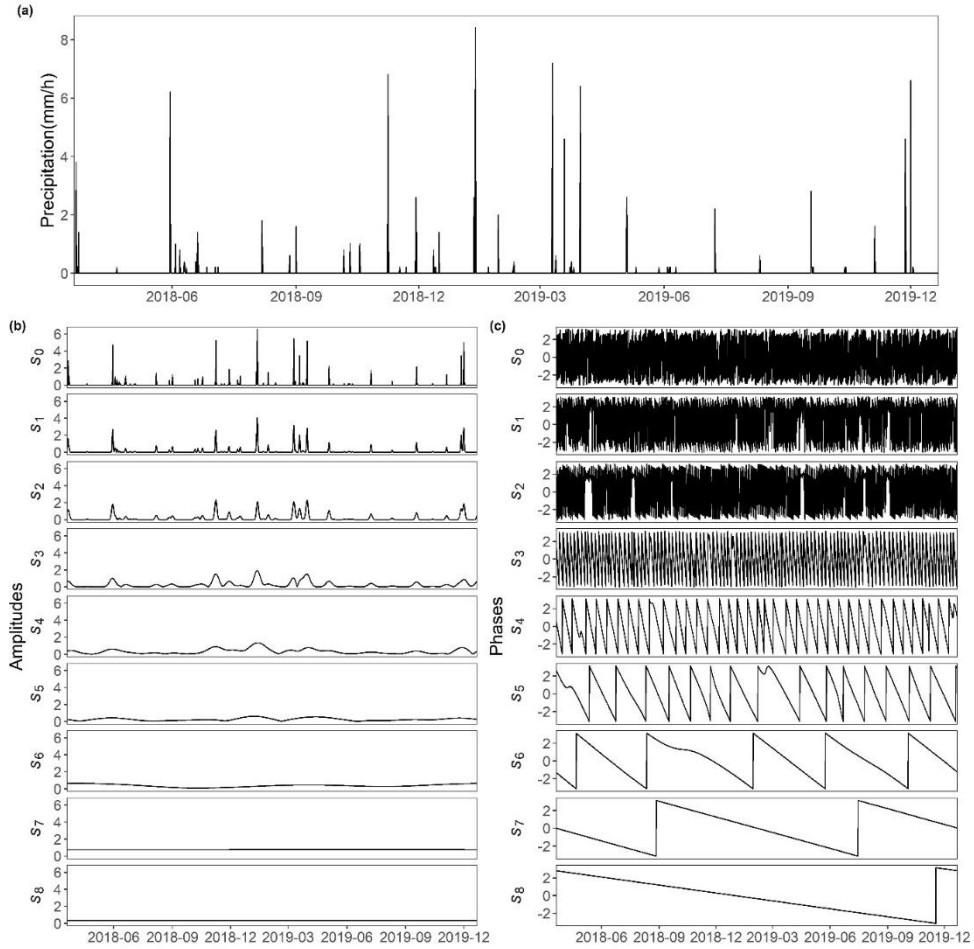

Fig. 3. Decomposition of an observed precipitation time series using continuous wavelet transform at selected wavelet scales. The original precipitation time series is given in panel (a) while its amplitudes and phases in the time-frequency domain are given in panel (b) and (c), respectively.

b. Amplitude correction using quantile mapping

As shown in Fig. 4(a), the decomposed amplitudes of observations (black lines) and raw forecasts across the spectrum from small scales (high-frequency information) to large scales (low-frequency information) are presented. According to the equations of scales in Equation (4), the equivalent periods of the investigated scales (s_0, s_{10},…, and s_{60}) are 2h, 4.8h, 11.3h, 26.9h, 64h, 152.2h, and 362h. The amplitudes are calibrated using the QM, which is given in the following formula (Cannon, Sobie, & Murdock, 2015):

$$\widehat{\boldsymbol{A}}_{\boldsymbol{j}} = F_y^{-1}[F_x(\boldsymbol{A}_{\boldsymbol{j}})] \tag{6}$$

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

where F_x is the CDF of raw model simulation, and F_y^{-1} is the inverse of the observation CDF.

With the QM applied to adjust the CDF of the amplitudes from raw forecasts at each frequency, their corresponding CDFs are shown in Fig. 4(b). As mentioned previously, precipitation is a highly random process, so the majority of variability is at short time scales (from hourly to daily), representing high-frequency information. With the corrected amplitudes, the next step is to generate its associated phases so that we can generate multiple realizations of precipitation forecasts and form the ensemble forecasts.

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

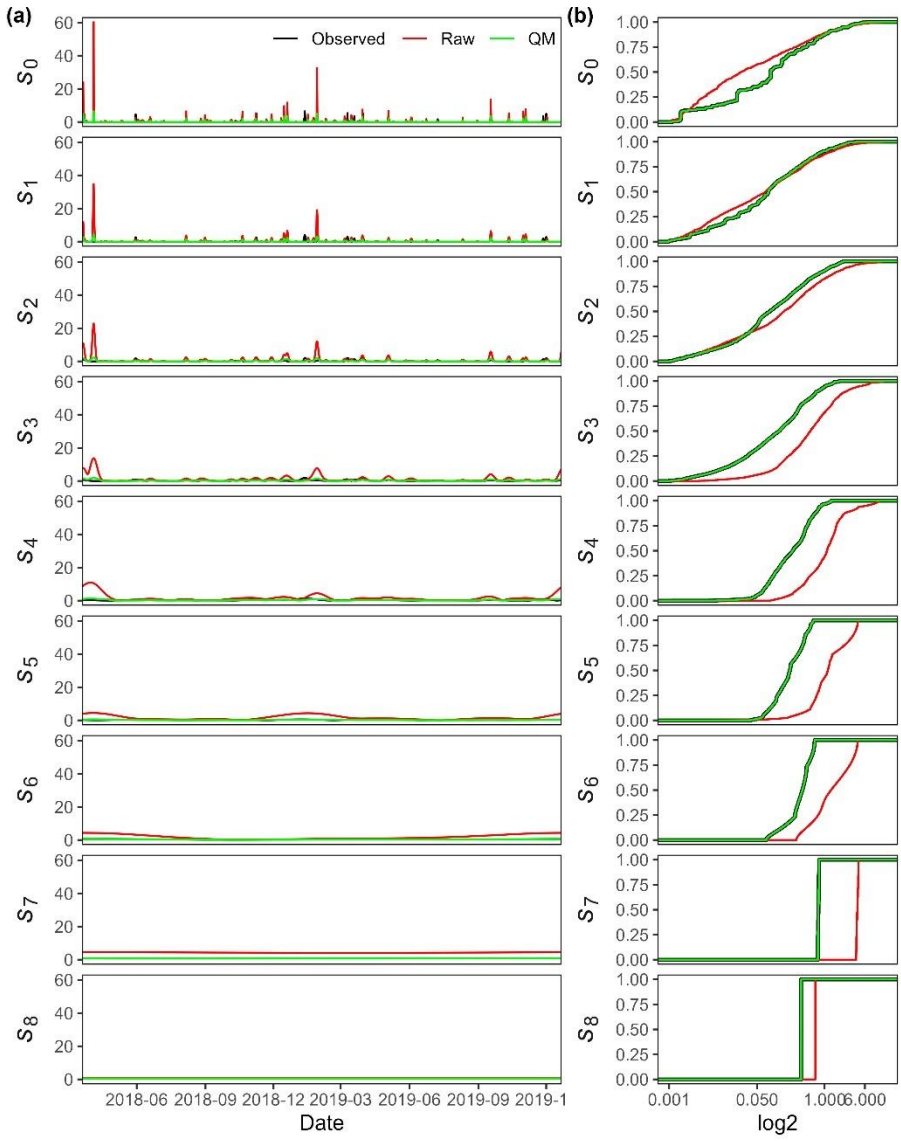

Fig. 4. Bias correction of amplitudes of wavelet decompositions across the frequency spectrum using quantile mapping: (a) Amplitudes of observations, raw and corrected forecasts; (b) CDF of amplitudes derived from observations, raw and corrected forecasts.

