The influence of 10–30-day boreal summer intraseasonal oscillation on the extended-range forecast skill of extreme rainfall over southern China

Zhiwei Zhu[1] · Junting Wu[2] · Hongjie Huang[1]

Abstract

Based on deterministic and probabilistic forecast verification, we investigated the performance of three subseasonal-to-seasonal (S2S) operational models, i.e., the model of the European Centre for Medium-Range Weather Forecasts (ECMWF) and two models of the China Meteorological Administration (CMA1.0 and CMA2.0), in the extended-range forecast of extreme rainfall over southern China (SCER) while considering the modulation of 10–30-day boreal summer intraseasonal oscillation (BSISO2). The Heidke Skill Score (HSS) of the SCER in the ECMWF, CMA2.0, and CMA1.0 models decreased to less than 0.1 at lead times of 13, 9, and 6 days, respectively. Similarly, the useful prediction skill of the BSISO2 index in the ECMWF, CMA1.0, and CMA2.0 models was up to 15, 13, and 8 days in advance, respectively. The BSISO2's phase error, rather than the amplitude error, determines its prediction skill. The HSS of the BSISO2 index is significantly correlated with that of SCER in all three S2S models, suggesting that the prediction skill of SCER is influenced by that of BSISO2. The ECMWF shows much higher skill than the two CMA models do in predicting the SCER probability changes under the influence of BSISO2 during Phases 5–7, with the useful prediction skill having up to a 10-day lead time. In contrast, CMA1.0 and CMA2.0 can only predict the modulation of BSISO2 on the SCER probability within a week. The prediction skill of BSISO2's modulation on SCER largely relies on moisture convergence, rather than on moisture advection. This study highlighted the importance of model's accurate representation of BSISO2 and its associated moisture convergence for improving extended-range forecast of SCER.

Keywords 10–30-day boreal summer intraseasonal oscillation · Extended-range forecast · Extreme rainfall over southern China

1 Introduction

In the context of global warming, extreme precipitation events have become China's most frequent high-impact weather events during the summertime (Li et al. 2016). The increased probability and intensity of extreme rainfall lead to agricultural failure, economic losses, and casualties. Traditional weather forecasts can provide relatively accurate information about extreme precipitation, but only within a very limited lead time (usually no longer than five days). Forecasting extreme rainfall at a lead time of 10–30 days (i.e., an "extended-range forecast") would make it possible to efficiently prepare for and adapt to upcoming adversity, which is important for disaster prevention and mitigation decision-making (White et al. 2017; He et al. 2020).

However, extended-range forecast is a challenging task, requiring accurate simulations of both the atmospheric initial conditions and boundary-layer forcing. The difficulty in creating extended-range forecasts is twofold. Firstly, the time range is so long that the contribution of the atmospheric initial conditions quickly declines with increased lead time, and forecast errors significantly increase, resulting in the upper

✉ Zhiwei Zhu
zwz@nuist.edu.cn

1   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, Ningliu Road 219, Nanjing 210044, China

2   Jinshan Meteorological Bureau of Shanghai, Shanghai 201508, China

limit of weather predictability being only within ~ 10 days. Secondly, this timescale is relatively short for boundary-layer forcing (e.g., sea-surface temperatures, snow cover, and soil moisture) to fully take effect (Liang and Lin 2018).

The boreal summer intraseasonal oscillation over the tropics (viz. BSISO) has been suggested to be the most significant source of extended-range forecast predictability. As an intraseasonal coupled mode of large-scale convection and circulation in the Asian monsoon region, the BSISO has complex spatiotemporal characteristics (Lee et al. 2013) and is distinct from the Madden–Julian Oscillation (MJO) in the boreal winter (Madden and Julian 1971). It propagates northeastward from the Indian Ocean to the western Pacific (Jiang et al. 2004) with a 30–90-day period (the first mode of the BSISO, BSISO1) and propagates from the tropical western Pacific northwestward with a shorter period of 10–30 days (the second mode of the BSISO, BSISO2). Both BSISO1 and BSISO2 have significant impacts on East Asian summer climate anomalies (Oh and Ha 2015), especially with respect to the rainfall anomalies over southern China (Ren et al. 2018). Based on the BSISO indices, Hsu et al. (2016) suggested that the probability of extreme rainfall occurring over southern China (SCER) during Phases 2–4 of BSISO1 and Phases 5–7 of BSISO2 is significantly increased compared with in other phases. Using this robust statistical relationship between BSISO and extreme rainfall, hindcast experiments have been conducted with approximately two-week lead time prediction skill of the BSISO's modulation of extreme rainfall over East Asia (Lee et al. 2017).

To what extent can current dynamic models simulate and predict the BSISO and its associated precipitation anomalies? Compared with the relatively longer lead time (up to 24.5 days) of the BSISO1 prediction skill, the BSISO2 prediction skill index from the S2S models is limited to within 14 days (Jie et al. 2017). Zhao et al. (2014) suggested that four Chinese general circulation models (GCMs) participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5) nicely reproduced the BSISO signals, but they all overestimated the intensity and duration of the BSISO-related precipitation over the tropics. Based on 27 GCMs, Neena et al. (2017) suggested that, although most models captured the propagation of BSISO, the bias is still considerable in the BSISO-related precipitation. Wang et al. (2020) suggested that the S2S models' prediction skill of the precipitation anomaly around the Maritime Continent was relatively high at lead times during the first week but dropped dramatically in the second week, and the prediction skill relied on the phase and intensity of the BSISO.

Previous studies have evaluated the models' ability to simulate either the BSISO modes or the mean precipitation anomalies related to the BSISO (Zhao et al. 2014; Wang et al. 2020), but less attention has been paid to the prediction skill of the BSISO modulation of extreme rainfall.

Influenced by both the southwest monsoon and mid-latitude synoptic systems, southern China (SC) is prone to extreme rainfall in boreal summer (Zhu et al. 2023; Li et al. 2023; Li et al. 2017), which often causes flood, mudslide, urban waterlogging, leading to causalities and property losses. As the probability of SCER is notably modulated by BSISO, unraveling the capacity of S2S models in forecasting BSISO and its influence on SCER is of great importance to disaster prevention and mitigation (Wu et al. 2023). Given that SCER is more closely associated with the quasi-biweekly oscillations that occur in a 10–30-day period (Hsu et al. 2016; Lee et al. 2017), it is necessary to unravel the S2S models' prediction skill in monitoring the BSISO2's modulation of SCER. This would provide the theoretical basis for extended-range forecast of SCER using BSISO2, and a point-of-reference for improving the dynamical prediction of SCER.

Using datasets from the three S2S models (ECMWF and two CMA models), this study will evaluate the prediction skill of BSISO2 and its influence on SCER based on both deterministic and probabilistic verification. The remainder of this paper is organized as follows. Section 2 describes the data, methods, and verification metrics used in the study. Section 3 provides the forecast verification of SCER and BSISO2. Section 4 presents the forecast verification of BSISO2's modulation on SCER, and the possible origin of the prediction error in BSISO2's modulation on SCER is revealed in Sect. 5. The final section provides a discussion and conclusions.

2 Data and methods

2.1 Observation and reanalysis data

The observation and reanalysis datasets employed in this study include: (1) the daily-mean precipitation data from gauge stations throughout China, which are gridded with a horizontal resolution of 0.25° × 0.25° (CN05.1) (Wu and Gao 2013); (2) the Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation (APHRODITE) gridded precipitation (Yatagai et al. 2012) with a 0.25° horizontal resolution; (3) the ERA-Interim reanalysis dataset (Dee et al. 2011), including the daily-mean wind and specific humidity with a 1.5° horizontal resolution; and (4) the real-time BSISO index (http://iprc.soest.hawaii.edu/users/jylee/bsiso/) (Lee et al. 2013). To reduce the uncertainty among different datasets, a simple arithmetic average of the two precipitation datasets is applied.

In the present study, southern China denotes the region of 18°–32.5° N, 105°–122° E, and the summer season is the extended boreal summer from May to August (hereafter,

MJJA). Ten summers during the period of 2005–2014 are selected for prediction verification.

2.2 S2S model data

The reforecast data of three operational models from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the China Meteorological Administration (CMA) are derived from the S2S database (http://s2s.cma.cn/index). Details on the reforecast data from the three S2S models are listed in Table 1.

In this study, the prediction skill for the boreal summer SCER and its probability modulated by BSISO2 during the overlapping period of the three models (2005–2014), are assessed. Because of the different reforecast frequencies (the ECMWF and CMA2.0 model are initialized twice weekly whereas the CMA1.0 model is daily initialized), to fairly assess the prediction skill of the three models, a data processing method proposed by Yang et al. (2018) is applied to rearrange the twice-weekly data of the ECMWF and CMA2.0 model to a daily interval. Briefly, the data in N-2 to N+2 days of each prediction is used as the results of the N-day lead reforecasts. An arithmetic average is then applied if there are two values for one specific lead time forecast. By applying this approach, the new data array contains a consecutive distribution of the 3-day lead to 42-day lead forecast at all lead times.

2.3 Methods

2.3.1 Definition of extreme rainfall

We adopted the percentile threshold relative to climatological rainfall distributions in the individual dataset to define the observed and modeled extreme rainfall. The reason for the option of definition is twofold. First, given that models generally produce systematic errors when representing extreme amplitudes, it is more comparable for observed and modeled extreme rainfall when percentile-based threshold is adopted. Second, because every grid has a percentile threshold, the percentile-based definition could lead to more evenly spatial-distributed extreme rainfall. Therefore, following the methods of previous studies (Zhang et al. 2011;

Xavier et al. 2014), a percentile-based threshold is applied for the definition of extreme rainfall.

Considering the remarkable annual cycle of extreme rainfall occurrence, the criterion of the 90th percentile is defined for each day during MJJA of 2005–2014. For each grid point, the precipitation of one selected day (excluding precipitation amounts < 0.1 mm and the missing value) and 90 adjacent days (91 days in total) in all years of 2005–2014 is firstly arranged in ascending order. Then, the 90th percentile value of the 910-days record is defined as the extreme rainfall threshold of the selected day in the specific grid. When the precipitation on a certain day is higher than the corresponding threshold at the grid point, this indicates that an extreme rainfall has occurred at the grid point on that day (Li and Wang 2018).

