Spatial mode-based calibration (SMoC) of forecast precipitation fields from numerical weather prediction models

Pengcheng Zhao*, Quan J. Wang, Wenyan Wu, Qichun Yang

Department of Infrastructure Engineering, The University of Melbourne, Parkville 3010, Australia

*: Corresponding author

E-mail address: pengcheng@student.unimelb.edu.au

ABSTRACT

Statistical calibration of precipitation forecasts from numerical weather prediction (NWP) models is routinely performed grid-cell by grid-cell, aiming to produce accurate and reliable ensemble forecasts for precipitation fields. Calibrated ensemble members from different grid-cells are then connected using ensemble reordering to form spatially structured ensemble forecasts. However, ensemble reordering approaches, such as the well-known Schaake shuffle method, are often criticized for not considering real physical atmospheric states of precipitation events. In this paper, we propose a spatial mode-based calibration (SMoC) model for post-processing forecast precipitation fields and producing ensemble forecasts with an inbuilt spatial structure, so that ensemble reordering is not required. The SMoC model is developed based on spatial modes derived from empirical orthogonal function (EOF) analyses of precipitation fields and linear regressions of derived EOF expansion coefficients of the first few dominant modes. Unlike conventional calibration models that are applied to forecast grid-cells individually, SMoC is applied to the whole forecast fields, and the spatial structure is inherently present in calibrated ensemble forecasts. There is therefore no need to rely on ensemble reordering for spatial reconstruction. The performance of SMoC is evaluated by applying it to NWP forecasts of substantive precipitation events over the Brisbane drainage basin in eastern Australia. Cross-validated results show that SMoC calibrated forecasts are of high quality at both grid-cell and basin scales. The spatial structure of precipitation is found to be well embedded in the ensemble members of the calibrated forecasts for the basin.

Keywords: forecast precipitation fields, numerical weather prediction, statistical calibration, spatial structure, empirical orthogonal function, spatial modes.

1. Introduction

Short-term precipitation forecasts are commonly produced using numerical weather prediction (NWP) models in major meteorological centers worldwide. However, current NWP models cannot fully simulate dynamic physics of the atmosphere due to various sources of uncertainty, including model structures, model initial and boundary conditions, and model parameters (Palmer et al., 2007). Consequently, raw forecasts from NWP models are often subject to systematic bias (Gneiting et al., 2007; Li et al., 2017) and can be even less skillful than naïve climatology forecasts (Zhao et al., 2020). Additionally, raw forecasts from ensemble NWP models are often subject to dispersion errors (Buizza, 2018). These deficiencies need to be addressed to improve the usefulness of raw precipitation forecasts.

A number of statistical calibration models have been developed in the past few decades to improve raw precipitation forecasts. Notable models include but are not limited to ensemble model output statistics (EMOS) (Gneiting et al., 2005; Scheuerer, 2014; Baran and Lerch, 2015; Taillardat et al., 2016), Bayesian model averaging (BMA) (Raftery et al., 2005; Sloughter et al., 2007; Schmeits and Kok, 2010; Wang et al., 2012a; Möller et al., 2013; Baran and Möller, 2015), the Bayesian Joint Probability (BJP) model (Robertson et al., 2013; Shrestha et al., 2015; Zhao et al., 2015; Wang et al., 2019a; Cattoën et al., 2020; W. Li et al., 2020; Y. Li et al., 2020; Li et al., 2022), the seasonally coherent calibration (SCC) model (Wang et al., 2019b; Zhao et al., 2020; Yang et al., 2021; Zhao et al., 2022), and machine learning based regression models (Dogulu et al., 2015). These calibration models are typically aimed at producing ensemble forecasts that are free from bias, reliable in ensemble spread, and as skillful as possible.

Current forecast calibration is mostly performed separately at individual grid-cells for forecast precipitation fields. However, the resulting calibrated ensemble members from different grid-cells are not spatially connected, and thus cannot reproduce the spatial structure of corresponding observed precipitation events (Shrestha et al., 2015; Schepen et al., 2020). Consequently, several ensemble reordering approaches have been proposed to connect calibrated ensemble members from different grid-cells, so that the spatial structure can be reconstructed in calibrated forecasts (Schefzik and Möller, 2018; Straaten et al., 2018). The two most popular ensemble reordering approaches are the Schaake shuffle (Clark et al., 2004; Scheuerer et al., 2017) and the ensemble copula coupling (Schefzik et al., 2013; Schefzik, 2017). However, both have drawbacks. The Schaake shuffle, which reorders calibrated ensemble members based on ranks of selected historical observations, is often criticized for not considering real physical atmospheric states of the forecasted precipitation events (Bellier et al., 2017; Scheuerer et al., 2017). Besides, Schaake shuffle becomes problematic when applied to precipitation forecasts, as there are often many zero values in observed precipitation that will result in tied ranks in constructing spatial dependence templates (Clark et al., 2004; Bellier et al., 2017; Wu et al., 2018). ECC takes raw ensembles as templates of spatial dependence, so it is only applicable to raw ensemble forecasts as there is no available template for raw deterministic forecasts. Its usefulness is therefore largely limited by the size of raw ensembles and the quality of raw ensemble structures.

In view of the above-mentioned limitations of existing methods of calibration of forecast precipitation fields, it would be highly valuable if the spatial structure can be inherently embedded in the calibration process, thereby eliminating the need for further spatial reconstruction using ensemble reordering approaches. However, precipitation fields are often high-dimensional in space, and therefore it is difficult to achieve the synchronous calibration for all of the grid-cells in the fields. One possible way is to first reduce the precipitation fields to only a few representative variables, and then calibrate these dimensionally reduced variables.

The empirical orthogonal function (EOF) analysis is the most widely used technique for dimensionality reduction and pattern extraction of spatial-temporal fields (Lorenz, 1956; Hannachi et al., 2007; Monahan et al., 2009; Schmidt et al., 2019). EOF can decompose high-dimensional data into a set of orthogonal spatial modes along with a set of associated uncorrelated expansion coefficients. These two features can give a spatial display and a temporal display of the large high-dimensional meteorological data and provide valuable insights into the main physical characteristics by focusing on a few dominant EOF modes. The EOF technique and its variants have been widely applied in a variety of studies, including meteorology, oceanography, and remote sensing. For example, Barnett (1978) used EOF to examine space-time variations of surface air temperature over the Northern Hemisphere. Servain and Legler (1986) applied EOF to analyze the seasonal variability of monthly sea surface temperature and wind stress within the tropical Atlantic region. Lintner (2002) identified the interannual variability of global carbon dioxide using the first few EOF modes. Wang et al. (2015) studied the Asian summer monsoon precipitation by employing an EOF variant with emphasis on seasonal predictability. And Feng et al. (2016) proposed an EOF-based data integration algorithm to improve the estimation of terrestrial latent heat flux from multiple remote sensing data sources.

In this study, we set out to develop a new approach based on the EOF analysis for post-processing forecast precipitation fields from NWP models and producing ensemble forecasts with an inbuilt spatial structure. For this purpose, we extract representative variables from forecasts and corresponding observations of precipitation fields based on spatial modes derived from the EOF analysis of long-term climatology of observations and establish calibration models for these extracted variables. We call this approach a spatial mode-based calibration (SMoC). In this paper, we introduce the rationale and mathematical formulations of the SMoC model in detail and evaluate its performance through a case study.

2. Model formulation

Here we consider daily deterministic NWP forecast precipitation fields at a single lead time. We aim to calibrate raw forecast fields and produce spatially structured ensemble forecasts that are of high quality at both grid-cell and field scales. The developed spatial mode-based calibration (SMoC) model in this paper involves data normalizations of precipitation data, empirical orthogonal function (EOF) analysis, and calibration of EOF variables. The framework of the SMoC formulation is illustrated in Figure 1 and detailed in Sections 2.1 and 2.2.

[Figure 1]

In this section, we first introduce the EOF analysis by demonstrating its application to long-term observations of precipitation fields. Thereafter, we explain how EOF is used to establish the SMoC model for raw precipitation forecasts and corresponding observations. Finally, we present how we apply SMoC to calibrate new forecast fields. For clear notation, we set up three time variables t, $\tilde{t}$, and $\tilde{\tilde{t}}$ to distinguish three different sets of precipitation data, as shown in Table 1. The long-term observation period ($t = 1,2, \dots, T$) may or may not overlap with the period of forecast data used for model training ($\tilde{t} = 1,2, \dots, \tilde{T}$), where T and $\tilde{T}$ denote the number of days in the long-term observation period and the period of forecast data, respectively. $\tilde{\tilde{t}}$ represents the time of a new forecast event that will be calibrated using the SMoC model in Section 2.3. And $\tilde{\tilde{t}}$ does not overlap with the long-term observation period or the forecast period.

[Table 1]

2.1 EOF analysis of long-term observation fields

We apply the EOF analysis to long-term observation fields to derive representative patterns with long-term statistical characteristics of climatology. Given a gridded precipitation field, long-term observations can be represented as

$$\boldsymbol{Y} = \begin{bmatrix} Y(1,1) & Y(1,2) & \cdots & Y(1,S) \\ Y(2,1) & Y(2,2) & \cdots & Y(2,S) \\ \vdots & \vdots & \vdots & \vdots \\ Y(T,1) & Y(T,2) & \cdots & Y(T,S) \end{bmatrix} \tag{1}$$

where T and S denote the number of days in the long-term observation period and the number of grid-cells in the gridded field, respectively; $Y(t,s)$ in the data matrix $\boldsymbol{Y}$ represents the observed precipitation at day t in grid-cell s ($t = 1,2,\dots,T$ and $s = 1,2,\dots,S$).

2.1.1 Data normalization

Statistical treatment on precipitation values is always challenging due to highly skewed distribution of precipitation data. It has therefore been a common practice to apply data transformation for normalizing precipitation data before formal analysis (Wang et al., 2012b; Wang et al., 2019b; Zhao et al., 2022). To ensure the spatial consistency of data normalizations among different grid-cells, we apply an approach of regionally optimized power transformation as recommended by Du et al. (2022) to transform the long-term precipitation observation fields from $\boldsymbol{Y}$ to $\boldsymbol{Y}_p$. Details of the power transformation can be found in Text S1 and Figure S1 of Supplementary Material S1.

2.1.2 EOF decomposition

As an integral part of the EOF analysis, we center precipitation values (i.e., subtracting the temporal average from the values) for each grid-cell. The centering process assures that the EOF technique only focuses on the variance within the precipitation data, rather than taking the overall average of the data as an important piece of extracted information. For $\boldsymbol{Y}_p$, temporal averages of precipitation values can be calculated as

$$\overline{\boldsymbol{Y}}_p = [\bar{Y}_p(1), \bar{Y}_p(2), \ldots, \bar{Y}_p(S)] \tag{2}$$

where for each grid-cell s, the precipitation average is calculated as

$$\bar{Y}_p(s) = \frac{1}{T}\sum_{t=1}^{T} Y_p(t,s) \tag{3}$$

where $Y_p(t,s)$ is the power transformed observed precipitation at day t in grid-cell s ($t = 1,2,\ldots,T$ and $s = 1,2,\ldots,S$). The centered precipitation field $\boldsymbol{Y}'_p$ can be produced by subtracting $\overline{\boldsymbol{Y}}_p$ from each row of $\boldsymbol{Y}_p$ and then decomposed using EOF as

$$\boldsymbol{Y}'_p = \boldsymbol{C}_{\boldsymbol{Y}}\boldsymbol{U}_{\boldsymbol{Y}} \tag{4}$$

where

$$\boldsymbol{C}_{\boldsymbol{Y}} = \begin{bmatrix} C_{\boldsymbol{Y}}(1,1) & C_{\boldsymbol{Y}}(1,2) & \cdots & C_{\boldsymbol{Y}}(1,K) \\ C_{\boldsymbol{Y}}(2,1) & C_{\boldsymbol{Y}}(2,2) & \cdots & C_{\boldsymbol{Y}}(2,K) \\ \vdots & \vdots & \vdots & \vdots \\ C_{\boldsymbol{Y}}(T,1) & C_{\boldsymbol{Y}}(T,2) & \cdots & C_{\boldsymbol{Y}}(T,K) \end{bmatrix} \tag{5}$$

$$\boldsymbol{U}_{\boldsymbol{Y}} = \begin{bmatrix} U_{\boldsymbol{Y}}(1,1) & U_{\boldsymbol{Y}}(1,2) & \cdots & U_{\boldsymbol{Y}}(1,S) \\ U_{\boldsymbol{Y}}(2,1) & U_{\boldsymbol{Y}}(2,2) & \cdots & U_{\boldsymbol{Y}}(2,S) \\ \vdots & \vdots & \vdots & \vdots \\ U_{\boldsymbol{Y}}(K,1) & U_{\boldsymbol{Y}}(K,2) & \cdots & U_{\boldsymbol{Y}}(K,S) \end{bmatrix} \tag{6}$$

where K is the number of derived EOF modes after the decomposition and generally equals to the smaller dimension of the original spatial-temporal patterns (i.e., the smaller of T and S). Equation (4) can also be written as

$$\boldsymbol{Y}'_p = \sum_{k=1}^{K} \boldsymbol{C}_{\boldsymbol{Y}}(t,k)\boldsymbol{U}_{\boldsymbol{Y}}(k,s) \tag{7}$$

where $\boldsymbol{C_Y}(t,k)$ $(t = 1,2,\ldots,T)$ is a column vector and varies with day t, representing expansion coefficients of the k^{th} EOF mode; $\boldsymbol{U_Y}(k,s)$ $(s = 1,2,\ldots,S)$ is a row vector and varies with grid-cell s, representing spatial modes of the k^{th} EOF mode. In this way, the original gridded long-term observation fields (i.e., $\boldsymbol{Y}$) can be represented using a set of temporal variables (expansion coefficients $\boldsymbol{C_Y}(t,k)$, $t = 1,2,\ldots,T$) and a set of spatial variables (spatial modes $\boldsymbol{U_Y}(k,s)$, $s = 1,2,\ldots,S$).

2.2 Establishing relationships between raw forecasts and observations

For a given k^{th} EOF mode, $\boldsymbol{C_Y}(t,k)$ $(t = 1,2,\ldots,T)$ is a function of day t, which is analogical to the fact that the occurrence of natural precipitation events varies with time in the field. We therefore choose to relate raw precipitation forecasts to corresponding observations by matching their respectively derived expansion coefficients.