c. Phase randomization and reshuffling

First, a random time series (Gaussian white noise with zero mean and unit standard deviation) of the same length as the input series was generated, and its phases were derived from the random time series using wavelet transform with a complex wavelet function, similar to the approach proposed by Chavez and Cazelles (2019) and the randomized phases are then reshuffled. Note that the reshuffling process is not part of the approach given in

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

Chavez and Cazelles (2019). A demonstration of the shuffling process is presented in Table 2 using ten randomly generated phases as an example. At time step 1, the rank of raw forecast is 4. This corresponds to a new value of -0.883 from the random time series, and in the randomized phase series, this new value will be shifted to time step 1. The same procedure will be applied to phases at other time steps until the new values from the random time series have the same rank. This is to ensure the pairwise relationship between amplitudes and phases from raw forecasts.

Table 2 Demonstration of the reshuffling process of phases

Time step	Phases from raw forecasts	Rank	Phases from random time series	Rank	Randomized phases	Rank
1	-1.208	4	0.785	6	-0.883	4
2	-1.523	3	2.401	10	-0.895	3
3	0.329	9	-1.380	2	1.650	9
4	-2.787	1	-0.638	5	-1.856	1
5	-0.198	6	1.650	9	0.785	6
6	-0.102	7	1.062	7	1.062	7
7	1.963	10	-1.856	1	2.401	10
8	-0.815	5	-0.895	3	-0.638	5
9	0.293	8	-0.883	4	1.196	8
10	-2.072	2	1.196	8	-1.380	2

It is clear that after reordering, the phases from the random time series which correspond with the phases of raw forecasts the randomized phases follow the pattern of raw forecasts. This example demonstrates the process at one frequency level; in the WQM method, this operation was carried out for each frequency level separately, and randomized phases (derived from the same random time series) at each decomposition level were reshuffled according to the rank of the raw forecasts. In addition to this, a block-wise shuffling is included to allow shifting forward or backward slightly of the forecasts, with separate shifts for each bias correction location. This is an optional step, with an additional parameter of the block size to be optimised, that ensures the proposed method can deal with a range of potential biases in NWP models in terms of timing and spatial structure errors. A simple illustration for the block wise shuffling is given in Table S2 of the Supplemental Material. In terms of the real example here (subset shown in Fig. 5 to highlight the differences resulting from the process), with the randomized phases from the random time series (red lines), we rank the phases according to the phases of raw forecasts at the corresponding time scale.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

After the reshuffling process, there is a greater similarity between the randomized phases (blue lines) and the phases from raw forecasts (black lines). Last but not least, phases from the same random time series are used for all the rainfall stations considered to retain spatial dependencies among those sites. To see the impact of phase randomization and reshuffling on the proposed method, please refer to Fig. S2 in the Supplemental Material.

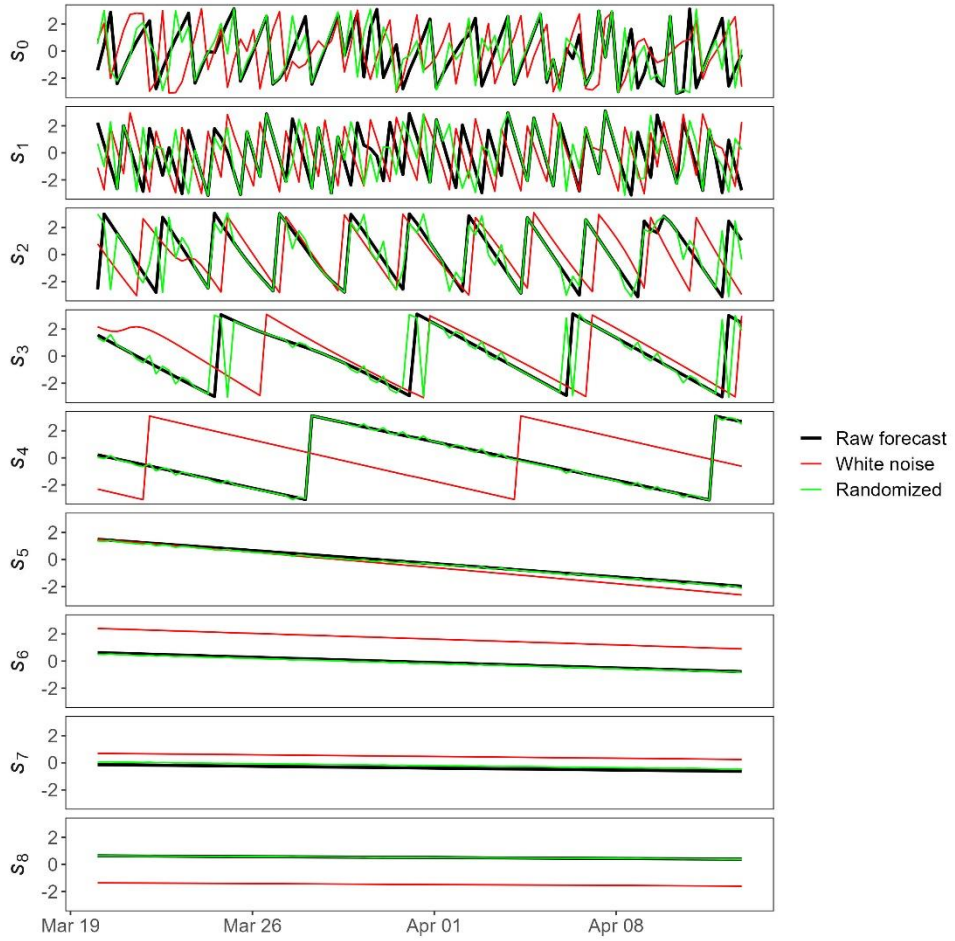

Fig. 5. Illustration of phase randomization and reshuffling using 100 data points at Station 210018. The phases from white noise (red line) change to randomized phases (blue line) according to the rank of phases from raw forecasts (black line). The full time series at Station 210018 can be found in Fig. S3 of the Supplemental Material.

d. Wavelet-based QM formulation

Processed precipitation forecasts $\hat{\boldsymbol{x}}_{\boldsymbol{n}}$ can be obtained by reconstructing the time series using corrected amplitudes ($\widehat{\boldsymbol{A}}_{\boldsymbol{j}}$) and randomized and shuffled phases ($\widehat{\boldsymbol{\phi}}_{\boldsymbol{j}}$). To summarize, the

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

wavelet-based QM procedure for post-processing NWP precipitation forecasts consists of five main steps:

1) Decompose the raw model simulation ($\boldsymbol{x_n}$) and observation ($\boldsymbol{y_n}$) into the frequency domain using continuous wavelet transform, extracting its amplitudes ($\boldsymbol{A_j}$) and phases ($\boldsymbol{\phi_j}$) across different time scales.

2) Adjust the CDF of amplitudes at each frequency from forecasts according to the CDF of amplitudes at the corresponding scales from the observations. Applying Equation (6), the calibrated amplitudes from forecasts $\widehat{\boldsymbol{A}}_{\boldsymbol{j}}$ is given by $\widehat{\boldsymbol{A}}_{\boldsymbol{j}} = F_y^{-1}[F_x(\boldsymbol{A_j})]$.

