A statistical-dynamical approach for probabilistic prediction of sub-seasonal precipitation anomalies over 17 hydroclimatic regions in China

5   Yuan LI[1], Kangning Xü[1], Zhiyong WU[1*], Zhiwei ZHU[2], Quan J. WANG[3]

[1] College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

[2] Key Laboratory of Meteorological Disaster, Ministry of Education (KLME)/ Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD)/Joint International Research Laboratory of Climate and Environment Change (ILCEC), Nanjing University of Information Science and Technology,

10   Nanjing 210044, China

[3] Department of Infrastructure Engineering, The University of Melbourne, Parkville 3010, Australia

Correspondence to: Zhiyong WU (wzyhhu@gmail.com)

Abstract. In this study, we develop a Spatial Temporal Projection based Calibration, Bridging, and Merging

15   (STP-CBaM) method to improve probabilistic sub-seasonal precipitation forecast skill over 17 hydroclimatic regions in China. The calibration model is established by post-processing ECMWF raw forecasts using the Bayesian Joint Probability (BJP) approach. The bridging models are built using large-scale atmospheric intraseasonal predictors (U200, U850, OLRA, H200, H500, and H850) defined by the Spatial-Temporal Projection method (STPM). The calibration model and bridging models are then merged through the Bayesian

20   Modeling Averaging (BMA) method. Our results indicate that the forecast skill of calibration model is higher compared to bridging models when the lead time is within 5-10 days. The U200 and OLRA-based bridging models outperform the calibration model in certain months and certain regions. The BMA merged forecasts take advantage of both calibration model and bridging models. Meanwhile, the BMA merged forecasts also show high reliability, especially at longer lead times. These findings demonstrate the great potential of

25   combining dynamical models and statistical models in improving sub-seasonal precipitation forecasts.

1. Introduction

Sub-seasonal forecasting (defined as the time range between 2 weeks and 2 months) bridges the gap between short-medium range weather forecasts and seasonal climate prediction (Vitart and Robertson, 2018; Liu et al., 2023). Skillful and reliable sub-seasonal precipitation forecasts are highly valuable for water resource management, flood disaster preparedness, and many other climate-sensitive sectors (White et al., 2022; Yan et al., 2023; Zhu et al., 2022). However, it is considered a difficult time range to generate skilful forecasts. The memory of atmospheric initial conditions is lost compared to short-medium range forecasts, while the variability of lower boundary conditions, such as sea surface temperature, is too slow to take effect (Vitart and Robertson, 2018). Statistical models, which use observational relationships between sub-seasonal precipitation and atmospheric intraseasonal oscillations, have been developed in recent years. The spatial-temporal projection model (STPM), which extracts the coupled patterns of preceding atmospheric intraseasonal oscillations and precipitation, has shown skill in predicting sub-seasonal precipitation. Zhu and Li (2017) constructed STPMs over different climatic regions during the boreal summer monsoon season, and their results indicated that the STPMs could generate skilful forecasts for intraseasonal precipitation patterns with lead time up to 20 days. Our previous study developed a spatial-temporal projection-based Bayesian hierarchical model (STP-BHM) to take the uncertainties in the relationships between atmospheric intraseasonal oscillations and sub-seasonal precipitation into account (Li et al., 2022). However, statistical models are highly reliant on stationary relationships between predictors and predictand. Seasonal changes in climatological conditions may lead to different relationships between atmospheric intraseasonal oscillations and precipitation. Liu and Lu (2022) suggested that the impacts of boreal summer intraseasonal oscillation (BSISO) on precipitation are different between early and late summers. Li et al. (2023a) found that the long-period BSISO event-affected region and associated precipitation anomalies are different compared to short-period BSISO events.

With a more comprehensive understanding and better representation of potential sources of predictability, there has been much improvement in dynamical models in recent years. Subseasonal-to-Seasonal Prediction Project (S2S) and the Subseasonal Experiment (SubX) project have been established to provide S2S forecasts from dynamical models. However, the sub-seasonal precipitation forecasts of GCMs are always of low accuracy (De Andrade et al., 2019; Li et al., 2022). The physical equations are simplified, while the small-scale processes, such as convections, cannot be well represented in most GCMs. In addition, insufficient data assimilation schemes, low capacity in capturing dynamic sources, misrepresentation of atmosphere-ocean

interactions and atmosphere-ocean interactions also contribute to the limited forecast skill (Wu et al., 2023; Zhang et al., 2021). Although post-processing methods have been proposed in recent years, the forecast skill after post-processing was still limited for lead time beyond 10-14 days (Li et al., 2021).

Despite the low forecast skill of sub-seasonal precipitation, the GCMs show much higher performance in predicting large-scale circulation patterns. Cui et al. (2021) evaluated the potential of GCMs for predicting intraseasonal surface air temperature over mid-high-latitude Eurasia. Their results indicated that the upper limit of the useful forecast skill ranged from ~10 to ~20 days. The BSISO is the predominant variability of the Asian summer monsoon, and most GCMs exhibit predictability on timescales of above 3 weeks for BSISO events (Chen and Zhai, 2017; Hsu et al., 2016; Ren et al., 2018). Lee et al. (2015) evaluated the prediction skill of BSISO indices using six coupled models in the Intraseasonal Variability Hindcast Experiment project (ISVHE), and their results suggested that skilful BSISO prediction was about 22 days in strong initial conditions. Shibuya et al. (2021) suggested that the overall useful prediction skill of the BSISO was approximately 24 days in a global Non-hydrostatic Icosahedral Atmospheric Model (NICAM) with explicit cloud microphysics. Similar results were also found by Wu et al. (2023), that the ECMWF model showed skilful prediction of BSISO1 index at 24-day lead time.

Given the strengths and weaknesses of both statistical models and dynamical models, there has been growing interest in developing hybrid prediction models that combine forecasts from both statistical and dynamical models (Slater et al., 2023). Schepen et al. (2014) used POAMA (Predictive Ocean Atmosphere Model for Australia) forecasts of seasonal climate indices as predictors to predict seasonal precipitation over Australia. Strazzo et al. (2019b) developed a hybrid statistical-dynamical system to predict seasonal temperature and precipitation over North America. Most previous statistical-dynamical models focus on seasonal predictions. Much fewer attempts have been made at sub-seasonal timescales. Specq and Batté (2020) proposed a statistical-dynamical post-processing scheme to improve the quality of sub-seasonal forecasts of weekly precipitation using Madden-Julian Oscillation (MJO) and El Niño Southern Oscillation (ENSO) indices as predictors. Wu et al. (2022) established a dynamical-statistical prediction model (DSPM) to improve sub-seasonal precipitation forecasts. Deep learning models were also proposed to predict sub-seasonal extreme rainfall events with the GCM predicted large-scale circulation patterns used as predictors (Xie et al., 2023). Zhu et al. (2023) developed a dynamical-statistical hybrid model using the novel indices of the zonal

displacements of the South Asia high and the western Pacific subtropical high to predict the Meiyu intraseasonal variation. Nevertheless, the relationships between large-scale circulation patterns and sub-seasonal precipitation are of high uncertainty. More sophisticated hybrid models are required to further improve probabilistic sub-seasonal precipitation forecast skill.

The calibration, bridging, and merging (CBaM) method, which employed Bayes-theorem based approaches to take advantage of both dynamical models and statistical models, have been proven to be able to generate skillful and reliable seasonal precipitation and temperature forecasts over different regions (Peng et al., 2014; Schepen et al., 2016; Schepen et al., 2014; Strazzo et al., 2019a). In calibration, the Bayesian joint probability (BJP) approach was used to post-process raw precipitation forecasts derived from GCMs. The BJP approach was also used to generate probabilistic forecasts using large-scale circulation patterns as predictors. This was also referred to as bridging. The calibrated forecasts and bridged forecasts were then merged through the Bayesian model averaging (BMA) method (Wang et al., 2012). Most previous studies used the CBaM method to generate seasonal forecasts. However, much less work has been done on sub-seasonal time scales for several reasons. Compared to seasonal forecasts, there are much fewer climate indices that can be used as predictors at sub-seasonal time scales. Moreover, the atmospheric intraseasonal oscillations may have different effects on precipitation anomalies in different months. As a consequence, it is much more difficult to establish bridging models for sub-seasonal precipitation forecasts. In addition, the evolution of intraseasonal variability of precipitation varies in different stages with different periods in different regions (Liu et al., 2020; Zhu and Li, 2017). The effectiveness of calibration models will be greatly affected if seasonality is not considered.

In this study, we develop a Spatial Temporal Projection based Calibration, Bridging, and Merging (STP-CBaM) method to improve probabilistic sub-seasonal precipitation forecast skill by combining the strengths of both dynamical models and statistical models. The ECMWF sub-seasonal precipitation forecasts are calibrated using the BJP approach for each month. The bridging models are then built using large-scale atmospheric intraseasonal predictors defined by the Spatial-Temporal Projection method (STPM). The calibration model and bridging models are merged through the BMA method to generate skillful and reliable sub-seasonal precipitation forecasts. The STP-CBaM method will be applied to predict sub-seasonal precipitation anomalies over each hydroclimatic region during the boreal summer monsoon from May to October. The accuracy and

reliability will be evaluated through a leave-one-year-out cross-validation strategy.

120

In the following two sections, data and methodology are introduced. The prediction skill and reliability of the STP-CBaM method are provided in Sect. 4. Sect. 5 discusses the forecast skill, limitations, and future work. Key findings are summarized in Sect. 6.