2.2.1 Expansion coefficients of raw forecast fields and corresponding observation fields

Raw forecast fields in the period of training forecast data (a.k.a. forecast period from hereon) can be represented as

$$\widetilde{\boldsymbol{X}} = \begin{bmatrix} \tilde{X}(1,1) & \tilde{X}(1,2) & \cdots & \tilde{X}(1,S) \\ \tilde{X}(2,1) & \tilde{X}(2,2) & \cdots & \tilde{X}(2,S) \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{X}(\tilde{T},1) & \tilde{X}(\tilde{T},2) & \cdots & \tilde{X}(\tilde{T},S) \end{bmatrix} \tag{8}$$

where $\tilde{T}$ and S denote the number of days in the forecast period and the number of grid-cells in the gridded field, respectively; $\tilde{X}(\tilde{t},s)$ in the data matrix $\widetilde{\boldsymbol{X}}$ represents the precipitation forecast at day $\tilde{t}$ in grid-cell s ($\tilde{t} = 1,2,\ldots,\tilde{T}$ and $s = 1,2,\ldots,S$). Similar to the processing of long-term observation fields, we apply the data normalization and the centering process to raw forecast fields and obtain $\widetilde{\boldsymbol{X}}'_p$. To derive expansion coefficients of raw forecasts, one way is to simply apply the

EOF decomposition to $\widetilde{\boldsymbol{X}}'_p$. However, two independent EOF analyses performed separately on raw forecasts and corresponding observations will produce different spatial modes. This makes it hard to match the two sets of derived expansion coefficients.

In this study, instead of repeating the EOF decomposition, we employ the spatial modes of the long-term observation fields (i.e., $\boldsymbol{U_Y}$) and a variant of Equation (4) to derive expansion coefficients from $\widetilde{\boldsymbol{X}}'_p$:

$$\boldsymbol{C}_{\widetilde{\boldsymbol{X}}} = \widetilde{\boldsymbol{X}}'_p \boldsymbol{U}_{\boldsymbol{Y}}^{-1} \tag{9}$$

where $\boldsymbol{U}_{\boldsymbol{Y}}^{-1}$ is the inverse of $\boldsymbol{U_Y}$; $\boldsymbol{C}_{\widetilde{\boldsymbol{X}}}$ represents the derived expansion coefficients of raw forecasts. Likewise, for corresponding observation fields in the forecast period

$$\widetilde{\boldsymbol{Y}} = \begin{bmatrix} \tilde{Y}(1,1) & \tilde{Y}(1,2) & \cdots & \tilde{Y}(1,S) \\ \tilde{Y}(2,1) & \tilde{Y}(2,2) & \cdots & \tilde{Y}(2,S) \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{Y}(\tilde{T},1) & \tilde{Y}(\tilde{T},2) & \cdots & \tilde{Y}(\tilde{T},S) \end{bmatrix} \tag{10}$$

where $\tilde{T}$ and S denote the number of days in the forecast period and the number of grid-cells in the gridded field, respectively; $\tilde{Y}(\tilde{t},s)$ in the data matrix $\widetilde{\boldsymbol{Y}}$ represents the observed precipitation at day $\tilde{t}$ in grid-cell s ($\tilde{t} = 1,2,\ldots,\tilde{T}$ and $s = 1,2,\ldots,S$), we apply the same data normalization and centering process to obtain $\widetilde{\boldsymbol{Y}}'_p$. It should be noted that we use the parameters of the data normalization and the centering process from long-term observation fields to normalize and center corresponding observation fields, as these parameters are derived from long-term statistical characteristics of climatology and are more sensible for use. We then derive expansion coefficients from $\widetilde{\boldsymbol{Y}}'_p$ using the spatial modes of the long-term observations:

$$\boldsymbol{C}_{\widetilde{\boldsymbol{Y}}} = \widetilde{\boldsymbol{Y}}'_p \boldsymbol{U}_{\boldsymbol{Y}}^{-1} \tag{11}$$

where $\boldsymbol{U}_{\boldsymbol{Y}}^{-1}$ is the inverse of $\boldsymbol{U}_{\boldsymbol{Y}}$; $\boldsymbol{C}_{\widetilde{\boldsymbol{Y}}}$ represents the derived expansion coefficients of corresponding observations. In this way, $\boldsymbol{C}_{\widetilde{\boldsymbol{X}}}$ and $\boldsymbol{C}_{\widetilde{\boldsymbol{Y}}}$ are both derived from the same spatial modes that represent long-term climatology statistics. This assures that these two expansion coefficients are comparable in a reasonable manner. Similarly, this use of the EOF analysis has been performed for hydrodynamic model inundation simulations in Fraehr et al. (2022), which derives expansion coefficients of low-resolution models using the spatial modes of high-resolution models.

In the EOF analysis, the first few modes are often selected to represent the whole set of original data, without compromising much of the explained data variance. The high-dimensional datasets can therefore be reduced to only a few representative variables. In practice, the number of the selected modes is often determined by specifying a certain amount of the total variance, e.g. 90%, and then finding the minimum number of modes needed to explain this amount of variance. In this study, we select expansion coefficients from the first M EOF modes, i.e., $\boldsymbol{C}_{\widetilde{\boldsymbol{X}}}(\tilde{t}, k)$ and $\boldsymbol{C}_{\widetilde{\boldsymbol{Y}}}(\tilde{t}, k)$ ($\tilde{t} = 1,2, \ldots, \tilde{T}$ and $k = 1,2, \ldots, M$), to represent the whole sets of expansion coefficients, where M is the number of the selected EOF modes. The actual value of M will vary with application. For example, in the case study later shown in Section 3.2, the first 10 modes are chosen, which account for 95% of the long-term observation data variance. In addition, detailed illustrations of the spatial modes and expansion coefficients from the first 10 EOF modes are also provided in Section 3.2 to further demonstrate the EOF analysis of the precipitation field data.

2.2.2 Seasonality removal in expansion coefficients

In establishing calibration models based on the two selected sets of expansion coefficients, we aim to produce ensembles that are not only free from bias, reliable in ensemble spread, and as skillful as possible, but also coherent with a seasonally varying climatology that exists in both forecasts and observations, as well as in their respective expansion coefficients. This is especially important for forecasts at long lead times when forecast skill becomes low and forecasts approach climatology (Wang et al., 2019b). However, it is challenging to produce coherent calibrated ensembles as NWP models tend to be updated regularly and available NWP forecasts to establish coherent calibration models are often limited. To resolve this problem, we choose to remove the seasonality in expansion coefficients and pool the data from all year round together to establish calibration models. When applying the established SMoC model for forecast calibration, we add the removed seasonality back to calibrated expansion coefficients and finally produce coherent calibrated forecasts (see Section 2.3 later).

In this study, we remove the seasonality by using daily climatological mean and daily climatological standard deviation to standardize expansion coefficients of forecasts and corresponding observations. For the k^{th} EOF mode ($k = 1,2,3, \dots, M$), we first employ the spectral method in Narapusetty et al. (2009) to estimate the daily climatological mean $\boldsymbol{\mu}_{\boldsymbol{C}_{\boldsymbol{Y}}(t,k)}$ ($t = 1,2,3, \dots, T$) for expansion coefficients of observations from the long-term period. We then use the daily climatological mean to calculate the squared error of expansion coefficients and employ the spectral method again to estimate the daily climatological standard deviation $\boldsymbol{\sigma}_{\boldsymbol{C}_{\boldsymbol{Y}}(t,k)}$ ($t = 1,2,3, \dots, T$) for expansion coefficients of observations from the long-term period. We apply these two estimated variables to produce daily climatological mean and daily climatological standard deviation for expansion coefficients of observations at day $\tilde{t}$ ($\tilde{t} = 1,2, \dots, \tilde{T}$) during the forecast period (i.e., $\mu_{C_{\tilde{\boldsymbol{Y}}}(\tilde{t},k)}$ and $\sigma_{C_{\tilde{\boldsymbol{Y}}}(\tilde{t},k)}$). The standardization goes as

$$C^*_{\tilde{\boldsymbol{X}}}(\tilde{t},k) = (C_{\tilde{\boldsymbol{X}}}(\tilde{t},k) - \mu_{C_{\tilde{\boldsymbol{Y}}}(\tilde{t},k)})/\sigma_{C_{\tilde{\boldsymbol{Y}}}(\tilde{t},k)} \quad (12)$$

$$C^*_{\tilde{\boldsymbol{Y}}}(\tilde{t},k) = (C_{\tilde{\boldsymbol{Y}}}(\tilde{t},k) - \mu_{C_{\tilde{\boldsymbol{Y}}}(\tilde{t},k)})/\sigma_{C_{\tilde{\boldsymbol{Y}}}(\tilde{t},k)} \quad (13)$$

where $C_{\widetilde{\boldsymbol{X}}}(\tilde{t},k)$ and $C_{\widetilde{\boldsymbol{Y}}}(\tilde{t},k)$ are expansion coefficients of forecasts and corresponding observations for the k^{th} mode at day $\tilde{t}$, respectively; $C^*_{\widetilde{\boldsymbol{X}}}(\tilde{t},k)$ and $C^*_{\widetilde{\boldsymbol{Y}}}(\tilde{t},k)$ are standardized $C_{\widetilde{\boldsymbol{X}}}(\tilde{t},k)$ and $C_{\widetilde{\boldsymbol{Y}}}(\tilde{t},k)$, respectively.

2.2.3 Linear regressions of expansion coefficients

After the standardization, we establish calibration models based on these two sets of expansion coefficients. As expansion coefficients from different EOF modes are independent from each other, we establish calibration models separately for the M modes. Specifically, we fit an ordinary least-square linear regression for each mode. For the k^{th} mode $(k = 1,2,3,\dots,M)$ at day $\tilde{t}$ $(\tilde{t} = 1,2,\dots,\tilde{T})$, the regression can be formulated as

$$C^*_{\widetilde{\boldsymbol{Y}}}(\tilde{t},k) = a(k) \times C^*_{\widetilde{\boldsymbol{X}}}(\tilde{t},k) + b(k) + \varepsilon(\tilde{t},k) \quad (14)$$

where $a(k)$ and $b(k)$ denote linear regression coefficients, and the residual term $\varepsilon(\tilde{t},k)$ follows a normal distribution:

$$\varepsilon(\tilde{t},k) \sim N(0,\sigma_\varepsilon^2(k)) \quad (15)$$

Estimates of $a(k)$, $b(k)$ and $\sigma_\varepsilon(k)$ can be found by using the method of maximum likelihood estimation.

2.3 Model use for calibrating new forecasts

Once parameters of the data normalization, data centering, spatial modes $\boldsymbol{U_Y}$, standardization of expansion coefficients, and linear regressions in Sections 2.1-2.2 are derived from training datasets, we can apply the established SMoC model to calibrate new forecasts. Given a new forecast field $\tilde{\tilde{\boldsymbol{X}}}(\tilde{\tilde{t}},s)$ $(s = 1,2,\dots,S)$ at day $\tilde{\tilde{t}}$, a calibrated ensemble forecast can be produced following the steps shown in Figure 2 and detailed as below.

(a) Apply the data normalization and the centering process to produce $\tilde{\tilde{\boldsymbol{X}}}'_p(\tilde{\tilde{t}}, s)$ $(s = 1,2, \dots, S)$.

(b) Derive expansion coefficients $\boldsymbol{C}_{\tilde{\tilde{\boldsymbol{X}}}}(\tilde{\tilde{t}}, k)$ $(k = 1,2, \dots, K)$ using Equation (9).

(c) Select the first M EOF modes of $\boldsymbol{C}_{\tilde{\tilde{\boldsymbol{X}}}}(\tilde{\tilde{t}}, k)$ $(k = 1,2, \dots, K)$ and remove the seasonality of expansion coefficients to give $\boldsymbol{C}^*_{\tilde{\tilde{\boldsymbol{X}}}}(\tilde{\tilde{t}}, k)$ $(k = 1,2, \dots, M)$ using Equations (12) and (13).

(d) For the k^{th} mode, draw a random $\varepsilon(\tilde{\tilde{t}}, k)$ from Equation (15). Then a predicted $C^*_{\tilde{\tilde{\boldsymbol{Y}}}}(\tilde{\tilde{t}}, k)$ can be obtained using $C^*_{\tilde{\tilde{\boldsymbol{X}}}}(\tilde{\tilde{t}}, k)$ and Equation (14). Results for the M EOF modes are pooled together to produce $\boldsymbol{C}^*_{\tilde{\tilde{\boldsymbol{Y}}}}(\tilde{\tilde{t}}, k)$ $(k = 1,2, \dots, M)$.

(e) Add the removed seasonality back to $\boldsymbol{C}^*_{\tilde{\tilde{\boldsymbol{Y}}}}(\tilde{\tilde{t}}, k)$ $(k = 1,2, \dots, M)$ to give $\boldsymbol{C}_{\tilde{\tilde{\boldsymbol{Y}}}}\left(\tilde{\tilde{t}}, k\right)$ $(k = 1,2, \dots, M)$ using variants of Equations (12) and (13).

(f) Derive $\tilde{\tilde{\boldsymbol{Y}}}'_p(\tilde{\tilde{t}}, s)$ $(s = 1,2, \dots, S)$ using Equation (4):

$$\tilde{\tilde{\boldsymbol{Y}}}'_p(\tilde{\tilde{t}}, s) = \boldsymbol{C}_{\tilde{\tilde{\boldsymbol{Y}}}}\left(\tilde{\tilde{t}}, k\right)\boldsymbol{U}_{\boldsymbol{Y}} \tag{16}$$

(g) Inverse the centering process and the data normalization to $\tilde{\tilde{\boldsymbol{Y}}}'_p(\tilde{\tilde{t}}, s)$ $(s = 1,2, \dots, S)$ and obtain a calibrated forecast field $\tilde{\tilde{\boldsymbol{Y}}}\left(\tilde{\tilde{t}}, s\right)$ $(s = 1,2, \dots, S)$ in the original spatial dimension for day $\tilde{\tilde{t}}$. Negative values in the inverse of data normalization are set to zero to give zero precipitation forecasts.

