ORIGINAL ARTICLE

Enhancing summer extreme precipitation prediction in the Yangtze River Basin through CWRF downscaling and its skillful multi-physics ensemble approach

Yang Zhao[1] · Xin-Zhong Liang[2,3]

Abstract

This study employs the regional Climate-Weather Research and Forecasting model (CWRF) downscaling and its skillful multi-physics ensemble approach to enhance summer extreme precipitation prediction in the Yangtze River Basin. The CWRF simulations at 30-km driven by the ECMWF Interim reanalysis during 1980–2015 are conducted using 28 model physics configurations. The prediction skill is evaluated for four standard indices: simple daily intensity, total extreme precipitation and its fraction to total precipitation, and maximum consecutive dry days. The control CWRF configuration outperforms the driving reanalysis, which underestimates all indices despite its utilization of surface data assimilation. Notably, the downscaling skills of CWRF exhibit substantial variability across diverse physics configurations, with cumulus parameterization being most influential. In particular, the ensemble cumulus parameterization (ECP) demonstrates remarkable proficiency in capturing both spatial patterns and interannual variations. This proficiency is further magnified through ECP's integration with Morrison or Morrison-aerosol microphysics and CCCMA, CAML or CAM radiation schemes, significantly enhancing overall skills. The ensemble average of these skill-enhanced physics configurations better reproduces observed geographic distributions and interannual anomalies for all four indices, surpassing the control CWRF's performance. This improvement can be primarily attributed to a more accurate portrayal of the East Asian Jet and its associated regional circulation patterns by the ensemble. These findings underscore a significant opportunity for enhancing predictions of extreme precipitation by further refining the physics representation within the climate system coupling, especially among cumulus, microphysics, and radiation processes. Optimizing the multi-physics ensemble approach holds substantial promise in this endeavor.

1 Introduction

Extreme precipitation can bring serious impacts on human health, economy, and ecosystems (Liang 2022). The losses caused by these extreme disasters are estimated to exceed 2.37% of the gross domestic product in China each year since 1990 (Jiang et al. 2015; Zhang and Zhou 2020). The Yangtze River Basin (YRB) is particularly vulnerable to extreme precipitation since it is home for nearly 40% of the Chinese population, and contributes about 40% of the total Chinese gross domestic product (Li and Lu 2017). The region is greatly affected by the East Asian monsoon and experiences a large interannual variability in summer precipitation, which have caused disastrous social and economic consequences through both frequent floods and droughts (Li and Lin 2015). For instance, the 1998 flood in the YRB affected 240 million people, caused more than 3000 deaths, and approximately $40 billion in damages (Zong and Chen 2000). The 2010 heavy rainfall events over East China affected 134 million people and costed nearly $18 billion (Murray and Ebi 2012; Zhao et al. 2012). Future risk would

✉ Xin-Zhong Liang
xliang@umd.edu

1   Climate, Environment and Sustainability Center, Nanjing University of Information Science and Technology, Nanjing 210044, China

2   Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20740, USA

3   Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA

increase, as the frequency and intensity of extreme events in the YRB have been detected increasing over the past decades (Wang and Zhou 2005; Zhai et al. 2005; Ma et al. 2015), and are projected to intensify under global warming (Piao et al. 2010; Sun et al. 2018; Jiang et al. 2021).

Global general circulation models (GCMs) are often used to simulate and predict extreme precipitation, however, to date, these models generally underestimate such events (Jiang et al. 2015; Dong and Dong 2021). For example, He et al. (2019) showed that almost all CMIP5 models fail to capture the observed spatial distribution of summer extreme precipitation over the YRB and South China, with the percentage of total rainfall from heavy events underestimated by 25–75%. Dong and Dong (2021) evaluated the performance of CMIP6 models in simulating seven extreme precipitation indices and showed that dry biases still exist in South China. Additionally, Xu et al. (2011) demonstrated that models are limited in reproducing the interannual variation of precipitation extremes in river basins over China. To overcome the problem, regional climate models (RCMs) have been developed for downscaling at higher spatial resolutions, which can represent more detailed regional-scale precipitation features and extreme events compared to their driving GCMs (Yang et al. 2016; Jiang et al. 2021). But resolution increases do not always improve simulated extreme precipitation (Chan et al. 2013; Kopparla et al. 2013; O'Brien et al. 2016). Even there is a tendency to overestimate the intensity magnitude of heavy rainfall over the YRB in convection permitting models with grid spacing of 1–5 km (Li et al. 2019; Dong et al. 2022). Hence, adequately representing finer-scale physical processes is the key to predicting extreme events (Sun and Liang 2020b; Jiang et al. 2021).

Physical parameterizations, particularly for cumulus convection, impact the extreme precipitation simulation. For example, Huang and Gao (2017) showed that the Kain-Fritsch scheme tends to overestimate summer extreme precipitation in the YRB, whereas the Grell scheme underestimates it, especially for the intensity and total amount. Zhaoye et al. (2022) found that Kain-Fritsch scheme performs better than the Grell-Devenyi and Bullock-Wang schemes for a rainstorm event simulation, but all underestimate precipitation intensity over the core affecting area in Northwest China. Some studies have also investigated the sensitivity of microphysics, radiation, planetary boundary layer, and land surface schemes (Kang et al. 2015; Gao et al. 2021; Kong et al. 2022; Merino et al. 2022). Given these studies have mostly concentrated on one or two particular processes for short-term or less than 10 years, more efforts are needed to construct a multi-physics ensemble of long integrations. In fact, Yuan et al. (2012) conducted 16 ensemble downscaling simulations with alternative microphysics, cumulus, land surface, and radiation schemes for 27 winters in China, and demonstrated the importance of radiation and

cumulus schemes in simulating extremes. Sun and Liang (2020b) found that the United States long-term extreme precipitation simulation was more sensitive to cumulus parameterization than the microphysics, aerosol, cloud, radiation, boundary layer, and surface schemes through all seasons. However, a comprehensive exploration into the relationship between model performance in simulating long-term summer extreme precipitation characteristics and diverse physics parameterization schemes across the YRB, along with the underlying mechanisms, has been seldomly attempted.

Recently, the regional Climate-Weather Research and Forecasting model (CWRF), developed by Liang et al. (2012), has shown advanced downscaling skills in reproducing monsoon rainbands, seasonal-interannual precipitation variations, and 95th percentile daily precipitation (P95) in China (Liang et al. 2019b; Li et al. 2020; Jiang et al. 2021). Most recently, Zhang et al. (2023) compared the sensitivity of CWRF downscaling seasonal P95 variations over China to five cumulus parameterization schemes and explored the physical processes and mechanisms underlying regional model biases. However, the performance of CWRF in downscaling extreme statistics across the YRB and its sensitivity to diverse physics (in addition to cumulus) parameterizations remain unclear. The main objective of this study is to enhance CWRF capabilities in downscaling YRB summer extreme precipitation spatiotemporal characteristics through exploration of varying model physics representations. The resulting multi-physics ensemble, derived from these refined CWRF configurations, is subsequently employed to enhance the overall performance. This approach is necessary as a single combination of selected parameterization schemes does not yield optimal results for all metrics in every region (Liang et al. 2012).