2. Data

125 **2.1 Precipitation dataset**

In this study, China is divided into 17 hydroclimatic regions on the basis of both climate classifications and watershed division standard (Figure 1). The precipitation data is derived from the latest Multi-Source Weighted-Ensemble Precipitation, version 2.8 (MSWEP V2.8) dataset. This dataset is developed by optimally merging precipitation data derived from gauge, satellite, and reanalysis datasets. It covers the period from 1979 to near

130 recent with a spatial resolution of $0.1° \times 0.1°$. Many studies have found that the MSWEP dataset is of high quality over China (Li et al., 2023b; Liu et al., 2019; Guo et al., 2023).
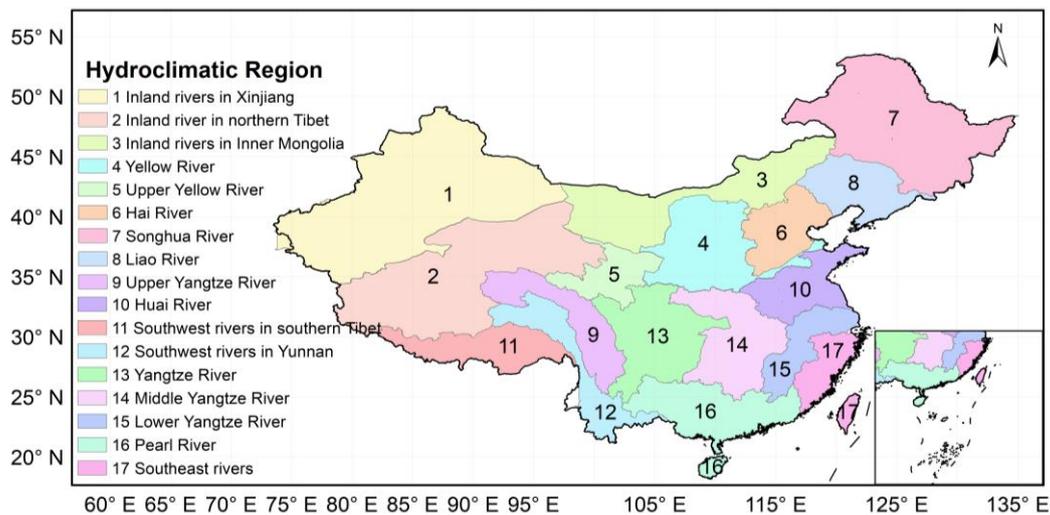

Figure 1. The 17 hydroclimatic regions over China.

2.2 Reanalysis dataset and OLR dataset

135 The daily mean geopotential height at 200 hPa, 500 hPa, and 850 hPa (H200, H500, H850), zonal wind at 200 hPa (U200), and 850 hPa (U850), are derived from the ERA5 (Hersbach et al., 2020) reanalysis dataset at https://cds.climate.copernicus.eu/. The daily mean outgoing longwave radiation (OLR) data is provided by the National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory (PSL), Boulder,

Colorado, USA, from their website at https://psl.noaa.gov. The OLR data is developed from high resolution
140   infrared radiation sounder instruments, and has been widely used over the globe. All daily mean data including
U200, U850, OLR, H200, H500, and H850 are bilinear interpolated onto a horizontal resolution of 2.5° × 2.5°
over the period of 2001-2020.

2.3 Hindcast dataset

145   The ECMWF hindcast data of precipitation, U850, U200, OLR, H200, H500, and H850 are retrieved from the
S2S database at http://apps.ecmwf.int/datasets/data/s2s/. Compared to other GCMs, the ECMWF model
shows highest forecast skill in various aspects (Jie et al., 2017; Wu et al., 2023; Zhang et al., 2021). In this
study, we choose hindcasts when the ECMWF model version dates are in the year 2021 from May to October.
Thus, the hindcasts cover the period of 2001-2020. The gridded precipitation hindcasts are area-weighted
150   averages through 17 hydroclimatic regions as the observational data. In addition, all atmospheric hindcast
fields including U200, U850, OLRA, H200, H500, and H850 are bilinear interpolated onto a horizontal
resolution of 2.5° × 2.5° as the reanalysis dataset.

3.   **Methodology**

3.1 Intraseasonal signal extraction

155   In this study, a non-filtering method is used to extract 10-60-day signals for both atmospheric variables (U200,
U850, OLRA, H200, H500, and H850) and precipitation (Hsu et al., 2015; Zhu et al., 2015). The climatological
annual cycle of observational data is first removed by subtracting 90-day low-pass filtered climatological
component. Lower-frequency signals are then removed by subtracting the last 30-day running mean. The
higher-frequency signals are then removed by taking a pentad mean. The so-derived variable represents the
160   observational 10-60-day signals of daily atmospheric field or precipitation. The daily intraseasonal signals are
then averaged to pentad data to further reduce the noise and improve the predictability. The pentad mean 10-
60-day intraseasonal precipitation is also referred to as pentad mean precipitation anomalies in the following
sections.

165   As for the hindcast fields of the ECMWF model, the model climatology of atmospheric variables (U200, U850,
OLRA, H200, H500, and H850) and precipitation is removed as a function of initial date and lead time. lower-

frequency signals longer than 60 days are then removed the same way as the observations by subtracting the running mean of last 30 days. In this process, the observed anomalies before the forecast initial date are used to make enough data for the running mean. The higher-frequency signals of predicted variables are then removed by taking a pentad mean. The so-derived variable represents the ECMWF model forecasted 10-60-day signals of daily atmospheric field or precipitation.

3.2 Model formulation

3.2.1 Predictor definition for bridging models

In this study, we establish the calibration model and bridging models for each hydroclimatic region, month, and lead time. For calibration model, the ensemble mean of ECMWF forecasted pentad mean precipitation anomalies are used as predictors. For bridging models, we define potential predictors using the Spatial-Temporal Projection method (STPM). Relevant areas of atmospheric fields that could affect 10-60-day precipitation variability are found by cell-wise correlation analysis. The effective degree of freedom is estimated following Livezey and Chen (1983).

The spatial-temporal coupled co-variance patterns are constructed for grid points where the correlation is statistically significant at the 5% level. The predictor is then defined by summing the product of the co-variance field derived from the observational data and the ECMWF model forecasted atmospheric intraseasonal signals,

$$cov(X_{i,p}, Y) = \frac{1}{T}\sum_{t=1}^{T}(y_t - E(y))(x_{i,p,t} - E(x_{i,p}))$$ (1)

$$X_p^* = \sum_{i=1}^{N} cov(X_{i,p}, Y) * X_{i,p}^*$$ (2)

where $X_{i,p}$ denotes the pentad mean 10-60-day signal of pth observational atmospheric field (U200, U850, OLRA, U200, H500, H850) where the correlation is statistically significant at the 5% level for grid i during the training period, $p = 1,2,\cdots,6$. Y denotes the corresponding pentad mean precipitation anomalies. $X_{i,p}^*$ denotes the pentad mean 10-60-day signal of pth hindcast atmospheric field derived from the ECMWF model for grid i. X_p^* denotes the pth predictor defined by the STPM method.

3.2.2 Calibration and bridging models

The calibration model and bridging models are established independently from each other, and each model has only one predictor and one predictand. Therefore, there is one calibration model and six bridging models

for each hydroclimatic region, month, and lead time.

Each calibration model or bridging model is established using the Bayesian Joint Probability (BJP) approach. The predictor X_k, $k = 1, \cdots, K$ and the corresponding predictand Y (pentad mean precipitation anomalies) are normalized to U_k and V using the Yeo-John transformation method,

$$u_k = \begin{cases} \dfrac{\left\{(x_k+1)^{\lambda_{x_k}}-1\right\}}{\lambda_{x_k}} & (x_k \geq 0, \lambda_{x_k} \neq 0) \\ \log(x_k + 1) & (x_k \geq 0, \lambda_{x_k} = 0) \\ -\dfrac{\left\{(-x_k+1)^{2-\lambda_{x_k}}-1\right\}}{2-\lambda_{x_k}} & (x_k < 0, \lambda_{x_k} \neq 2) \\ \log(-x_k + 1) & (x_k < 0, \lambda_{x_k} = 2) \end{cases} \tag{3}$$

$$v = \begin{cases} \dfrac{\left\{(y+1)^{\lambda_y}-1\right\}}{\lambda_y} & (y \geq 0, \lambda_y \neq 0) \\ \log(y + 1) & (y \geq 0, \lambda_y = 0) \\ -\dfrac{\left\{(-y+1)^{2-\lambda_y}-1\right\}}{2-\lambda_y} & (y < 0, \lambda_y \neq 2) \\ \log(-y + 1) & (y < 0, \lambda_y = 2) \end{cases} \tag{4}$$

where λ_{x_k} and λ_y are the unknown transformation parameters for predictor X_k and predictand Y. The matrix $Z^T = [U_k \, V]$ is then assumed to follow a bivariate normal distribution,

$$Z \sim \mathbf{N}(\boldsymbol{\mu}, \textstyle\sum) \tag{5}$$

where $\boldsymbol{\mu}$ and $\sum$ are the mean vector and covariance matrices to be estimated,

$$\boldsymbol{\mu}^T = [\mu_{U_k} \, \mu_V] \tag{6}$$

$$\textstyle\sum = \boldsymbol{\sigma R \sigma}^T \tag{7}$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{R}$ are the standard deviation vector and correlation coefficient matrix, respectively:

$$\boldsymbol{\sigma}^T = [\sigma_{U_k} \, \sigma_V] \tag{8}$$

$$\mathbf{R} = \begin{bmatrix} 1 & r_{U_k V} \\ r_{V U_k} & 1 \end{bmatrix} \tag{9}$$

Note that the correlation coefficient matrix $\mathbf{R}$ is symmetric. Thus, there are only 5 unknown parameters. Here, we denote the 5 unknown parameters of the joint distribution as $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \textstyle\sum\}$.

Given a data series of $\mathbf{D_0} = \{(x_{k,t} \, y_t), \, t = 1, \cdots, n\}$, We apply the SCE-UA (shuffled complex evolution method developed at The University of Arizona) method to estimate transformation parameters that maximize the log-likelihood function (Duan et al., 1994). The data series of $\mathbf{D_0}$ is then normalized to $\mathbf{D} = \{(u_{k,t} \, v_t), \, t = 1, \cdots, n\}$. The posterior distribution of $\boldsymbol{\theta}$ is estimated using a Bayesian framework,

$$p(\boldsymbol{\theta}|\mathbf{D}(U_k, V)) \propto p(\boldsymbol{\theta})p(\mathbf{D}(U_k, V)|\boldsymbol{\theta}) \tag{10}$$

where $p(\boldsymbol{\theta})$ is the prior distribution of parameters, and $p(\mathbf{D}(U_k, V)|\boldsymbol{\theta})$ is the likelihood. As the posterior

distributions of parameters $\boldsymbol{\theta}$ are not standard distributions, analytical integration is difficult. To overcome this problem, we use the new Gibbs sampling algorithm proposed by Wang et al. (2019) to draw a sample of 1000 sets of parameter values. A more detailed description of the sampling strategy can be found in Li et al. (2021).