(h) Repeat steps (d)-(g) N times to produce a calibrated ensemble forecast field $\tilde{\tilde{\boldsymbol{Y}}}\left(\tilde{\tilde{t}}, s, n\right)$ $(s = 1,2, \dots, S$ and $n = 1,2, \dots, N)$ for day $\tilde{\tilde{t}}$, N being the size of the ensemble.

[Figure 2]

Following above steps, raw forecast precipitation fields are calibrated as a whole, not grid-cell by grid-cell. The calibrated ensemble forecast field for one precipitation event contains N ensemble members, each being a forecast ensemble member for the whole precipitation field. The spatial structure is inherently embedded in calibrated ensemble members and therefore there is no need to further apply conventional ensemble reordering approaches for spatial reconstruction.

3. Case study

3.1 Research basin and data

We apply the SMoC model to calibrate NWP forecasts from a case study to demonstrate the model performance. The Brisbane drainage basin located in eastern Australia is selected as the research basin (shown in Figure 3(a)). The basin area is 13,549.2 km^2, and the average annual precipitation in the basin is 866 mm per year.

[Figure 3]

Observed precipitation data are obtained from the Australian Water Availability Project's (AWAP) (Jones et al., 2009) climate datasets, which are commonly considered the best available reference climate data for Australia. AWAP datasets have a horizontal grid spacing of 0.05° x 0.05° and are produced by interpolating rain gauge observations across Australia. Daily observations for a period of 30 years from 3 November 1988 to 2 November 2018 are used as long-term observations for EOF analysis and estimation of long-term statistics of seasonality. Daily observations for a period of 3 years from 3 November 2018 to 2 November 2021 are used as reference data for forecast calibration and forecast evaluation.

Raw precipitation forecasts are generated from the Australian Community Climate and Earth-System Simulator Global 3 (ACCESS-G3) model developed by the Australian Bureau of Meteorology. ACCESS-G3 model produces deterministic NWP forecasts with a horizontal grid spacing of 0.18° (longitude) by 0.12° (latitude). With a forecasting horizon of 10 days, ACCESS-G3 forecasts are issued at 0000 UTC, 0600 UTC, 1200 UTC and 1800 UTC every day and are available at an hourly temporal resolution. ACCESS-G3 forecasts have been operational since 25 July 2019. To minimize possible impacts of sample size on calibration performance, we integrate experimental forecasts (from 3 November 2018 to 24 July 2019) and operational forecasts (from 25 July 2019 to 2 November 2021) to obtain 3 years of forecasts.

To match the daily AWAP data, we adjust the spatial and temporal resolutions of ACCESS-G3 forecasts. We apply bilinear interpolation to re-grid the forecasts to the horizontal grid spacing of observations. And we aggregate the hourly forecasts to daily values by accumulating forecast values according to the daily period on which the AWAP data are produced. Consequently, precipitation forecasts and observations are both provided at a grid spacing of 0.05° x 0.05° and on a daily basis for this study. Brisbane drainage basin covers 493 grid-cells at the 0.05° x 0.05° scale, and average annual precipitation values of the 33 years of AWAP data at these grid-cells are shown in Figure 3(b).

3.2 SMoC model setup

The SMoC model is set up for application to 30 years of long-term observations and 3 years of forecasts and corresponding observations for the case study. The total number of days in the long-term period and the forecast period (i.e., T and $\tilde{T}$) are therefore 10957 and 1096, respectively. The total number of grid-cells in the research basin (i.e., S) is 493. The size of the generated ensemble of calibrated forecasts (i.e., N) is set as 1,000 members. The model is applied to forecasts of day 1 ahead, which represent the best performing forecasts among different lead times.

In the SMoC model establishment, the number of the first few EOF modes (i.e., M) is set as 10. From our case study, the first 10 modes are found to account for 96% of the long-term data variance (shown in Figure S7 from Supplementary Material S1). The large number of no and little precipitation days in the research basin (as typically in most regions) means that there are numerous very small values of expansion coefficients (shown in Figure S6 from Supplementary Material S1). When all these values are used to derive the linear relationships between forecasts and observations, the model fit will be heavily geared towards the no and little precipitation events, neglecting the model fit to the substantive precipitation events, which the public cares about the most (Lerch et al., 2017). Therefore, we choose only to fit linear regressions to expansion coefficients of substantive precipitation events whose corresponding basin average of raw forecasts is beyond a pre-defined threshold in the 3-year forecast period. The threshold is set as 90% quantile of raw forecast basin averages and equals to 5.47 mm per day. And the resulting number of chosen substantive precipitation events is 110.

To help readers understand the EOF analysis and the SMoC model more intuitively, we produce a number of figures (Figures S2-S5 in Supplementary Material S1) to illustrate the spatial modes and expansion coefficients of the precipitation data used in our case study. Specifically, we apply the EOF analysis to decompose 30 years of normalized and centered observations to obtain a set of spatial modes $\boldsymbol{U_Y}$ along with a set of expansion coefficients $\boldsymbol{C_Y}$. The first 10 spatial modes (i.e., the first 10 rows of $\boldsymbol{U_Y}$) are shown in Figure S2, each representing one spatial pattern across the Brisbane drainage basin over the 30-year period. The first 10 expansion coefficients (i.e., the first 10 columns of $\boldsymbol{C_Y}$) are shown in Figure S3, each representing one temporal pattern in the 30-year period for its corresponding spatial mode.

The first 10 expansion coefficients of 3 years of normalized and centered forecasts as well as corresponding observations (i.e., the first 10 columns of $\boldsymbol{C_{\tilde{X}}}$ and $\boldsymbol{C_{\tilde{Y}}}$ that are derived using $\boldsymbol{U_Y}$) are shown in Figures S3-S4. We also select five precipitation events of different magnitudes in the 3-

year period (shown in Figures S4-S5) and investigate if original forecasts and observations can be reconstructed using the first 10 EOF modes of the spatial modes $\boldsymbol{U_Y}$ and expansion coefficients $\boldsymbol{C_{\tilde{X}}}$ and $\boldsymbol{C_{\tilde{Y}}}$. Results in Figure S5 show that the EOF analysis can reconstruct the precipitation fields reasonably well, even with only the first 10 EOF modes, both for light and heavy precipitation events. Therefore, it is suggested that the calibration of raw forecast precipitation fields can be represented by the calibration of raw forecast expansion coefficients using linear regressions in Section 2.2.

3.3 Forecast evaluation

In this study, the performance of SMoC calibrated forecasts is evaluated using a leave-one-event-out cross validation. To assess different aspects of forecast quality over the precipitation field, we use continuous ranked probability score (CRPS) (Hersbach, 2000), probability integral transform (PIT) (Renard et al., 2010), and structure-amplitude-location (SAL) spatial verification (Wernli et al., 2008; Radanovics et al., 2018) to perform a comprehensive evaluation. Furthermore, we implement the forecast evaluation at both grid-cell and basin scales, with the evaluation at grid-cell scale assessing model calibration performance for individual grid-cells and the evaluation at basin scale for the whole basin. Forecasts and corresponding evaluation metrics are presented in Table 2 and detailed in Sections 3.3.1 and 3.3.2.

[Table 2]

3.3.1 Forecast evaluation at grid-cell scale

We apply CRPS and PIT to evaluate forecast skill and forecast reliability of forecasts at each of the 493 grid-cells. CRPS measures the difference between ensemble forecast cumulative distributions and corresponding observations. The average CRPS for calibrated ensemble forecasts at day $i = 1,2, \ldots, I$ is evaluated as

$$\overline{CRPS} = \frac{1}{I} \sum_{i=1}^{I} \int \left\{ F(i,x) - H\big(x - y(i)\big) \right\}^2 dx \qquad (17)$$

where $F(i,x)$ is the forecast cumulative density function, and $y(i)$ is the observation at day i; H is the Heaviside step function that equals 1 if $x - y(i) \geq 0$ and equals 0 otherwise; and I is the number of the evaluated precipitation events.

In practice, we often use CRPS skill score to evaluate the relative forecast skill improvement of forecasts compared to reference forecasts. In this study, we produce a reference ensemble climatology forecast for each substantive precipitation event. For this purpose, we first fit a normal distribution in power transformed space for 30 years of AWAP observations of the month in which the event occurred. We then draw a random sample from the distribution and inverse the power transformation to generate an ensemble climatology forecast for the event. The sample size is also set as 1,000. Then we calculate the average CRPS of reference climatology ensemble forecasts ($\overline{CRPS_{ref}}$) for the evaluated precipitation events, and give a CRPS skill score:

$$CRPS\ skill\ score = \frac{\overline{CRPS_{ref}} - \overline{CRPS}}{\overline{CRPS_{ref}}} \times 100(\%) \qquad (18)$$

The CRPS skill score is positively oriented, with a maximum skill score of 100% indicating perfect forecasts and a skill score of 0% indicating that forecasts have comparable errors to the reference forecasts. A negative skill score indicates that forecasts are less skillful than reference forecasts.

For deterministic precipitation forecasts, CRPS is equivalent to the mean absolute error (MAE):

$$MAE = \frac{1}{I} \sum_{i=1}^{I} |x(i) - y(i)| \qquad (19)$$

where $x(i)$ and $y(i)$ are the forecast and corresponding observation at day i, respectively; and I is the number of the evaluated precipitation events. Similarly, we use CRPS skill score to demonstrate the performance of raw forecasts relative to reference climatology forecasts. To have a fair evaluation, we use the same reference CRPS for raw deterministic forecasts and calibrated ensemble forecasts.

We use PIT to evaluate the reliability of ensemble forecasts (ensemble spread not too narrow or too wide). Statistically, forecast reliability represents the consistency between ensemble forecast probability distributions and the observed frequency of corresponding observations. The PIT of an ensemble forecast at day i is calculated as

$$\pi(i) = F\big(i, x = y(i)\big) \tag{20}$$

where $F(i,x)$ is the forecast cumulative density function, and $y(i)$ is the corresponding observation. For a set of reliable forecasts, $\pi(i)$ follows a uniform distribution, and the uniformity can be examined by the well-known PIT alpha index:

$$PIT\ alpha\ index = 1 - \frac{2}{I}\sum_{i=1}^{I}\left|\pi^*(i) - \frac{i}{I+1}\right| \tag{21}$$

where $\pi^*(i)$ is the sorted $\pi(i)$ in an increasing order; and I is the number of the evaluated precipitation events. PIT alpha index is also positively oriented, and ranges from 0 (poorest reliability) to 1 (perfect reliability).

3.3.2 Forecast evaluation at basin scale

We evaluate basin average forecasts to give an overall assessment of forecast precipitation fields. For raw forecasts, we calculate the basin average of forecasts and corresponding observations. Then we calculate CRPS skill score of raw basin average forecasts. For calibrated ensemble forecasts, we calculate the basin average of each ensemble member to produce a basin average ensemble forecast for each evaluated precipitation event. Then we calculate CRPS skill score of calibrated basin average ensemble forecasts. Likewise, the reference ensemble climatology forecast used for calculating CRPS skill score of each event is sampled from the distribution fitted for 30-year basin average observations of the month in which the event occurred. Besides calculating the PIT alpha index of calibrated basin average ensemble forecasts, we also apply the PIT uniform probability plot to visualize the forecast reliability, where sorted PIT values $\pi^*(i)$ of all evaluated precipitation events are plotted against corresponding theoretical quantiles of the uniform distribution. Reliable forecasts will display a scatter plot close to the 1:1 line.

To indicate whether SMoC calibrated ensemble forecasts are spatially correlated in an appropriate way across the grid-cells, we evaluate reliability of ensemble spreads of basin average precipitation forecasts. The rationale is as follows. When forecast spatial correlation is low, the ensemble spreads of basin average forecasts will be narrow because much of the randomness across the grid-cells will be averaged out. On the contrary, when forecast spatial correlation is high, the ensemble spreads of basin average precipitation forecasts will be wide because there is less randomness across the grid-cells in individual ensemble members. Therefore, when ensemble spreads of forecasts of individual grid-cells and basin average are both reliable, appropriate spatial correlation is strongly indicated.

We also apply SAL spatial verification to evaluate the structure ("S"), amplitude ("A"), and location ("L") of precipitation forecasts for the research basin. SAL is an object-based and dimensionless quality measure. Formulas for calculating these three SAL components are presented in Appendix A. "S" measures the size and shape of precipitation objects, assessing whether the predicted precipitation objects are too large (positive "S") or too small (negative "S").

"S" ranges from -2 to 2 and a "S" value of 0 indicates perfect forecasts. "A" measures the relative deviation of the basin average forecasts from corresponding observations. Positive (negative) "A" values indicate that forecasts overestimate (underestimate) the basin average precipitation. Similarly, "A" ranges from -2 to 2 and an "A" value of 0 indicates perfect forecasts. "L" measures the displacement of the center of mass of predicted precipitation field and the error in the weighted-average distance of the individual precipitation objects from the total precipitation field's center of mass. "L" ranges from 0 to 2, with 0 indicating perfect forecasts and positive values indicating spatial forecasts that have location errors.

In addition, we plot spatial precipitation maps for several selected ensemble members of calibrated ensemble forecasts to check whether calibrated ensemble members are capable of capturing the intensity and spatial features of observed precipitation in the research basin. Specifically, we select three extreme precipitation events in terms of either raw forecasts or observations or both, to investigate the SMoC model performance in different forecasting situations. To facilitate visualization, we choose to plot 10 ensemble members for each calibrated ensemble forecast, corresponding to 10 equally divided quantiles from 5% to 95% based on basin averages of the 1,000 calibrated ensemble members for the whole research basin.

4. Results

4.1 Forecast evaluation at grid-cell scale

4.1.1 Forecast skill

Results of CRPS skill score for raw forecasts and calibrated ensemble forecasts at individual grid-cells of the research basin are shown in Figure 4. To have a fair comparison, we also provide calibrated ensemble median forecasts as deterministic-style forecasts for comparing with raw deterministic forecasts, as MAE is found to work better for ensemble median than for ensemble mean (Gneiting, 2011). It can be seen from the figure that raw forecasts overall perform the worst among these three forecasts in terms of forecast skill. There are negative CRPS skill score values of raw forecasts across most grid-cells, especially for grid-cells in the middle and southeast areas of the basin. By contrast, the SMoC model markedly improves the forecast skill of raw forecasts, with the average of CRPS skill scores across all of the grid-cells increasing from -7.21% to 19.25% (calibrated ensemble median forecasts) and 40.29% (calibrated ensemble forecasts). The calibrated forecasts have positive CRPS skill score values at all of the grid-cells, suggesting that they are more skillful than reference climatology forecasts.