3) (i) Derive wavelet decompositions of the generated Gaussian white noise, $\boldsymbol{\varepsilon_n}$, to extract the phases, $\boldsymbol{\phi_j}$, and reorder these phases to get the adjusted phases as follows:

$$\widehat{\boldsymbol{\phi}}_{\boldsymbol{j}} = R_x^{-1}[R_\varepsilon(\boldsymbol{\phi_j})] \tag{7}$$

where R_ε is the rank of the random time series, and R_x^{-1} is the inverse of the rank of raw precipitation forecasts (referring to Table 2 for a simple illustration). Furthermore, a block-wise shuffling is included to allow shifting in the timing of the events.

(ii) Combine the corrected amplitudes ($\widehat{\boldsymbol{A}}_{\boldsymbol{j}}$) and randomized and shuffled phases ($\widehat{\boldsymbol{\phi}}_{\boldsymbol{j}}$) to obtain adjusted time-frequency decompositions $\widehat{\boldsymbol{W}}_{\boldsymbol{n}}(\boldsymbol{s})$, and then post-processed $\widehat{\boldsymbol{x}}_{\boldsymbol{n}}$ is obtained by using the inverse wavelet transform in Equation (**3**). It should be noted that negative values may appear after the inverse wavelet transform is applied and we use a threshold (in this study 0.1mm precipitation as used by Shahrban, Walker, Wang et al. (2016) for Australia precipitation forecasts), below which any values are set to zero.

4) Repeat step 3 using a newly generated Gaussian white noise to generate another ensemble member of post-processed precipitation forecasts. If the procedure is repeated n_r times, there will be n_r realizations of forecasts for each station.

5) Form an ensemble forecast including one set of post-processed forecasts with corrected amplitudes only (without phase randomization) and additional n_r realizations of post-processed forecasts from phase randomizations.

The above process was used to create an ensemble of post-processed precipitation forecasts for one location from the single deterministic raw forecast. To extend it to other

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

locations and preserve the spatial dependence, the above process is repeated using the same set of random time series generated for the first station.

e. Demonstration of Wavelet-based QM at two sample stations

Fig. 6 compares the performance of the proposed method (WQM and five realizations derived from randomized phases, r1, r2, …, and r5) to the raw and post-processed precipitation forecasts using QM at two selected rainfall stations in northern and south of NSW, Australia (seen in Fig. 1). Note that WQM presents the post-processed precipitation with corrected amplitudes only (i.e., without phase randomization), and this ensemble member is presented in Fig. 2 as the first processed forecast. The largest number of wavelet scales (J) used for results presented hereafter was 8. However, additional assessments of post-processed forecasts with $J = 17$ are provided in Supplemental Material from Fig. S5-S11.

As shown in the figure, QM-based approaches are able to adjust the distribution of raw simulation matching it with the observations (indicated by the PDF skill score in Fig. 6(a)), and it also corrects the extreme values simultaneously leading to the improved bias (the difference between forecasts and observations) and RMSE. However, the standard QM approach performs poorly in correlation coefficients while the proposed approach improves them substantially. The skill of the postprocessed rainfall forecasts using wavelet-based QM approach is significantly better than the conventional QM method (HG test p-values << 0.05) at both Station 210018 and 41000269. Full test statistics and associated p-values from different ensemble members are given in Table S1 of the Supplemental Material. It is noted that the correlation between forecasts and observations is low at the hourly time step while it increases when aggregating to a longer time step. More details regarding the characteristics of bias in ACCESS NWP can be found in Section 4.a.

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

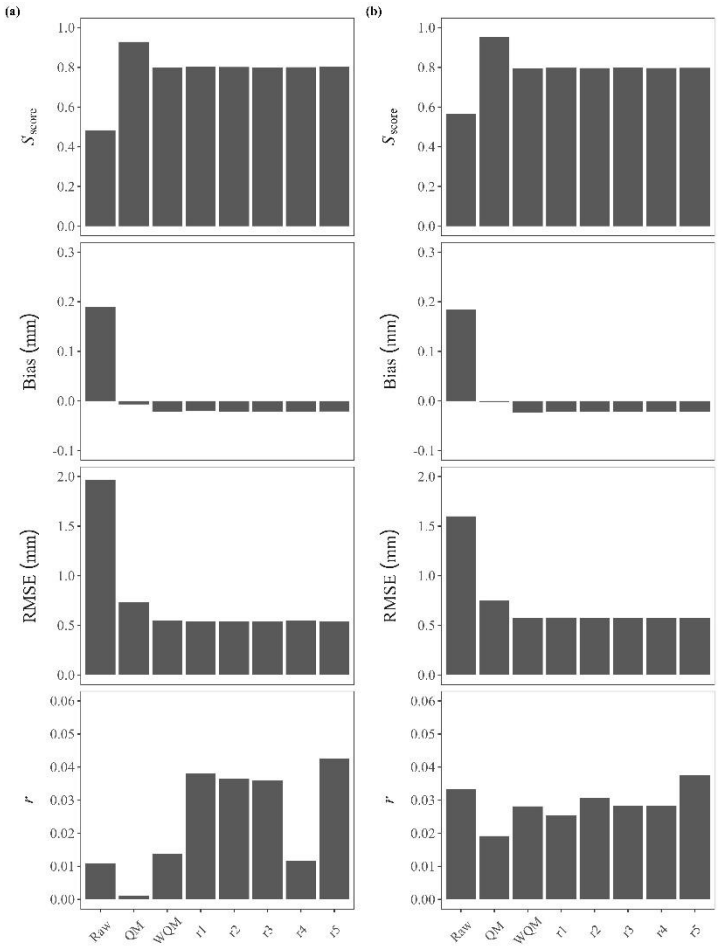

Fig. 6. Performance of post-processed precipitation forecasts using the proposed approach, including WQM and five realizations derived from phase randomization and reshuffling, r1, r2, …, and r5. WQM stands for the post-processed precipitation with corrected amplitudes and raw phases from the raw forecasts. Comparison is made against raw forecasts as well as adjusted precipitation forecasts using QM at two selected stations: (a) Station No. 210018 in the Hunter River catchment; (b) Station No. 41000269 in the Murrumbidgee River catchment.

The number of wavelet scales depends on the δj, and it should be chosen carefully in practical applications. A smaller number of wavelet scales is preferred in post-processing precipitation forecasts, since a larger number of wavelet scales providing finer frequency resolution would lead to overfitting and poor performance. In addition, a large number of wavelet scales in wavelet transforms requires greater computational cost (power, time, etc) when decomposing precipitation at a finer frequency resolution. In Fig. S1 of the Supplemental Material, results of a sensitivity analysis on the impact of the number of wavelet scales are presented.

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

4. Results and discussions

a. Characterization of bias in precipitation forecasts from ACCESS NWP

Fig. 7 shows the NWP precipitation bias at various lead times, and raw hourly forecasts have larger errors at lead 1 hour. When aggregated over longer durations, RMSE increases as expected, with lead 1 hour forecasts still having the largest RMSE across lead times. Correlation between observed and predicted precipitation decreases when lead time increases, and this pattern remains unchanged when aggregating to larger durations. Correlation is substantially improved over longer aggregations (Cuo, Pagano, & Wang, 2011; Shrestha, Robertson, Wang et al., 2013). According to error decomposition and the relationship between RMSE and correlation (Gupta, Kling, Yilmaz et al., 2009), the larger RMSE of forecasts at short lead times is likely due to the difference in standard deviation and mean between observed and predicted forecasts,

$$\mathrm{RMSE}^2 = 2\sigma_x\sigma_y(1-r) + \left(\sigma_x - \sigma_y\right)^2 + (\mu_x - \mu_y)^2 \quad (8)$$

where σ_x and σ_y represent the standard deviation of forecasts and observations, respectively, and μ_x and μ_y are the mean of forecasts and observations. r is the correlation coefficient between the forecasts and observations (seen in Table 1). As a result, the characterization of bias in ACCESS NWP forecasts suggests that there is a need to post-process precipitation forecasts to match their mean and standard deviation across multiple time scales, especially at short lead times. Previous studies on ACCESS NWP post-processing also found that the RMSE of precipitation forecast varies at different lead times with the lowest RMSE for lead times of 13-24h, and there were larger errors due to spin-up problems in the shorter lead times (Shahrban et al., 2016). This was observed in other NWP systems (Huang & Luo, 2017). In the following sections, precipitation forecasts at lead 1 hour were used to evaluate the performance of the proposed method in correcting biases across multiple time scales.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