2.3.2 BSISO2 and its phases

To capture BSISO activities, two BSISO indices based on multivariate empirical orthogonal function (MV-EOF) analysis of daily anomalies of outgoing longwave radiation (OLR) and 850-hPa zonal wind over the Asian monsoon sector (10°S–40°N, 40°E–160°E) (Lee et al. 2013) are employed in this study. The first mode of BSISO (BSISO1) activities is defined by the first two principal components (PCs) of the MV-EOF, and the second mode of BSISO (BSISO2) are represented by the third and fourth PCs (PC3 and PC4). In the present study, we only focused on BSISO2. By constructing the phase diagram of the corresponding time series of PC3 and PC4, the life cycle of the BSISO2 can be divided into eight phases. An active BSISO2 case is identified when its amplitude is greater than 1 (i.e., $\sqrt{PC3^2 + PC4^2} > 1.0$), whereas the amplitude less than 1 is considered as the non-BSISO2 period ($P_{non-BSISO2}$).

2.3.3 Verification metrics

To quantitatively assess the prediction skill of the models, two types of metrics are used to evaluate the performance of the S2S models: deterministic and probabilistic metrics. For the deterministic metrics, to compare the similarity and deviation for the spatial patterns between the observed and forecast fields, the pattern correlation coefficient (PCC) and normalized root-mean-square error (NRMSE) are applied

Table 1 Description of the ECMWF and CMA models

Model	Model version	Time range (days)	Model resolution	Reforecast frequency	Available period	Ensemble size
ECMWF	CY47R1	0–46	Tco639/Tco319, L91	2/week	1997–2016	11
CMA1.0	BCC-CPS-S2Sv1	0–60	T106, L40	Daily	1995–2014	4
CMA2.0	BCC-CPS-S2Sv2	0–60	T266, L56	2/week	2005–2019	4

(Lee and Wang 2014). NRMSE is calculated using the standard deviation in the observations.

The bivariate temporal anomaly correlation coefficient (ACC) is used to evaluate the models' prediction skills in reproducing the BSISO2 index at different lead times (Lin et al. 2008; Gottschalck et al. 2010).

$$ACC = \frac{\sum_{t=1}^{T}[F_{1t}O_{1t} + F_{2t}O_{2t}]}{\sqrt{\sum_{t=1}^{T}[F_{1t}^2 + F_{2t}^2]}\sqrt{\sum_{t=1}^{T}[O_{1t}^2 + O_{2t}^2]}} \quad (1)$$

F and O denote forecasted and observed BSISO2 index. Here, t is time and T indicates the total number of forecast times. The subscripts 1 and 2 refer to different variables (such as PC3 and PC4).

Because errors in the amplitude or phase of the BSISO2 could result in decreased prediction skill, the BSISO2 index is rewritten in the form of polar coordinates as F (f, θ), and O (o, φ) (Wang et al. 2019) to distinguish the relative contributions of the amplitude and phase to the prediction skill (Wu et al. 2023), which is:

$$ACC = \frac{\sum_{t=1}^{T} f_t \cdot o_t \cdot \cos(\theta_t - \varphi_t)}{\sqrt{\sum_{t=1}^{T} f_t^2}\sqrt{\sum_{t=1}^{T} o_t^2}} \quad (2)$$

Here, f and o are amplitude, and θ and φ refer to phase angles for predictions and observations, respectively. This expression makes it clear to separate the contributions of amplitude and phase to ACC skill.

If the phase is perfectly forecasted, i.e., $\cos(\theta_t - \varphi_t) = 1$, ACC is completely determined by amplitude, which is:

$$ACC_a = \frac{\sum_{t=1}^{T} f_t \cdot o_t}{\sqrt{\sum_{t=1}^{T} f_t^2}\sqrt{\sum_{t=1}^{T} o_t^2}} \quad (3)$$

If there is no amplitude error, i.e., the linear correlation between f_t and o_t equals to 1, ACC is then only depend on the phase relation between the predictions and observations:

$$ACC_p = \frac{\sum_{t=1}^{T} o_t^2 \cdot \cos(\theta_t - \varphi_t)}{\sum_{t=1}^{T} o_t^2} \quad (4)$$

We use the BSISO2 time series (PC3 and PC4) during summers of 2005–2014 between the three models' predictions and observations at all leads to calculate the ACCs. At each lead time, 2460 samples are used to calculate the ACC.

From the perspective of probabilistic prediction, the SCER prediction skill of the S2S models is measured by the Heidke Skill Score (HSS) (Heidke 1926). The HSS is usually applied in evaluating the accuracy of model prediction after removing accidental events. In this study, it reflects the performance of the S2S models in reproducing

the occurrence of extreme precipitation events. The value range of HSS is (− ∞, 1). A negative HSS denotes a forecast worse than a random forecast, while a HSS of zero indicates no skill. A higher positive HSS indicates a better forecast for concurrent extremes.

3 Forecast verification of SCER and BSISO2

3.1 Prediction skill of the climatological mean and variation of summer rainfall

Figure 1 shows the observed and forecast spatial distribution of the climatological mean and the daily standard deviation of precipitation in the summer. Owing to the influence of the East Asian summer monsoon, the maxima of the climatological mean and standard deviation of precipitation both appear in southern China (Fig. 1a, f). Three S2S models are all able to generally capture the spatial precipitation distribution, with the precipitation decreasing from southeastern to northwestern China for the climatological and daily variation of precipitation. The systematic errors of the S2S models mainly occur in the magnitude of the mean and standard deviation of the precipitation. The ECMWF (CMA1.0) overestimates (underestimates) the mean and standard deviation of the precipitation in southern China, while the CMA2.0 has the best performance. At a forecast lead time of 20-days (Fig. 1e, j), the ECMWF shows the best performance for the spatial distribution of climatology (PCC = 0.82, NRMSE = 0.48) and standard deviation (PCC = 0.86, NRMSE = 0.35) of precipitation.

Consistent with the results of the spatial distribution, the areal averages of the prediction skill for the mean and standard deviation of precipitation over southern China from the ECMWF and CMA2.0 are still excellent, with PCCs larger than 0.5 and NRMSEs smaller than 1.0 throughout all lead times (Fig. 2a, b). CMA1.0 shows no skill beyond a 3-days lead time. Compared with CMA2.0, the ECMWF has better performance within a lead time of 12-days, but it decreases rapidly at a 12–18-days lead time (Fig. 2a).

3.2 Prediction skill of the SCER

Consistent with the biases in the climatological mean precipitation (Fig. 1a–e), the ECMWF (CMA1.0) model systematically overestimates (underestimates) the regionally averaged mean of the 90th percentile threshold over southern China (Fig. 2c). CMA2.0 has the best performance in simulating the 90th percentile threshold, which is much closer to the observational values at all 5–30-days lead times. Based on the threshold at each grid in each model, the prediction skill of the three S2S models for the occurrence of SCER is calculated. Figure 2d shows that the HSS of the SCER
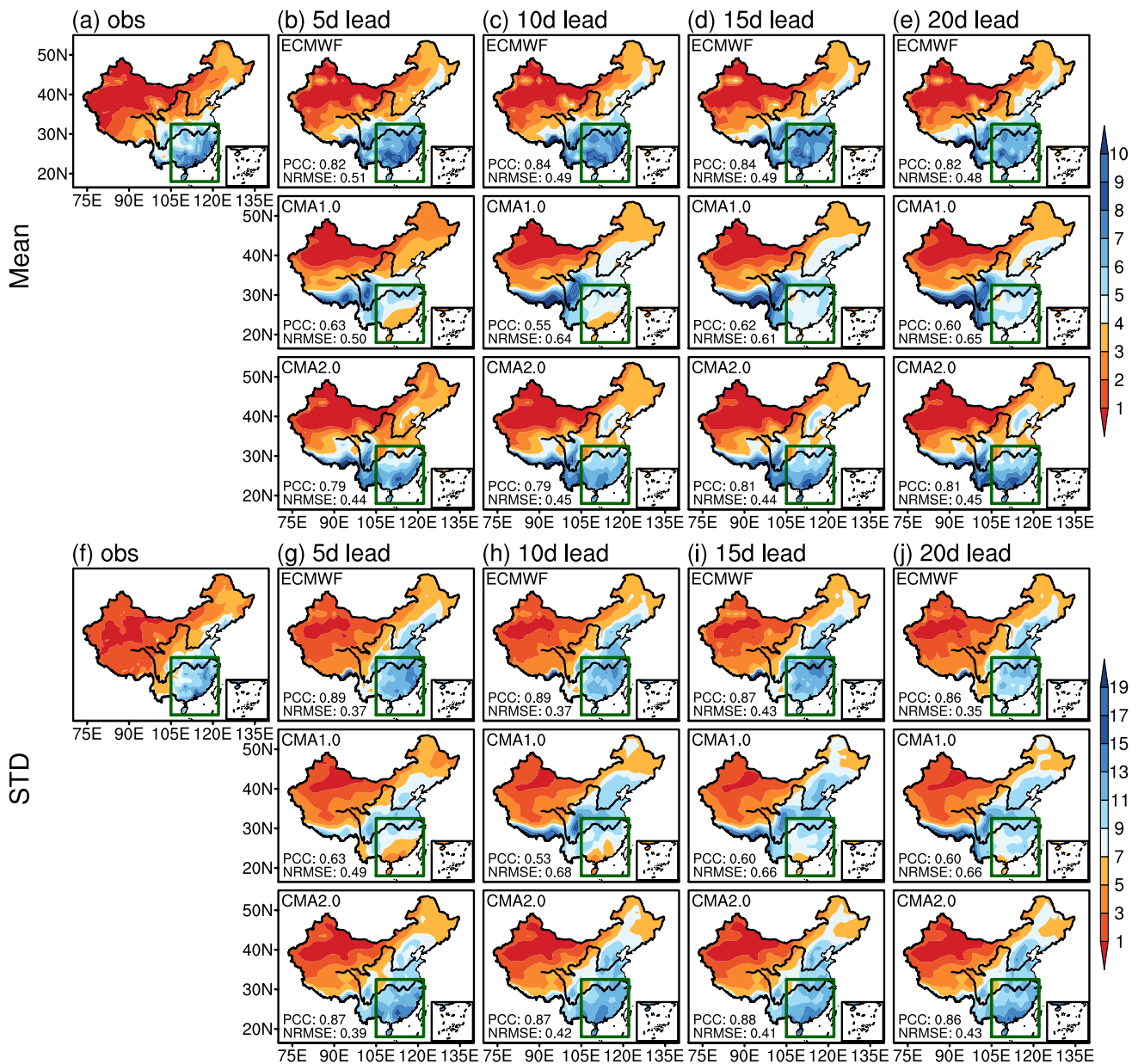
Fig. 1 Distribution of the **a** observed and three S2S model **b** 5-days, **c** 10-days, **d** 15-days, and **e** 20-days lead time forecast climatological summer mean precipitation (shading, units: mm day^{-1}) during 2005–2014. **f–j** are same as **a–e** but for daily standard deviation of the precipitation. The PCC and NRMSE skills over the mainland China are shown in the bottom-left of each panel. The green box delineates the domain of southern China (18°–32.5° N, 105°–122° E). The black lines outline the Yellow River and Yangtze River

decreases rapidly within a 15-days lead time for both the individual members and multi-member ensemble (MME). However, note that the MME always outperforms the individual members, especially for the ECMWF. Taking an HSS of 0.1 as the threshold for prediction skill, the MME of CMA1.0 and CMA2.0 can predict the SCER occurrence 6 and 9 days in advance, respectively, while the ECMWF has the best skill up to a 13-days lead time (Fig. 2d).