The posterior predictive distribution of $v_k(t^*)$ is given by

$$f_{V_k}(v_k(t^*)) = \int p(v_k(\text{t}^*)|u_k(t^*), \boldsymbol{\theta})p(\boldsymbol{\theta})p(\mathbf{D}(U_k, V)|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{11}$$

where $x_k(t^*)$ is the new forecast value.

Again, the Gibbs Sampling algorithm is used to obtain 1000 samples of $f_{V_k}(v_k(t^*))$. The samples are then back-transformed to produce the calibrated or bridged predictive density $f_k(y|x_k)$ using parameter λ_y.

3.2.3 Combining models

The merging forecasts are carried out by the Bayesian model averaging (BMA) method proposed by Wang et al. (2012). Given all candidate models, $f_k(y|x_k)$, $k = 1, \cdots, K$, and the corresponding model weights, w_k, $k = 1, \cdots, K$, the predictive density of the BMA probabilistic forecasts can be represented as

$$f_{BMA}(y|x_1, \cdots, x_K) = \sum_{i=1}^{K} w_k f_k(y|x_k) \tag{12}$$

where x_k is the predictor and y is the corresponding predictand.

To encourage even weights among the models, the prior of model weights is assumed to follow a symmetric Dirichlet distribution, given as

$$p(w_k, k = 1, \cdots, K) \propto \prod_{k=1}^{K}(w_k)^{\alpha-1} \tag{13}$$

where α is the concentration parameter slightly over 1 and more specifically, $\alpha = 1 + \alpha_0/K$ and α_0=0.5. The posterior distribution of model weights given $t = 1, \cdots, T$ events is as follows:

$$p(w_k, k = 1, \cdots, K|x_k^T, y^T, f_k(y|x_k), k = 1, \cdots, K) \propto \prod_{k=1}^{K}(w_k)^{\alpha-1} \prod_{t=1}^{T} \sum_{k=1}^{K} w_k f_k^{(t)}(y^t|x_k^t) \tag{14}$$

where $f_k^{(t)}(y^t|x_k^t)$ is the cross-validated predictive density. This indicates that the weights are assigned by the model predictive ability rather than fitting ability. An Expectation-Maximization (EM) algorithm is then used to estimate the weights by maximizing the likelihood function. Initially, all weights are equal. The EM algorithm is then iterated until the likelihood function converges.

3.3 Evaluation

In this study, a leave-one-year-out cross-validation strategy is used to avoid any bias in skill, including predictor selection, data normalization, model building, parameter inference, and verification.

The temporal correlation coefficient (TCC) is used to evaluate the performance of the ECMWF model for predicting atmospheric intraseasonal oscillations. We should note that the ECMWF model has an initial frequency of twice a week on Tuesday and Thursday. Therefore, 160 or 180 initial dates are found for each month during the period of 2001-2020. As the atmospheric variables are autocorrelated, the effective degree of freedom is estimated following Livezey and Chen (1983).

The continuous ranked probability score (Matheson and Winkler, 1976) is used to evaluate the accuracy of probabilistic forecasts for a given lead time t:

$$\text{CRPS} = \frac{1}{N} \sum_{i=1}^{N} \int [F_{i,t}(y) - H(y - o_{i,t})]^2 dy \tag{15}$$

where $F^{i,t}()$ is the cumulative distribution function of the probabilistic forecasts for case i at lead time t; and $H()$ is the Heaviside step function defined as:

$$H(y - o_{i,t}) = \begin{cases} 0 & y < o_{i,t} \\ 1 & y \geq o_{i,t} \end{cases} \tag{16}$$

where $o_{i,t}$ is the corresponding observation.

A CRPS skill score is then calculated by comparing the CRPS of STP-CBaM forecasts with the CRPS of reference forecasts:

$$\text{CRPS}_{\text{SS}} = \frac{\text{CRPS}_{\text{REF}} - \text{CRPS}}{\text{CRPS}_{\text{REF}}} \times 100\% \tag{17}$$

The reference forecasts are generated using the BJP approach to fit the observations used in the training dataset. When the CRPS skill score is 100%, the probabilistic forecasts are the same as the observations. Whereas, a skill score of 0% indicates that the probabilistic forecasts show similar accuracy compared to the cross-validated climatology. Negative skill score means that the probabilistic forecasts are inferior to the cross-validated climatology.

The forecast reliability is evaluated using the α-index (Renard et al., 2010). The Probability Integral Transform (PIT) values of probabilistic forecasts for each case i at lead time t are given as,

$$\pi_{i,t} = F^{i,t}(o_{i,t}) \tag{18}$$

where $F^{i,t}()$ is the cumulative distribution function of probabilistic forecasts, and $o_{i,t}$ is the corresponding observations. If the ensemble forecasts are reliable, $\pi_{i,t}$ should be uniformly distributed. The π_t values are then summarized into an α-index,

$$\alpha = 1.0 - \frac{2}{N}\sum_{i=1}^{N}\left|\pi_{i,t}^* - \frac{i}{N+1}\right| \tag{19}$$

where $\pi_{i,t}^*$ is the sorted $\pi_{i,t}$ in increasing order. The α-index ranges from 0 to 1, and a higher α-index indicates higher reliability.

4. Results

4.1 Correlation analysis between atmospheric intraseasonal oscillation and precipitation anomalies

Figure 2 presents the correlation between pentad mean 10–60 d signals of U200 and precipitation over Region 1 (inland rivers in Xinjiang) from May to October. The U200 signals near Mongolian Plateau have a positive impact on precipitation anomalies over Region 1 in May, while the impact of U200 signals near eastern Tibetan Plateau is negative. In June and July, the U200 signals in West Siberian Plain and Mongolian Plateau show positive correlations with precipitation anomalies. The spatial patterns of correlations between U200 signals and precipitation anomalies are similar in August, September, and October. The U200 signals near Barents sea and Iranian plateau have positive impacts on precipitation anomalies over Region 1. In comparison, U200 signals over West Siberian Plain show strong negative correlations with precipitation anomalies in these months. The OLRA signals show similar wave patterns as other atmospheric variables (Fig. 3). The spatial patterns of correlations between U850, U200, OLRA, H200, H500, H850 and precipitation anomalies are different for each month as well (Fig. S1~S4).
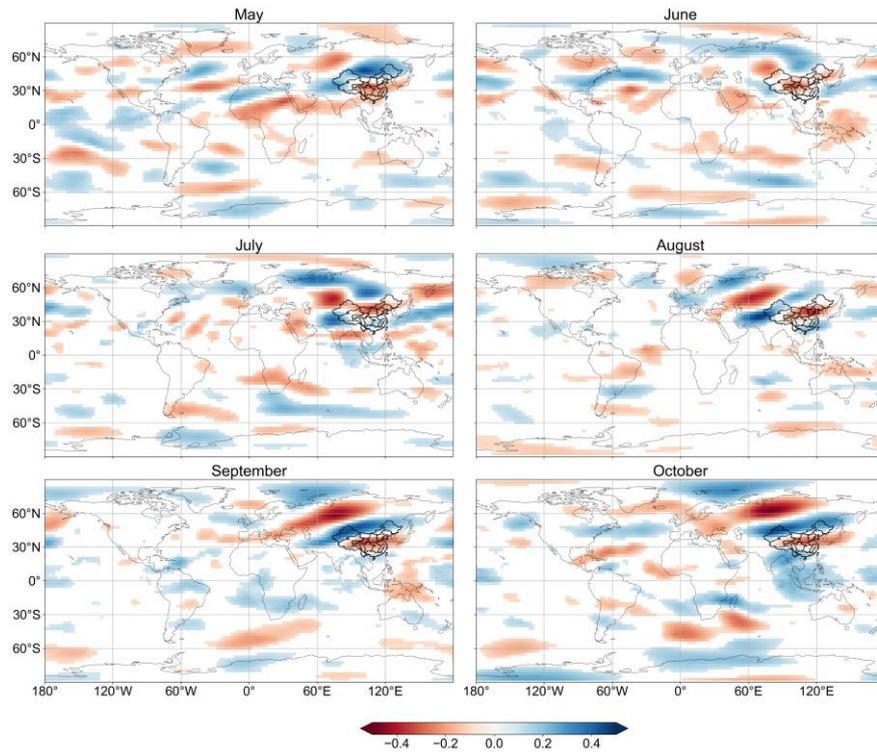
Figure 2. Correlation coefficient between pentad mean 10–60 d signals of U200 and precipitation over Region 1 (inland rivers in Xinjiang) in different months. Correlation coefficients that are statistically significant at the 5 % level are shaded.
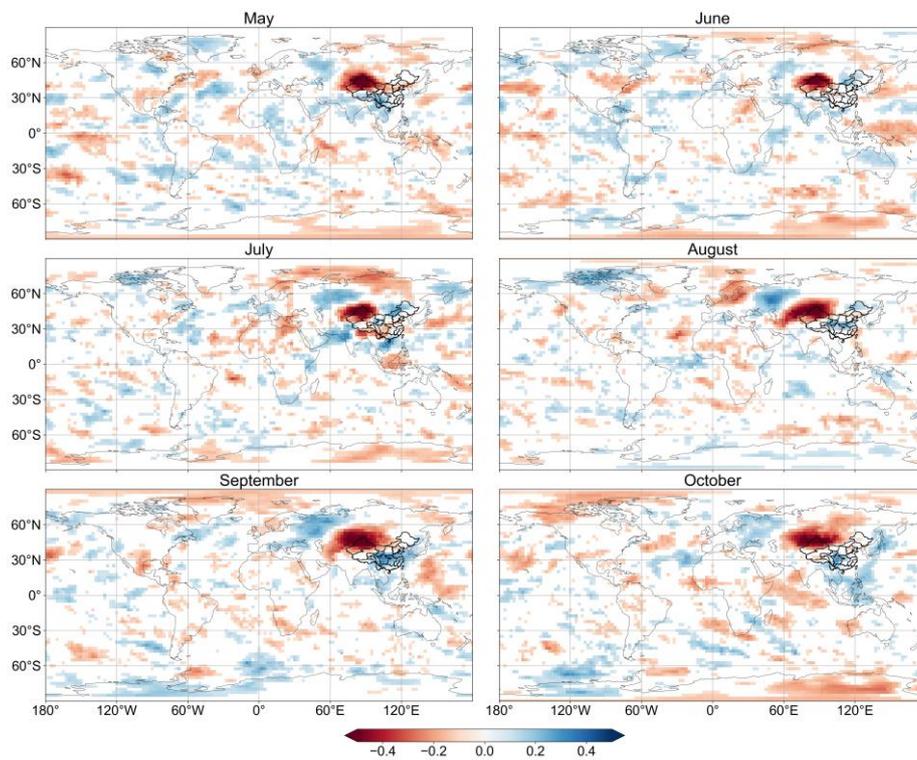

Figure 3. Same as Figure 2, but for OLRA.