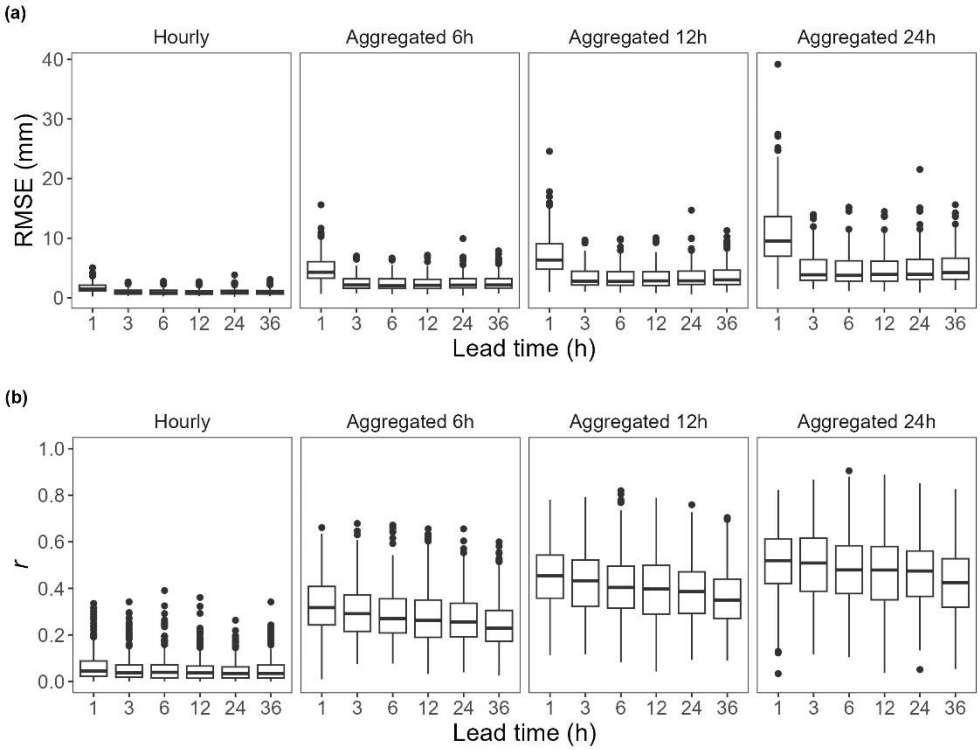

Fig. 7. Characterization of errors in raw precipitation forecasts from ACCESS NWP across all stations over the Sydney domain: (a) RMSE; (b) *r*.

b. Evaluation of the proposed method from a temporal perspective

Fig. 8 shows the bias, presented as the difference between the forecasts and observations, of both raw and post-processed forecasts. The raw forecasts consistently overestimate precipitation at all locations, with slightly larger biases in the austral summer (Fig. 8b). After post-processing, the biases are substantially reduced in both seasons and for all time aggregations. However there is a tendency for a slight underestimation following bias correction (bias values tend to be slightly less than zero), and the underestimation of extreme rainfall in post-processed forecasts is likely the cause. Fig. 9 compares the raw and post-processed forecasts using RMSE. In both periods, forecasts at the majority of the rain gauge sites are improved following post-processing. When aggregating to a longer duration, the RMSE of the raw precipitation forecasts is larger, and the improvement following post-processing is thus also larger.

The temporal correlations of the raw and post-processed forecasts are compared in Fig. 10. Correlations for the hourly precipitation are improved at almost all stations after post-processing using the proposed WQM approach (p-value << 0.05, paired t-test). At longer

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

time aggregations, however, only a small percentage of stations show improvement in the corrected forecasts with most stations having lower correlations at daily aggregation (Fig. 10(a)). The ensemble mean of post-processed forecasts (including the forecasts from phase randomization and reshuffling) do not correctly capture drizzle and tend to have a higher percentage of zero values compared to the raw forecasts. On average, the post-processed forecasts had around ten percent more zero values not matching with observations than the raw forecasts for 12 and 24-hour aggregation times.

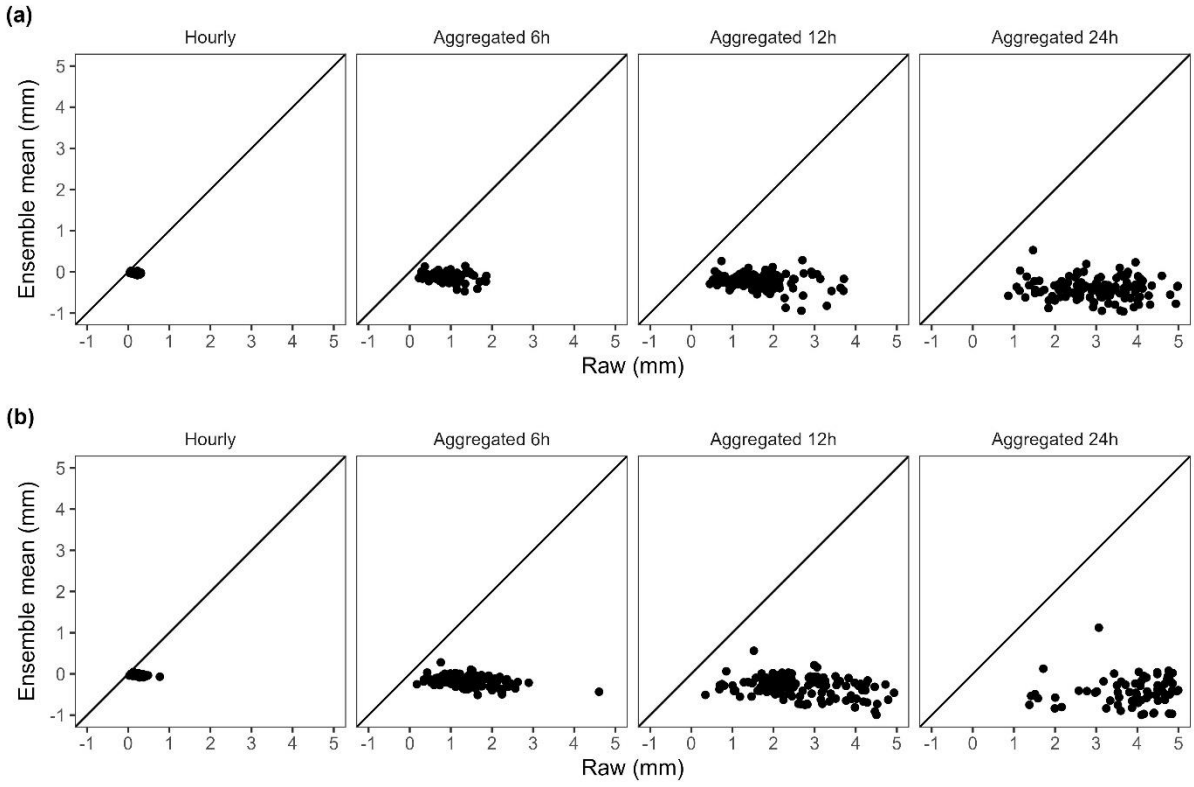

Fig. 8. Bias of post-processed forecasts at different time scales across all stations (a) March-August; (b) September-February.