The paper is arranged as follows. Section 2 introduces the model, experimental design, observations, and analysis methods. Section 3 presents the main results, including (1) a comparison of CWRF capabilities in simulating YRB summer extreme precipitation spatial patterns and interannual variations among diverse physics parameterizations, alongside the driving reanalysis; (2) an evaluation of overall model performance rankings and optimal multi-physics ensemble skill enhancements; and (3) a process understanding of model biases and skill enhancements. Section 4 gives conclusions and discussion.

2 Data and methods

The CWRF has been developed from the Weather Research and Forecasting model (WRF) (Skamarock et al. 2008) to extend applications for climate research, incorporating numerous advancements in system interactions among land, atmosphere, ocean; convection, microphysics; and

cloud, aerosol, radiation processes (Liang et al. 2012). Prominently, CWRF incorporates a state-of-the-art Conjunctive Surface–Subsurface Process model (CSSP) to realistically represent terrestrial hydrology, land surface, and atmosphere processes (Choi et al. 2007, 2013; Choi and Liang 2010; Yuan and Liang 2011; Xu et al. 2014); a built-in Cloud-Aerosol-Radiation ensemble model (CAR) to fully estimate interactions among cloud properties, aerosol properties, and radiation transfers (Liang and Zhang 2013; Zhang et al. 2013). Furthermore, a built-in Ensemble Cumulus Parameterization (ECP) significantly improves simulation performances of precipitation climatology and extremes in the United States (Liang et al. 2012; Qiao and Liang 2015, 2016, 2017; Sun and Liang 2020a, b) and China (Liu et al. 2008; Zeng et al. 2008; Liang et al. 2019b; Li et al. 2020; Jiang et al. 2021). The CWRF integration of various parameterization schemes for each major physical process enables a comprehensive sensitivity analysis of extreme precipitation simulation in relation to physics representations (Sun and Liang 2020b).

The model computational domain centers at (35.18°N, 110°E) based on the Lambert conformal map projection with horizontal grid spacing of 30 km and has 36 vertical terrain-following levels with the top at 50 hPa. The buffer zone includes 14 grids in each lateral boundary of the four domain edges. We conducted an ensemble of 28 CWRF simulations using different combinations of key physics parameterization schemes (Table 1). This includes one control configuration (CTL, Liang et al. 2019a) and 27 physics configurations that swap CTL with one alternate parameterization scheme for cumulus, microphysics, radiation, boundary layer, surface and cloud processes. Given available computing resources, we primarily focus on the first four processes with multiple parameterization schemes; for both the surface and cloud processes, we choose one alternative: NOAH land model (Ek et al. 2003) and prognostic cloud scheme (Wilson et al. 2008), which are compared respectively with the control CSSP and diagnostic cloud scheme (Xu and Randall 1996). This list represents a trade-off for a diverse range of physics representations between availability built in CWRF, suitability for climate modeling, and relative performance based

Table 1 Summary of the CWRF control and sensitivity configurations with different physics parameterization schemes

Physics	Control configuration	Sensitivity configuration	Parameterization scheme
Cumulus (CU)	ECP penetrative convection plus UW shallow convection	E1	KFeta
		E2	BMJ
		E3	Grell
		E4	Tiedtke
		E5	NSAS
		E6	Donner
		E7	Emanuel
Microphysics (MP)	GSFCGCE	E8	Lin
		E9	WSM6
		E10	Etamp new
		E11	Thompson
		E12	Thompson-aero
		E13	Morrison
		E14	Morrison + 3d aerosol
Radiation (RA)	GSFCLXZ	E15	CCCMA
		E16	CAWCR
		E17	CAM
		E18	CAML
		E19	FuLiou
		E20	RRTMG
Boundary layer (BL)	CAM3	E21	YSU
		E22	MYNN
		E23	Boulac
		E24	ACM
		E25	UW
Surface (SF)	CSSP land plus UOM ocean	E26	NOAH
Cloud (CL)	Xu-Randall Diagnostic	E27	Prognostic

on initial testing over a thousand combinations (Liang et al. 2012, 2019a, b; Liang and Zhang 2013; Zhang et al. 2013). More details of the model physics schemes can be found in Liang et al. (2012) and Sun and Liang (2020b). All simulations are integrated over the period of October 1, 1979 to December 31, 2015 (the beginning two months for spin-up), driven by 6-hourly lateral boundary conditions from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim reanalysis (ERI) at ~80-km grid spacing (Dee et al. 2011). Following Liang et al. (2019b), this study adopts the subregion divisions of distinct climate regimes and geographical conditions for result analysis, focusing on those relevant to the YRB.

For model validation, observed daily precipitation data is from the CN05.1 dataset at 0.25° horizontal grid spacing based on an objective analysis of rain gauge measurements at 2416 meteorological stations in Mainland China during 1980–2015 by the China Meteorological Administration (Wu and Gao 2013). In addition, the driving ERI precipitation is used to evaluate the CWRF downscaling ability. The newly released fifth generation ECMWF reanalysis (ERA5) with 31 km horizontal grid spacing (Hersbach et al. 2019) is chosen as the best proxy for the observed atmospheric circulation characteristics, as it more realistically represents interactions between precipitation, land, and atmospheric processes (Sun and Liang 2020a, b). All these data are interpolated onto the CWRF 30-km grid using the conservative mapping method for direct comparison.

This study uses four major extreme precipitation indices (Table S1), selected due to their wide adoption in climate research (Peterson 2005; Cui et al. 2019; Tang et al. 2021). Each index is calculated from daily rainfall data, providing insights into different aspects of extreme precipitation events. The simple daily intensity index (SDII) quantifies the average precipitation intensity over the rainy days, in which rainfall exceeds 1 mm. The total extreme precipitation (R95P) measures the cumulative amount of all daily rainfalls that surpass the long-term 95th percentile of rainy days (P95), reflecting the total magnitude of extreme precipitation events. The ratio of R95P over the total precipitation (R95T) depicts the relative contribution from extreme wet days (exceeding P95), offering an insight into the frequency and regularity of extreme events within the overall rainfall distribution. The maximum number of consecutive dry days (CDD) is the longest duration of continuous dry conditions with daily rainfall below 1 mm, measuring the pattern of droughts and the persistence of dry spells.