4.2 Skill of ECMWF model in forecasting atmospheric intraseasonal oscillations

300   The forecast skill of bridging models is reliant on the forecast skill of atmospheric intraseasonal oscillations derived from dynamical models. The temporal correlation coefficient (TCC) between the ensemble mean of ECMWF forecasted U200 intraseasonal signals and the observations in May are shown in Figure 4. The ECMWF model shows high forecast skill in predicting U200 intraseasonal signals when the lead time is within 10 days, and the correlation coefficients are mostly over 0.7 over the globe. Although the forecast skill

305   decreases as lead time increases, there are still regions where the forecasted U200 signals are significantly correlated with the observations. The forecast skill of OLRA intraseasonal oscillations is lower than that of the U200 signals (Fig. 5). High forecast skill is mostly observed near the equator from 30°N to 30°E when the lead time is beyond 10 days. Similar results are also found for U850, H200, H500, and H850, where significant correlations are found mostly near the equator at longer lead times (Fig. S5 to S8). This suggests that the

310   forecast skill of sub-seasonal precipitation can be potentially improved by taking advantages of both skillful prediction of atmospheric intraseasonal oscillations and stable relationships between precipitation and large-scale circulations, especially for tropical regions.
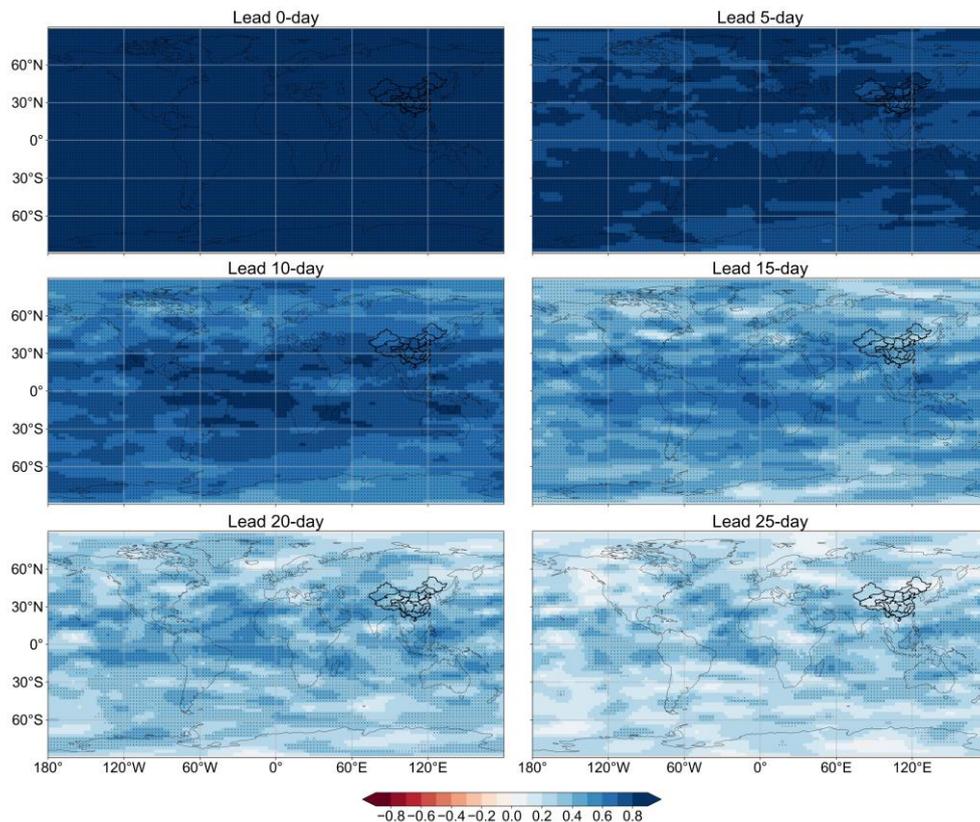
Figure 4. Temporal correlation coefficient (TCC) of the ensemble mean of U200 intraseasonal signals derived

315   from the ECMWF model compared to the ERA5 reanalysis data in May. Correlation coefficients that are statistically significant at the 5 % level are shaded.
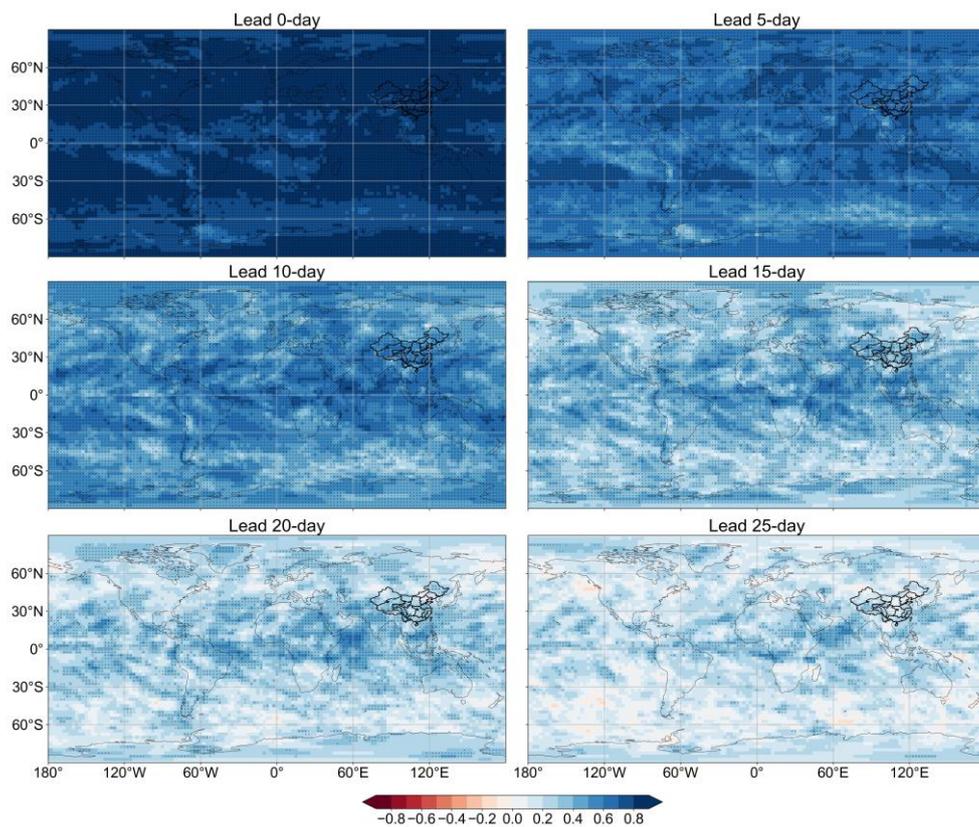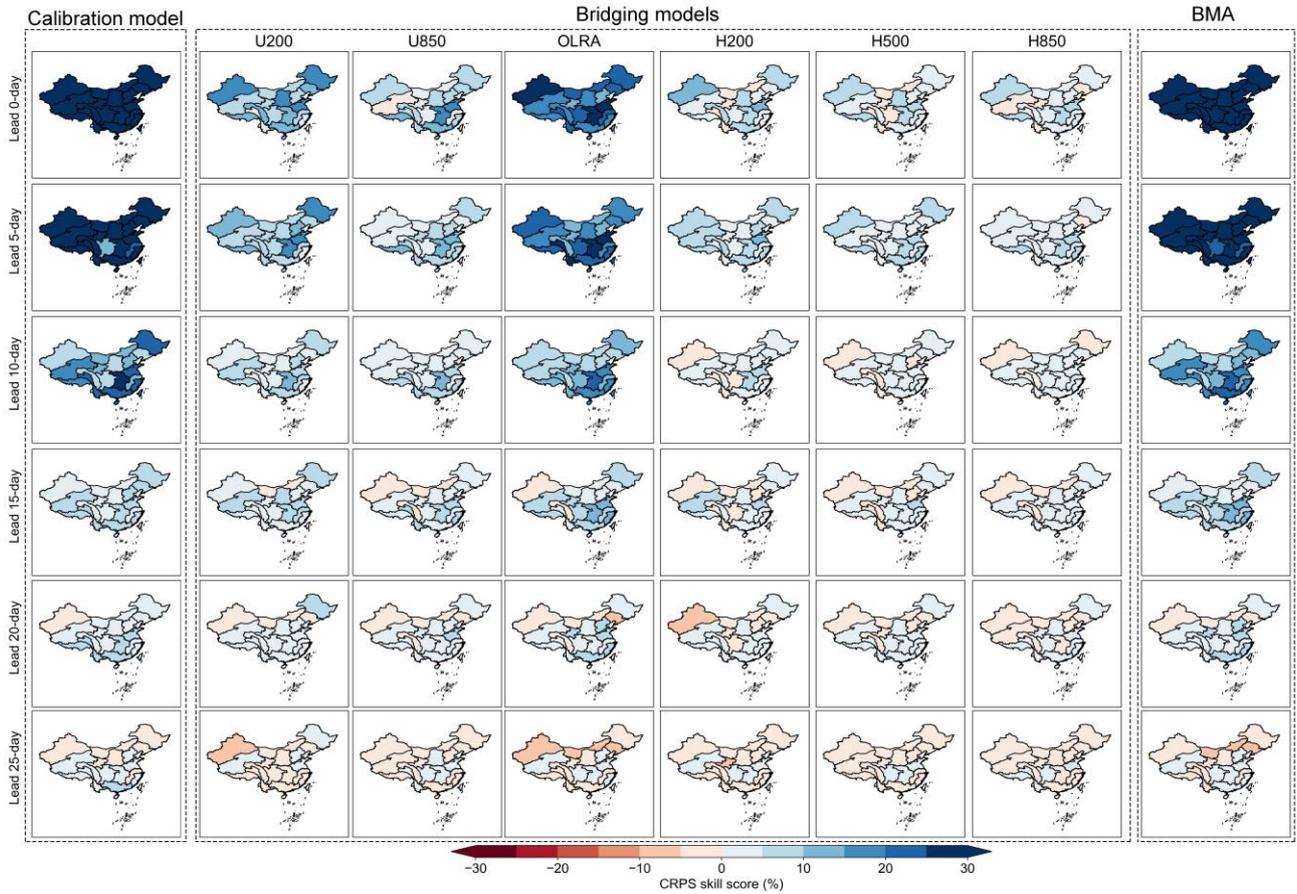
13

Figure 5. Same as Figure 4, but for OLRA.