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

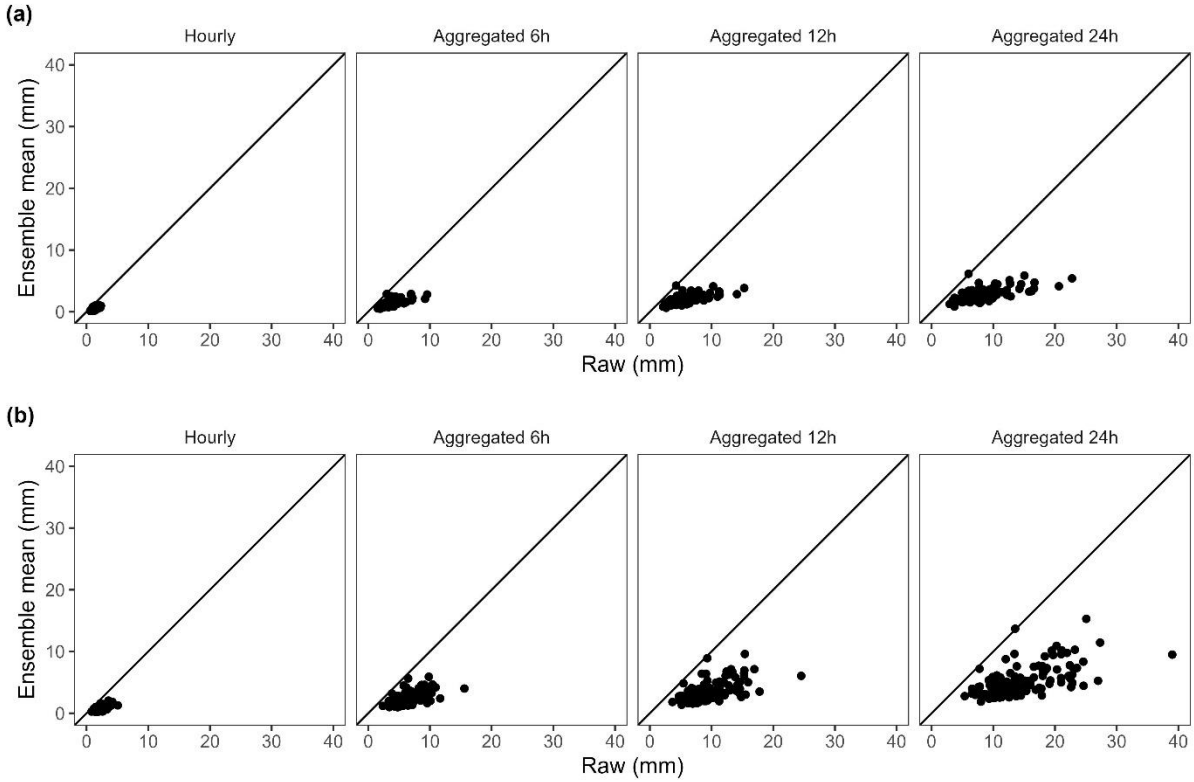

Fig. 9. RMSE of post-processed forecasts at different time scales across all stations (a) March-August; (b) September-February.

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

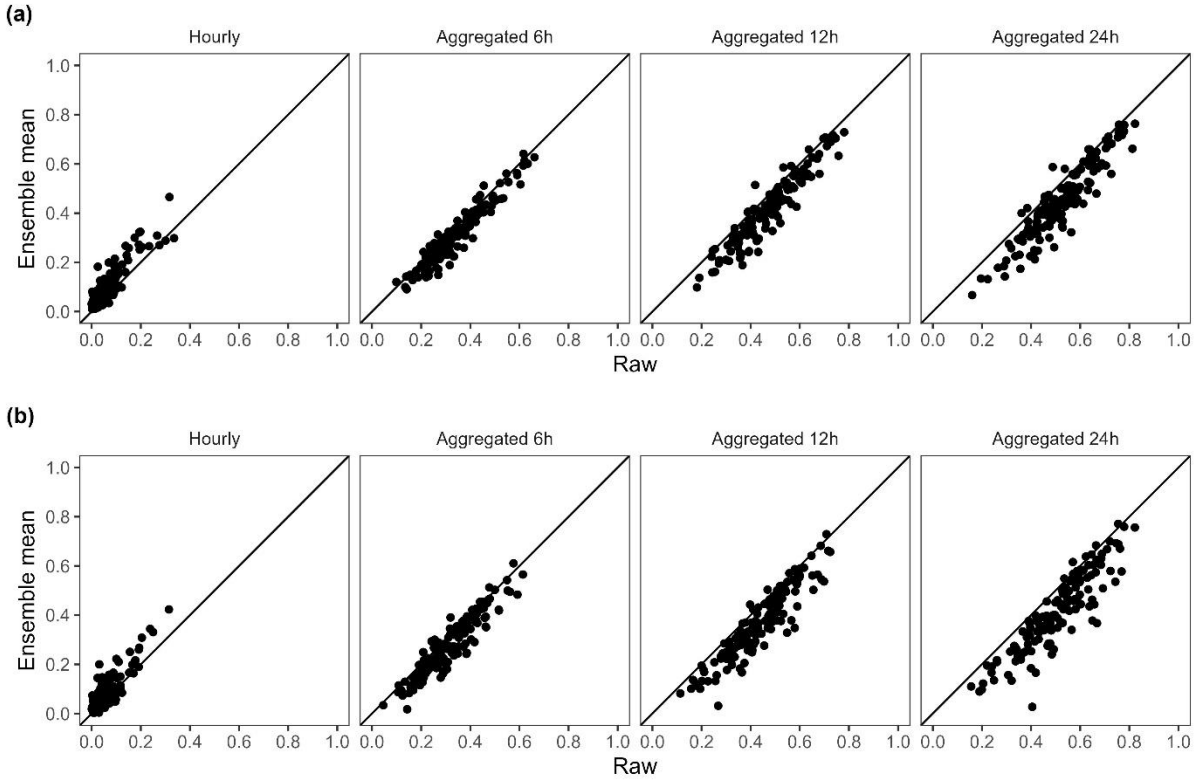

Fig. 10. Temporal correlation of post-processed forecasts at different time scales across all stations (a) March-August; (b) September-February.

c. Evaluation of the proposed method from a distributional perspective

Following post-processing, most sites have improved PDF skill scores across the range of time aggregations (Fig. 11). The improvement is the most evident for the hourly time scale since the raw forecasts had a very low PDF skill score. Our method for correcting distributional characteristics of precipitation has been found to be highly effective. Specifically, 83% of the stations had significant improvements (paired t-test at the 5% level) for hourly precipitation, 58% for aggregated 6-hour precipitation, 78% for 12-hour precipitation, and 91% for 24-hour precipitation.

Since quantile correction has been applied at different time scales, the WQM approach leads to improved representation of the precipitation distributions and lower bias. Note that the results presented so far in Section 4 are derived from the mean of the ensemble forecast consisting of 20 realizations. Results would differ when a different number of ensemble members was adopted, but the improvements from the WQM approach are consistent across the temporal and distributional metrics investigated.