Figure 3 indicates the spatial distribution of the HSS for SCER in the three S2S models. The two CMA models show relatively lower HSS skill than the ECMWF does at all lead times, and they have no skill in predicting the occurrence of the SCER beyond a lead time of 14 days. However, the CMA2.0 has an HSS prediction skill comparable with that of the ECMWF at 5-days and 10-days lead times. All three S2S models have poor prediction skill for SCER beyond a 15-days lead time.
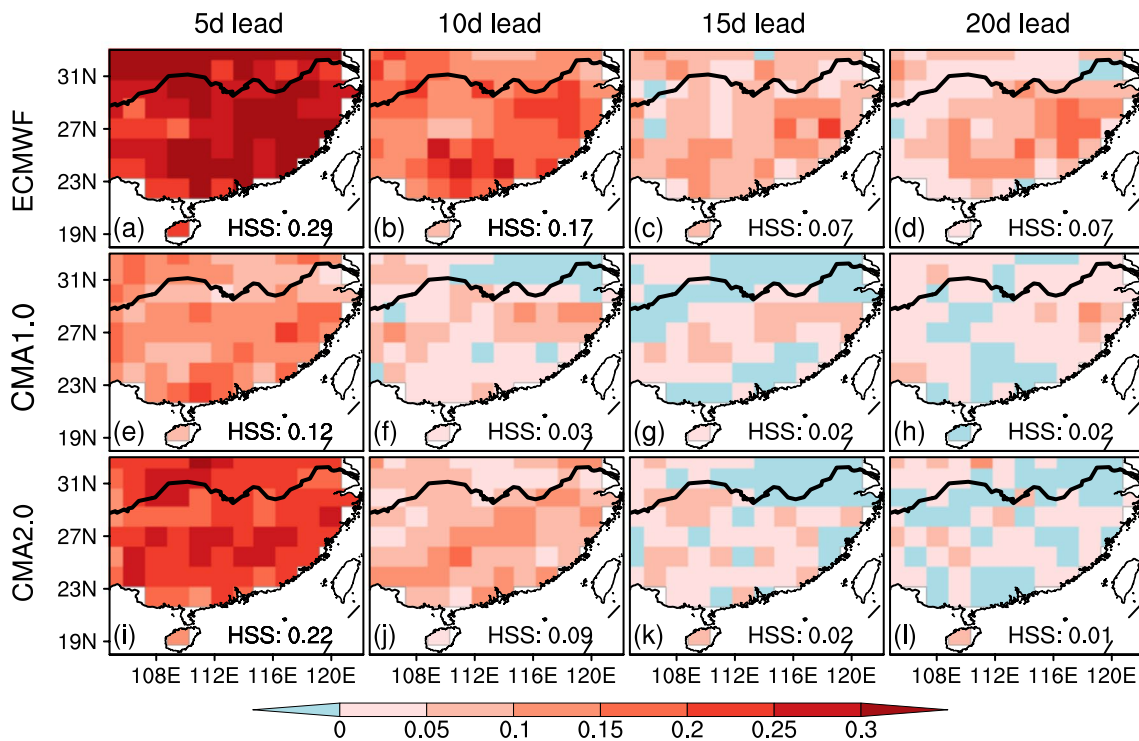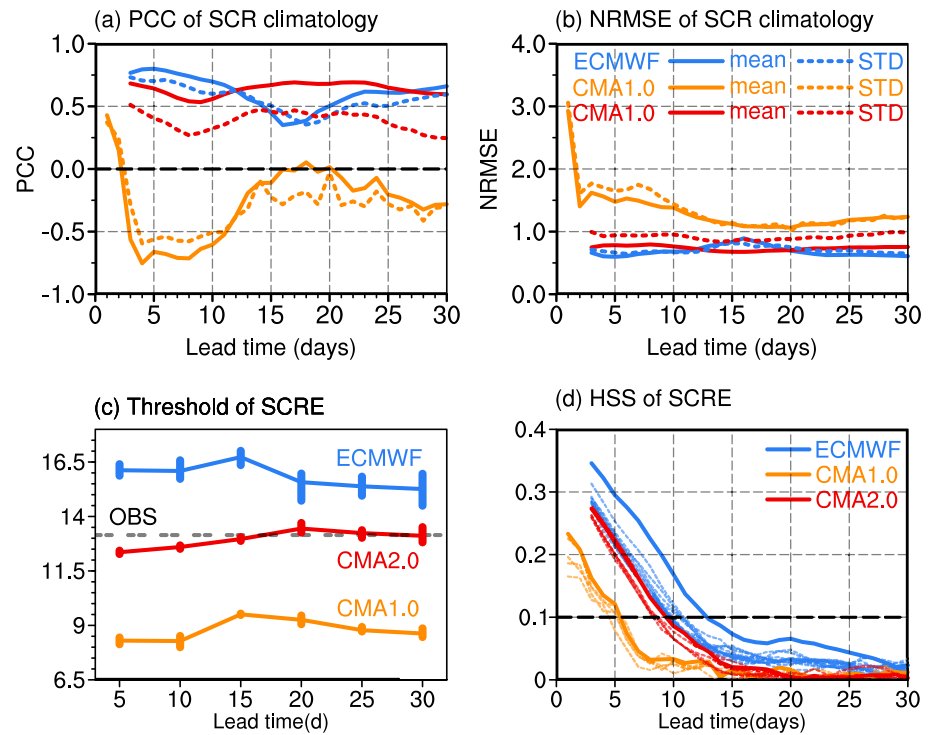
Fig. 3 HSS distribution (shading) of SCER at **a** 5-days, **b** 10-days, **c** 15-days, and **d** 20-days lead times in ECMWF. **e–h** and **i–l** are same as **a–d** but for CMA1.0, and CMA2.0, respectively. The areal-mean HSS over SC is shown in each panel. The black line outlines the Yangtze River
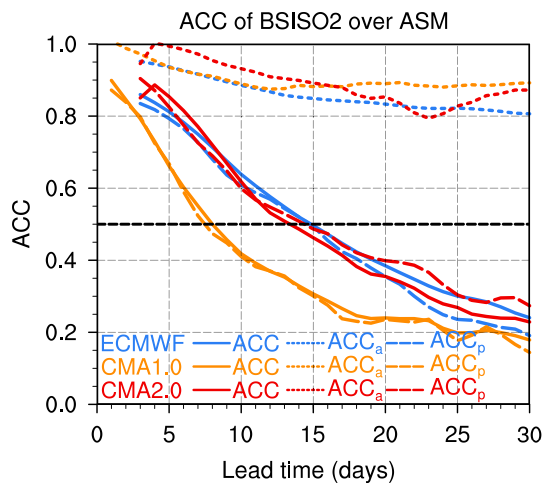
Fig. 4 The bivariate anomaly correlation coefficient (ACC) for the forecast BSISO2 indices from the ECMWF (solid blue line), CMA1.0 (solid orange line), and CMA2.0 (solid red line) during the summers (MJJA of 2005–2014 as a function of the forecast lead time) (in days). The short and long dashed lines represent the ACC with the assumption of perfect phase (ACCa) and perfect amplitude (ACCp) predictions, respectively

3.3 Prediction skill of the BSISO2 index

Given that the BSISO activities play an important role in modulating the intraseasonal variation of precipitation in southern China (Hsu et al. 2016), the prediction skill of BSISO may directly relate to that of SCER. Therefore, in this section, the models' performances in predicting the

BSISO2 index are evaluated. The BSISO2 index is obtained from projecting the model-predicted OLR and 850-hPa zonal wind anomalies onto the observed empirical orthogonal function (EOF) spatial pattern from 2005 to 2014.

Figure 4 shows that the ACC skill of the BSISO2 index from the three S2S models decreases with increased lead time. Taking an ACC of 0.5 as the threshold for useful prediction skill (Liebmann and Smith 1996; Xiang et al. 2015), the ECMWF and CMA2.0 can accurately reproduce the BSISO2 index at 15-days and 13-days lead times, respectively. CMA1.0 can only reproduce the index at an 8-days lead time. The ACC skill values with perfect amplitudes show results similar to those in the ECMWF and CMA1.0 models, and a 2-days improvement can be seen in the CMA2.0 model. If the phase of the BSISO2 can be perfectly predicted, the ACCs are always higher than 0.8 up to 30 days in advance. These results indicate that the amplitude error of the BSISO2 does not influence the final prediction skill, while the accuracy in predicting the phase of the BSISO2 is crucial.