4.3 Skill of calibration model, bridging models, and merged forecasts

320   Figure 6 presents the spatial distribution of CRPS skill score of calibration model, bridging models, and merged forecasts at different lead times in May. The calibration model shows highest forecast skill compared to bridging models at short lead times. The forecast skill of calibrated forecasts decreases rapidly, and the CRPS skill scores are mostly below 10% when the lead time is beyond 10 days. The forecast skill of bridging models is higher than the calibration model in Region 10 (Huai River), Region 14 (Middle Yangtze River), and Region 17

325   (Southeast rivers) at a lead time of 15 days when the OLRA is used as predictor. The forecast skill of bridging models is higher in Region 6 (Hai River) and Region 7 (Songhua River) when the OLRA and U200 signals are used as predictors at a lead time of 20 days, separately. The merged forecasts take advantages of both calibration model and bridging models. Figure 7 shows the boxplots of CRPS skill scores of calibration model, bridging models (U200, U850, OLRA, H200, H500, H850), and merged forecasts (BMA) at different lead times

330   in May. The distribution of CRPS skill score of merged forecasts is similar to the calibration model at a lead time of 0-day. The minimum CRPS skill score of merged forecasts is over 20% in Region 13 (Yangtze River) at a lead time of 5-day, higher than both the calibration model and bridging models. The bridging models, which use the U200 and OLRA as predictors, show higher minimum CRPS skill score compared to the calibration

14

model and other bridging models at a lead time of 10~15 days. The distributions of CRPS skill score of

335 calibration model, bridging models, and the BMA merged forecasts are similar at longer lead times.
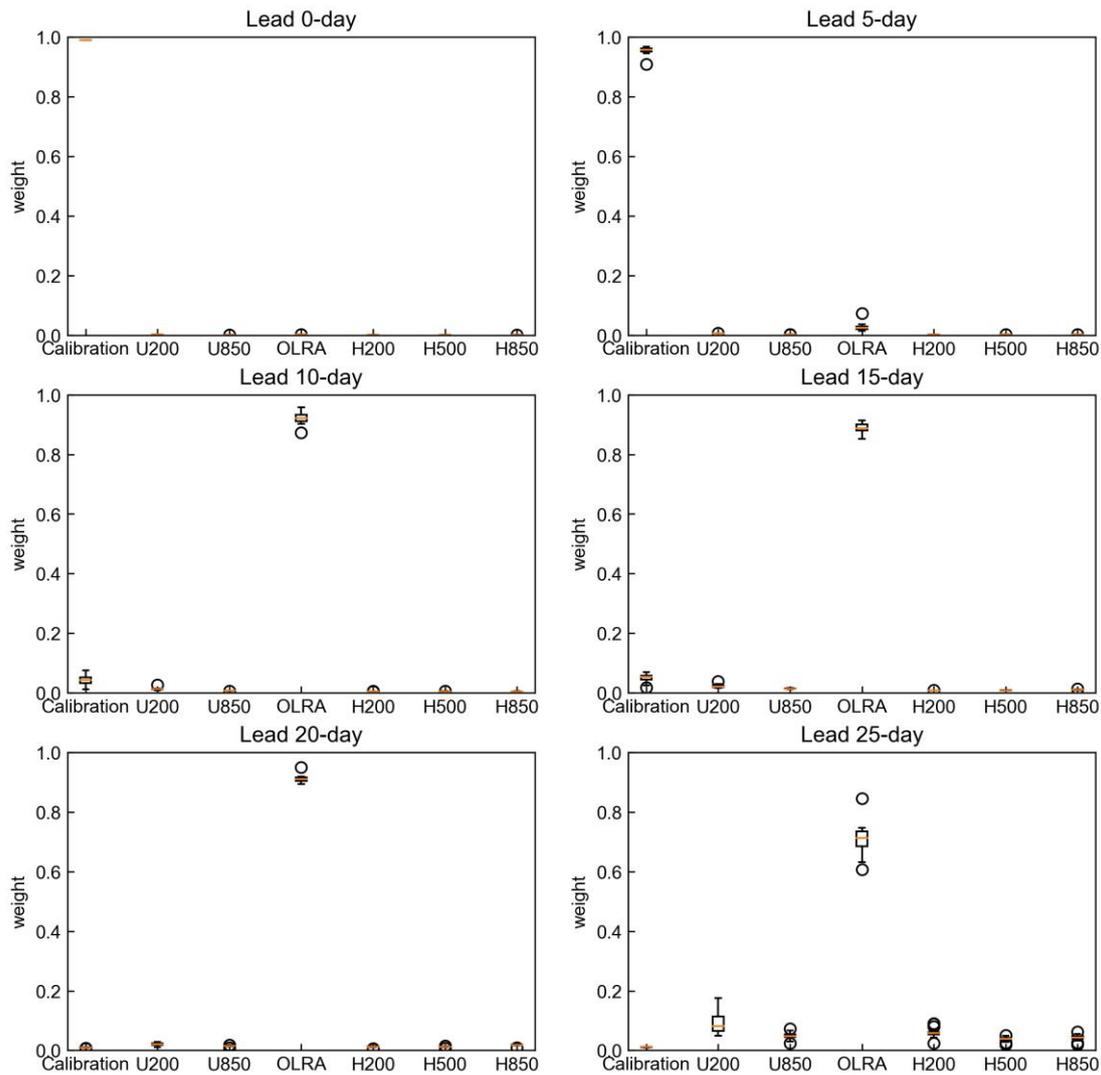

Figure 6. Spatial distribution of CRPS skill score of calibration model, bridging models (U200, U850, OLRA, H200, H500, H850), and merged forecasts (BMA) at different lead times in May.
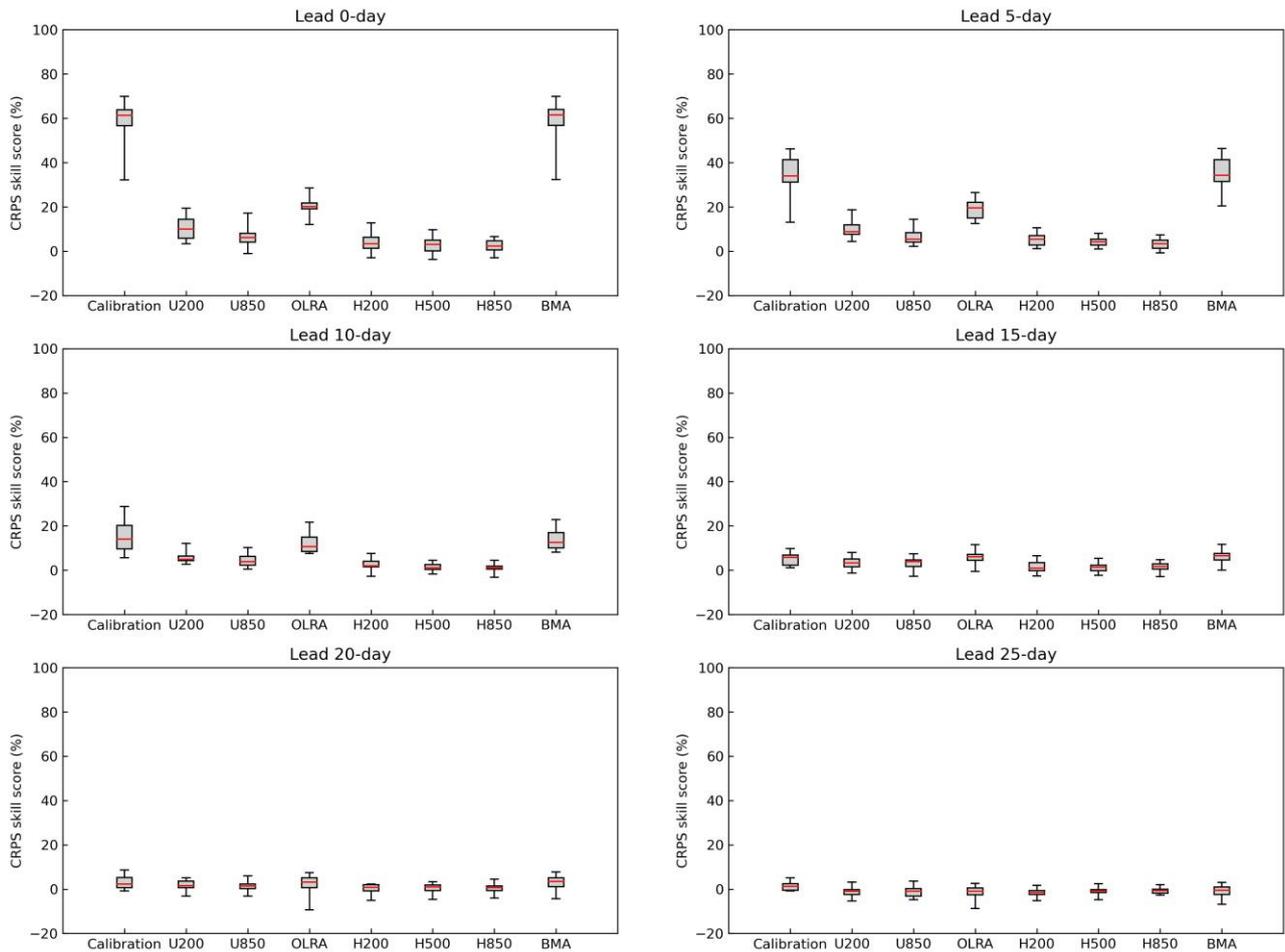
Figure 7. Box plots of CRPS skill score of calibration model, bridging models (U200, U850, OLRA, H200, H500, H850), and merged forecasts (BMA) at different lead times in May. The red lines are the 50th percentiles, top and bottom of each box are the 75th and 25th percentiles, and whiskers are the maximum and minimum skill scores.