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

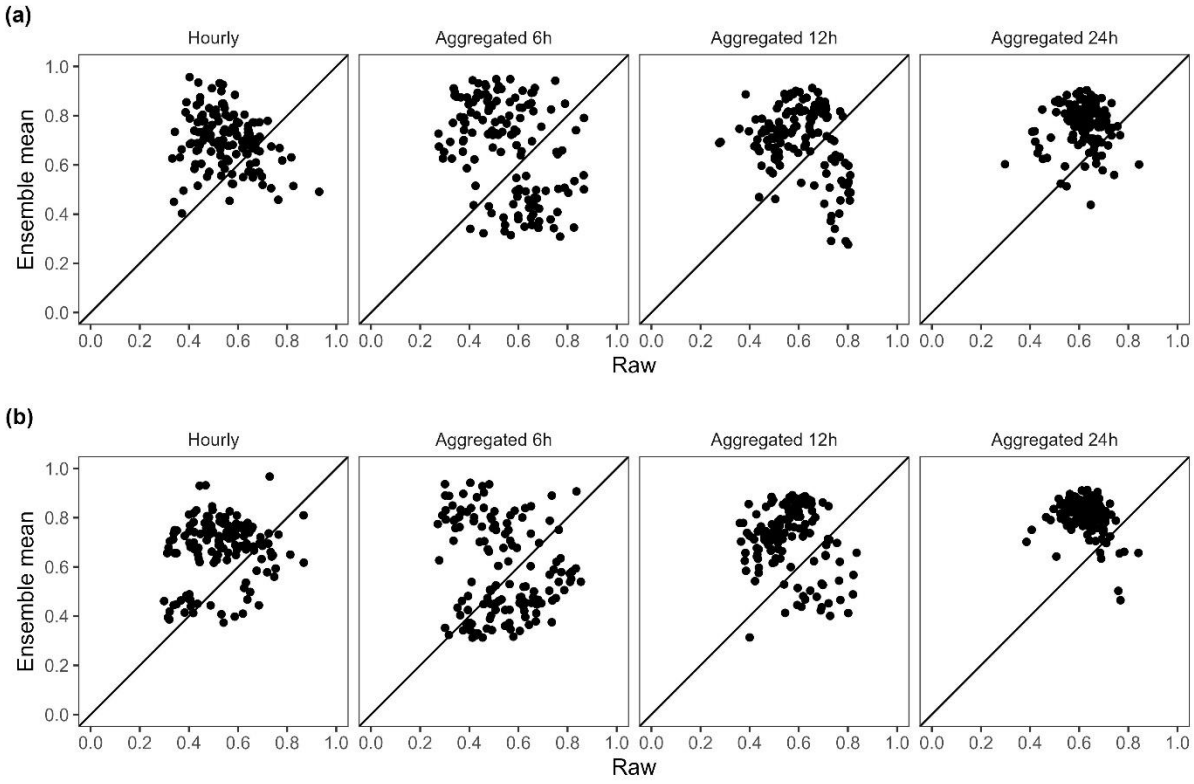

Fig. 11. PDF skill score of post-processed forecasts at different time scales across all stations (a) March-August; (b) September-February.

d. Evaluation of the proposed method from a spatial perspective

The gauges cover a large area and given the often, small spatial structure of storms leading to short-duration precipitation events, the correlations between those stations over the entire domain varies substantially.

Spatial correlations between all station pairs plotted against distance are shown in Fig. 12. For clarity, only the fitted power models from different forecasts are shown. Correlation with distance increases when precipitation is aggregated to longer time scales. The spatial structure of the raw forecasts (red line) has the largest bias for the longest time scales. WQM post-processed precipitation forecasts improve the representation of spatial structures of precipitation, with the biggest improvements at the longest time scales where the raw errors are largest.

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

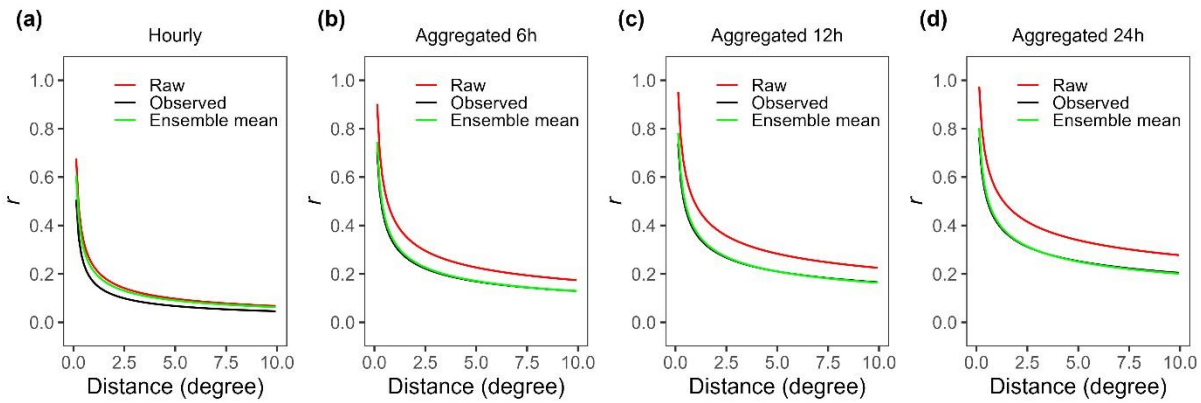

Fig. 12. Spatial correlation of post-processed NWP precipitation forecasts against their distance between all stations over the entire study domain, and subplots represent the performance of post-processed forecasts from different time scales.

We focus our attention on two river catchments, where correctly representing spatial correlation is vital for accurate flood forecasting. Fig. 13 shows the spatial correlations of post-processed precipitation forecasts between stations within the Hunter and Murrumbidgee River catchments. The observed correlation between stations against distance is lower than the raw simulation from ACCESS NWP, reflecting model structural errors (Jankov, Gregory, Ravela et al., 2021). The WQM approach corrects those biases and better matches the dependence across all the stations. Using the ensemble of 20 realizations (in grey color), the observed spatial correlation structure is within the band of these randomized precipitation forecasts. The post-processed forecasts still have more structure (higher correlation with distance) than the observations and this is likely due to the same randomization phases (from a random Gaussian white noise) being used to generate the different realizations (Brunner & Gilleland, 2020; Chavez & Cazelles, 2019).

File generated with AMS Word template 2.0

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

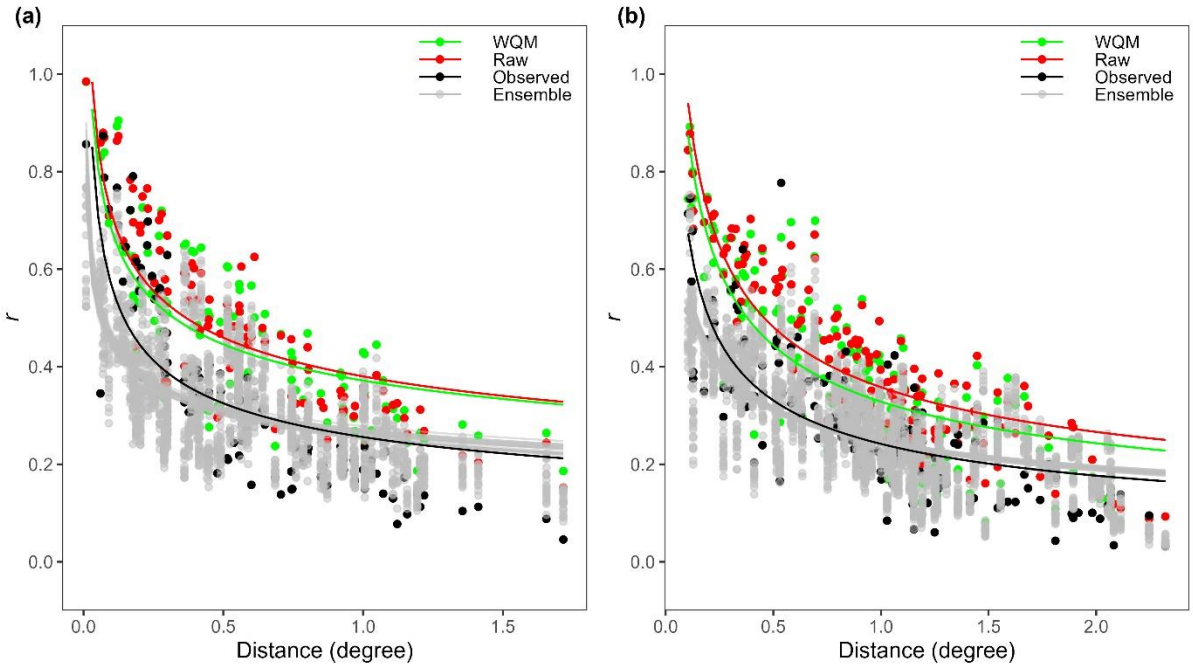

Fig. 13. Spatial correlations of post-process NWP precipitation forecasts between stations (solid points) and their fitted power model (solid lines) within two catchments: (a) Hunter River catchment; (b) Murrumbidgee River catchment.