Our analysis first evaluates the impacts of model physics configurations on spatial distributions of extreme precipitation climatological means in Sect. 3.1 and then on their interannual variations in Sect. 3.2. We employ a comprehensive suite of metrics to assess model performance, including spatial and temporal correlation, root-mean-square-error

(RMSE), standard deviation, and bias of each extreme index as well as mean absolute relative bias (MARB) across all indices over the YRB. The MARB is defined as the average of the absolute values of relative biases across all indices (see equation S1 for the calculation detail), eliminating the compensating effect between positive and negative biases in the indices (e.g., Jiang et al. 2015). Furthermore, in Sect. 3.3, we determine the overall model skill based on a comprehensive ranking metric (MR) that integrates relative scores of correlation, deviation, and RMSE among all indices. This evaluation strategy using the multifaceted metrics enables a comprehensive assessment of the models' relative strengths and weaknesses across various dimensions. Finally, in Sect. 3.4, we explore the linkages of the model biases to regional circulation patterns, adding depth to our physical understanding of the model performance from a climate system perspective. This evaluation approach ensures a nuanced and comprehensive understanding of the model's capabilities.
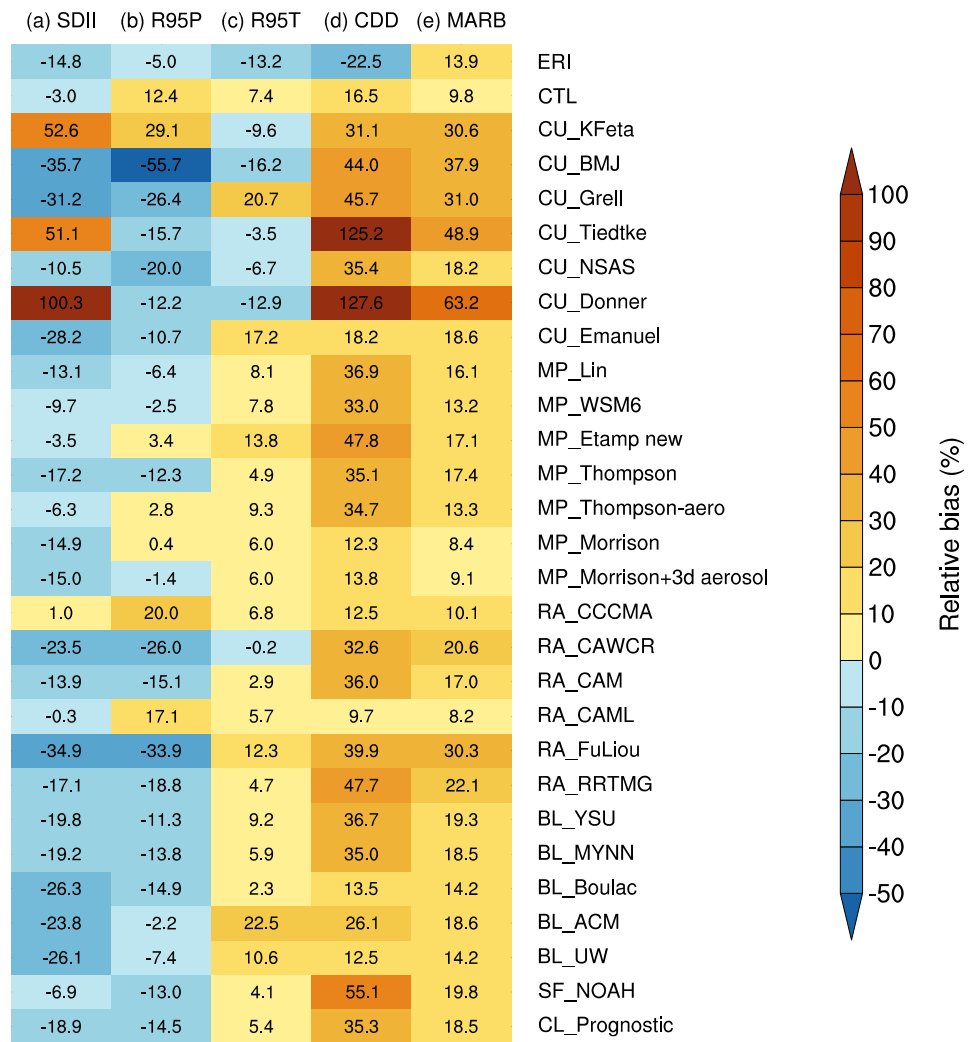
3 Results

3.1 Impact of physics configurations on extreme precipitation mean spatial patterns

First, the general abilities in capturing observed summer YRB extreme precipitation characteristics are compared among the 28 CWRF physics configurations along with the driving ERI. Figure 1 compares the 36-year (1980–2015) mean relative biases (simulated vs observed) for the four indices averaged over the YRB. The ERI largely underestimates all four indices, especially for CDD by about 22.5%, which is associated with overestimating the total number of rainy days due to its significant drizzling problem (Sun and Liang 2020b). Compared to ERI, CWRF CTL generally reduces the magnitude of these biases, with MARB decreased from 13.9% to 9.8%. The CTL captures more realistic extreme precipitation characteristics, except for overestimating R95P by 12.4%, in contrast to ERI's underestimation by 5.0%.

All CWRF members overestimate CDD and hence tend to underestimate the number of wet days. This tendence seems to result from a model bias against low-intensity rain events and thus may be affected by the standard 1 mm/day threshold used for defining rainy days (Peterson 2005). The threshold has been widely adopted in climate analyses, including regional climate model simulations at a 30-km resolution (e.g., Sun and Liang 2020a; Jiang et al. 2021). However, our test with a lower threshold of 0.1 mm/day showed minimal impact on our results (Figure S1), implying that other factors like physics representation may cause this tendence.