4 Forecast verification of BSISO2's modulation of SCER

Figure 5 shows the observed percentage changes of the occurrence probability of extreme rainfall during each phase (Phases 1–8) of BSISO2 compared with weak/no BSISO2 activities (non-BSISO). The distribution of the probability
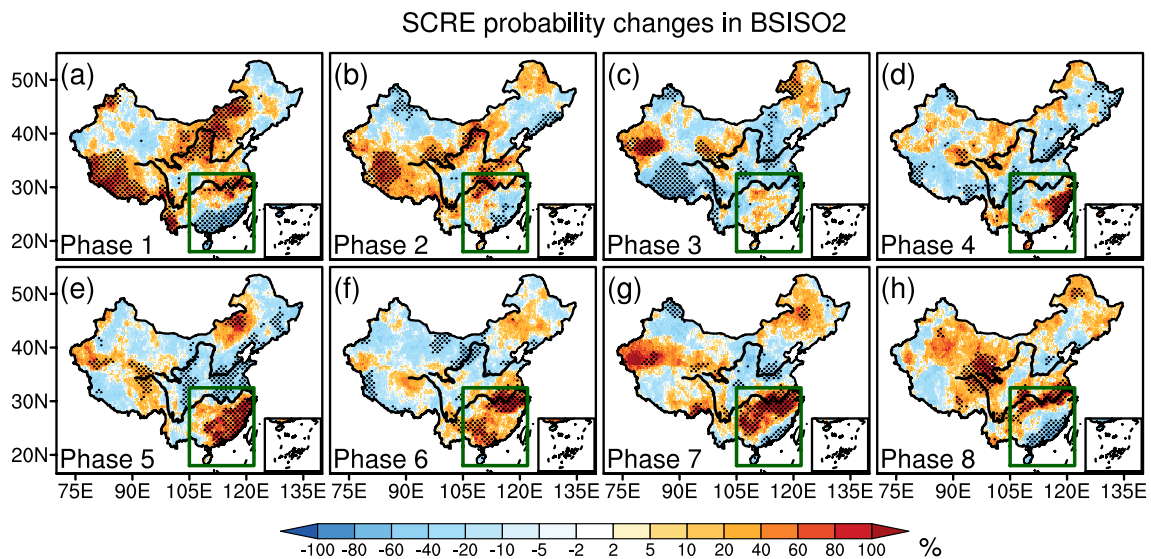

Fig. 5 The percentage changes (%) in extreme rainfall occurences probability in China from Phase 1 to Phase 8 (**a**–**h**) of the BSISO2 with respect to the non-BSISO2 period. The percentage change in the probability of extreme rainfall occurrence during each of the BSISO2 phases is calculated as $[(P_X - P_{non-BSISO2})/P_{non-BSISO2}]$, where $P_{non-BSISO2}$ and P_X represents the probability of extreme rainfall during the non-BSISO2 period and Phase X. Changes exceeding the 95% confidence level are dotted. The green box delineates the domain of southern China (18°–32.5° N, 105°–122° E). The black lines outline the Yellow River and Yangtze River

of extreme rainfall over China shows significant differences with BSISO2's phases. In general, the most significant changes in the probability of extreme precipitation appear over southern China during Phases 5–7 of BSISO2 (Fig. 5e-g). The increased probability (more than 40%) of extreme rainfall propagates from the southeast coast (Fig. 5e) to the middle and lower reaches of the Yangtze River Basin (Fig. 5g) from Phase 5 to Phase 7, corresponding to the propagation features of BSISO2 (Yang et al. 2010; Lee et al. 2013).

With the observed evident modulation of BSISO2 on SCER probability, a natural question arises: does the prediction skill of BSISO2 contribute to the prediction skill of SCER? To examine their relationship, the linear correlations between the HSS of the BSISO2 index and the SCER in three S2S models are calculated. Figure 6 shows that the HSS of the BSISO2 index is significantly correlated with the HSS of SCER in all three S2S models. The correlation coefficients are 0.97 in ECMWF and 0.94 in both CMA1.0 and CMA2.0, passing the 99% confidence level. Although the three models have different capacities in capturing the SCER and BSISO2 (see triangle symbols for the averaged HSS in each model in Fig. 6), they all present a good linear relationship between the HSS of the BSISO2 and the SCER, suggesting that the prediction skill of BSISO2 in the three S2S models may indeed influence that of SCER.
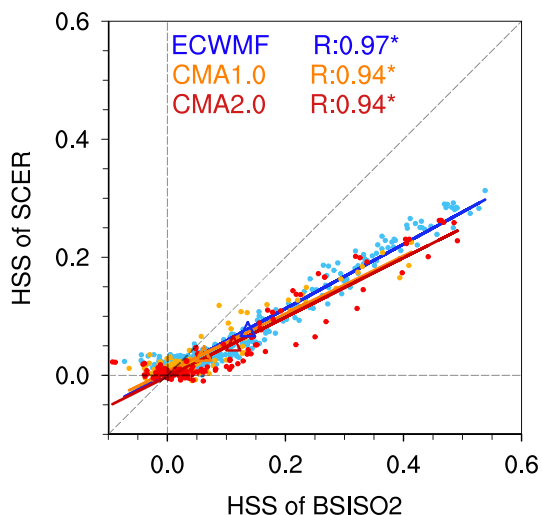

Fig. 6 Scatter diagrams for the areal-mean HSS skill of SCER (y-axis) against the BSISO2 indices (x-axis) for all individual members at all forecast lead times from the three models. The linear-fit curves for the ECMWF (308 blue dots), CMA1.0 (120 orange dots), and CMA2.0 (120 red dots) are in blue, orange, and red, and the blue/orange/red triangle symbols are the averaged HSS for the ECMWF/CMA1.0/CMA2.0, respectively. The correlation coefficients between the HSS of SCER and BSISO2 are given in the upper corner, and asterisks indicate coefficients exceeding the 99% confidence level

Given that the most significant changes in the probability of extreme precipitation appear over southern China during Phases 5–7 of BSISO2 (Hsu et al. 2016; Ren et al. 2018), whether the models are able to reproduce the modulation of BSISO2 on SCER probability during Phases 5–7 of BSISO2 contributed to the prediction skills of SCER. Could the models capture the BSISO2's modulation on the SCER probability during Phases 5–7? Fig. 7 shows the probability of SCER during Phases 5–7 of the BSISO2 in the three S2S models at 5-days, 10-days, 15-days, and 20-days lead times. The ECMWF model could predict the increased extreme precipitation probability over the southeast coast in Phase 5, over southwest of southern China and the lower reaches of the Yangtze River Basin in Phase 6, and in the middle and lower reaches of the Yangtze River Basin in Phase 7 at lead times within 20 days. However, beyond 10-d lead times, the ECMWF model overestimates the changes in extreme precipitation probability in the Yangtze River Basin during Phases 6–7. In contrast, the two CMA models show poor SCER prediction skill in Phases 5–7 of the BSISO2. During Phase 5, the two CMA models both underestimate the change in extreme rainfall over the southeast coast even at a 10-days lead time. During Phase 6, the ECMWF underestimates the probability over the lower reaches of the Yangtze River Basin at a 5-days lead time and overestimates it beyond a 5-days lead time; the two CMA models demonstrate lower skills in predicting the distribution of the probability, especially CMA1.0. In Phase 7, while the ECMWF can still nicely capture the pattern of the probability of extreme rainfall over the Yangtze River Basin, both CMA1.0 and CMA2.0 underestimate the probability of extreme rainfall (Fig. 7).

Figure 8 shows the evolution of the area-averaged deterministic prediction skill (i.e., PCC and NRMSE) of the SCER probability under Phases 5–7 of the BSISO2. In general, both the PCC and NRMSE decrease with increased lead time. In Phase 5 (Fig. 8a), the ECMWF obviously outperforms the CMA models. In Phase 6 (Fig. 8b), none of these models have PCC larger than 0.5, and the NRMSE is also quite large, due to the complex pattern of the observed probability changes with two separate maximum centers. Owing to overestimation of the probability changes, the NRMSE in the ECMWF is especially large (Fig. 8b). In Phase 7 (Fig. 8c), the ECMWF has excellent PCC skills up to a 30-days lead time, while the two CMA models show no skill for the distribution of the SCER probability for any lead time. Note that the spreads are evident in the PCC skill among the members in each model, suggesting the large uncertainty in predicting the SCER probability change.