5. Implications for post-processing NWP

The treatment of systematic timing and spatial errors in raw precipitation forecasts is challenging and new methods are needed that can simultaneously address improve temporal, spatial, or intervariable dependence. In addition, precipitation forecast postprocessing often does not account for the bias across multiple time scales, especially at short lead times. Postprocessing methods that can address the multi-scale issue while preserving spatio-temporal or intervariable dependency are needed. The proposed method makes use of the wavelet transform with a complex wavelet function, which can provide additional information on phases representing the timing of the precipitation occurrence. It overcomes the issue of representativeness of reordering in Schaake shuffle by using the randomized phases from Gaussian white noise. Meanwhile, using the same set of randomized phases used across all stations investigated preserves the spatial-temporal dependency.

As such, instead of building statistical relationships between observations and raw forecasts for a particular location and a specific forecast time, the proposed method accounts for the timing and spatial errors, and it can capture the true variability (distributional

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

similarity) of the observations across multiple time scales. The WQM method provides a practical and effective way to correct these errors by incorporating spatio-temporal neighborhood information through the frequency domain. With timing and spatial errors removed, precipitation forecasts will be more skillful, and hydrological modeling using the post-processed forecasts can provide higher accuracy of streamflow estimation.

6. Conclusions

This paper has presented a new method for postprocessing NWP precipitation forecasts. The novelty of the approach is to transfer a traditional quantile mapping method into the frequency domain. This method makes use of the power of wavelet transform using the complex wavelet functions, which give us the opportunities to adjust both amplitudes and phases of precipitation forecasts. We assessed the new approach using forecasts from an NWP model covering the Sydney region. The advantage of the approach over the QM approach is demonstrated using two demonstration stations in detail, and the performance of the post-processed forecasts has been assessed from temporal, distributional as well as spatial perspectives. The results have shown clear improvements in a range of metrics, including RMSE, temporal and spatial correlation, and PDF skill scores. More importantly, the proposed method has shown great potential in correcting bias in short-lead time forecasts across multiple time scales. This is particularly important for early flood warnings assisting in providing adequate time for emergency responses. The preservation of the spatio-temporal dependency from post-processed precipitation forecasts can further improve the accuracy of flood forecasts for spatially-distributed hydrologic systems.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

Acknowledgments.

This research was funded by the Australian Research Council linkage grant (LP150100548) and Crown lands & Water Division, Department of Industry, NSW, Australia. Fiona Johnson is supported by the Australian Research Council Training Centre in Data and Analytics for Resources and Environments (DARE) IC190100031. We are thankful to three anonymous referees and the editor for their constructive comments.

Data Availability Statement.

Observed rainfall station data were obtained from the Bureau of Meteorology, Australia. It also can be accessed through the website of WaterNSW: https://realtimedata.waternsw.com.au/. ACCESS NWP data from APS2 and APS3 is available through the NCI THREDDS Data Server, and data is available from the following websites: https://dapds00.nci.org.au/thredds/catalogs/lb4/access/access.html for Australian APS2 and https://dapds00.nci.org.au/thredds/catalogs/wr45/ops_aps3/ops_aps3.html for the latest APS3. Data and code are publicly available on figshare: https://doi.org/10.6084/m9.figshare.21903033. This includes a wrapped R package, as well as the code used to generate the results and figures presented in the paper.

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

REFERENCES

Bogner, K., Liechti, K., & Zappa, M. (2016). Post-Processing of Stream Flows in Switzerland with an Emphasis on Low Flows and Floods. *Water, 8*(4). doi:10.3390/w8040115

Bogner, K., & Pappenberger, F. (2011). Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resources Research, 47*(7). doi:https://doi.org/10.1029/2010WR009137

Brunner, M. I., & Gilleland, E. (2020). Stochastic simulation of streamflow and spatial extremes: a continuous, wavelet-based approach. *Hydrol. Earth Syst. Sci., 24*(8), 3967-3982. doi:10.5194/hess-24-3967-2020

Bureau of Meteorology. (2010). *Operational implementation of the ACCESS numerical weather prediction systems*. Retrieved from Bureau of Meteorology, Melbourne, Australia: NMOC Operations Bulletin No. 83, available at: http://www.bom.gov.au/australia/charts/bulletins/apob83.pdf

Bureau of Meteorology. (2012). *APS1 upgrade of the ACCESS-G numerical weather prediction system.* Retrieved from Bureau of Meteorology, Melbourne, Australia: NMOC Operations Bulletin No. 93, available at: http://www.bom.gov.au/australia/charts/bulletins/apob93.pdf

Buschow, S., & Friederichs, P. (2021). SAD: Verifying the scale, anisotropy and direction of precipitation forecasts. *Quarterly Journal of the Royal Meteorological Society, 147*(735), 1150-1169. doi:https://doi.org/10.1002/qj.3964

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate, 28*(17), 6938-6959. doi:10.1175/jcli-d-14-00754.1

Chavez, M., & Cazelles, B. (2019). Detecting dynamic spatial correlation patterns with generalized wavelet coherence and non-stationary surrogate data. *Sci Rep, 9*(1), 7389. doi:10.1038/s41598-019-43571-2

Clark, M. P., & Hay, L. E. (2004). Use of Medium-Range Numerical Weather Prediction Model Output to Produce Forecasts of Streamflow. *Journal of Hydrometeorology, 5*(1), 15-32. doi:10.1175/1525-7541(2004)005<0015:UOMNWP>2.0.CO;2

Cuo, L., Pagano, T. C., & Wang, Q. J. (2011). A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting. *Journal of Hydrometeorology, 12*(5), 713-728. doi:10.1175/2011JHM1347.1

Daubechies, I. (1992). *Ten lectures on wavelets*. Philadelphia, Pa.: Philadelphia, Pa. : Society for Industrial and Applied Mathematics.

de Roo, A. P. J., Gouweleeuw, B., Thielen, J., Bartholmes, J., Bongioannini‐Cerlini, P., Todini, E., . . . Sprokkereef, E. (2003). Development of a European flood forecasting system. *International Journal of River Basin Management, 1*(1), 49-59. doi:10.1080/15715124.2003.9635192

Farge, M. (1992). Wavelet Transforms and their Applications to Turbulence. *Annual Review of Fluid Mechanics, 24*, 395-457.