Fig. 1 The 1980–2015 mean relative biases (from observations, %) of the four extreme precipitation indices (**a–d**) and their absolute average (MARB, **e**) over the YRB for summer (JJA) as assimilated (ERI) and simulated by all CWRF physics configurations: **a** SDII, **b** R95P, **c** R95T, **d** CDD, and **e** MARB. The number listed in each grid cell represents the corresponding relative bias or MARB, with the colored scale on the right side

(a) SDII	(b) R95P	(c) R95T	(d) CDD	(e) MARB	
-14.8	-5.0	-13.2	-22.5	13.9	ERI
-3.0	12.4	7.4	16.5	9.8	CTL
52.6	29.1	-9.6	31.1	30.6	CU_KFeta
-35.7	-55.7	-16.2	44.0	37.9	CU_BMJ
-31.2	-26.4	20.7	45.7	31.0	CU_Grell
51.1	-15.7	-3.5	125.2	48.9	CU_Tiedtke
-10.5	-20.0	-6.7	35.4	18.2	CU_NSAS
100.3	-12.2	-12.9	127.6	63.2	CU_Donner
-28.2	-10.7	17.2	18.2	18.6	CU_Emanuel
-13.1	-6.4	8.1	36.9	16.1	MP_Lin
-9.7	-2.5	7.8	33.0	13.2	MP_WSM6
-3.5	3.4	13.8	47.8	17.1	MP_Etamp new
-17.2	-12.3	4.9	35.1	17.4	MP_Thompson
-6.3	2.8	9.3	34.7	13.3	MP_Thompson-aero
-14.9	0.4	6.0	12.3	8.4	MP_Morrison
-15.0	-1.4	6.0	13.8	9.1	MP_Morrison+3d aerosol
1.0	20.0	6.8	12.5	10.1	RA_CCCMA
-23.5	-26.0	-0.2	32.6	20.6	RA_CAWCR
-13.9	-15.1	2.9	36.0	17.0	RA_CAM
-0.3	17.1	5.7	9.7	8.2	RA_CAML
-34.9	-33.9	12.3	39.9	30.3	RA_FuLiou
-17.1	-18.8	4.7	47.7	22.1	RA_RRTMG
-19.8	-11.3	9.2	36.7	19.3	BL_YSU
-19.2	-13.8	5.9	35.0	18.5	BL_MYNN
-26.3	-14.9	2.3	13.5	14.2	BL_Boulac
-23.8	-2.2	22.5	26.1	18.6	BL_ACM
-26.1	-7.4	10.6	12.5	14.2	BL_UW
-6.9	-13.0	4.1	55.1	19.8	SF_NOAH
-18.9	-14.5	5.4	35.3	18.5	CL_Prognostic

Relative bias (%) scale: 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 0, -10, -20, -30, -40, -50

In contrast, most members tend to underestimate SDII and R95P, but overestimate R95T. Biases in the extreme precipitation *amount* is likely influenced by precipitation intensity (Sun and Liang 2020b), while biases in the extreme precipitation *ratio* is related to the tendency that models with less extreme precipitation often have less total precipitation amount (Tang et al. 2021). In general, the relative biases of R95P and R95T are smaller than those of SDII and CDD, capturing extreme precipitation events better than mean conditions. This aligns partially with Jiang et al. (2015), indicating less pronounced biases for SDII and R95T than those for total precipitation and CDD in CMIP5 models across eastern China. CWRF has been demonstrated with superior performance in simulating extreme precipitation (Liang et al. 2019a, b; Sun and Liang et al. 2020a, b), due to its advanced physics representation, especially cumulus parameterization. The variability in the biases across indices depicts the inherent complexity of climate modeling, emphasizing the need to consider regional and multiple factors in interpreting results.

The performance of CWRF in capturing observed extreme precipitation characteristics varies significantly among different physics configurations. The most influential factor is cumulus parameterization, where the spread of mean relative biases among the eight schemes spans between [− 35.7, 100.3] or 136% for SDII, [− 55.7, 29.1] or 84.8% for R95P, [− 16.2, 20.7] or 36.9% for R95T, and [16.5, 127.6] or 111.1% for CDD. The influence is moderate by radiation parameterization, where the spread of mean relative biases among the seven schemes is 35.9%, 53.9%, 12.5%, and 38% for SDII, R95P, R95T, and CDD, respectively. On the other hand, the sensitivity to boundary layer, microphysics, surface, or cloud parameterization is relatively weak, where the spread of mean relative biases ranges around 16–24% for all indices. These results suggest that cumulus parameterization plays the dominant role for CWRF's ability to simulate YRB summer extreme precipitation. According to the MARB score (Fig. 1e), CAML radiation and Morrison or Morrison plus 3d aerosol microphysics schemes further improve over CTL, reducing the overall bias magnitude from 9.8%

to 8.2% and 8.4% or 9.1%, respectively. The CCCMA radiation scheme also performs well, with a MARB of 10.1%. However, Tiedtke and Donner cumulus schemes perform poorly, with substantially large MARBs of 48.9% and 63.2%.

Figure 2 compares the overall performances among ERI and 28 CWRF physics configurations in capturing observed geographic distributions of summer mean extreme precipitation indices over the YRB, including spatial pattern correlation, normalized standard deviation, and centered pattern RMSE. For SDII (Fig. 2a), ERI shows a small negative pattern correlation (− 0.10) and substantially underestimates standard deviation (0.6), whereas CWRF CTL performs better, having a much higher correlation (0.36) albeit an overestimated deviation (1.48). As discussed later, the negative correlation of ERI with observations is identified with incorrect spatial pattern and large local underestimation. The greatest discrepancy among the CWRF members is identified with those using different cumulus schemes. In particular, NSAS most strongly correlates with the observed pattern (0.47) and slightly overestimates deviation (1.17). Compared

to the control ECP, KFeta simulates a higher correlation (0.38) but an excessive deviation (2.12). Other cumulus schemes generally have lower correlations (0.06–0.32), or abnormally high deviations (e.g., 2.48–2.87 by Donner and Tiedtke). When ECP is combined with the alternate schemes in other physical processes, CWRF performs similarly as in CTL, producing correlations between 0.35 and 0.50 and deviations between 1.13 and 1.67. One exception is that the ACM boundary layer scheme has a much lower correlation (0.19) than the control CAM3.