[Figure 4]

In addition, calibrated ensemble forecasts have clearly greater CRPS skill scores than calibrated ensemble median forecasts. This is due to the benefits of the ensemble spread information included in calibrated ensemble forecasts. And this highlights the advantages of forecast calibration models that can produce ensembles even when taking raw deterministic forecasts as model inputs (Schaake et al., 2007; Shrestha et al., 2015).

4.1.2 Forecast reliability

Results of PIT alpha index for calibrated ensemble forecasts at individual grid-cells of the research basin are shown in Figure 5. It is obvious that PIT alpha index values at almost all of the grid-cells are close to 1, with the average across all of the grid-cells being 0.9491. Therefore, the ensemble spread of calibrated ensemble forecasts at grid-cell scale is overall reliable in the research basin. This indicates that the assumed normal distributions of residuals in the linear regressions can reliably quantify the uncertainty of predicted expansion coefficients as well as the uncertainty of calibrated ensemble forecasts.

[Figure 5]

In addition, we calculate the root mean square errors of SMoC ensemble mean forecasts and plot them against the square root values of average SMoC ensemble variance over the 110 substantive precipitation events, both for grid-cell and basin scales, as shown in Figure S11 from Supplementary Material S1. It is found that SMoC calibrated ensemble forecasts in grid-cell scale perform well in terms of the well-known "spread-error correlation" (Whitaker and Loughe, 1998; Fortin et al., 2014; Van Schaeybroeck and Vannitsem, 2016) as well as the ensemble forecast reliability. However, it should be noted that the ensemble forecasts tend to be over-dispersed, especially for grid-cells with large ensemble forecast variance.

4.2 Forecast evaluation at basin scale

For forecasts at basin scale, we first evaluate the performance of basin average forecasts. Figure 6 shows 90% and 50% ensemble intervals, and ensemble medians of calibrated basin average ensemble forecasts against raw basin average forecasts for substantive precipitation events in the research basin. Corresponding observations are also provided for visual forecast verification. The width of calibrated basin average ensemble intervals generally increases with raw basin average forecasts. This is in line with the well-known consensus that forecast uncertainty on precipitation events increases with precipitation amounts. Calibrated ensemble medians are shown to be overall closer to corresponding observations than raw forecasts, especially for the several extremely substantive precipitation events on the far right of Figure 6. Statistically, the distribution of observations is visually consistent with the shown 90% and 50% ensemble forecast intervals. To further assess the quality of basin average forecasts, we provide evaluation results on forecast skill and forecast reliability in Section 4.2.1.

[Figure 6]

4.2.1 Forecast skill and forecast reliability of basin average forecasts

Results of CRPS skill score for basin average forecasts are summarized in Table 3. Similar to the evaluation at grid-cell scale, in the order of raw forecasts, calibrated ensemble median forecasts, and calibrated ensemble forecasts, CRPS skill scores are found to increase successively, indicating gradual forecast skill improvements. For reliability of calibrated basin average ensemble forecasts, PIT alpha index is calculated as 0.9692 and PIT uniform probability plot is close to the uniform distribution (Figure 7), showing that these ensemble forecasts have good performance in forecast reliability. The "spread-error correlation" results in Figure S11 from Supplementary Material S1 also show that SMoC calibrated basin average ensemble forecasts in basin scale perform well in forecast reliability, although there is a potential over-dispersion issue of ensemble forecasts.

Together with the forecast reliability results presented in Section 4.1.2, it can be concluded that calibrated ensemble forecasts are reliable at both grid-cell and basin scales. This indicates that calibrated ensemble members from different grid-cells are connected in an appropriate way so that basin average forecasts are reliable in ensemble spread. Spatial correlation is well embedded in SMoC calibrated forecasts of the whole research basin.

[Table 3]

[Figure 7]

4.2.2 SAL verification of forecast precipitation fields

Results of SAL verification for substantive precipitation forecasts of the whole research basin are shown in Figure 8. In addition to raw forecasts and calibrated ensemble forecasts, we also evaluate the SAL of climatology forecasts as a reference. Forecasts of high quality are illustrated as small dots in the center of the SAL diagram. Most climatology forecasts are located in the third quadrant of the diagram where both structure and amplitude are underestimated. This is because climatology forecasts are produced based on historical precipitation events with all magnitudes of precipitation amounts. When such climatology forecasts are used to predict substantive precipitation events, they tend to provide light forecasts and therefore underestimate the size of precipitation objects and the capacity of precipitation amounts.

[Figure 8]

Most of raw forecasts and calibrated ensemble forecasts, however, are found in the first quadrant of the diagram where both structure and amplitude are overestimated. Raw forecasts tend to overestimate the precipitation amounts by producing too large precipitation objects, while calibrated ensemble forecasts, though with a similar tendency, have weaker overestimation issues, especially for structure. For dots in top-right corner of the diagram, substantive precipitation is predicted in terms of structure and amplitude, but little precipitation is observed. These cases are generally considered as false alarms. By contrast, there are few blue and green dots in the bottom-left corner of the diagram, indicating that the raw and calibrated forecasts rarely miss the prediction of substantive precipitation events.

Furthermore, calibrated ensemble forecasts also have smaller location errors than raw forecasts and climatology forecasts. Dots with the largest sizes are mainly found at the top-right corner of the diagram. Green dots have overall smaller sizes than blue dots, indicating that calibrated ensemble forecasts perform better than raw forecasts in terms of the largest location errors. To sum up, calibrated ensemble forecasts have the best forecast quality among these three forecasts in terms of the SAL verification.

4.2.3 Spatial visualization plots of forecast examples

Figures 9-11 provide visualization of forecast examples from three selected extreme precipitation events, corresponding to three different forecast performances of calibrated ensemble forecasts. Each example contains a raw forecast, the corresponding observation, and 10 selected calibrated ensemble members. Overall, ensemble members from Figures 9-11 have smooth precipitation surfaces across the research basin, meaning that there are similar predicted precipitation amounts in adjacent grid-cells. This is an important feature in correctly structured precipitation forecasts for a spatial field.

[Figure 9]

In Figure 9, extreme precipitation is forecasted by raw forecast but not observed in the basin. Although the raw forecast clearly overestimates precipitation amounts, we still obtain several calibrated ensemble members (e.g., quantiles of 25% and 35%) that can predict a similar amount of precipitation to the observation for the whole basin. However, calibrated ensemble members with large precipitation (i.e., high quantiles) should not be considered useless, as extreme precipitation events might occur with any magnitude of forecasted precipitation amounts, especially with high magnitudes predicted in raw forecasts. This will be further illustrated with the following two forecast examples.

[Figure 10]

In Figure 10, extreme precipitation is not forecasted by raw forecast but observed in the research basin. It can be seen that there is intense observed precipitation in the northeast of the research basin. However, raw forecast fails to capture the large amount of observed precipitation and its spatial distribution. This is a typical case that raw forecasts cannot successfully predict extreme precipitation events, which might lead to severe flooding events when people have no precautions. Fortunately, SMoC model can generate more forecast possibilities even when only relatively light precipitation is predicted in raw forecasts. For example, the two calibrated ensemble members of 85% and 95% quantiles are approximately able to predict the intensity of the observed precipitation event. Indeed, it is crucial for forecast users to be aware of possible extreme precipitation events in flooding risk management (Wu et al., 2020).

[Figure 11]

In Figure 11, extreme precipitation is forecasted by raw forecast and also observed in the research basin. However, there are differences between the raw forecast and the observation in terms of the intensity and spatial distribution of precipitation. For example, the raw forecast has intense precipitation amounts in the west where the intensity is much lower for the observed event. By contrast, many calibrated ensemble members (e.g., quantiles of 35%, 45%, and 55%) are shown to be capable of predicting the intensity of the observed precipitation. This is especially the case for the ensemble member of 35% quantile, which has similar spatial patterns to the observed precipitation event.

In addition to the displayed quantiles (from 5% to 95%) of basin averages of SMoC calibrated ensemble members, we provide a few more calibrated ensemble members for forecast visualization by using animation, as shown in Animations S1-S3 from Supplementary Material S1. Apart from the basin-scale quantiles, we also produce spatial precipitation plots based on the grid-cell-scale quantiles of calibrated ensemble members for the three selected forecast examples, as shown in Figures S8-S10 from Supplementary Material S1. The quantiles are also set to 10 equally divided values from 5% to 95% but applied to the 1,000 calibrated ensemble members from each grid-cell. Different from Figures 9-11 that display multiple quantiles of calibrated ensemble members for the whole research basin, Figures S8-S10 offer insight into the forecast uncertainty information present in individual grid-cells. SMoC calibrated ensemble members are found to contain diverse forecast possibilities at both basin and grid-cell scales.

5. Discussion

In the case study, we select the first 10 EOF modes that account for over 95% of the data variance to represent all the 493 modes. To investigate the influence of the number of selected EOF modes on the SMoC model performance, we conduct further experiments using some other numbers of EOF modes. Results of CRPS skill score and PIT alpha index for grid-cell-scale calibrated forecasts and calibrated basin average forecasts are shown in Figure S12 (from Supplementary Material S1). It can be seen that CRPS skill scores do not change much with the number of EOF modes used, especially for numbers larger than 2. By contrast, PIT alpha index values increase dramatically with the increase in the number of EOF modes used initially, and then have only small fluctuations when the number of EOF modes is larger than 9. These results indicate that we can achieve the highest possible forecast skill with only a few EOF modes. However, a relatively large number of EOF modes is required to reliably quantify the forecast uncertainty.

In our analysis, we standardize the EOF expansion coefficients prior to linear regressions by removing the embedded seasonality, so that we can pool values from different months together to establish the SMoC model and produce coherent calibrated forecasts. As a comparison, we also establish a calibration model without the standardization of expansion coefficients. Results of CRPS skill score and PIT alpha index for grid-cell-scale calibrated forecasts and calibrated basin average forecasts are summarized in Table S1 (from Supplementary Material S1). It can be found that the standardization of expansion coefficients does not have a significant impact on forecast reliability but leads to higher forecast skill of calibrated ensemble forecasts, at both grid-cell and basin scales. The removal of seasonality also helps demonstrate the true correspondence between expansion coefficients of forecasts and corresponding observations (Zhao et al., 2020). This facilitates the establishment of a robust calibration model and improves the performance of calibrated forecasts.

The EOF analysis is well known as an effective tool to perform dimension reduction for climate data, typically for Gaussian-distributed data such as temperature variables (Barnett, 1978; Servain and Legler, 1986). However, for climate data that are not Gaussian-distributed such as the daily precipitation investigated in this study, the use of EOF for dimension reduction may lead to loss of useful data information (Lim et al., 2012). To alleviate this issue, we apply power transformations to normalize precipitation data before the EOF analysis and standardize EOF expansion coefficients before the Gaussian-based linear regressions. The EOF analysis is found to perform well for this study according to the fairly good representation of precipitation field data (shown in Figure S5 and Figure S7 from Supplementary Material S1) and the remarkable calibration results of the SMoC model. Despite these performances, it would be of interest to further investigate the possible effects of EOF on the loss of precipitation data information. It would also be worthwhile to try out other dimension reduction methods that are suitable for non-Gaussian-distributed data, such as the independent component analysis (ICA) (Lim et al., 2012).

We acknowledge that the SMoC model is only used to calibrate forecasts of substantive precipitation events in our case study. Theoretically, if we have a long record of NWP forecasts, it will be more sensible to establish multiple calibration models respectively for events with different magnitudes of precipitation (Li et al., 2019; Wang et al., 2019b). However, the archived record of forecasts is commonly short as NWP models are frequently updated and there are only a limited number of NWP models producing hindcasts. The developed SMoC model might be further improved using longer archived NWP forecasts and could be extended to calibrate events with other magnitudes of precipitation. Indeed, extending the SMoC model for calibration of non-substantive precipitation events in the research basin will be investigated in our future work.

SMoC calibrated ensemble forecasts are found to be reliable in ensemble spread. However, it should be noted that the produced ensemble members tend to be over-dispersed to some extent, as shown in Figure S11 from Supplementary Material S1. In other words, these ensemble forecasts are not sharp in terms of ensemble spread. According to the well-known paradigm of forecast calibration in Gneiting et al. (2007), the sharpness of predictive ensembles should be maximized subject to forecast reliability. Indeed, this issue could potentially be alleviated by using a longer record of archived NWP forecasts, in which case we will be able to establish a calibration model specifically for those extremely substantive events. In this way, the residuals of linear regressions for expansion coefficients can exclusively apply to the fitted extreme precipitation events, which helps to reduce the forecast uncertainty information and thus enlarges the sharpness of calibrated ensemble forecasts.

In this study, the ordinary least-square linear regression is employed to establish statistical calibration models for forecast expansion coefficients and is found to perform well in producing high-quality calibrated ensemble forecasts. Indeed, this classical linear regression, also named model output statistics (MOS), has been extended to incorporate many complexities for accommodating a variety of objectives of forecast calibration (Van Schaeybroeck and Vannitsem, 2011). For example, the Error-in-Variable MOS (EVMOS) method (Vannitsem, 2009) was developed to take into account the errors in both observations and forecasts, unlike the original MOS method that only considers the errors in observations. Consequently, EVMOS was found to produce more appropriate variabilities for calibrated forecasts, even at long lead times. Other extended linear regression methods include the Time-Dependent Tikhonov Regularization (TDTR) (Golub and Loan, 1996), the Total least-square (TLS) (Huffel and Vandewalle, 1991; Golub and Loan, 1996), and the Geometric Mean (GM) (Teissier, 1948). It would be valuable to further investigate if the uses of these extended linear regression methods and some other methods like machine learning (Dogulu et al., 2015) can produce improved model performance compared to the ordinary least-square linear regression used in the current SMoC model.