Figure 9 shows the spatial distribution of the SCER HSS score in Phases 5–7 of BSISO2 at 5-, 10-, 15-, and 20-days lead times in the three models. The HSS skill in some regions drops dramatically after a 15-days lead time
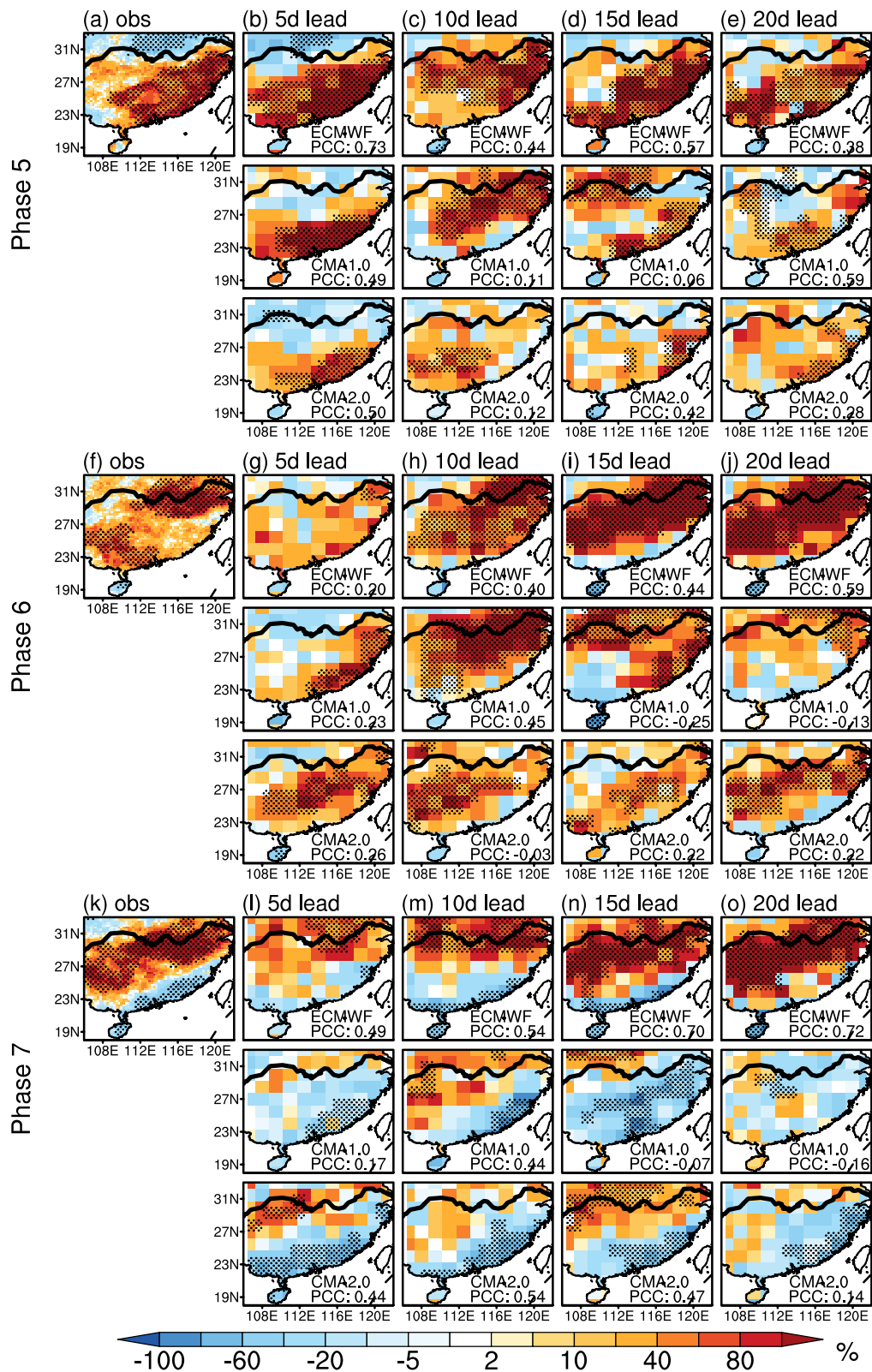
Fig. 7 Probability changes of SCER of **a** observed and S2S model **b** 5-days, **c** 10-days, **d** 15-days, and **e** 20-days lead time forecasts in Phase 5 of BSISO2 with respect to the non-BSISO2 period during the summers of 2005–2014. **f–j** and **k–o** are same as **a–e** but for Phase 6 and Phase 7, respectively. The PCC skills are shown in the bottom of each panel. Changes exceeding the 95% confidence level are dotted
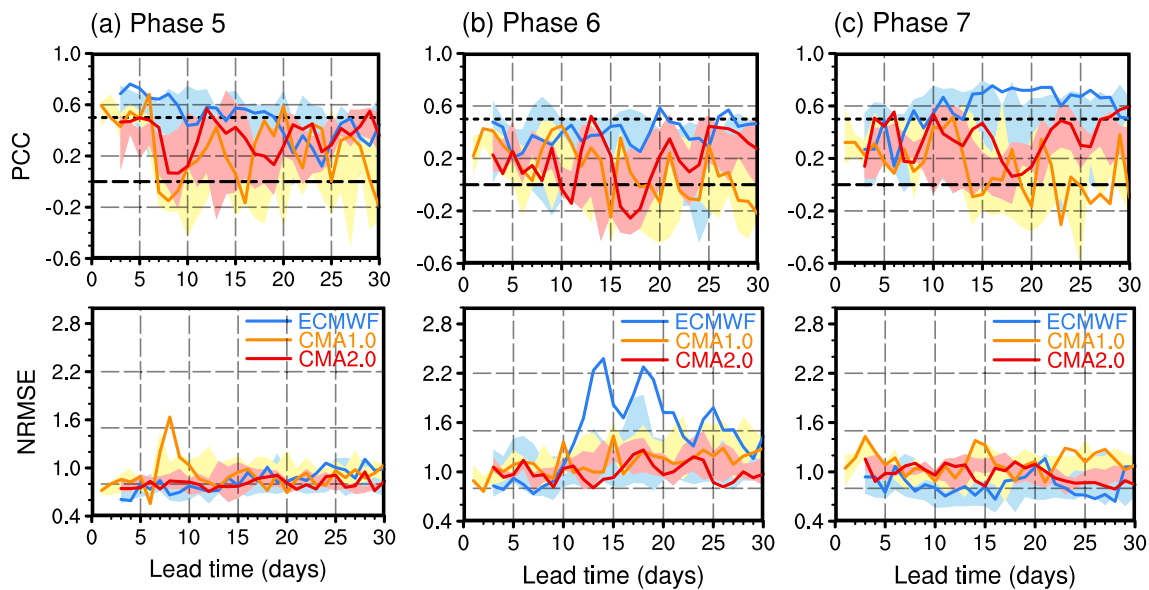
Fig. 8 The PCC and NRMSE skills for the percentage changes in the SCER probability in **a** Phase 5, **b** Phase 6, and **c** Phase 7 of BSISO2 with respect to the non-BSISO2 period as a function of lead time (in days). The curves for the ECMWF, CMA1.0, and CMA2.0 are in blue, orange, and red, along with inter-member spreads shown by the corresponding-colored shadings

(Fig. 9c). The averaged HSS of the ECMWF exceeds 0.1 up to a 10-days lead time for southern China during BSISO2 Phases 5–7, but it decreases quickly beyond a 10-days lead time (Fig. 9g, k). The HSS skill of the two CMA models is significantly lower than that of the ECMWF, especially for CMA1.0, with the HSS less than 0.1 only at a 5-days lead time (Fig. 9a, e) and even negative at a 10-days lead time (Fig. 9j). These results indicate that the CMA models still have large biases in accurately predicting SCER during Phases 5–7 of BSISO2.

From the perspective of the areal-mean HSS over southern China (Fig. 10), the ECMWF model could capture the SCER probability in Phases 5–7 up to a 13-days lead time. While CMA2.0 is comparable with the ECMWF in Phases 6–7, it can only reproduce the SCER probability with a 5-days lead time in Phase 5. CMA1.0 has the lowest HSS, with useful prediction skill only up to a 5-days lead time for all three phases of BSISO2.

The three S2S models have some capacity to predict SCER and BSISO2's modulation on the SCER probability. However, large biases exist in predicting both the intensity and pattern of the SCER probability. From the perspective of deterministic prediction, the ECMWF can effectively reproduce the probability of SCER up to a 20-days lead time, while CMA1.0 and CMA2.0 show prediction skill only within a 10-days lead time in Phase 5. In Phase 7, the ECMWF has useful prediction skill up to a 30-days lead time, while CMA1.0 and CMA2.0 show no skill at any lead time. The three models all show limited skill in Phase 6. From the perspective of probability prediction, the ECMWF

could nicely predict the SCER probability in Phases 5–7 of BSISO2 up to a 13-days lead time, while the CMA1.0 model has no skill beyond a 5-days lead time. CMA2.0 could effectively predict the SCER about 6 and 10 days in advance in Phases 5 and 6 of BSISO2 and 16 days in advance in Phase 7 of BSISO2.

5 The origin of the prediction error in BSISO2's modulation on SCER

While the S2S models show some capacity to predict the SCER probability under the BSISO2's modulation, the overall prediction skill is still limited. What are the reasons for the prediction error in BSISO2's modulation? Answering this question can provide useful hints for improving the performance of the S2S model in predicting extreme rainfall. Given that the occurrence of the extreme rainfall largely depends on the corresponding dynamic circulation and moisture conditions, the water vapor flux convergence $-\nabla \cdot (q\vec{V})$ (q is the scalar specific humidity and $\vec{V}$ is the wind velocity, including zonal and meridional winds) is often considered an important factor causing extreme precipitation under the influence of the BSISO (O'Gorman and Schneider 2009; Hsu et al. 2016; Loriaux et al. 2016). The water vapor flux convergence $-\nabla \cdot (q\vec{V})$ can be further divided into two terms: the moisture convergence term ($-q\nabla \cdot \vec{V}$) and moisture advection term ($-\vec{V} \cdot \nabla q$). In this section, the observed and forecasted water vapor flux convergence over East Asia during Phases 5–7 of the BSISO2 will be diagnosed, and

Fig. 9 HSS distribution (shading) in S2S model **a** 5-days, **b** 10-days, **c** 15-days, and **d** 20-days lead time forecasts in Phase 5 of BSISO2 during the summers of 2005–2014. **e–h** and **i–l** are same as **a–d** but for Phase 6 and Phase 7, respectively. The areal-mean HSS over SC is shown in the bottom of each panel
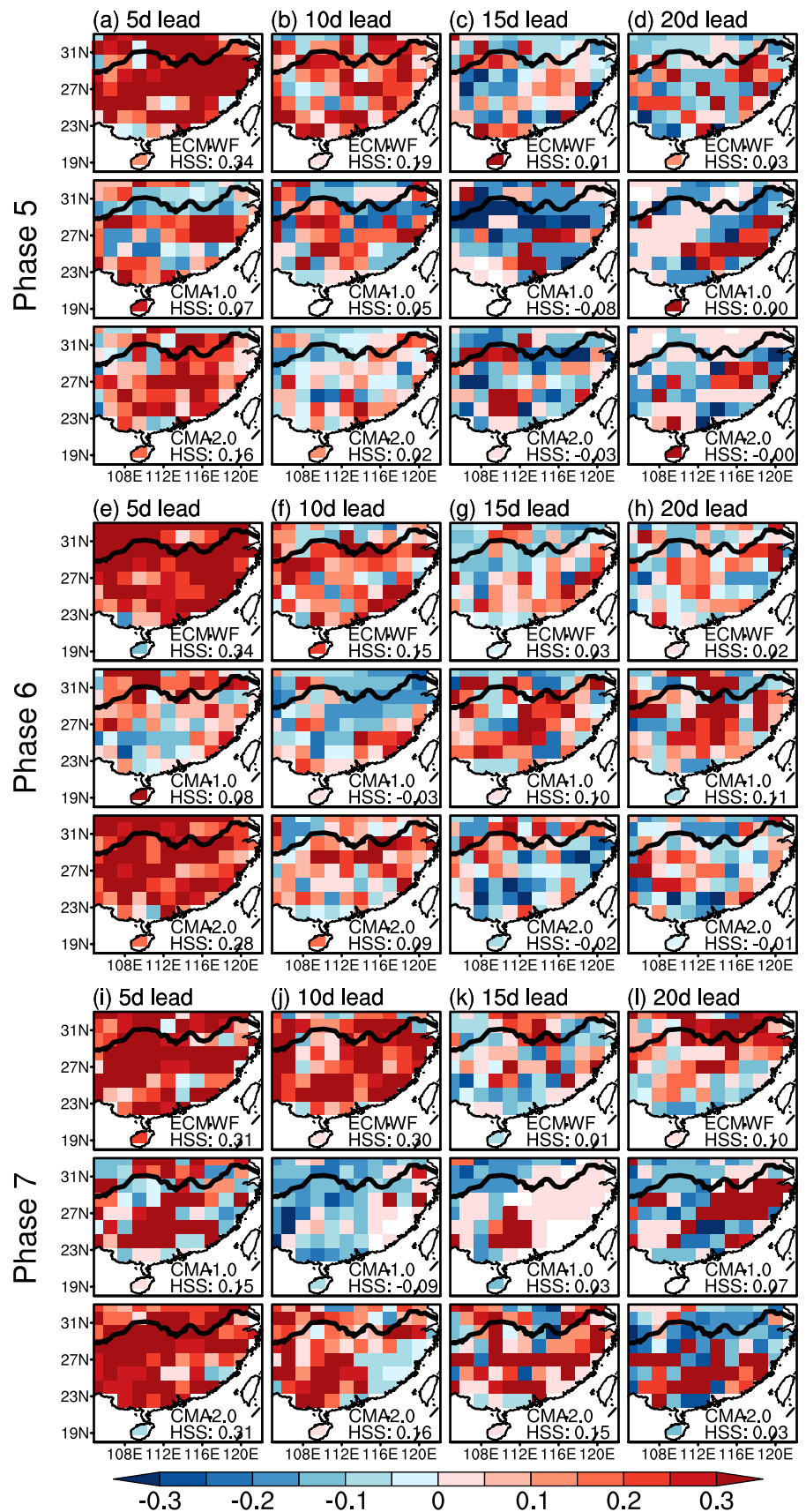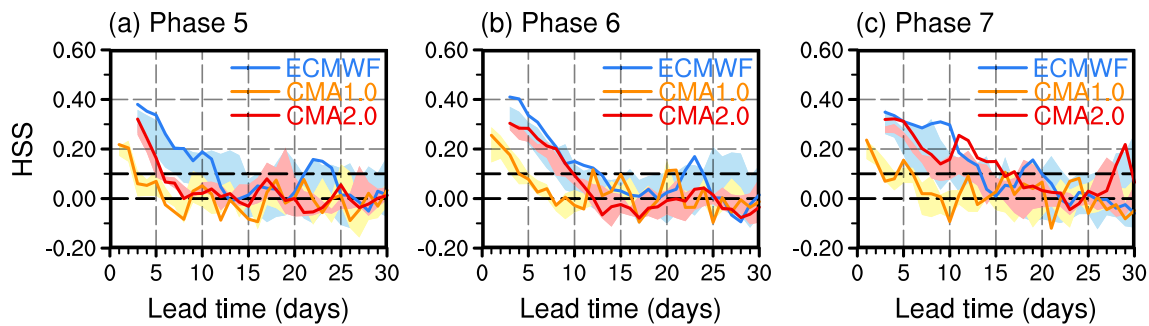
Fig. 10 As in Fig. 8 but for the areal-mean HSS