For R95P (Fig. 2b), ERI correlates more highly (than SDII) with the observed pattern (0.38), but still underestimates the deviation (0.78). In contrast, CWRF CTL produces even higher correlation (0.52) but larger deviation (1.70). As discussed in Liang et al. (2019b), this increased spatial variability may partly result from the coarse reference data that cannot represent actual distribution details. Of all the cumulus schemes, the control ECP performs best overall. KFeta correlates with observations more strongly (0.63), but also overestimates more deviation (1.98). Other
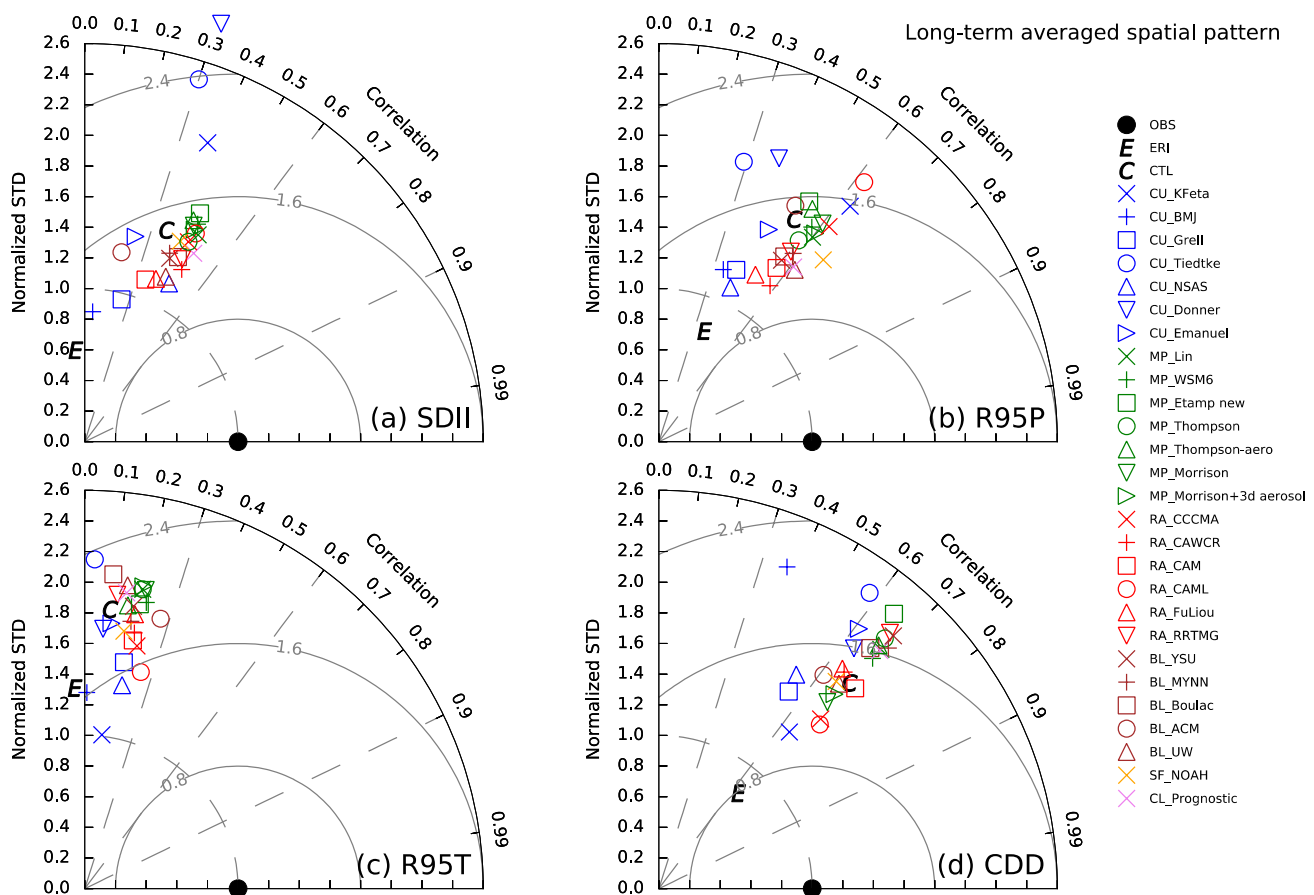

Fig. 2 Taylor diagrams of the performance among ERI and all CWRF physics configurations in simulating 1980–2015 mean summer four indices geographic distributions over the YRB: **a** SDII, **b** R95P, **c** R95T, and **d** CDD. Shown are the corresponding spatial correlation (azimuthal) and normalized standard deviation (radius) against observations. The distance of the simulation to observation indicates the root-mean-square error. The black dot (OBS) represents the perfect score with a unit correlation and deviation

cumulus schemes have systematically lower scores than ECP, with smaller pattern correlations (0.29–0.46) or larger spatial deviations (e.g., 1.91–2.01 by Donner and Tiedtke). While ECP is combined with the alternate schemes in other physical processes, CWRF skills resemble CTL, producing generally higher correlations (0.50–0.67) and comparable deviations (1.25–1.85) in R95P than SDII. One exception is that the CAML radiation scheme substantially overestimates spatial deviation (2.16).

For R95T (Fig. 2c), ERI has a small negative pattern correlation with observations (-0.05) and substantially overestimates spatial deviation (1.31). As compared with SDII and R95P, all CWRF ECP members systematically decrease pattern correlations (0.09–0.27) and more significantly overestimate spatial deviations (1.46–2.06). The result indicates that it is more challenging to capture R95T. The BMJ, Tiedtke, and Donner cumulus schemes produce lower correlations (0.01–0.07) or greater deviation (e.g., 2.15 by Tiedtke), whereas the KFeta, Grell, NSAS, and Emanuel schemes increase correlations (0.10–0.18) and reduce the overestimation of deviations (1.01–1.74).

For CDD (Fig. 2d), as compared with SDII, R95P and R95T, all CWRF ECP members significantly increase pattern correlations (0.61–0.70), albeit producing spatial deviations with a wide range (1.50–2.36). Thus, ECP has a higher skill in simulating consecutive dry days than precipitation intensity and extremes. The ECP has a slightly higher pattern correlation (0.68) than ERI (0.64) and overestimates spatial deviation (1.83) opposite to ERI's systematic underestimation (0.81). Other cumulus schemes generally show less skills than ECP, with systematically lower pattern correlations (0.37–0.65) or substantially overestimated deviations (e.g., 2.26–2.37 by Tiedtke and BMJ).

The above comparisons show large sensitivity to cumulus parameterization schemes in simulating spatial pattern of extreme events, which is consistent with previous studies (Sun and Liang 2020a, 2023a, b). The sensitivity also differs among extreme precipitation indices—the CWRF downscaling ability is generally more skillful in capturing R95P and CDD than SDII and R95T. To quantify overall skills in reproducing long-term averaged spatial patterns of extreme precipitation characteristics, Fig. 3 displays the ranks among ERI and 28 CWRF members on each extreme precipitation index. The ranking is based on the comprehensive rating metrics (MR) defined in the Supplementary Information equation (S2) following Jiang et al. (2015). The MR measures the composite performance of three key statistics (pattern correlation, spatial deviation, RMSE) in the Taylor diagram. It is arranged in the increasing order such that a smaller rank number (more red boxes) indicates an overall higher skill. The result highlights a few CWRF physics configurations that notably improve the overall skill over CTL. In particular, the CAML, CCCMA, and CAM radiation schemes and the NSAS cumulus scheme are overall more skillful when they replace GSFCLXZ and ECP respectively in the CTL configuration. Thus, there is still large room for further improvement in simulating YRB extreme precipitation by refining model physics representation or through optimizing the multi-physics ensemble.