The SMoC model is innovatively developed to calibrate forecast precipitation fields as a whole, rather than calibrating univariate forecasts separately for each grid-cell as in conventional calibration models. The main advantage is that calibrated ensemble forecasts are produced with an inbuilt spatial structure, and there is no need to further employ ensemble reordering to connect calibrated ensemble members from different grid-cells. And forecast post-processing no longer suffers from the aforementioned drawbacks of ensemble reordering approaches (e.g., the Schaake shuffle and ECC). As SMoC calibrated ensemble forecasts in our case study have shown good performance, it would be highly valuable to compare the SMoC model to conventional calibration models along with ensemble reordering approaches in future study.

In addition to the spatial structure across different grid-cells, calibrated ensemble forecasts should also have a correct temporal structure across different forecasting lead times (Clark et al., 2004), especially for adjacent ones. Substantive precipitation events often occur at consecutive days, and so does light precipitation events. In practice, meteorological centers often issue precipitation forecasts for several lead times. Raw forecasts are produced from different temporal phases of the evolution of simulated atmospheric states in NWP models and are therefore temporally correlated. The SMoC model is currently capable of calibrating raw forecasts of different lead times individually (e.g., forecasts of day 1 ahead in our case study). In our future work, we will apply SMoC to calibrate raw precipitation forecasts of other lead times, especially long lead times, and further investigate how to connect SMoC calibrated ensemble members from different lead times in a sensible way.

6. Summary and conclusions

In this paper, a spatial mode-based calibration (SMoC) model for calibrating forecast precipitation fields from NWP models is presented. We aim to produce calibrated ensemble forecasts that are of high quality at both grid-cell and field scales, and crucially, with a correct spatial structure. The SMoC model is developed based on the EOF analysis and linear regressions. Representative EOF variables are derived from forecasts and corresponding observations using the same spatial modes that represent long-term climatology statistics, so that forecasts and observations are related in a sensible manner. Linear regression models are established for calibrating the extracted EOF variables, with distributions of regression residuals provided as uncertainty information to produce ensemble forecasts.

SMoC is an innovative model developed to post-process forecasts from multiple grid-cells as a whole and produce calibrated ensemble forecasts with an inbuilt spatial structure. By contrast, conventional post-processing of forecast fields typically requires two steps, i.e., statistical calibration of raw forecasts separately for individual grid-cells and ensemble reordering of calibrated ensemble members from different grid-cells. The SMoC model turns the two-step post-processing to only one step and avoids a few well-known drawbacks of ensemble reordering approaches.

The performance of SMoC is evaluated by applying it to forecasts of substantive precipitation events over the Brisbane drainage basin in eastern Australia. A number of evaluation measures are used for verification of calibrated ensemble forecasts, including CRPS, PIT, SAL verification, and also spatial visualization plots. Cross-validated results show that calibrated forecasts are reliable in ensemble spread and have much improved forecast skill compared to raw forecasts, at both grid-cell and basin scales. Calibrated ensemble members from different grid-cells are shown to be spatially correlated appropriately. Calibrated forecasts are also found to outperform both raw forecasts and climatology forecasts in terms of the structure, amplitude, and location of substantive precipitation forecasts for the whole research basin.

Future work of the SMoC model is being undertaken on the improvement of SMoC using other dimension reduction and statistical regression methods, the application of SMoC for calibrating forecasts of non-substantive precipitation events and forecasts of long lead times, the maximization of the sharpness of calibrated ensemble forecasts, the comparison between SMoC and conventional calibration models along with ensemble reordering approaches, the construction of the temporal structure among calibrated ensemble forecasts from different lead times, and the extension of SMoC to calibrate forecast fields of other weather variables.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

Data used in this study are produced by the Australian Bureau of Meteorology and accessed via the National Computational Infrastructure system. Please contact the Bureau of Meteorology to request data access.

Acknowledgments

This work is funded by an Australian Research Council Linkage Project (Grant No. LP170100922) and a collaborative project (Grant No. TP707466) between the University of Melbourne and Australian Bureau of Meteorology. Wenyan Wu acknowledges support from the Australian Research Council via the Discovery Early Career Researcher Award (DE210100117). We thank the Australian Bureau of Meteorology for supplying the ACCESS-G3 and AWAP precipitation data. We gratefully acknowledge the editor and the two reviewers for their thorough reviews and constructive comments.

Appendix A. Structure-amplitude-location (SAL) spatial verification

We employ an object-based spatial verification, the SAL method (Wernli et al., 2008; Radanovics et al., 2018), to compare the structure (“S”), amplitude (“A”), and location (“L”) components between precipitation forecasts and corresponding observations. Considering a precipitation field that contains a number of grid-cells, precipitation objects are identified as contiguous grid-cells where precipitation amount is above a given threshold. The identification of objects is implemented separately for forecast and observed precipitation fields. Following an approach in Wernli et al. (2009), we determine the threshold as $1/15$ of the 95th percentile of all grid-cell precipitation amounts in the field larger than 0.2 mm/day. Here we display the formulas of SAL for deterministic forecasts of the precipitation field.

In this study, given a forecast precipitation field $\widetilde{\boldsymbol{X}}(\tilde{t},s)$ $(s = 1,2,\dots,S)$ and a corresponding observed precipitation field $\widetilde{\boldsymbol{Y}}(\tilde{t},s)$ $(s = 1,2,\dots,S)$ at day $\tilde{t}$, the three components “A”, “L”, and “S” are calculated in order from easy to difficult as follows.

(1) The amplitude component (“A”)

The amplitude component corresponds to the relative difference in field average precipitation between the forecast field and the observation field.

$$A = 2\left(\overline{\widetilde{\boldsymbol{X}}(\tilde{t},s)} - \overline{\widetilde{\boldsymbol{Y}}(\tilde{t},s)}\right)/\left(\overline{\widetilde{\boldsymbol{X}}(\tilde{t},s)} + \overline{\widetilde{\boldsymbol{Y}}(\tilde{t},s)}\right) \tag{A1}$$

where the overbar represents the field average precipitation of $\widetilde{\boldsymbol{X}}(\tilde{t},s)$ and $\widetilde{\boldsymbol{Y}}(\tilde{t},s)$. “A” provides a simple evaluation on the quantitative accuracy of total precipitation amount in the field.

(2) The location component (“L”)

The location component consists of two parts: “L_1” and “L_2”. “L_1” measures the relative distance between the centers of mass of the forecast field and the observation field.

$$L_1 = \left| c\left(\widetilde{\boldsymbol{X}}(\tilde{t},s)\right) - c\left(\widetilde{\boldsymbol{Y}}(\tilde{t},s)\right)\right| /d \tag{A2}$$

where $c(\cdot)$ denotes the center of mass of the precipitation in the field and d represents the largest distance between any two grid-cells within the field. "L_1" gives a first-order evaluation on the overall distribution accuracy of the precipitation in the field.

However, a zero value of "L_1" does not guarantee a perfect forecast in terms of the location error, as more detailed distribution features can be very different between the forecast field and the observation field. "L_2" is therefore introduced to capture such difference by considering the weighted average distance between the center of mass of the precipitation field and identified precipitation objects. Specifically, assuming the number of the precipitation objects is J for the forecast precipitation field $\widetilde{\boldsymbol{X}}(\tilde{t}, s)$, the weighted average distance is calculated as

$$r\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right) = \left(\sum_{j=1}^{J} P_j \left| c\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right) - c\left(\widetilde{\boldsymbol{X}}_j\right) \right|\right) / \sum_{j=1}^{J} P_j \tag{A3}$$

where P_j is total precipitation amount that integrates all grid-cell precipitation amounts in the j^{th} precipitation object $\widetilde{\boldsymbol{X}}_j$ $(j = 1,2, \dots, J)$. The same calculation is applied to the observation precipitation field $\widetilde{\boldsymbol{Y}}(\tilde{t}, s)$ to obtain $r\left(\widetilde{\boldsymbol{Y}}(\tilde{t}, s)\right)$. "$L_2$" is then calculated as

$$L_2 = 2\left| r\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right) - r\left(\widetilde{\boldsymbol{Y}}(\tilde{t}, s)\right) \right| / d \tag{A4}$$

and "L" is finally obtained as

$$L = L_1 + L_2 \tag{A5}$$

(3) The structure component ("S")

The structure component aims to measure the size and shape of precipitation objects. For this purpose, a scaled volume for the j^{th} forecast precipitation object $\widetilde{\boldsymbol{X}}_j$ $(j = 1,2, \dots, J)$ is calculated as

$$V_j = P_j / \widetilde{\boldsymbol{X}}_j^{max} \tag{A6}$$

where $\widetilde{\boldsymbol{X}}_j^{max}$ denotes the grid-cell maximum precipitation amount in the precipitation object $\widetilde{\boldsymbol{X}}_j$. A weighted average of scaled volumes for all forecast precipitation objects is then produced by

$$V\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right) = \sum_{j=1}^{J} P_j V_j \,/ \sum_{j=1}^{J} P_j \tag{A7}$$

The same calculation is applied to the observation precipitation field $\widetilde{\boldsymbol{Y}}(\tilde{t}, s)$ to obtain $V\left(\widetilde{\boldsymbol{Y}}(\tilde{t}, s)\right)$. "S" is then defined as the relative difference in weighted average of scaled volumes between the forecast field and the observation field, which is analogous to the component "A" (Equation (A1)):

$$S = 2\left(V\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right) - V\left(\widetilde{\boldsymbol{Y}}(\tilde{t}, s)\right)\right) / \left(V\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right) + V\left(\widetilde{\boldsymbol{Y}}(\tilde{t}, s)\right)\right) \tag{A8}$$

SAL for ensemble forecast precipitation fields has similar forms to the introduced deterministic SAL. Taking an ensemble forecast field as an example, each ensemble member is treated as a deterministic forecast and the $\overline{\widetilde{\boldsymbol{X}}(\tilde{t}, s)}$, $c\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right)$, and $V\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right)$ calculation results of all ensemble members are averaged to be used in Equations (A1), (A2), and (A8) to derive "A", "L_1", and "S", respectively. The calculation of "L_2" is defined as 2 times the continuous ranked probability score (CRPS) of $r\left(\widetilde{\boldsymbol{X}}(\tilde{t}, s)\right)$ calculation results of all ensemble members compared to $r\left(\widetilde{\boldsymbol{Y}}(\tilde{t}, s)\right)$ divided by d. This relationship between the deterministic "L_2" and the ensemble "L_2" is similar to the relationship between the mean absolute error (MAE) of deterministic forecasts and the CRPS of ensemble forecasts introduced in Section 3.3. Further details of the deterministic SAL and ensemble SAL can be found in Radanovics et al. (2018).

References

Baran, S., Lerch, S., 2015. Log-normal distribution based Ensemble Model Output Statistics models for probabilistic wind-speed forecasting. Quarterly Journal of the Royal Meteorological Society, 141(691), 2289-2299. https://doi.org/10.1002/qj.2521.

Baran, S., Möller, A., 2015. Joint probabilistic forecasting of wind speed and temperature using Bayesian model averaging. Environmetrics, 26(2), 120-132. https://doi.org/10.1002/env.2316.

Barnett, T. P., 1978. Estimating Variability of Surface Air Temperature in the Northern Hemisphere. Monthly Weather Review, 106(9), 1353-1367. https://doi.org/10.1175/1520-0493(1978)106<1353:EVOSAT>2.0.CO;2.

Bellier, J., Bontron, G., Zin, I., 2017. Using Meteorological Analogues for Reordering Postprocessed Precipitation Ensembles in Hydrological Forecasting. Water Resources Research, 53(12), 10085-10107. https://doi.org/10.1002/2017WR021245.

Buizza, R., 2018. Chapter 2 - Ensemble Forecasting and the Need for Calibration, in: Vannitsem, S., Wilks, D. S. and Messner, J. W. (Eds.), Statistical Postprocessing of Ensemble Forecasts. Elsevier, pp. 15-48. https://doi.org/10.1016/B978-0-12-812372-0.00002-9.

Cattoën, C., Robertson, D. E., Bennett, J. C., Wang, Q. J., Carey-Smith, T. K., 2020. Calibrating Hourly Precipitation Forecasts with Daily Observations. Journal of Hydrometeorology, 21(7), 1655-1673. https://doi.org/10.1175/jhm-d-19-0246.1.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., Wilby, R., 2004. The Schaake Shuffle: A Method for Reconstructing Space–Time Variability in Forecasted Precipitation and Temperature Fields. Journal of Hydrometeorology, 5(1), 243-262. https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2.

Dogulu, N., López López, P., Solomatine, D. P., Weerts, A. H., Shrestha, D. L., 2015. Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. Hydrology and Earth System Sciences, 19(7), 3181-3201. https://doi.org/10.5194/hess-19-3181-2015.

Du, Y., Wang, Q. J., Wu, W., Yang, Q., 2022. Power transformation of variables for post-processing precipitation forecasts: Regionally versus locally optimized parameter values. Journal of Hydrology, 610, 127912. https://doi.org/10.1016/j.jhydrol.2022.127912.

Feng, F., Li, X., Yao, Y., Liang, S., Chen, J., Zhao, X., Jia, K., Pintér, K., McCaughey, J. H., 2016. An Empirical Orthogonal Function-Based Algorithm for Estimating Terrestrial Latent Heat Flux from Eddy Covariance, Meteorological and Satellite Observations. PLOS ONE, 11(7), e0160150. https://doi.org/10.1371/journal.pone.0160150.

Fortin, V., Abaza, M., Anctil, F., Turcotte, R., 2014. Why Should Ensemble Spread Match the RMSE of the Ensemble Mean? Journal of Hydrometeorology, 15(4), 1708-1713. https://doi.org/10.1175/JHM-D-14-0008.1.

Fraehr, N., Wang, Q. J., Wu, W., Nathan, R., 2022. Upskilling low-fidelity hydrodynamic models of flood inundation through spatial analysis and Gaussian Process learning. Water Resources Research, 58, e2022WR032248. https://doi.org/10.1029/2022WR032248.

Gneiting, T., 2011. Making and Evaluating Point Forecasts. Journal of the American Statistical Association, 106(494), 746-762. https://doi.org/10.1198/jasa.2011.r10138.

Gneiting, T., Fadoua, B., Raftery, A. E., 2007. Probabilistic Forecasts, Calibration and Sharpness. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(2), 243-268. https://doi.org/10.1111/j.1467-9868.2007.00587.x.