origins of the prediction error in BSISO2's modulation on SCER will be further clarified.

Firstly, the moisture convergence term $(-q\nabla \cdot \vec{V})$ is examined (Fig. 11). In Phase 5 of BSISO2, influenced by an anomalous low-level anticyclone over the Philippine Sea and an anomalous cyclone to the east of Japan (Fig. 11a), strong moisture convergence is observed in southern China (south of 30° N), corresponding to an increased probability of extreme rainfall south of the Yangtze River (Fig. 7a). The ECMWF could nicely reproduce the locations of the two critical circulation systems and the moisture convergence up to a 20-days lead time, resulting in good performance in the PCC skill of the BSISO2's modulation on SCER (Fig. 11b, c). The ECMWF underestimates the intensity of the anomalous circulation and the moisture convergence, especially beyond a 10-days lead time (Fig. 11d, e). The CMA1.0 predicts a much stronger low-level anticyclone, leading to a northward location bias of the SCER probability, while CMA2.0 underestimates the low-level cyclone to the east of Japan, leading to an underestimation of the intensity of the SCER probability (Fig. 7a–e).

In Phase 6 of BSISO2, as the anomalous low-level anticyclone moves northward, the center of the moisture convergence moves northward to the lower reaches of the Yangtze River Basin, and the southeastern coast of southern China is no longer controlled by moisture convergence. Meanwhile, the India–Burma trough leads to moisture convergence over the southwest of southern China (Fig. 11f). It is interesting that all three models could reproduce the anomalous low-level anticyclone to some extent, even up to a 20-d lead time, but they missed the low-level cyclone to the east of Japan in both intensity and location, leading to biases in the moisture convergence and poor performance of the SCER probability under Phase 6 of BSISO2 (Figs. 8b and 10b).

In Phase 7 of BSISO2, the anomalous low-level anticyclone moves to the north. The southwesterly wind encounters the northwesterly wind, resulting in moisture convergence over the Yangtze River Basin. However, because of the

weakened India–Burma trough, the moisture convergence over the southwest of southern China in Phase 6 of BSISO2 disappears. The ECMWF could nicely reproduce the location and amplitude of the southwesterly wind and the northeasterly wind over northern China up to a 20-days lead time, resulting in relatively high PCC skill in Phase 7 of BSISO2 up to a longer lead time (Fig. 8c). Otherwise, both CMA models underestimate the southwesterly and northwesterly winds over northern China, leading to weakened moisture convergence in the Yangtze River Basin (Fig. 11i–o). Therefore, the CMA shows relatively low prediction skill for the BSISO2's modulation on SCER compared with that of the ECMWF (Figs. 7i–o and 8c).

Figure 12 shows the spatial distribution of moisture advection $(-\vec{V} \cdot \nabla q)$ in Phases 5–7 of BSISO2. Compared with the moisture convergence (Fig. 11), the intensity of the moisture advection is much smaller, and it shows negative anomalies over southern China in both observations and predictions, indicating that the moisture advection has little contribution to the total moisture flux convergence and the probability of SCER.

To further quantify the influence of the moisture convergence/advection prediction skill on the SCER probability prediction skill under Phases 5–7 of BSISO2, the linear relationship between the PCCs of moisture convergence/advection and the probability of SCER derived from every forecast lead time and every individual member in the three S2S models are calculated. As shown in the scatter plot of Fig. 13, the ECMWF had a significant positive correlation between the PCC skill of moisture convergence and the SCER probability in all phases (Fig. 13a–c); in the CMA models, except for CMA1.0 in Phases 5 and 7 (Fig. 13a, c), no significant relationship could be found, suggesting that the CMA model still have a long way to go in capturing this physical process.