Figure 8 shows the distribution of model weights at different lead times for Region 1 (Inland rivers in Xinjiang) in May. The weights are rather stable at short lead times, which more than 90% of the total weights are assigned to the calibration model. Similar results are also found in other regions and months (not shown). The weights of calibration model decrease rapidly when the lead time is beyond 10 days. More weights are assigned to U200 and OLRA at longer lead times. This is mostly consistent with the distribution of CRPS skill scores shown in Figure 7. The CRPS skill scores of U200 and OLRA based bridging models are higher than the calibration model and other bridging models, especially when the lead time is between 10 - 20 days. This indicates that the U200 and OLRA signals are more useful in predicting sub-seasonal precipitation anomalies compared to other large-scale atmospheric circulation variables.

16

Figure 8. Box plots showing the distribution of model weights at different lead times in cross-validation for Region 1 (Inland rivers in Xinjiang) in May.

Figure 9 presents the CRPS skill score of merged forecasts at different lead times from May to October. In general, the forecast skill shows regional, monthly, and lead time-dependent patterns. The merged forecasts

360 show higher skill in predicting sub-seasonal precipitation anomalies over Region 16 (Pearl River) than other regions. The CRPS skill scores are positive for all months at a lead time of 0-20 days. This is mainly owing to the higher prediction skill of OLRA in these regions as shown in Figure 5. In addition, the merged forecasts show highest skill in October, which positive skill scores are found over 14 hydroclimatic regions for all lead times except Region 1 (Inland rivers in Xinjiang), Region 7 (Songhua River), and Region 8 (Liao River). In

365 comparison, positive skill scores are found only over 3 hydroclimatic regions for all lead times (Region 1, Inland rivers in Xinjiang, Region 13, Yangtze River, and Region 14, Middle Yangtze River) in July.
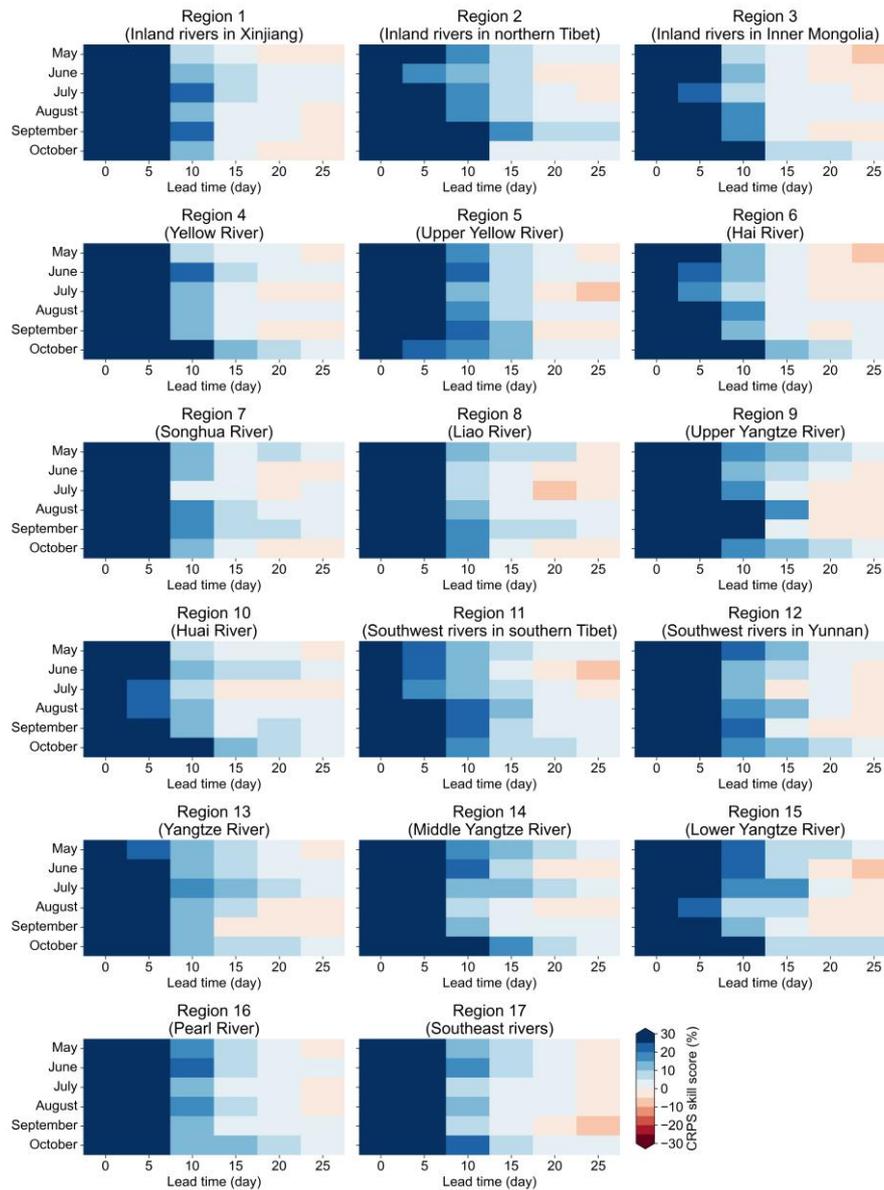
Figure 9. CRPS skill score of merged forecasts at different lead times from May to October.

The α-index of merged forecasts at different lead times are shown in Figure 10. The α-index is around 0.7 at short lead times, suggesting that the forecasts are of relatively low reliability. The quantile ranges of merged forecasts suggest that the merged forecasts are too narrow (too confident) at short lead times (not shown). The α-index is over 0.85 when the lead time is beyond 10 days for all hydroclimatic regions and lead times. This indicates that the merged forecasts are of higher reliability at longer lead times.
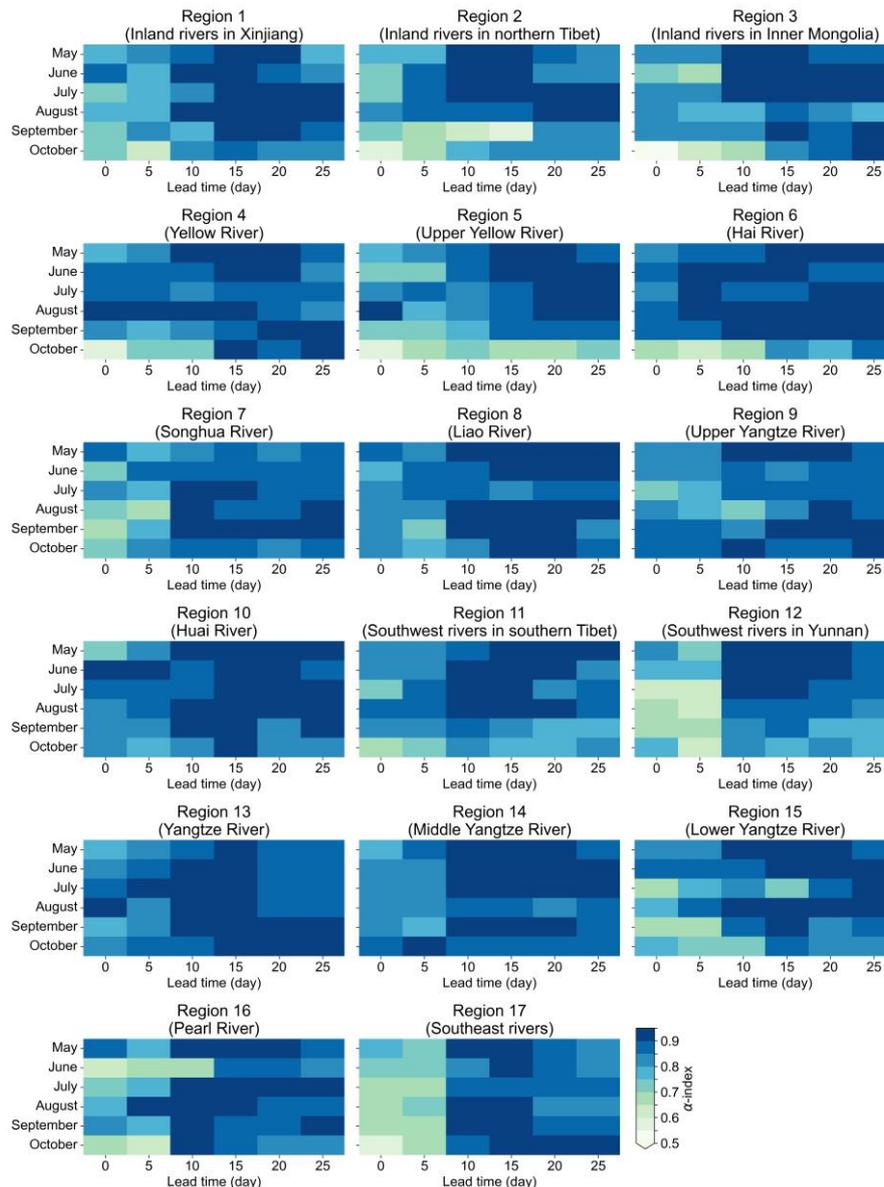
**Figure 10.** Same as Figure 9, but for α-index.

5. Discussion

5.1 Discussion of forecast skill

Through the STP-CBaM model display a good ability to generate skillful and reliable sub-seasonal precipitation forecasts over China, the forecast skill shows great diversity at different regions, months, and lead times. The calibration model shows highest forecast skill compared to bridging models for all regions and all months when the lead time is within 5 days (Fig S9~S13). The U200 and OLRA based bridging models outperform the calibration model and other bridging models when the lead time is beyond 10 days in certain months and certain regions. This may be explained by the strong relationship between U200, OLRA and precipitation

19

385   anomalies, and forecast skill of U200 and OLRA in the ECMWF model in these regions (Fig. S5~Fig. S8).