3.2 Impact of physics configurations on extreme precipitation interannual variations
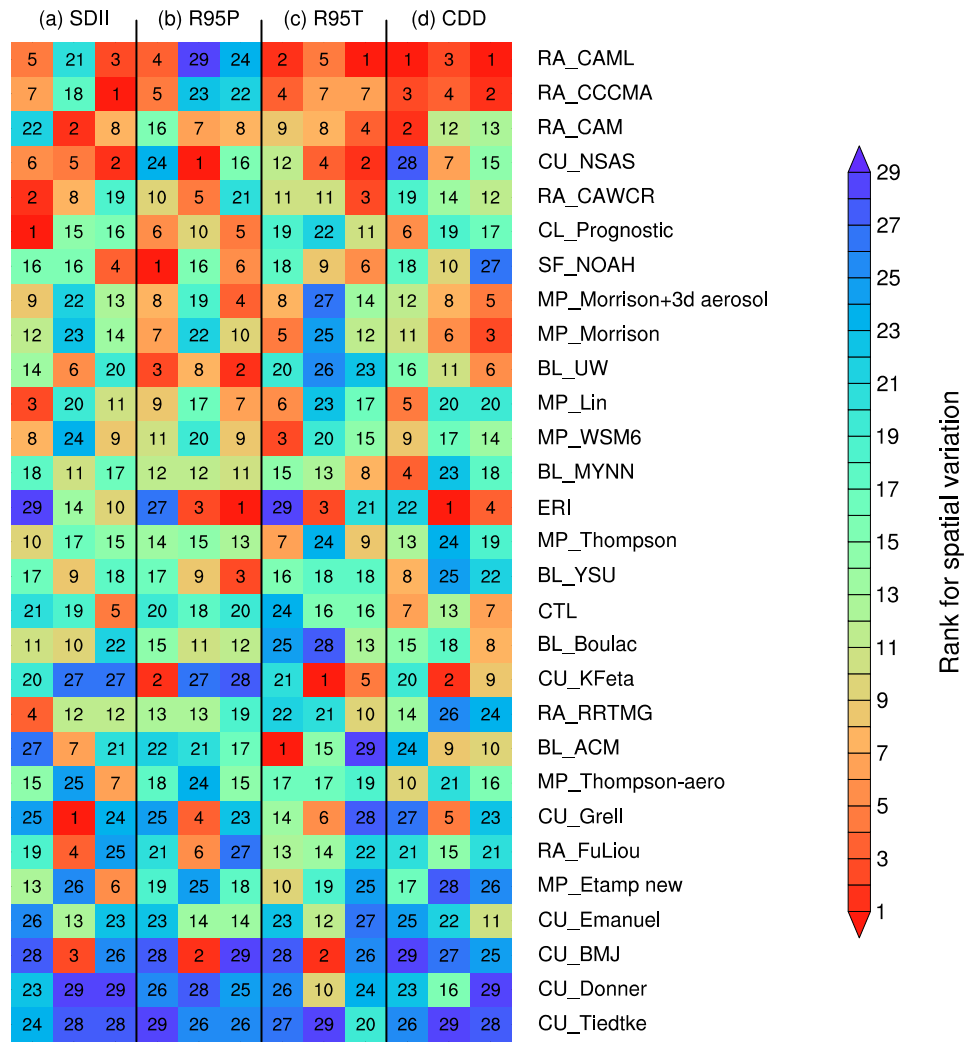
Figure 4 compares the performances among ERI and 28 CWRF physics configurations in capturing observed YRB regional mean interannual variations of summer extreme precipitation indices during 1980–2015. For all the extreme precipitation indices, ERI produces good correlations (0.52–0.79) and reasonable deviations (0.76–1.21). This is expected since ERI has assimilated pseudo rainfall observations on a daily basis that should have contained most authentic temporal features (Liang et al. 2019b; Sun and Liang 2020a). In contrast, CWRF CTL produces smaller interannual correlations with observations (0.41–0.64) and larger temporal deviations (1.32–2.46). The CWRF downscaling ability is still remarkable as compared to other models (Liang et al. 2019b).

Among all eight cumulus schemes, ECP exhibits an overall outstanding performance. In contrast, the KFeta scheme simulates comparable interannual correlations (0.40–0.57) and less overestimates temporal deviations (1.28–1.70) for R95T and CDD. In addition, NSAS produces a slightly smaller correlation (0.54) and more realistic deviation (1.18) than ECP for SDII. Other cumulus schemes perform systematically worse, having much lower correlations and substantially underestimated (e.g., BMJ for SDII and R95P) or overestimated (e.g., Tiedtke and Donner for SDII and CDD) deviations. In particular, BMJ fails completely, with the worst or even negative correlations for SDII and R95T.

Other physical processes' schemes are relatively clustered for SDII, R95P, and R95T, producing interannual correlations between 0.45 and 0.67 and deviations between 0.85 and 1.60. They are more scattered for CDD, producing lower correlations (0.25–0.46) and excessively high deviations (1.34–3.26). A few members, such as the Etamp_new microphysics scheme, the YSU boundary layer scheme, and the NOAH surface scheme, perform persistently worse than the majority, having systematically lower correlation (0.23–0.50) or larger variability (1.16–3.26). These results indicate that CWRF well reproduces observed SDII, R95P and R95T interannual variations in the YRB, although it is more difficult to capture CDD. Most models have limited skills in simulating CDD interannual variations during the summer monsoon (Jiang et al. 2015).

Figure 5 compares the performance ranks among ERI and 28 CWRF members on simulating interannual variations of

Fig. 3 The rank on the ability to simulate the climatological mean spatial distribution for each extreme precipitation index over the YRB, in terms of corresponding spatial correlation (left), normalized standard deviation (center), and root-mean-square error (right): **a** SDII, **b** R95P, **c** R95T, and **d** CDD. The number listed in each grid cell represents the respective rank. The ordering models names from top to bottom follow their averaged ranking across all indices