The positive correlation between the PCCs of moisture convergence and the SCER probability in the ECMWF indicates that the prediction error of the SCER probability
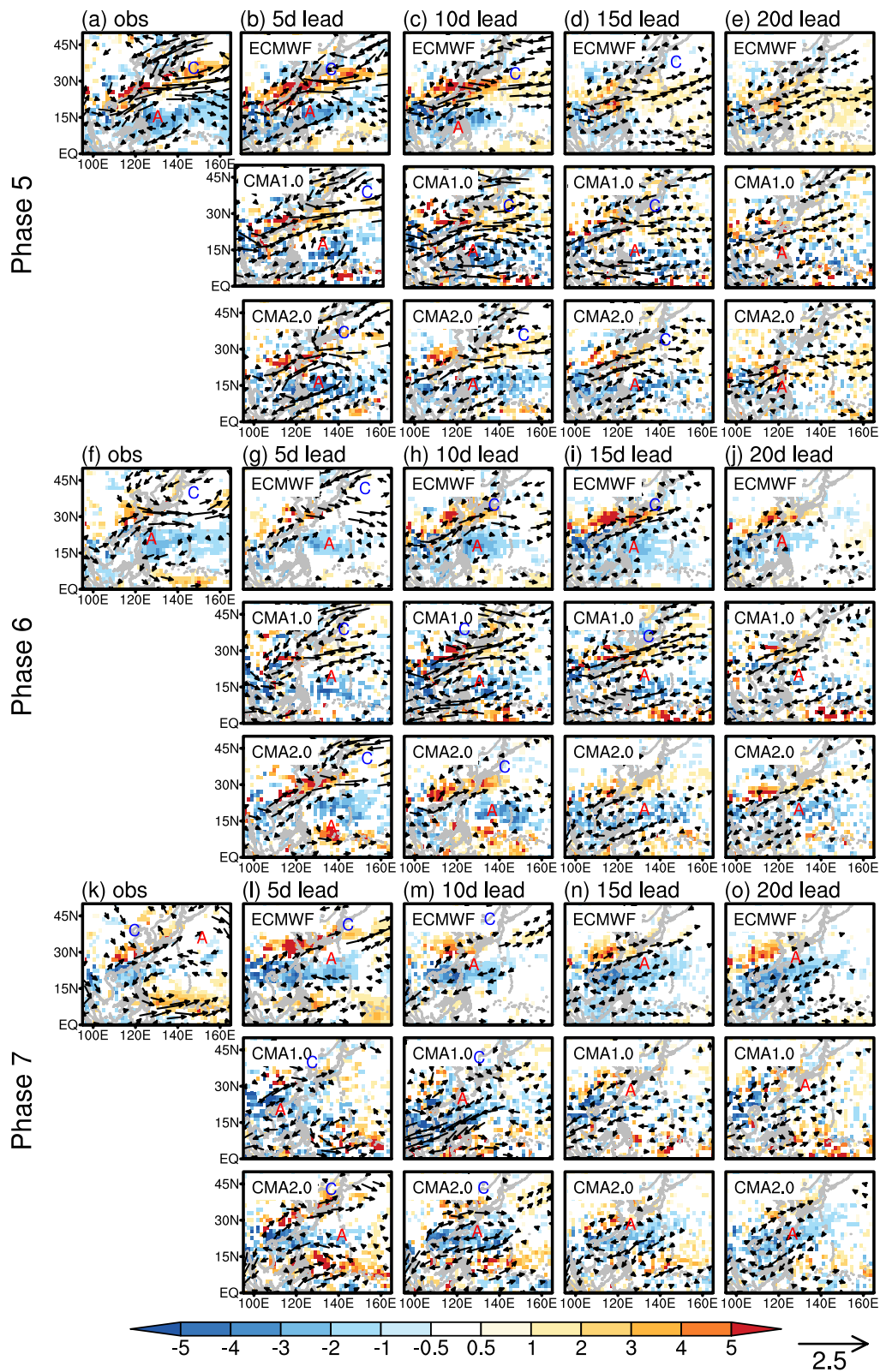
Fig. 11 Composite of the integrated moisture convergence (shading, unit: 10^{-5} m s^{-2}) and 850-hPa wind field (vector, unit: m s^{-1}) anomalies of **a** observed and S2S model **b** 5-days, **c** 10-days, **d** 15-days, and **e** 20-days lead time forecasts in Phase 5 during the summers of 2005–2014. **f–j** and **k–o** are same as **a–e** but for Phase 6 and Phase 7, respectively. The letters "A" and "C" represent the centers of the anomalous anticyclone or cyclone, respectively
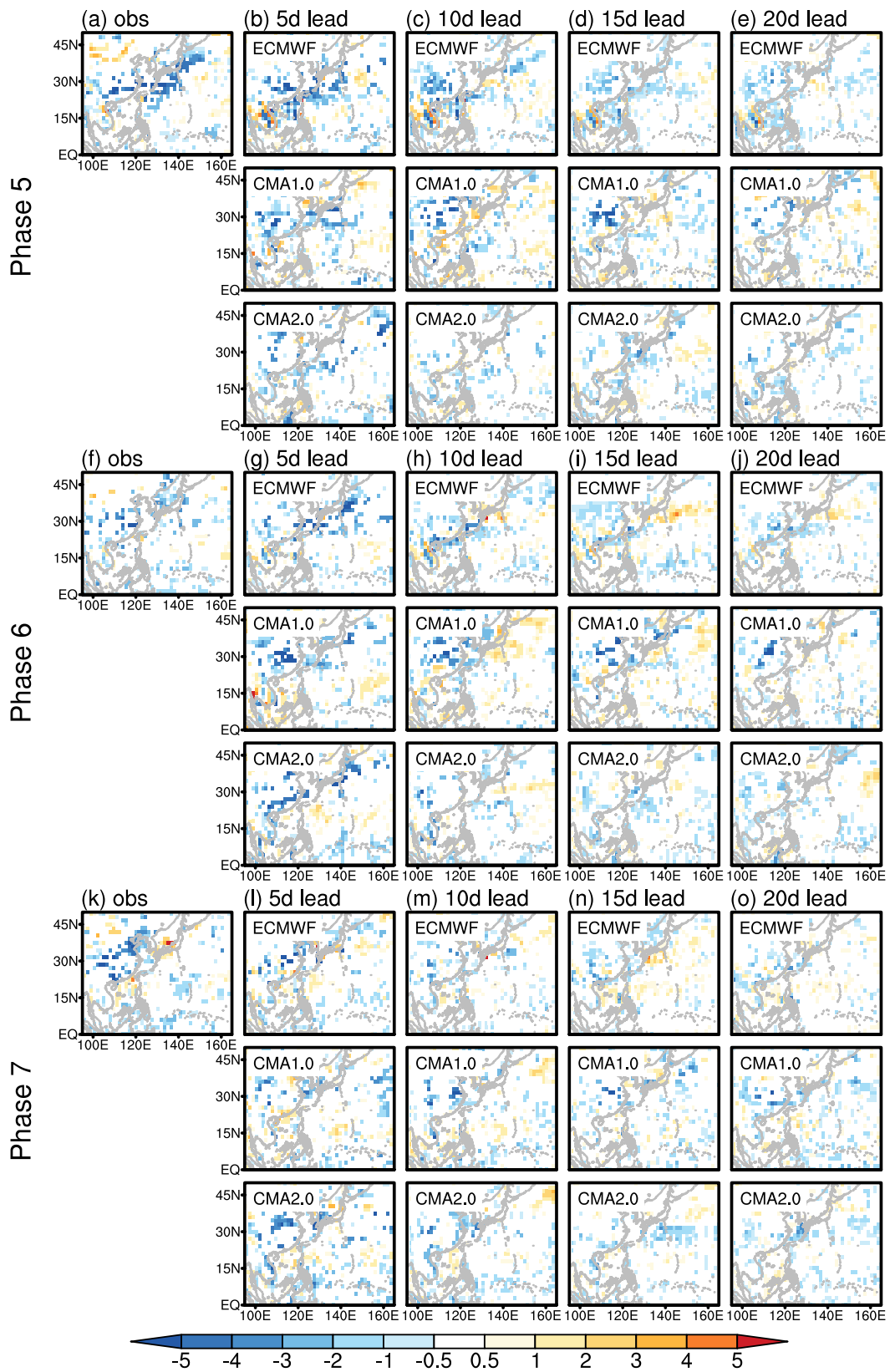
Fig. 12 As in Fig. 11, but for the integrated moisture advection (shading, unit: 10^{-5} m s^{-2}) anomalies
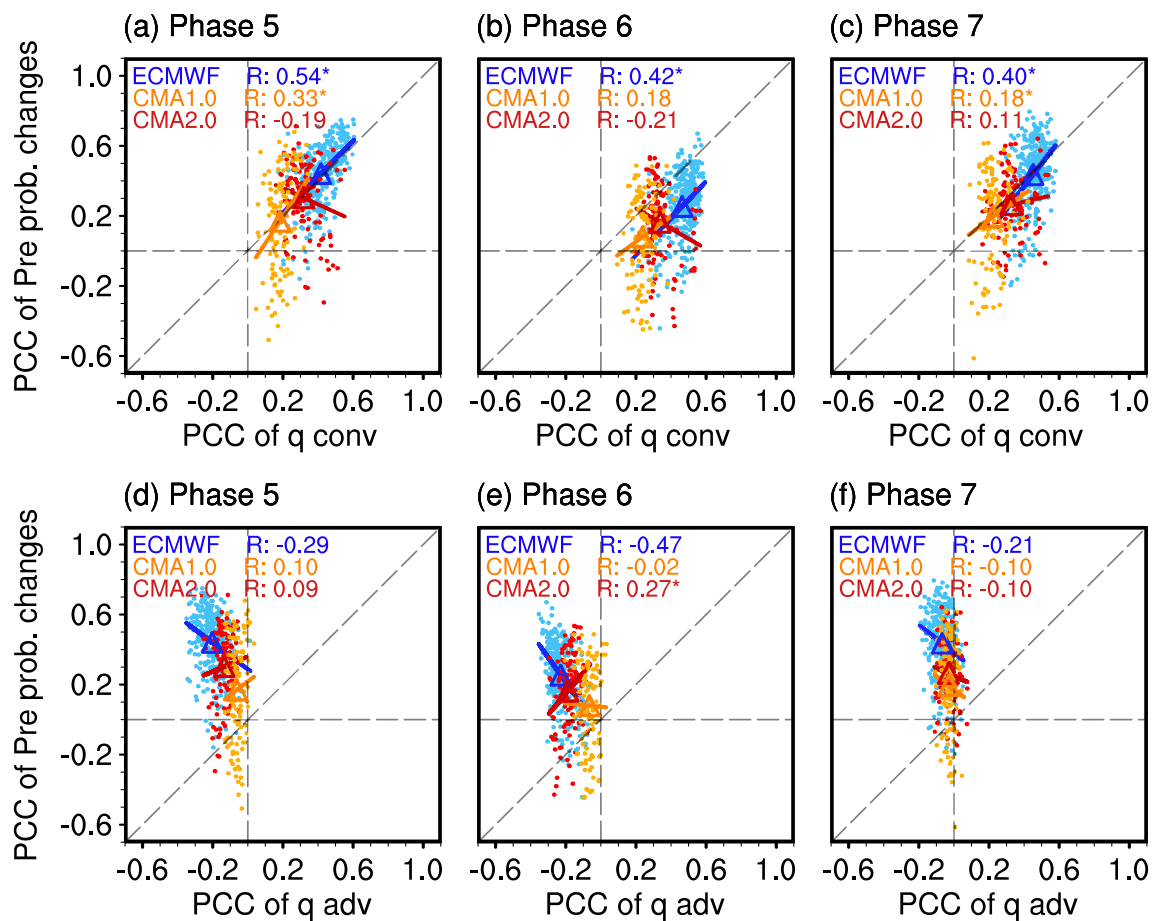
Fig. 13 Scatter diagrams for the PCC skill of the SCER probability changes (y-axis) against the integrated moisture convergence (x-axis) over East Asia (0°–50° N, 95°–165° E) in **a** Phase 5, **b** Phase 6, and **c** Phase 7 of BSISO2 for all individual members at all forecast lead times from the three models. **d–f** as in **a–c**, but the x-axis represents the PCC skill of the column-integrated moisture advection. There are linear-fit curves for the ECMWF (308 blue dots), CMA1.0 (120 orange dots), and CMA2.0 (120 red dots), and the triangles are the averaged PCC values for the ECMWF, CMA1.0, and CMA2.0. The correlation coefficients (R) between the PCC skill of the percentage changes in the SCER probability and those of the column-integrated moisture convergence (advection) are shown, and asterisks indicate the R is significant at the 95% confidence level

mainly comes from biases in predicting the large-scale moisture convergence in Phases 5–7 of the BSISO2. In contrast to the results of the moisture convergence term, no model shows a significant positive correlation between the PCCs of the moisture advection term and the SCER probability changes, except the results of CMA2.0 in Phase 6 of BSISO2. Therefore, the SCER probability prediction skill depends on the prediction skill of the moisture convergence rather than that of the moisture advection. Given that moisture convergence is the key predictability source of the extended-range forecast of the SCER, improving its prediction skill is key to improving the model performance in predicting SCER probability.

6 Discussion and conclusion

Densely populated southern China is prone to extreme precipitation during the boreal summer. Revealing the predictability sources of extended-range forecast and the origins of bias in the operational dynamical models are the keys to improving their prediction skill and extending the potential forecast lead time. Using the prediction database of three S2S models in the period of 2005–2014, from both deterministic and probabilistic perspectives, this study evaluated the models' performances in predicting summertime extreme rainfall over southern China (SCER), the 10–30-d boreal summer intraseasonal oscillation (BSISO2), and the

modulation of BSISO2 on the probability of SCER. The possible origins of the forecast bias in the three models were then further analyzed. The main conclusions are as follows:

(1) The three S2S models have different deviations in their forecasts of the mean summertime precipitation and the intensity of the daily precipitation variability. While the ECMWF model slightly overestimates the summer mean precipitation and its daily variation over southern China, CMA1.0 significantly underestimates both. CMA2.0 shows comparable capacity in predicting the mean and standard deviation of precipitation over southern China, and it performs best in predicting the threshold of extreme precipitation.

(2) From the perspective of probabilistic verification, the ensemble predictions of the ECMWF model and CMA2.0 can provide useful prediction skill up to a 13-days and 9-days lead time, respectively, while the ensemble mean of CMA1.0 has little skill in predicting the SCER beyond 4 days in advance.

(3) The ECMWF, CMA2.0, and CMA1.0 could effectively predict the BSISO2 index up to a 15-days, 13-days, and 8-days lead time, respectively. The HSS of all three models shows significant correlation between the SCER and BSISO2 index, indicating the significant influence of the prediction skill of the BSISO2 on the SCER probability prediction skill.

(4) The ECMWF can reproduce the probability of SCER in any phase of the BSISO2 up to a 20-days lead time, while CMA1.0 and CMA2.0 show prediction skill only within a 10-days lead time in Phase 5. In Phase 7, the ECMWF has useful prediction skill up to a 30-days lead time, while CMA1.0 and CMA2.0 show no skill at any lead time. The three models all show limited skill in Phase 6.

(5) The positive correlation between the PCCs of moisture divergence and SCER probability in the ECMWF indicates that the origin of the SCER probability prediction error mainly comes from the prediction error in the large-scale moisture convergence in Phases 5–7 of BSISO2.