However, we also note that there are several regions where the forecast skill of calibration model is higher than the bridging models at longer lead times. This may be caused by the autocorrelations of sub-seasonal precipitation anomalies defined in this study. In our data processing section, the observed anomalies before the forecast initial date are used to make enough data for the running mean. Thus, the predictand is not purely

390   based on the ECMWF raw forecasts. The observational data is also introduced. The preceding observed precipitation anomalies may provide useful forecast information when the autocorrelations are high.

In addition, limited forecast skill of large-scale circulations at mid-high latitudes in dynamical models may contribute to limited forecast skill of bridging models as well. To figure out the potential skill of the STP-CBaM

395   method for predicting sub-seasonal precipitation anomalies, we use the pth 10-60-day signal of atmospheric field derived from ERA5 reanalysis dataset as predictor for the bridging model, instead of the atmospheric field derived from the ECMWF model. Thus, the potential forecast skill is based on ERA5 reanalysis data, while the practical forecast skill is based on ECMWF model. Figure 11 compares the potential CRPS skill score (based on the ERA5 reanalysis) and practical CRPS skill score (based on the ECMWF model) of merged forecasts

400   over China. The potential CRPS skill scores are similar to the practical CRPS skill scores as the precipitation forecasts derived from the ECMWF model are of high accuracy at short lead times. The potential CRPS skill scores are much higher than the practical CRPS skill scores at longer lead times. This indicates that the forecast skill will be greatly improved when the atmospheric field is well predicted in the GCMs.

405
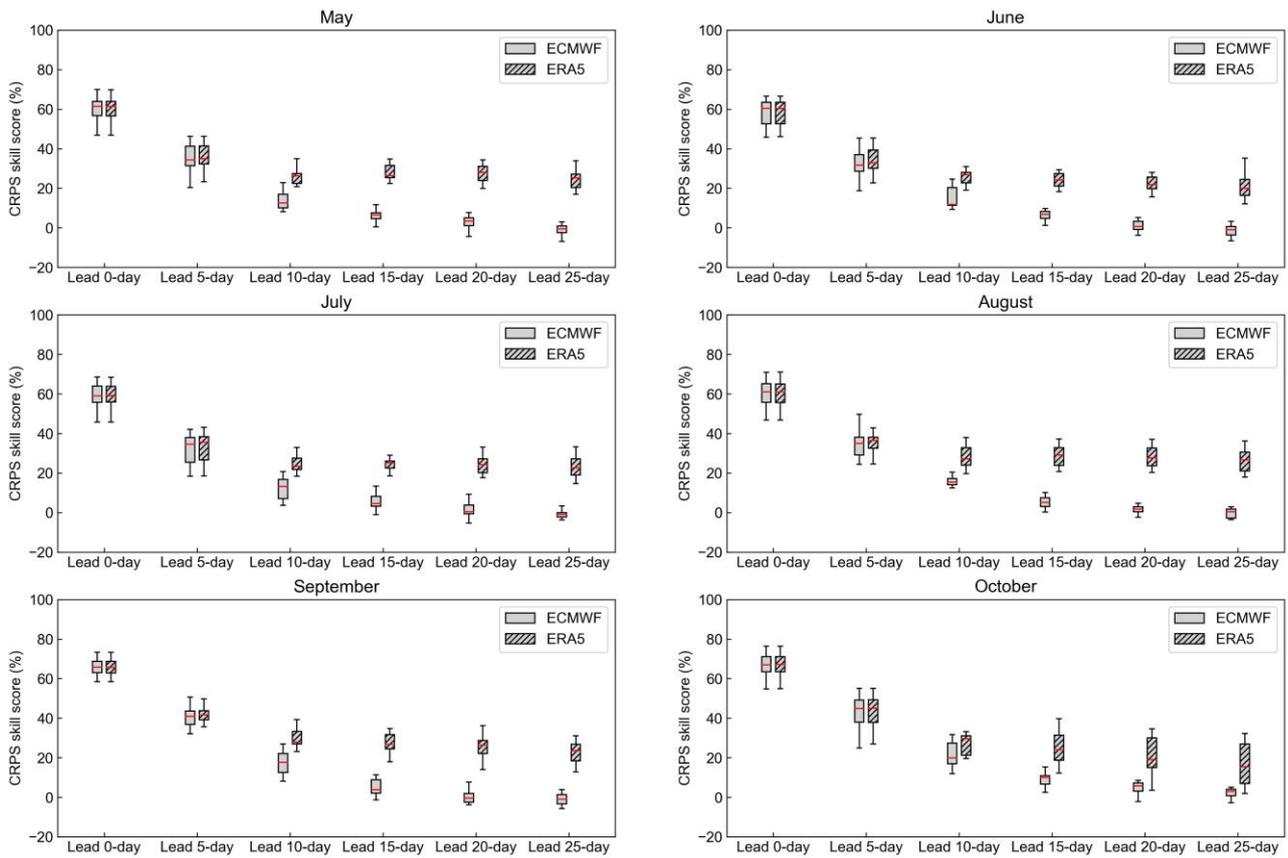
Figure 11. Practical CRPS skill score of merged forecasts based on the ECMWF model (solid) and potential CRPS skill score based on ERA5 reanalysis (hatched). The red lines are the 50th percentiles, top and bottom of each box are the 75th and 25th percentiles, and whiskers are the maximum and minimum skill scores.

410

We also note that the weights are not always match the skill patterns. In this study, the posterior distributions of model weights are assigned by the model predictive ability rather than fitting ability. Indeed, there is much literature in support of using predictive performance measures for model choice and combination based on the idea that a model is only as good as its predictions (Stock and Watson, 2006; Eklund and Karlsson, 2007). 415 Thus, the CRPS skill score is not used when inferring model weights. This may lead to the discrepancy between model weights and forecast skill score, especially when none of the models show high predictive skill.

5.2 Limitations and future work

In this study, we aim at investigating the capability of dynamical models for improving the forecast skill of sub- 420 seasonal precipitation anomalies using large-scale circulations as predictors. The bridging models are built based on the concurrent relationships between atmospheric intraseasonal oscillations and precipitation

anomalies. Thus, the forecast skill of bridging models is highly reliant on the forecast skill of atmospheric intraseasonal oscillations derived from dynamical models. In the future, the lagged relationships between atmospheric intraseasonal oscillations and precipitation anomalies will be considered to further improve sub-

425   seasonal precipitation forecast skill. Although the forecast skill of calibration model is high at short lead times, the results also suggest that the calibrated forecasts are too narrow (too confident). We would like to focus on improving the forecast reliability especially at short lead times in the future.

Meanwhile, we define the predictors using the STPM method for each month and each hydroclimatic region.

430   Intraseasonal climate indices, such as the MJO index and BSISO index, have not been considered yet. Recently, Zhu et al. (2023) proposed two sets of novel indices based on the compound zonal displacements of the South Asia high (SAH) and the western Pacific subtropical high (WPH) to monitor and predict the intraseasonal variation of Meiyu. These climate indices will be introduced in the bridging models to investigate the potential improvement for forecast skill.

435

In addition, we mainly focus on the prediction of intraseasonal (10-60 day) precipitation anomalies in this study. However, previous studies suggested that the intraseasonal component may only account for 7% of the total variability in northeastern China, while the seasonal component accounted for nearly 70%. Thus, the relationships between seasonal precipitation anomalies and large-scale circulation patterns should also be

440   investigated in these regions in the future.

6. Conclusions

In this study, we develop a STP-CBaM method to improve probabilistic sub-seasonal precipitation forecast skill over 17 hydroclimatic regions in China. The STP-CBaM method takes advantage of both dynamical models and statistical models. The calibration model is built by calibrating pentad mean precipitation anomalies derived

445   from the ECMWF model. Bridging models are built by defining potential predictors using the Spatial-Temporal Projection method (STPM) method. The calibration model and bridging models are merged through the Bayesian Modeling Averaging (BMA) method. Our results suggest that the forecast skill of calibration model is higher compared to bridging models when the lead time is within 5-10 days. The U200 and OLRA based bridging models outperform the calibration model when the lead time is beyond 10 days in certain months and

450   certain regions. The BMA merged forecasts take advantage of both calibration model and bridging models.

The BMA weights are rather stable at short lead times, which over 90% of the total weights are assigned to the calibration model. More weights are assigned to U200 and OLRA based bridging models when the lead time is beyond 10 days. The results of α-index suggest that BMA merged forecasts are reliable, especially at longer lead times.

455

Data availability

The ERA5 dataset can be sourced from https://cds.climate.copernicus.eu/, and the precipitation dataset is derived from http://www.gloh2o.org/mswep/. The outgoing Longwave radiation (OLR) dataset is sourced at https://www.ncdc.noaa.gov/news/new-outgoing-longwave-radiation-climate-data-record. The ECMWF

460   hindcast data can be retrieved from the S2S database at http://apps.ecmwf.int/datasets/data/s2s/.

Author contribution

Y.L.: conceptualization, methodology, and writing (original draft preparation, review, and editing). K.N. Xü: methodology and writing (review and editing). Z.Y. WU: conceptualization, funding acquisition, supervision, and writing (review and editing). Z.W. ZHU: conceptualization, review and editing. Q.J. WANG:

465   conceptualization, supervision, and writing (review and editing).

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

References

Chen, Y. and Zhai, P.: Simultaneous modulations of precipitation and temperature extremes in Southern parts of China by the boreal summer intraseasonal oscillation, Climate Dynamics, 49, 3363-3381, 2017.

Cui, J., Yang, S., and Li, T.: How well do the S2S models predict intraseasonal wintertime surface air

23

475   temperature over mid-high-latitude Eurasia?, Climate Dynamics, 57, 503-521, 2021.

de Andrade, F. M., Coelho, C. A., and Cavalcanti, I. F.: Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models, Climate Dynamics, 52, 5451-5475, 2019.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, Journal of hydrology, 158, 265-284, https://doi.org/10.1016/0022-
480   1694(94)90057-4, 1994.

Eklund, J. and Karlsson, S.: Forecast Combination and Model Averaging Using Predictive Measures, Econometric Reviews, 26, 329-363, 10.1080/07474930701220550, 2007.