(a) SDII			(b) R95P			(c) R95T			(d) CDD			
5	21	3	4	29	24	2	5	1	1	3	1	RA_CAML
7	18	1	5	23	22	4	7	7	3	4	2	RA_CCCMA
22	2	8	16	7	8	9	8	4	2	12	13	RA_CAM
6	5	2	24	1	16	12	4	2	28	7	15	CU_NSAS
2	8	19	10	5	21	11	11	3	19	14	12	RA_CAWCR
1	15	16	6	10	5	19	22	11	6	19	17	CL_Prognostic
16	16	4	1	16	6	18	9	6	18	10	27	SF_NOAH
9	22	13	8	19	4	8	27	14	12	8	5	MP_Morrison+3d aerosol
12	23	14	7	22	10	5	25	12	11	6	3	MP_Morrison
14	6	20	3	8	2	20	26	23	16	11	6	BL_UW
3	20	11	9	17	7	6	23	17	5	20	20	MP_Lin
8	24	9	11	20	9	3	20	15	9	17	14	MP_WSM6
18	11	17	12	12	11	15	13	8	4	23	18	BL_MYNN
29	14	10	27	3	1	29	3	21	22	1	4	ERI
10	17	15	14	15	13	7	24	9	13	24	19	MP_Thompson
17	9	18	17	9	3	16	18	18	8	25	22	BL_YSU
21	19	5	20	18	20	24	16	16	7	13	7	CTL
11	10	22	15	11	12	25	28	13	15	18	8	BL_Boulac
20	27	27	2	27	28	21	1	5	20	2	9	CU_KFeta
4	12	12	13	13	19	22	21	10	14	26	24	RA_RRTMG
27	7	21	22	21	17	1	15	29	24	9	10	BL_ACM
15	25	7	18	24	15	17	17	19	10	21	16	MP_Thompson-aero
25	1	24	25	4	23	14	6	28	27	5	23	CU_Grell
19	4	25	21	6	27	13	14	22	21	15	21	RA_FuLiou
13	26	6	19	25	18	10	19	25	17	28	26	MP_Etamp new
26	13	23	23	14	14	23	12	27	25	22	11	CU_Emanuel
28	3	26	28	2	29	28	2	26	29	27	25	CU_BMJ
23	29	29	26	28	25	26	10	24	23	16	29	CU_Donner
24	28	28	29	26	26	27	29	20	26	29	28	CU_Tiedtke

Rank for spatial variation
29 27 25 23 21 19 17 15 13 11 9 7 5 3 1

extreme precipitation indices. The Morrison plus 3d aerosol or Morrison microphysics scheme improves CWRF skills when replacing the GSFCGCE scheme in the control CWRF. Immediately following CTL, the CAML, CCCMA and CAM radiation schemes produce good skills as the GSF-CLXZ scheme. Note that these radiation schemes are also identified earlier as top-skilled in reproducing the long-term average spatial distributions of extreme precipitation. They are preferred for CWRF to consistently capture both spatial pattern and interannual variability of extreme precipitation over the YRB.

3.3 Overall model skill and optimal multi-physics ensemble

Figure 6 shows the scattering relationship in terms of the MR ranks between spatial distributions and interannual variations among ERI and 28 CWRF physics configurations. The ranks among all models in capturing observed spatial distributions and interannual variations over the YRB are correlated with a coefficient of 0.63, which is statistically significant at the 5% significance level. Thus, the models that more accurately capture the spatial pattern tend to more realistically reproduce the interannual variation of regional extreme precipitation, and vice versa. Similar significant correspondences were reported among CMIP5 models in simulating extreme precipitation over eastern China (Jiang et al. 2015). It is interesting to note that ERI ranks much higher for interannual variation (0.84) than spatial pattern (0.53). So do CWRF CTL and the configurations using Morrison or Morrison plus 3d aerosol microphysics scheme.

The ranks differ largely among the CWRF physics configurations. Among six major physical processes, the highest sensitivity is identified to cumulus schemes which show the largest scattering range of the ranks. Among all eight cumulus schemes, ECP is overall superior in capturing both spatial pattern and interannual variability, having balanced highest spatial and temporal MR values. Although NSAS ranks higher than ECP for the spatial distribution, it ranks significantly lower for the interannual variation, producing less
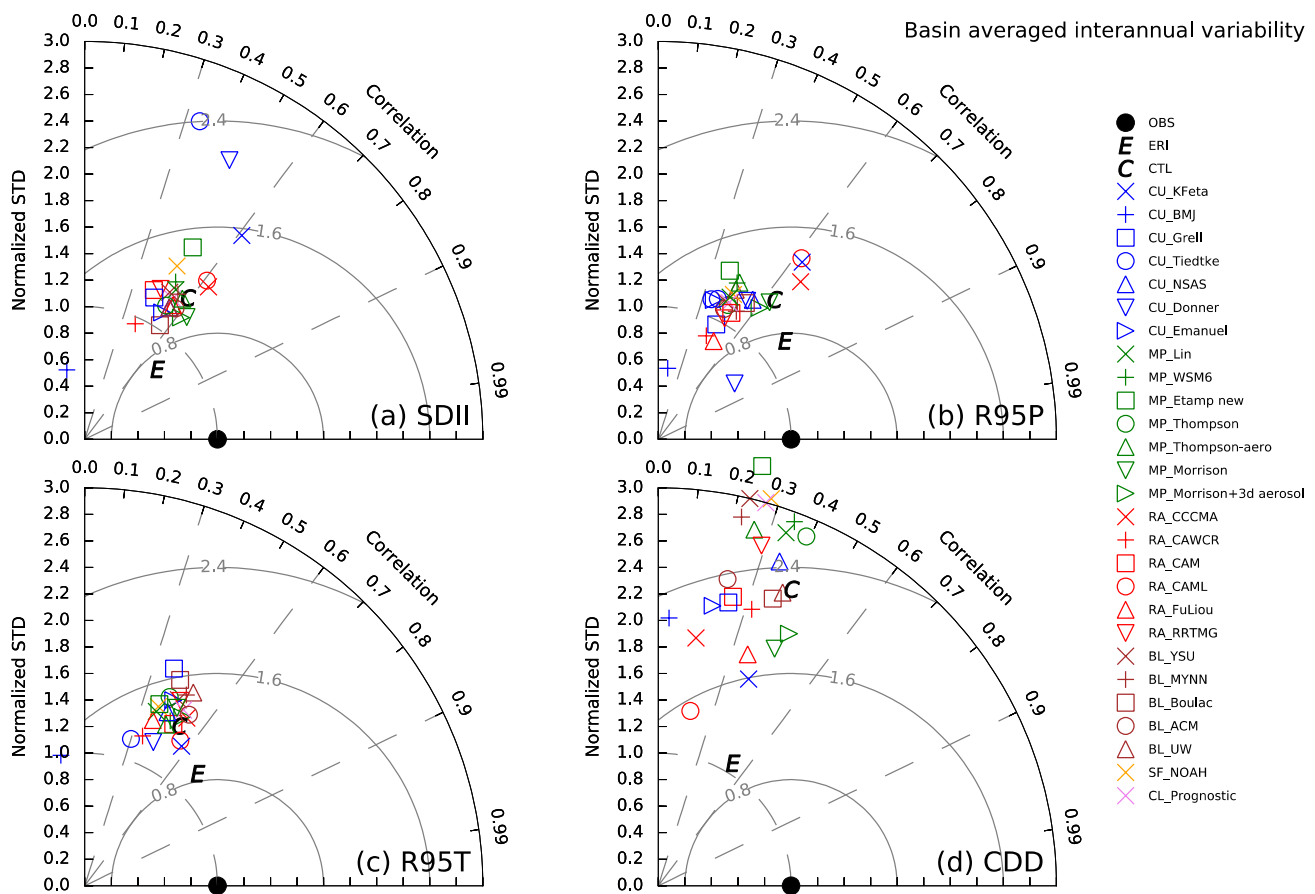
Fig. 4 Same as Fig. 2 except for simulating interannual variability of summer four indices during 1980–2015 averaged over the YRB. Shown are the corresponding interannual correlations (azimuthal) and normalized standard deviations (radius) compared with observations