Gneiting, T., Raftery, A. E., III, A. H. W., Goldman, T., 2005. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. Monthly Weather Review, 133(5), 1098-1118. https://doi.org/10.1175/MWR2904.1.

Golub, G. H., Loan, C. F. V., 1996. Matrix computations, 3rd ed. Johns Hopkins University Press. https://dl.acm.org/doi/10.5555/248979.

Hannachi, A., Jolliffe, I. T., Stephenson, D. B., 2007. Empirical orthogonal functions and related techniques in atmospheric science: A review. International Journal of Climatology, 27(9), 1119-1152. https://doi.org/10.1002/joc.1499.

Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. Weather and Forecasting, 15(5), 559-570. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Huffel, S. V., Vandewalle, J., 1991. The total least squares problem : computational aspects and analysis. Philadelphia : Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971002.

Jones, D., Wang, W., Fawcett, R., 2009. High-quality spatial climate data-sets for Australia. Australian Meteorological and Oceanographic Journal, 58(2009), 233-248. https://doi.org/10.22499/2.5804.003.

Lerch, S., Thorarinsdottir, T. L., Ravazzolo, F., Gneiting, T., 2017. Forecaster's Dilemma: Extreme Events and Forecast Evaluation. Statistical Science, 32(1), 106-127. https://doi.org/10.1214/16-STS588.

Li, W., Duan, Q., Ye, A., Miao, C., 2019. An improved meta-Gaussian distribution model for post-processing of precipitation forecasts by censored maximum likelihood estimation. Journal of Hydrology, 574, 801-810. https://doi.org/10.1016/j.jhydrol.2019.04.073.

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

Li, W., Pan, B., Xia, J., Duan, Q., 2022. Convolutional neural network-based statistical post-processing of ensemble precipitation forecasts. Journal of Hydrology, 605, 127301. https://doi.org/10.1016/j.jhydrol.2021.127301.

Li, W., Wang, Q. J., Duan, Q., 2020. A Variable-Correlation Model to Characterize Asymmetric Dependence for Postprocessing Short-Term Precipitation Forecasts. Monthly Weather Review, 148(1), 241-257. https://doi.org/10.1175/MWR-D-19-0258.1.

Li, Y., Wang, Q. J., He, H., Wu, Z., Lu, G., 2020. A method to extend temporal coverage of high quality precipitation datasets by calibrating reanalysis estimates. Journal of Hydrology, 581, 124355. https://doi.org/10.1016/j.jhydrol.2019.124355.

Lim, Y., Jo, S., Lee, J., Oh, H.-S., Kang, H.-S., 2012. Prediction of East Asian summer precipitation via independent component analysis. Asia-Pacific Journal of Atmospheric Sciences, 48(2), 125-134. https://doi.org/10.1007/s13143-012-0012-8.

Lintner, B. R., 2002. Characterizing global CO2 interannual variability with empirical orthogonal function/principal component (EOF/PC) analysis. Geophysical Research Letters, 29(19), 27-1-27-4. https://doi.org/10.1029/2001GL014419.

Lorenz, E. N., 1956. Empirical orthogonal functions and statistical weather prediction. Massachusetts Institute of Technology, Department of Meteorology, Cambridge.

McLachlan, G. J., Krishnan, T., 2008. Extensions of the EM Algorithm, in, The EM Algorithm and Extensions. pp. 159-218. https://doi.org/10.1002/9780470191613.ch5.

Möller, A., Lenkoski, A., Thorarinsdottir, T. L., 2013. Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. Quarterly Journal of the Royal Meteorological Society, 139(673), 982-991. https://doi.org/10.1002/qj.2009.

Monahan, A., Fyfe, J., Ambaum, M., Stephenson, D., North, G., 2009. Empirical Orthogonal Functions: The Medium is the Message. Journal of Climate, 22(24), 6501–6514. https://doi.org/10.1175/2009JCLI3062.1.

Narapusetty, B., DelSole, T., Tippett, M. K., 2009. Optimal Estimation of the Climatological Mean. Journal of Climate, 22(18), 4845-4859. https://doi.org/10.1175/2009JCLI2944.1.

Palmer, T. N., Roberto, B., Martin, L., Renate, H., Jung, T., Mark, R., Frédéric, V., Berner, J., Hágel, E., Lawrence, A. R., Florian, P., Park, Y. Y., Bremen, L. v., Gilmour, I. (2007) 'The Ensemble Prediction System - Recent and Ongoing Developments'. ECMWF Technical Memorandum 540. ECMWF: Reading, UK. Available at: https://www.ecmwf.int/node/12527.

Radanovics, S., Vidal, J.-P., Sauquet, E., 2018. Spatial Verification of Ensemble Precipitation: An Ensemble Version of SAL. Weather and Forecasting, 33(4), 1001-1020. https://doi.org/10.1175/waf-d-17-0162.1.

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. https://doi.org/10.1175/mwr2906.1.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S. W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research, 46(5). https://doi.org/10.1029/2009WR008328.

Robertson, D. E., Shrestha, D. L., Wang, Q. J., 2013. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. Hydrology and Earth System Sciences, 17(9), 3587-3603. https://doi.org/10.5194/hess-17-3587-2013.

Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., Seo, D. J., 2007. Precipitation and temperature ensemble forecasts from single-value forecasts. Hydrology and Earth System Sciences, 2007(4), 655-717. https://doi.org/10.5194/hessd-4-655-2007.

Schefzik, R., 2017. Ensemble calibration with preserved correlations: unifying and comparing ensemble copula coupling and member-by-member postprocessing. Quarterly Journal of the Royal Meteorological Society, 143(703), 999-1008. https://doi.org/10.1002/qj.2984.

Schefzik, R., Möller, A., 2018. Chapter 4 - Ensemble Postprocessing Methods Incorporating Dependence Structures, in: Vannitsem, S., Wilks, D. S. and Messner, J. W. (Eds.), Statistical Postprocessing of Ensemble Forecasts. Elsevier, pp. 91-125. https://doi.org/10.1016/B978-0-12-812372-0.00004-2.

Schefzik, R., Thorarinsdottir, T. L., Gneiting, T., 2013. Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling. Statistical Science, 28(4), 616-640. https://doi.org/10.1214/13-STS443.

Scheuerer, M., 2014. Probabilistic quantitative precipitation forecasting using Ensemble Model Output Statistics. Quarterly Journal of the Royal Meteorological Society, 140(680), 1086-1096. https://doi.org/10.1002/qj.2183.

Scheuerer, M., Hamill, T. M., Whitin, B., He, M., Henkel, A., 2017. A method for preferential selection of dates in the Schaake shuffle approach to constructing spatiotemporal forecast fields of temperature and precipitation. Water Resources Research, 53(4), 3029-3046. https://doi.org/10.1002/2016WR020133.

Schmeits, M. J., Kok, K. J., 2010. A Comparison between Raw Ensemble Output, (Modified) Bayesian Model Averaging, and Extended Logistic Regression Using ECMWF Ensemble Precipitation Reforecasts. Monthly Weather Review, 138(11), 4199-4211. https://doi.org/10.1175/2010mwr3285.1.

Schmidt, O. T., Mengaldo, G., Balsamo, G., Wedi, N. P., 2019. Spectral Empirical Orthogonal Function Analysis of Weather and Climate Data. Monthly Weather Review, 147(8), 2979-2995. https://doi.org/10.1175/MWR-D-18-0337.1.

Servain, J., Legler, D., 1986. Empirical orthogonal function analyses of Tropical Atlantic sea surface temperature and wind stress : 1964-1979. Journal of Geophysical Research Atmospheres, 91(C12), 14181– 14191. https://doi.org/10.1029/JC091iC12p14181.

Shrestha, D. L., Robertson, D. E., Bennett, J. C., Wang, Q. J., 2015. Improving Precipitation Forecasts by Generating Ensembles through Postprocessing. Monthly Weather Review, 143(9), 3642-3663. https://doi.org/10.1175/MWR-D-14-00329.1.

Sloughter, J. M. L., Raftery, A. E., Gneiting, T., Fraley, C., 2007. Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. Monthly Weather Review, 135(9), 3209-3220. https://doi.org/10.1175/mwr3441.1.

Straaten, C. v., Whan, K., Schmeits, M., 2018. Statistical Postprocessing and Multivariate Structuring of High-Resolution Ensemble Precipitation Forecasts. Journal of Hydrometeorology, 19(11), 1815-1833. https://doi.org/10.1175/JHM-D-18-0105.1.

Taillardat, M., Mestre, O., Zamo, M., Naveau, P., 2016. Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics. Monthly Weather Review, 144(6), 2375-2393. https://doi.org/10.1175/MWR-D-15-0260.1.

Teissier, G., 1948. La Relation D'Allometrie sa Signification Statistique et Biologique. Biometrics, 4(1), 14-53. https://doi.org/10.2307/3001695.

Van Schaeybroeck, B., Vannitsem, S., 2011. Post-processing through linear regression. Nonlinear Processes in Geophysics, 18(2), 147-160. https://doi.org/10.5194/npg-18-147-2011.

Van Schaeybroeck, B., Vannitsem, S., 2016. A Probabilistic Approach to Forecast the Uncertainty with Ensemble Spread. Monthly Weather Review, 144(1), 451-468. https://doi.org/10.1175/mwr-d-14-00312.1.

Vannitsem, S., 2009. A unified linear Model Output Statistics scheme for both deterministic and ensemble forecasts. Quarterly Journal of the Royal Meteorological Society, 135(644), 1801-1815. https://doi.org/10.1002/qj.491.

Wang, B., Lee, J.-Y., Xiang, B., 2015. Asian summer monsoon rainfall predictability: a predictable mode analysis. Climate Dynamics, 44(1), 61-74. https://10.1007/s00382-014-2218-1.

Wang, Q. J., Schepen, A., Robertson, D. E., 2012a. Merging Seasonal Rainfall Forecasts from Multiple Statistical Models through Bayesian Model Averaging. Journal of Climate, 25(16), 5524-5537. https://doi.org/10.1175/JCLI-D-11-00386.1.

Wang, Q. J., Shao, Y., Song, Y., Schepen, A., Robertson, D. E., Ryu, D., Pappenberger, F., 2019a. An evaluation of ECMWF SEAS5 seasonal climate forecasts for Australia using a new forecast calibration algorithm. Environmental Modelling & Software, 122, 104550. https://doi.org/10.1016/j.envsoft.2019.104550.

Wang, Q. J., Shrestha, D. L., Robertson, D. E., Pokhrel, P., 2012b. A log-sinh transformation for data normalization and variance stabilization. Water Resources Research, 48(5). https://doi.org/10.1029/2011WR010973.

Wang, Q. J., Zhao, T., Yang, Q., Robertson, D., 2019b. A Seasonally Coherent Calibration (SCC) Model for Postprocessing Numerical Weather Predictions. Monthly Weather Review, 147(10), 3633-3647. https://doi.org/10.1175/mwr-d-19-0108.1.

Wernli, H., Hofmann, C., Zimmer, M., 2009. Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique. Weather and Forecasting, 24(6), 1472-1484. https://doi.org/10.1175/2009WAF2222271.1.

Wernli, H., Paulat, M., Hagen, M., Frei, C., 2008. SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. Monthly Weather Review, 136(11), 4470-4487. https://doi.org/10.1175/2008mwr2415.1.

Whitaker, J. S., Loughe, A. F., 1998. The Relationship between Ensemble Spread and Ensemble Mean Skill. Monthly Weather Review, 126(12), 3292-3302. https://doi.org/10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2.

Wu, L., Zhang, Y., Adams, T., Lee, H., Liu, Y., Schaake, J., 2018. Comparative Evaluation of Three Schaake Shuffle Schemes in Postprocessing GEFS Precipitation Ensemble Forecasts. Journal of Hydrometeorology, 19(3), 575-598. https://doi.org/10.1175/JHM-D-17-0054.1.

Wu, W., Emerton, R., Duan, Q., Wood, A. W., Wetterhall, F., Robertson, D. E., 2020. Ensemble flood forecasting: Current status and future opportunities. WIREs Water, 7(3), e1432. https://doi.org/10.1002/wat2.1432.

Yang, Q., Wang, Q. J., Hakala, K., 2021. Achieving effective calibration of precipitation forecasts over a continental scale. Journal of Hydrology: Regional Studies, 35, 100818. https://doi.org/10.1016/j.ejrh.2021.100818.

Zhao, P., Wang, Q. J., Wu, W., Yang, Q., 2020. Which precipitation forecasts to use? Deterministic versus coarser-resolution ensemble NWP models. Quarterly Journal of the Royal Meteorological Society, 147(735), 900-913. https://doi.org/10.1002/qj.3952.

Zhao, P., Wang, Q. J., Wu, W., Yang, Q., 2022. Extending a joint probability modelling approach for post-processing ensemble precipitation forecasts from numerical weather prediction models. Journal of Hydrology, 605, 127285. https://doi.org/10.1016/j.jhydrol.2021.127285.

Zhao, T., Wang, Q. J., Bennett, J. C., Robertson, D. E., Shao, Q., Zhao, J., 2015. Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model. Journal of Hydrology, 528, 329-340. https://doi.org/10.1016/j.jhydrol.2015.06.043.

Tables

Table 1. Time variables set up for precipitation data.

Time variables	Descriptions
t	Time of precipitation data in long-term observation period ($t = 1,2, \dots, T$)
$\tilde{t}$	Time of forecast data used for model training ($\tilde{t} = 1,2, \dots, \tilde{T}$)
$\tilde{\tilde{t}}$	Time of a new forecast precipitation event

Table 2. Forecasts and corresponding evaluation metrics at grid-cell and basin scales.

Evaluation scales	Forecasts	Evaluation metrics
Grid-cell	Forecasts of individual grid-cells	CRPS, PIT
Basin	Basin average forecasts	CRPS, PIT
Basin	Forecasts of the whole precipitation basin	SAL

Table 3. CRPS skill score verifications of basin average forecasts.