Feldmann, K., Scheuerer, M., & Thorarinsdottir, T. L. (2015). Spatial Postprocessing of Ensemble Forecasts for Temperature Using Nonhomogeneous Gaussian Regression. *Monthly weather review, 143*(3), 955-971. doi:10.1175/MWR-D-14-00210.1

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

Fossa, M., Dieppois, B., Massei, N., Fournier, M., Laignel, B., & Vidal, J. P. (2021). Spatiotemporal and cross-scale interactions in hydroclimate variability: a case-study in France. *Hydrol. Earth Syst. Sci., 25*(11), 5683-5702. doi:10.5194/hess-25-5683-2021

Gilleland, E., Hering, A. S., Fowler, T. L., & Brown, B. G. (2018). Testing the Tests: What Are the Impacts of Incorrect Assumptions When Applying Confidence Intervals or Hypothesis Tests to Compare Competing Forecasts? *Monthly weather review, 146*(6), 1685-1703. doi:10.1175/mwr-d-17-0295.1

Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. (2005). Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly weather review, 133*(5), 1098-1118. doi:10.1175/MWR2904.1

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrol. Earth Syst. Sci., 16*(9), 3383-3390. doi:10.5194/hess-16-3383-2012

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology, 377*(1-2), 80-91. doi:10.1016/j.jhydrol.2009.08.003

Hamill, T. M. (2001). Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly weather review, 129*(3), 550-560. doi:10.1175/1520-0493(2001)129<0550:Iorhfv>2.0.Co;2

Hamill, T. M., Engle, E., Myrick, D., Peroutka, M., Finan, C., & Scheuerer, M. (2017). The U.S. National Blend of Models for Statistical Postprocessing of Probability of Precipitation and Deterministic Precipitation Amount. *Monthly weather review, 145*(9), 3441-3463. doi:10.1175/MWR-D-16-0331.1

Hamill, T. M., & Scheuerer, M. (2018). Probabilistic Precipitation Forecast Postprocessing Using Quantile Mapping and Rank-Weighted Best-Member Dressing. *Monthly weather review, 146*(12), 4079-4098. doi:10.1175/mwr-d-18-0147.1

He, Y., Wetterhall, F., Cloke, H. L., Pappenberger, F., Wilson, M., Freer, J., & McGregor, G. (2009). Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteorological Applications, 16*(1), 91-101. doi:10.1002/met.132

Hu, W., & Si, B. (2021). Technical Note: Improved partial wavelet coherency for understanding scale-specific and localized bivariate relationships in geosciences. *Hydrol. Earth Syst. Sci., 25*(1), 321-331. doi:10.5194/hess-25-321-2021

Huang, L., & Luo, Y. (2017). Evaluation of quantitative precipitation forecasts by TIGGE ensembles for south China during the presummer rainy season. *Journal of Geophysical Research: Atmospheres, 122*(16), 8494-8516. doi:https://doi.org/10.1002/2017JD026512

Jankov, I., Gregory, S., Ravela, S., Toth, Z., & Peña, M. (2021). Partition of Forecast Error into Positional and Structural Components. *Advances in Atmospheric Sciences, 38*(6), 1012-1019. doi:10.1007/s00376-021-0251-7

Johnson, F., & Sharma, A. (2012). A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Water Resources Research, 48*(1).

Kusumastuti, C., Jiang, Z., Mehrotra, R., & Sharma, A. (2021). A Signal Processing Approach to Correct Systematic Bias in Trend and Variability in Climate Model Simulations. *Geophysical Research Letters, 48*(13), e2021GL092953. doi:https://doi.org/10.1029/2021GL092953

Kusumastuti, C., Jiang, Z., Mehrotra, R., & Sharma, A. (2022). Correcting Systematic Bias in Climate Model Simulations in the Time-Frequency Domain. *Geophysical Research Letters, 49*(19), e2022GL100550. doi:https://doi.org/10.1029/2022GL100550

Lane, S. N. (2007). Assessment of rainfall-runoff models based upon wavelet analysis. *Hydrological Processes, 21*(5), 586-607. doi:https://doi.org/10.1002/hyp.6249

Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., & Di, Z. (2017). A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *WIREs Water, 4*(6), e1246. doi:https://doi.org/10.1002/wat2.1246

Mehrotra, R., & Sharma, A. (2015). Correcting for systematic biases in multiple raw GCM variables across a range of timescales. *Journal of Hydrology, 520*, 214-223. doi:10.1016/j.jhydrol.2014.11.037

Nguyen, H., Mehrotra, R., & Sharma, A. (2016). Correcting for systematic biases in GCM simulations in the frequency domain. *Journal of Hydrology, 538*, 117-126.

Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis* (Vol. 4): Cambridge university press.

Perkins, S., Pitman, A., Holbrook, N., & McAneney, J. (2007). Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *Journal of Climate, 20*(17), 4356-4376.

Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly weather review, 133*(5), 1155-1174. doi:10.1175/MWR2906.1

Robertson, D. E., Shrestha, D. L., & Wang, Q. J. (2013). Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrol. Earth Syst. Sci., 17*(9), 3587-3603. doi:10.5194/hess-17-3587-2013

Roff, G., Bermous, I., Dietachmayer, G., Fernon, J., Fraser, J., Lu, W., . . . Xiao, Y. (2022). APS2-ACCESS-C2: the first Australian operational NWP convection-permitting model. *Journal of Southern Hemisphere Earth Systems Science, 72*(1), 1-18.

Sang, Y. F. (2013). A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric Research, 122*, 8-15. doi:10.1016/j.atmosres.2012.11.003

Schaefli, B., Maraun, D., & Holschneider, M. (2007). What drives high flow events in the Swiss Alps? Recent developments in wavelet spectral analysis and their application to hydrology. *Advances in Water Resources, 30*(12), 2511-2525. doi:https://doi.org/10.1016/j.advwatres.2007.06.004

Accepted for publication in *Monthly Weather Review*. DOI 10.1175/MWR-D-22-0217.1

Scheuerer, M., & Büermann, L. (2014). Spatially adaptive post-processing of ensemble forecasts for temperature. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 63*(3), 405-422. doi:https://doi.org/10.1111/rssc.12040

Scovell, R. W. (2020). Applications of directional wavelets, Universal Multifractals and anisotropic scaling in ensemble nowcasting; a review of methods with case-studies. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 2066-2095. doi:https://doi.org/10.1002/qj.3780

Shahrban, M., Walker, J. P., Wang, Q. J., Seed, A., & Steinle, P. (2016). An evaluation of numerical weather prediction based rainfall forecasts. *Hydrological Sciences Journal, 61*(15), 2704-2717. doi:10.1080/02626667.2016.1170131

Shrestha, D. L., Robertson, D. E., Wang, Q. J., Pagano, T. C., & Hapuarachchi, H. A. P. (2013). Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose. *Hydrol. Earth Syst. Sci., 17*(5), 1913-1931. doi:10.5194/hess-17-1913-2013

Themeßl, M. J., Gobiet, A., & Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change, 112*(2), 449-468. doi:10.1007/s10584-011-0224-4

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society, 79*(1), 61-78. doi:10.1175/1520-0477(1998)079<0061:Apgtwa>2.0.Co;2

Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., . . . Ylhaisi, J. (2021). Statistical Postprocessing for Weather Forecasts: Review, Challenges, and Avenues in a Big Data World. *Bulletin of the American Meteorological Society, 102*(3), E681-E699. doi:10.1175/bams-d-19-0308.1

Verkade, J. S., Brown, J. D., Reggiani, P., & Weerts, A. H. (2013). Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology, 501*, 73-91. doi:https://doi.org/10.1016/j.jhydrol.2013.07.039

Veronica, J. B., Adrian, E. R., & Tilmann, G. (2008). Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *The Annals of Applied Statistics, 2*(4), 1170-1193. doi:10.1214/08-AOAS203

Wang, Q. J., Zhao, T., Yang, Q., & Robertson, D. (2019). A Seasonally Coherent Calibration (SCC) Model for Postprocessing Numerical Weather Predictions. *Monthly weather review, 147*(10), 3633-3647. doi:10.1175/MWR-D-19-0108.1