Guo, B., Xu, T., Yang, Q., Zhang, J., Dai, Z., Deng, Y., and Zou, J.: Multiple Spatial and Temporal Scales Evaluation of Eight Satellite Precipitation Products in a Mountainous Catchment of South China,
485   10.3390/rs15051373,   2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz‐Sabater, J., Nicolas, J., Peubey, C., Radu, R., and Schepers, D.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999-2049, 2020.

Hsu, P.-C., Li, T., You, L., Gao, J., and Ren, H.-L.: A spatial–temporal projection model for 10–30 day rainfall
490   forecast in South China, Climate Dynamics, 44, 1227-1244, 10.1007/s00382-014-2215-4, 2015.

Hsu, P. C., Lee, J. Y., and Ha, K. J.: Influence of boreal summer intraseasonal oscillation on rainfall extremes in southern China, Int. J. Climatol., 36, 1403-1412, 10.1002/joc.4433, 2016.

Jie, W. H., Vitart, F., Wu, T. W., and Liu, X. W.: Simulations of the Asian summer monsoon in the sub-seasonal to seasonal prediction project (S2S) database, Quarterly Journal of the Royal Meteorological Society, 143,
495   2282-2295, 10.1002/qj.3085, 2017.

Lee, S.-S., Wang, B., Waliser, D. E., Neena, J. M., and Lee, J.-Y.: Predictability and prediction skill of the boreal summer intraseasonal oscillation in the Intraseasonal Variability Hindcast Experiment, Climate Dynamics, 45, 2123-2135, 10.1007/s00382-014-2461-5, 2015.

Li, W., Yang, X.-Q., Fang, J., Tao, L., and Sun, X.: Asymmetric boreal summer intraseasonal oscillation events
500   over the western North Pacific and their impacts on East Asian precipitation, Journal of Climate, 36, 2645-2661, 2023a.

Li, Y., Wu, Z., He, H., and Yin, H.: Probabilistic subseasonal precipitation forecasts using preceding atmospheric intraseasonal signals in a Bayesian perspective, Hydrol. Earth Syst. Sci., 26, 4975-4994, 10.5194/hess-26-4975-2022, 2022.

Li, Y., Wu, Z., He, H., Wang, Q. J., Xu, H., and Lu, G.: Post-processing sub-seasonal precipitation forecasts at various spatiotemporal scales across China during boreal summer monsoon, Journal of Hydrology, 598, 125742, 2021.

Li, Y., Pang, B., Zheng, Z., Chen, H., Peng, D., Zhu, Z., and Zuo, D.: Evaluation of Four Satellite Precipitation Products over Mainland China Using Spatial Correlation Analysis, Remote Sensing, 15, 1823, 2023b.

Liu, B., Zhu, C., Ma, S., Yan, Y., and Jiang, N.: Subseasonal processes of triple extreme heatwaves over the Yangtze River Valley in 2022, Weather and Climate Extremes, 40, 100572, 2023.

Liu, F., Ouyang, Y., Wang, B., Yang, J., Ling, J., and Hsu, P.-C.: Seasonal evolution of the intraseasonal variability of China summer precipitation, Climate Dynamics, 54, 4641-4655, 2020.

Liu, J. and Lu, R.: Different Impacts of Intraseasonal Oscillations on Precipitation in Southeast China between Early and Late Summers, Advances in Atmospheric Sciences, 39, 1885-1896, 2022.

Liu, J., Shangguan, D., Liu, S., Ding, Y., Wang, S., and Wang, X.: Evaluation and comparison of CHIRPS and MSWEP daily-precipitation products in the Qinghai-Tibet Plateau during the period of 1981–2015, Atmospheric Research, 230, 104634, https://doi.org/10.1016/j.atmosres.2019.104634, 2019.

Livezey, R. E. and Chen, W. Y.: Statistical Field Significance and its Determination by Monte Carlo Techniques, Monthly Weather Review, 111, 46-59, 10.1175/1520-0493(1983)111<0046:SFSAID>2.0.CO;2, 1983.

Matheson, J. E. and Winkler, R. L.: Scoring rules for continuous probability distributions, Management science, 22, 1087-1096, https://doi.org/10.1287/mnsc.22.10.1087, 1976.

Peng, Z., Wang, Q., Bennett, J. C., Schepen, A., Pappenberger, F., Pokhrel, P., and Wang, Z.: Statistical calibration and bridging of ECMWF System4 outputs for forecasting seasonal precipitation over China, Journal of Geophysical Research: Atmospheres, 119, 7116-7135, 2014.

Ren, P., Ren, H. L., Fu, J. X., Wu, J., and Du, L.: Impact of boreal summer intraseasonal oscillation on rainfall extremes in southeastern China and its predictability in CFSv2, Journal of Geophysical Research: Atmospheres, 123, 4423-4442, 2018.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46, W05521, https://doi.org/10.1029/2009WR008328, 2010.

Schepen, A., Wang, Q., and Everingham, Y.: Calibration, bridging, and merging to improve GCM seasonal temperature forecasts in Australia, Monthly Weather Review, 144, 2421-2441, 2016.

Schepen, A., Wang, Q. J., and Robertson, D. E.: Seasonal Forecasts of Australian Rainfall through Calibration

535   and Bridging of Coupled GCM Outputs, Monthly Weather Review, 142, 1758-1770, 10.1175/MWR-D-13-00248.1, 2014.

Shibuya, R., Nakano, M., Kodama, C., Nasuno, T., Kikuchi, K., Satoh, M., Miura, H., and Miyakawa, T.: Prediction Skill of the Boreal Summer Intra-Seasonal Oscillation in Global Non-hydrostatic Atmospheric Model Simulations with Explicit Cloud Microphysics, Journal of the Meteorological Society of Japan. Ser.
540   II, 99, 973-992, 10.2151/jmsj.2021-046, 2021.

Slater, L. J., Arnal, L., Boucher, M. A., Chang, A. Y. Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., and Zappa, M.: Hybrid forecasting: blending climate predictions with AI models, Hydrol. Earth Syst. Sci., 27, 1865-1889, 10.5194/hess-27-1865-2023, 2023.

Specq, D. and Batté, L.: Improving subseasonal precipitation forecasts through a statistical–dynamical
545   approach : application to the southwest tropical Pacific, Climate Dynamics, 55, 1913-1927, 10.1007/s00382-020-05355-7, 2020.

Stock, J. H. and Watson, M. W.: Chapter 10 Forecasting with Many Predictors, in: Handbook of Economic Forecasting, edited by: Elliott, G., Granger, C. W. J., and Timmermann, A., Elsevier, 515-554, https://doi.org/10.1016/S1574-0706(05)01010-4, 2006.

550 Strazzo, S., Collins, D. C., Schepen, A., Wang, Q., Becker, E., and Jia, L.: Application of a hybrid statistical–dynamical system to seasonal prediction of North American temperature and precipitation, Monthly Weather Review, 147, 607-625, 2019a.

Strazzo, S., Collins, D. C., Schepen, A., Wang, Q. J., Becker, E., and Jia, L.: Application of a Hybrid Statistical?Dynamical System to Seasonal Prediction of North American Temperature and Precipitation,
555   Monthly Weather Review, 147, 607-625, 10.1175/MWR-D-18-0156.1, 2019b.

Vitart, F. and Robertson, A. W.: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events, npj Climate Atmospheric Science, 1, 1-7, 2018.

Wang, Q., Schepen, A., and Robertson, D. E.: Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging, Journal of climate, 25, 5524-5537, 2012.

560 Wang, Q., Shao, Y., Song, Y., Schepen, A., Robertson, D. E., Ryu, D., and Pappenberger, F.: 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, 2019.

White, C. J., Domeisen, D. I., Acharya, N., Adefisan, E. A., Anderson, M. L., Aura, S., Balogun, A. A., Bertram, D., Bluhm, S., and Brayshaw, D. J.: Advances in the application and utility of subseasonal-to-seasonal

565   predictions, Bulletin of the American Meteorological Society, 103, E1448-E1472, 2022.

Wu, J., Li, J., Zhu, Z., and Hsu, P.-C.: Factors determining the subseasonal prediction skill of summer extreme rainfall over southern China, Climate Dynamics, 60, 443-460, 10.1007/s00382-022-06326-w, 2023.

Wu, J., Ren, H.-L., Zhang, P., Wang, Y., Liu, Y., Zhao, C., and Li, Q.: The dynamical-statistical subseasonal prediction of precipitation over China based on the BCC new-generation coupled model, Climate

570   Dynamics, 59, 1213-1232, 2022.

Xie, J., Hsu, P.-C., Hu, Y., Ye, M., and Yu, J.: Skillful extended-range forecast of rainfall and extreme events in East China based on deep learning, Weather and Forecasting, 2023.

Yan, Y., Zhu, C., and Liu, B.: Subseasonal predictability of the July 2021 extreme rainfall event over Henan China in S2S operational models, Journal of Geophysical Research: Atmospheres, 128, e2022JD037879,

575   2023.

Zhang, K., Li, J., Zhu, Z., and Li, T.: Implications from Subseasonal Prediction Skills of the Prolonged Heavy Snow Event over Southern China in Early 2008, Advances in Atmospheric Sciences, 38, 1873-1888, 10.1007/s00376-021-0402-x, 2021.

Zhu, C., Liu, B., Li, L., Ma, S., Jiang, N., and Yan, Y.: Progress and Prospects of Research on Subseasonal to

580   Seasonal Variability and Prediction of the East Asian Monsoon, Journal of Meteorological Research, 36, 677-690, 2022.

Zhu, Z. and Li, T.: The statistical extended-range (10–30-day) forecast of summer rainfall anomalies over the entire China, Climate Dynamics, 48, 209-224, 10.1007/s00382-016-3070-2, 2017.

Zhu, Z., Li, T., Hsu, P.-c., and He, J.: A spatial–temporal projection model for extended-range forecast in the

585   tropics, Climate Dynamics, 45, 1085-1098, 10.1007/s00382-014-2353-8, 2015.

Zhu, Z., Zhou, Y., Jiang, W., Fu, S., and Hsu, P.: Influence of compound zonal displacements of the South Asia high and the western Pacific subtropical high on Meiyu intraseasonal variation, Climate Dynamics, https://doi.org/10.1007/s00382-023-06726-6, 2023.