	Basin average forecasts		
	Raw forecasts	Calibrated ensemble median forecasts	Calibrated ensemble forecasts
CRPS skill score	19.62%	33.83%	52.14%

Figures

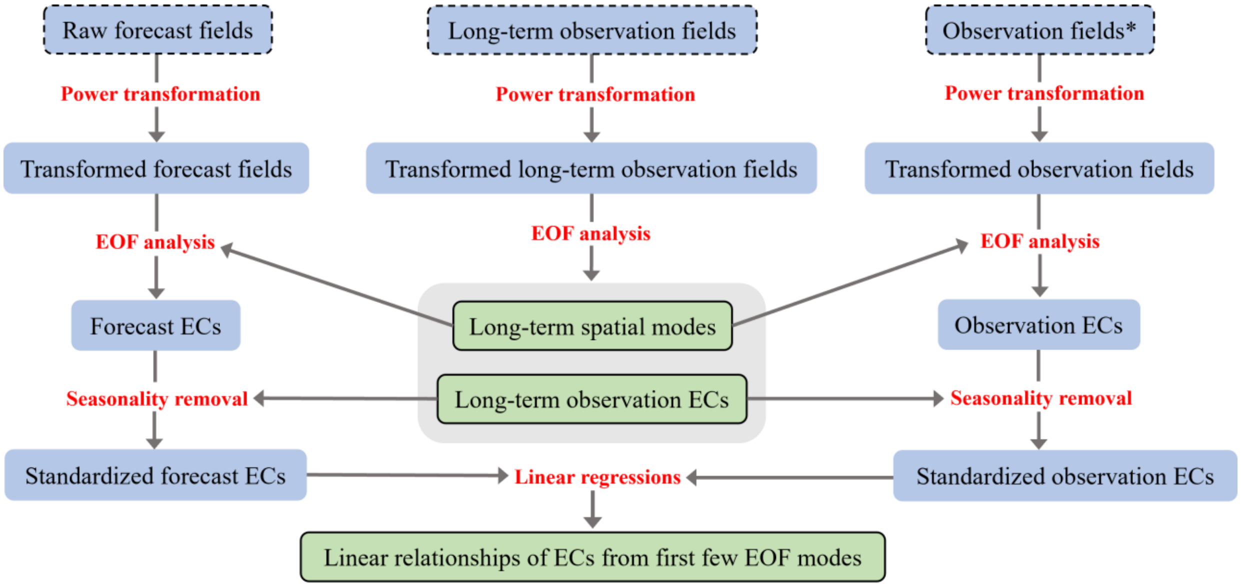

Figure 1. The modelling process of the SMoC model. Observation fields* represent observations in the same period as raw forecasts used in model training. Power transformation is used for normalizing precipitation data, and ECs are derived expansion coefficients from the EOF analysis. Red texts denote processing steps between two neighboring datasets (shown in shaded text boxes). Boxes with dashed boundaries and solid boundaries respectively represent inputs and outputs of the modelling process. Green boxes will be used for calibrating new forecasts as in Figure 2 later.

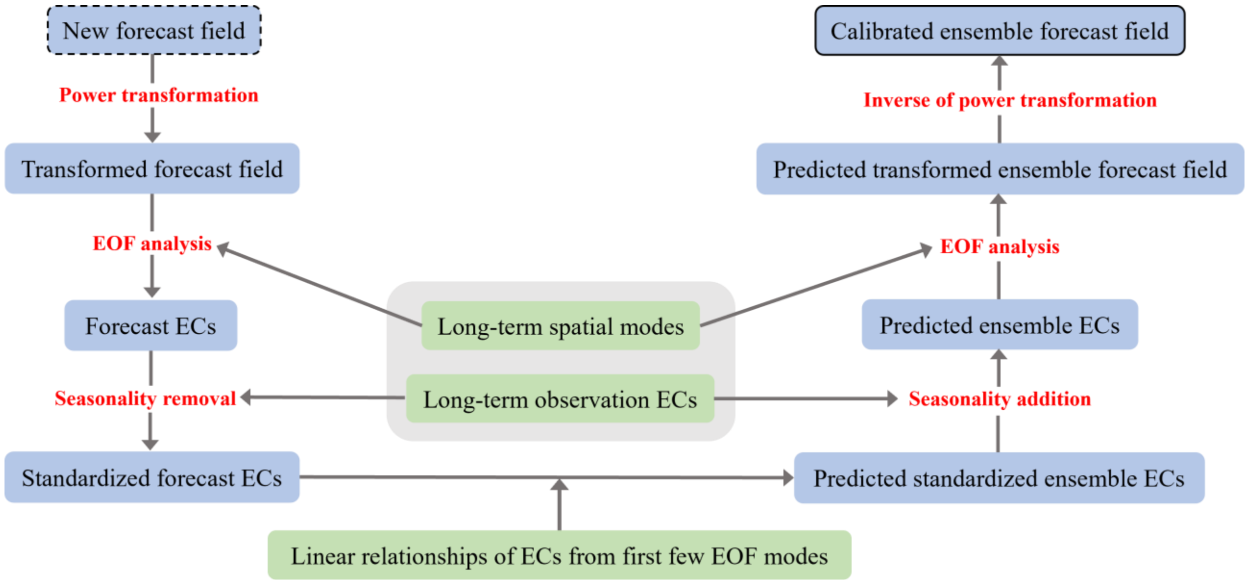

Figure 2. Calibration of new forecast fields using the SMoC model. Power transformation is used for normalizing precipitation data, and ECs are derived expansion coefficients from the EOF analysis. Red texts denote processing steps between two neighboring datasets (shown in shaded boxes). Boxes with a dashed boundary and a solid boundary respectively represent the input and output of the calibration process. Green boxes are obtained from the modelling process of the SMoC model.

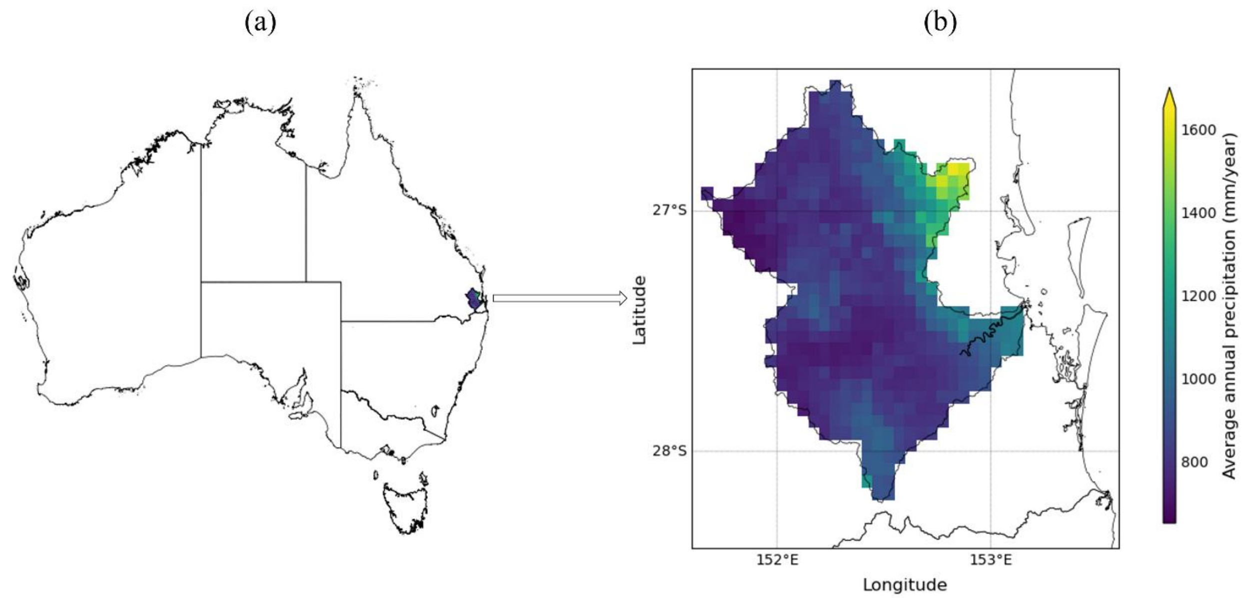

Figure 3. Location (a) and average annual precipitation map (b) of the Brisbane drainage basin.

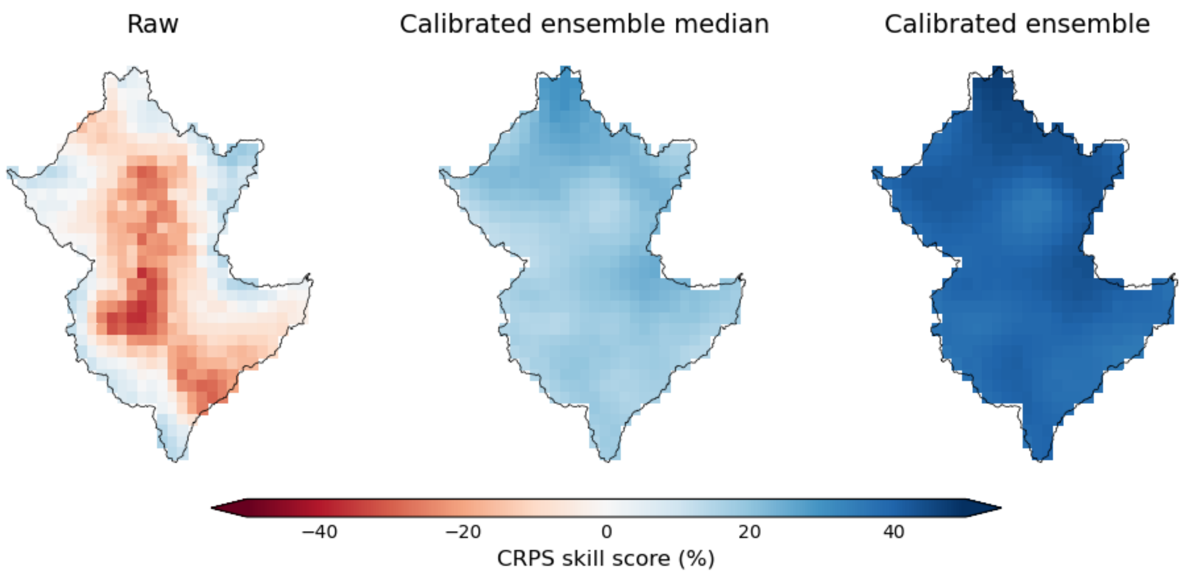

Figure 4. CRPS skill score verification at grid-cell scale of raw forecasts, calibrated ensemble median forecasts, and calibrated ensemble forecasts for substantive precipitation events in the research basin during the 3-year forecast period. The averages of CRPS skill scores across all of the grid-cells for raw forecasts, calibrated ensemble median forecasts, and calibrated ensemble forecasts are respectively calculated as -7.17%, 19.16%, and 40.27%. A positive (negative) CRPS skill score indicates that forecasts are more (less) skillful than reference climatology forecasts. The arrow of the left (right) color bar indicates that values can be below (above) the displayed minimum (maximum) CRPS skill score.

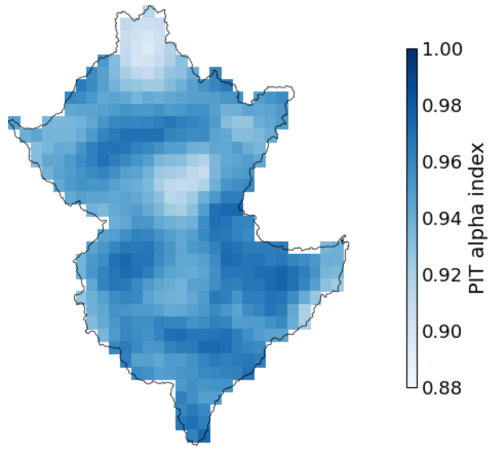

Figure 5. PIT alpha index verification at grid-cell scale of calibrated ensemble forecasts for substantive precipitation events in the research basin during the 3-year forecast period. The average of PIT alpha index values across all of the grid-cells calibrated ensemble forecasts is calculated as 0.9491. PIT alpha index values close to 1 indicate good forecast reliability.

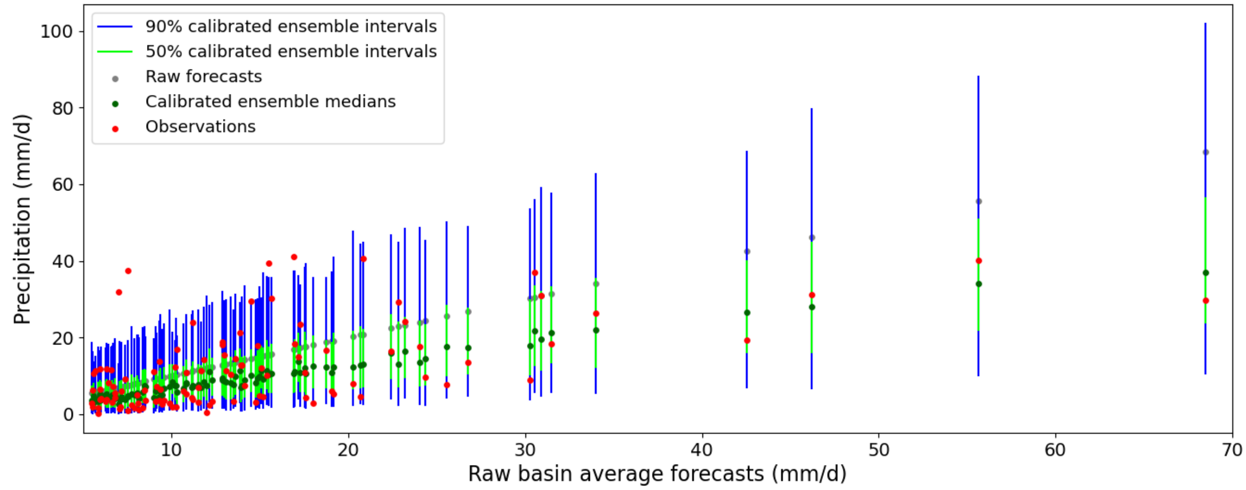

Figure 6. 90% and 50% ensemble intervals, and ensemble medians of calibrated basin average ensemble forecasts against raw basin average forecasts for substantive precipitation events in the research basin during the 3-year forecast period. Each vertical line represents a calibrated basin average ensemble forecast for one substantive precipitation event, with blue and green segments respectively representing 90% and 50% of ensemble intervals. Grey, dark green, and red dots represent raw forecasts, calibrated ensemble medians, and corresponding observations, respectively.