This study indicates the essential role of moisture convergence on the S2S models' performance in the prediction of extreme rainfall over southern China. Compared to the significant positive correlation between the PCC skill of moisture convergence and the SCER probability changes from the ECMWF in all phases, no significant relationship could be found in CMA models, suggesting that the forecasting error may also relate to the data assimilation scheme (given that the three data sets have totally different

data assimilation schemes) and capacity in capturing related physical processes (dynamic circulation and moisture conditions) (Zhang et al. 2021).

Previous studies have suggested that the prediction skill of BSISO2 can be extended by 2–3 days by introducing more accurate atmospheric and oceanic initial conditions (Bo et al. 2020) or by increasing the frequency of nudging the observed sea surface temperatures during initialization of the coupled model (i.e., from monthly to weekly or daily) (Zhu et al. 2021). This study only emphasized the role of horizontal moisture convergence under the influence of BSISO, but it has been suggested that vertical moisture advection also provides favorable conditions for the occurrence of extreme precipitation (Ren et al. 2018). Meanwhile, because the diabatic heating caused by precipitation would result in feedback to the atmospheric circulation, extreme precipitation itself may also cause local ascending motion and vertical moisture advection (Lu and Lin 2009). Therefore, the physical mechanism of the increase in the extreme precipitation probability under the modulation of BSISO merits further investigation. Resolution improvements and upgrades to the initialization scheme are also important for improving the extended-range forecasts of extreme rainfall.

Note that, although both BSISO1 (Wu et al. 2023) and BSISO2 significantly modulate summer extreme precipitation in southern China, their modulation is dependent on phase and region. To improve SCER probability prediction, the combined effect of the two BSISO modes should be considered, and a more refined statistical–dynamical hybrid model should be developed to advance the prediction of extreme rainfall over southern China.

at https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-interim. The reforecast data from the operational S2S models can be downloaded from https://confluence.ecmwf.int/display/S2S.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Bo Z, Liu X, Gu W et al (2020) Impacts of atmospheric and oceanic initial conditions on boreal summer intraseasonal oscillation forecast in the BCC model. Theor Appl Climatol 142:303–406. https://doi.org/10.1007/s00704-020-03312-2

Dee D, de Uppala P et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597. https://doi.org/10.1002/qj.828

Gottschalck J, Wheeler M, Weickmann K et al (2010) A framework for assessing operational Madden-Julian oscillation forecasts: a CLIVAR MJO Working Group project. Bull Am Meteorol Soc 91:1247–1258. https://doi.org/10.1175/2010BAMS2816.1

He H, Yao S, Huang A, Gong K (2020) Evaluation and error correction of the ECMWF subseasonal precipitation forecast over eastern China during summer. Adv Meteorol 2020:1–20. https://doi.org/10.1155/2020/1920841

Heidke P (1926) Berechnung des Erfolges und der Gute der Windstarkevorhersagen im Sturmwarnungsdienst. Geogr Ann 8:301–349. https://doi.org/10.2307/519729

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

Jiang X, Li T, Wang B (2004) Structures and mechanisms of the northward propagating boreal summer intraseasonal oscillation. J Clim 17(5):1022–1039. https://doi.org/10.1175/1520-0442(2004)017<1022:SAMOTN>2.0.CO;2

Jie W, Vitart F, Wu T, Liu X (2017) Simulations of the Asian summer monsoon in the sub-seasonal to seasonal prediction project (S2S) database. Q J R Meteorol Soc 143:2282–2295. https://doi.org/10.1002/qj.3085

Lee J-Y, Wang B (2014) Future change of global monsoon in CMIP5. Clim Dyn 42:101–119. https://doi.org/10.1007/s00382-012-1564-0

Lee J-Y, Wang B, Wheeler MC et al (2013) Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. Clim Dyn 40:493–509. https://doi.org/10.1007/s00382-012-1544-4

Lee S-S, Moon J-Y, Wang B, Kim H-J (2017) Extended-range forecast of extreme precipitation over Asia: boreal summer intraseasonal oscillation perspective. J Clim 30:2849–2865. https://doi.org/10.1175/JCLI-D-16-0206.1

Li J, Wang B (2018) Predictability of summer extreme precipitation days over eastern China. Clim Dyn 51:4543–4554. https://doi.org/10.1007/s00382-017-3848-x

Li W, Jiang Z, Xu J, Li L (2016) Extreme precipitation indices over China in CMIP5 models. Part II: Probabilistic projection. J Clim 29:8989–9004. https://doi.org/10.1175/JCLI-D-16-0377.1

Li J, Zheng C, Yang Y, Lu R, Zhu Z (2023) Predictability of spatial distribution of pre-summer extreme precipitation days over southern China revealed by the physical-based empirical model. Clim Dyn. https://doi.org/10.1007/s00382-023-06681-2

Li J, Zhu Z, Dong W (2017) Assessing the uncertainty of CESM-LE in simulating the trends of mean and extreme temperature and precipitation over China. Int J Climatol 37(4): 2101–2110. https://doi.org/10.1002/joc.4837

Liang P, Lin H (2018) Sub-seasonal prediction over East Asia during boreal summer using the ECCC monthly forecasting system. Clim Dyn 50:1–16. https://doi.org/10.1007/s00382-017-3658-1

Liebmann B, Smith C (1996) Description of a complete (interpolated) outgoing longwave radiation dataset. Bull Am Meteorol Soc 77:1275–1277

Lin H, Brunet G, Derome J (2008) Forecast skill of the Madden-Julian oscillation in two Canadian atmospheric models. Mon Weather Rev 136:4130–4149. https://doi.org/10.1175/2008MWR2459.1

Loriaux J, Lenderink G, Siebesma AP (2016) Peak precipitation intensity in relation to atmospheric conditions and large-scale forcing at midlatitudes. J Geophy Res 121:5471–5487. https://doi.org/10.1002/2015JD024274

Lu R, Lin Z (2009) Role of subtropical precipitation anomalies in maintaining the summertime meridional teleconnection over the western North Pacific and East Asia. J Clim 22:2058–2072. https://doi.org/10.1175/2008JCLI2444.1

Madden RA, Julian PR (1971) Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. J Atmos Sci 28:702–708. https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2

Neena JM, Waliser D, Jiang X (2017) Model performance metrics and process diagnostics for boreal summer intraseasonal variability. Clim Dyn 48:1661–1683. https://doi.org/10.1007/s00382-016-3166-8

O'Gorman P, Schneider T (2009) The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. PNAS 106:14773–14777. https://doi.org/10.1073/pnas.0907610106

Oh H, Ha K-J (2015) Thermodynamic characteristics and responses to ENSO of dominant intraseasonal modes in the East Asian summer monsoon. Clim Dyn 44:1751–1766. https://doi.org/10.1007/s00382-014-2268-4

Ren P, Ren H-L, Fu J-X et al (2018) Impact of boreal summer intraseasonal oscillation on rainfall extremes in southeastern China and its predictability in CFSv2. J Geophys Res 123:4423–4442. https://doi.org/10.1029/2017JD028043

Wang S, Sobel AH, Tippett MK, Vitart F (2019) Prediction and predictability of tropical intraseasonal convection: seasonal dependence and the Maritime Continent prediction barrier. Clim Dyn 52(9):6015–6031. https://doi.org/10.1007/s00382-018-4492-9

Wang Y, Ren H, Zhou F et al (2020) Multi-model ensemble subseasonal forecasting of precipitation over the maritime continent in boreal summer. Atmosphere 11:515. https://doi.org/10.3390/atmos11050515

White CJ, Carlsen H, Robertson AW et al (2017) Potential applications of subseasonal-to-seasonal (S2S) predictions. Meteorol Appl 24:315–325. https://doi.org/10.1002/met.1654

Wu J, Li J, Zhu Z, Hsu P-C (2023) Factors determining the extended-range forecast skill of summer extreme rainfall over southern China. Clim Dyn 60:443–460. https://doi.org/10.1007/s00382-022-06326-w

Wu J, Gao XJ (2013) A gridded daily observation dataset over China region and comparison with the other datasets. Chin J Geophys 56(4):1102–1111. https://doi.org/10.6038/cjg20130406

Xavier P, Rahmat R, Cheong W, Wallace E (2014) Influence of Madden-Julian oscillation on Southeast Asia rainfall extremes—observations and predictability. Geophys Res Lett 41:4406–4412. https://doi.org/10.1002/2014GL060241

Xiang B, Zhao M, Jiang X et al (2015) The 3–4-week MJO prediction skill in a GFDL coupled model. J Clim 28:5351–5364. https://doi.org/10.1175/JCLI-D-15-0102.1

Yang J, Wang B, Bao Q (2010) Biweekly and 21–30-day variations of the subtropical summer monsoon rainfall over the lower reach of the Yangtze River Basin. J Clim 23:1146. https://doi.org/10.1175/2009JCLI3005.1

Yang J, Zhu T, Gao M et al (2018) Late-July barrier for subseasonal forecast of summer daily maximum temperature over Yangtze River Basin. Geophys Res Lett 45:610–615. https://doi.org/10.1029/2018GL080963

Yatagai A, Kamiguchi K, Arakawa O et al (2012) APHRODITE: constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. Bull Am Meteorol Soc 93:1401–1415. https://doi.org/10.1175/BAMS-D-11-00122.1

Zhang X, Alexander L, Hegerl G et al (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. Wiley Interdiscip Rev Clim Change 2:851–870. https://doi.org/10.1002/wcc.147

Zhang K, Li J, Zhu Z, Li T (2021) Subseasonal prediction skill of the persistent snowstorm event over southern China during early 2008 in ECMWF and CMA S2S prediction models. Adv Atmos Sci 38(11):1873–1888. https://doi.org/10.1007/s00376-021-0402-x

Zhao C, Zhou T, Song L, Ren H (2014) The boreal summer intraseasonal oscillation simulated by four Chinese AGCMs participating in the CMIP5 project. Adv Atmos Sci 31:1167–1180. https://doi.org/10.1007/s00376-014-3211-7

Zhu Z, Feng Y, Jiang W, Lu R, Yang Y (2023) The compound impacts of sea surface temperature modes in the Indian and North Atlantic oceans on the extreme precipitation days in the Yangtze River. Basin Clim Dyn. https://doi.org/10.1007/s00382-023-06733-7

Zhu X, Liu X, Huang A et al (2021) Impact of the observed SST frequency in the model initialization on the BSISO prediction. Clim Dyn 57:1097–1117. https://doi.org/10.1007/s00382-021-05761-5