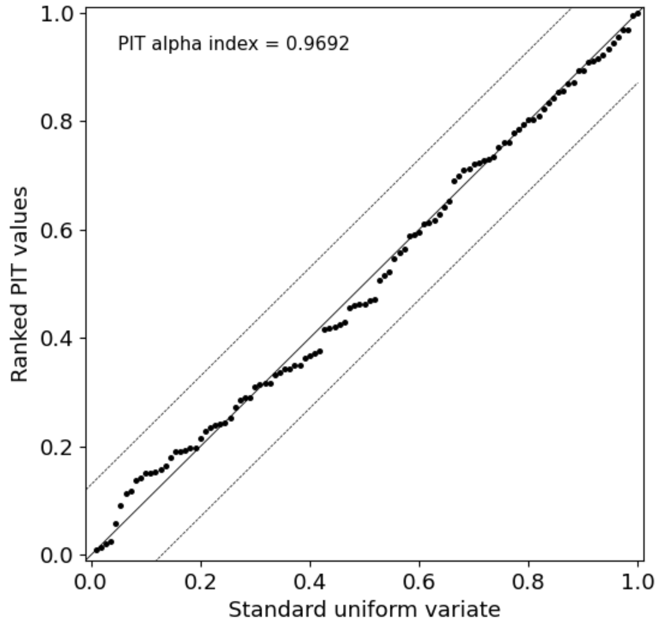

Figure 7. PIT uniform probability plot of calibrated basin average ensemble forecasts for substantive precipitation events in the research basin during the 3-year forecast period. Dots are sorted PIT values of observed precipitation events; solid line represents 1:1 uniform distribution and dashed line is Kolmogorov 5% significance band.

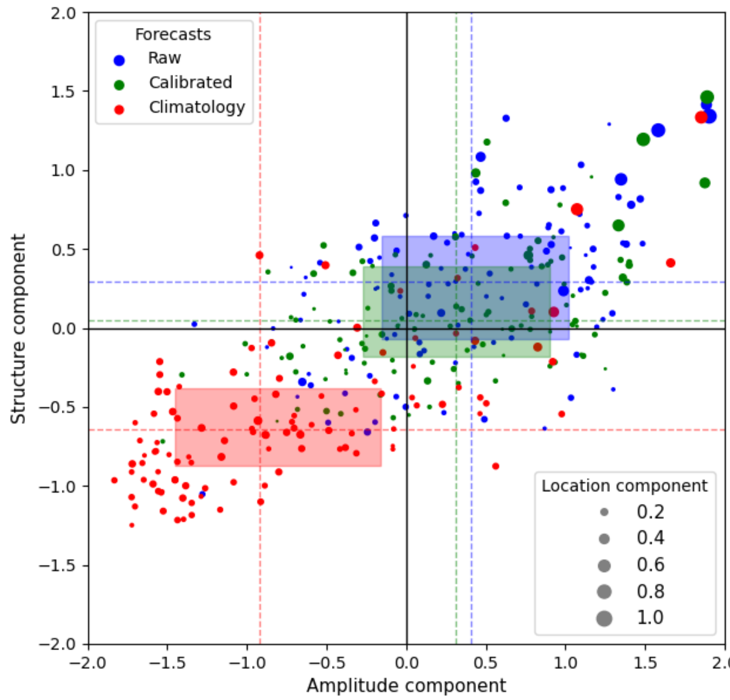

Figure 8. SAL verification diagram for raw forecasts, calibrated ensemble forecasts, and climatology forecasts of substantive precipitation events in the research basin during the 3-year forecast period. Every dot shows values of the three components of SAL for a particular substantive precipitation event. The values of the amplitude and structure components are shown along the x and y axes. The value of the location component is indicated by the size of the dots. The median values of location are 0.1163 (climatology forecasts), 0.0843 (raw forecasts), and 0.0693 (calibrated ensemble forecasts). Median values for structure and amplitude components are shown as dashed lines, and colored boxes extend from the 25th to the 75th percentile of the distributions of structure and amplitude. It should be noted that the differences among forecasts in the amplitude do not reflect the differences in forecast bias. Indeed, biases of these three forecasts are calculated as -9.49 mm/day (climatology forecasts), 2.79 mm/day (raw forecasts), and 0.26 mm/day (calibrated ensemble forecasts). This indicates that the linear regressions inside the SMoC model are capable of correcting the bias of raw forecast expansion coefficients and consequently the bias of raw precipitation forecasts.

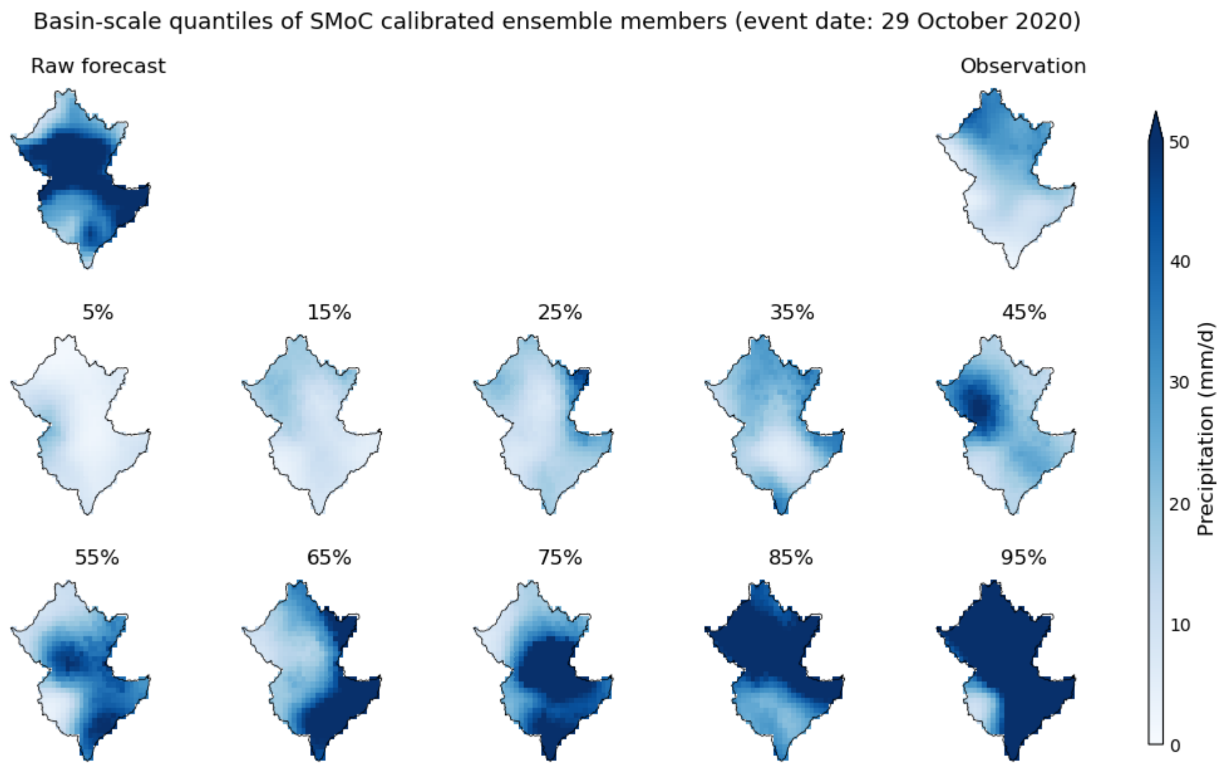

Figure 9. Spatial precipitation plots of calibrated ensemble members selected based on a set of quantiles of basin averages of SMoC calibrated ensemble members for a substantive precipitation event on 29 October 2020. Extreme precipitation is present in the raw forecast but not in the observation for this event.

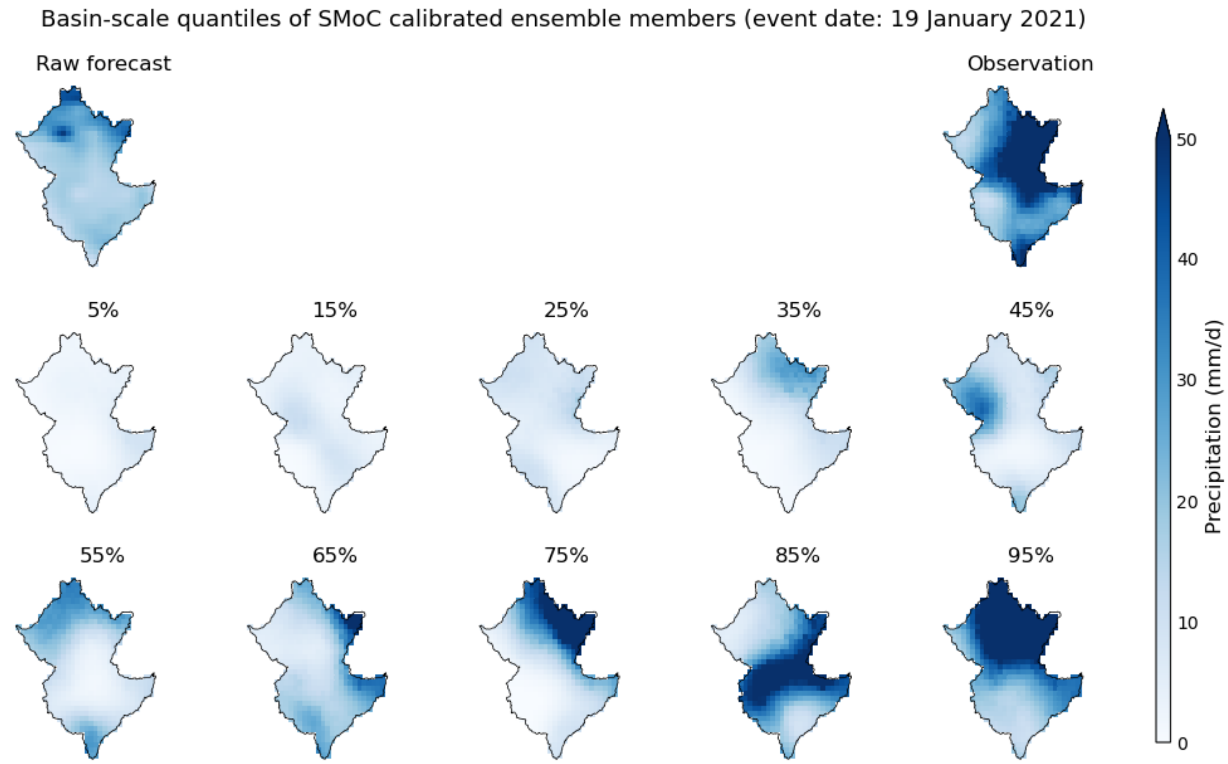

Figure 10. Spatial precipitation plots of calibrated ensemble members selected based on a set of quantiles of basin averages of SMoC calibrated ensemble members for a substantive precipitation event on 19 January 2021. Extreme precipitation is present in the observation but not in the raw forecast for this event.

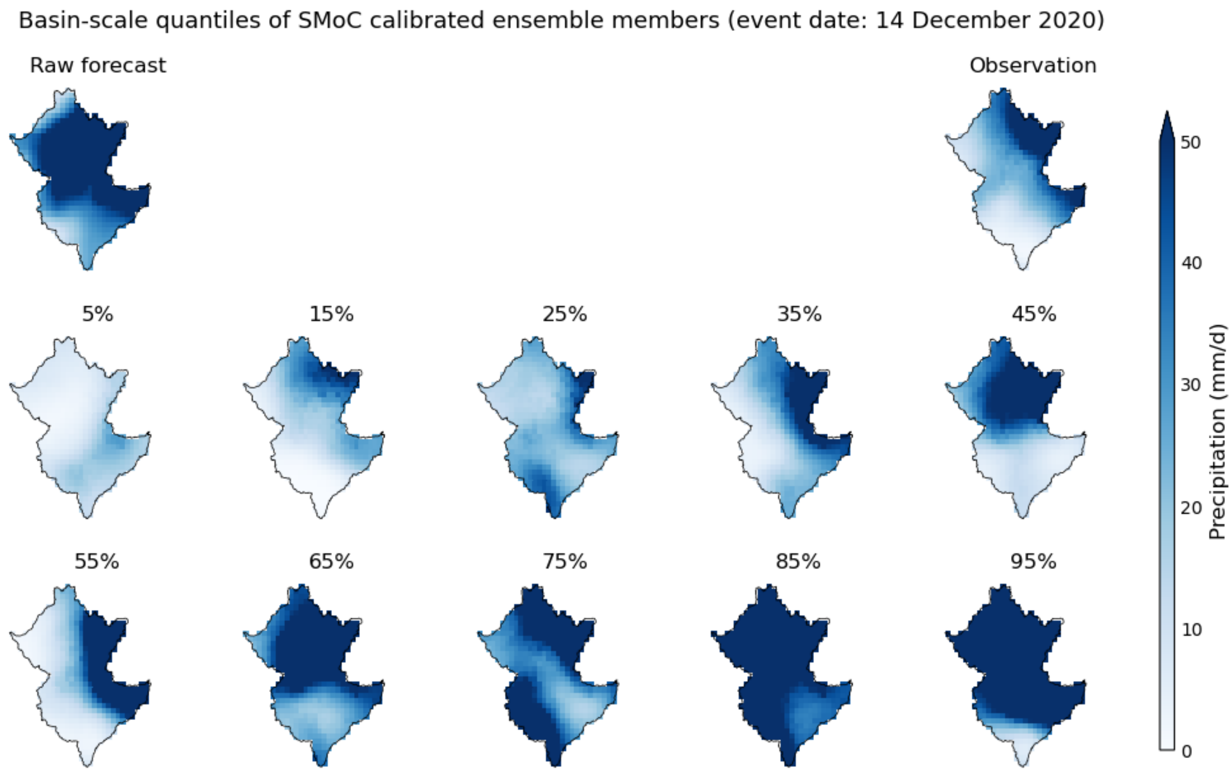

Figure 11. Spatial precipitation plots of calibrated ensemble members selected based on a set of quantiles of basin averages of SMoC calibrated ensemble members for a substantive precipitation event on 14 December 2020. Extreme precipitation is present in both the raw forecast and the observation for this event.