intense precipitation and more consecutive dry days (Fig. 1). Other cumulus schemes rank much worse. Especially, BMJ, Tiedtke, and Donner schemes perform the worst, with MR values systematically less than 0.3. The result is consistent with the conclusion of Zhang et al. (2023) that ECP overall best represents P95 spatial distribution in China, while the other four schemes either overestimated (KFeta, Tiedtke) or underestimated (BMJ, NSAS) it. They showed that, in Central China (YRB), summer P95 interannual departures simulated by ECP are mainly associated with positive moisture convergence (27%) and negative convective available potential energy (18%) departures. The ECP better captures the balance of the two opposite factors for a more realistic P95 simulation in the YRB.

We can easily identify the five top-ranked CWRF configurations using the Morrison and Morrison plus 3d aerosol microphysics schemes and the CCCMA, CAML and CAM radiation schemes, which produce the spatial and temporal MR values larger than 0.56 and 0.53 respectively. As coupled with the ECP cumulus scheme, these radiation and microphysics schemes significantly enhance the CWRF ability to capture observed spatiotemporal variations of extreme

precipitation over the YRB. The result implies that realistic extreme precipitation simulations require improved system coupling especially among cumulus, microphysics and radiation processes (Sun and Liang 2020a, b).

Given the superior performance of ECP to other cumulus schemes, our best multi-physics ensemble mean (BMPE) integrates ECP with the Morrison and Morrison plus 3d aerosol microphysics schemes as well as the CCCMA, CAML, and CAM radiation schemes. This BMPE is designed to enhance CWRF's ability to accurately capture extreme precipitation characteristics in the YRB. The BMPE construction is detailed in the Supplementary Information. Figure 7 compares the geographic distributions of the four extreme precipitation indices among observations, ERI, CWRF CTL, and BMPE. Also shown are the corresponding spatial pattern correlation, RMSE, and bias over the YRB between each simulation and observations.

The observed rainfall intensity (SDII) shows maxima of more than 16 mm day^{-1} over broad areas along the Yangtze River, especially in the middle and lower reaches of the basin (Fig. 7a). ERI totally misses this intensity core, with peaks smaller than 13 mm day^{-1} and only in the upper reach,

Fig. 5 Same as Fig. 3 except for interannual variability for each extreme precipitation index averaged over the YRB, in terms of corresponding interannual correlation (left), normalized standard deviation (center), and root-mean-square error (right)
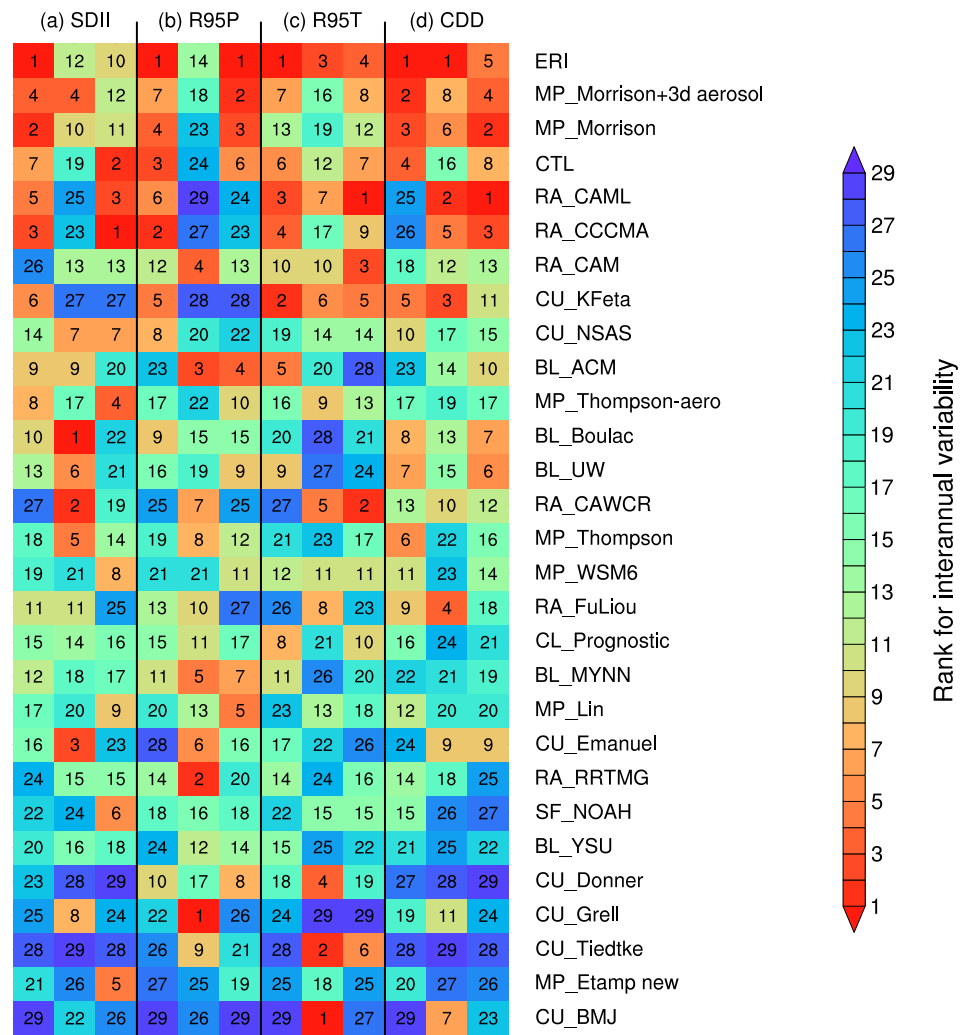
leading to a negative pattern correlation and a large RMSE with a systematic underestimation by 1.6 mm day^{-1} as averaged over the YRB. In contrast, CWRF CTL realistically captures the core with a sufficient intensity and reasonable distribution, notably increases the pattern correlation by 0.46 and reduces RMSE by 10% with a slight underestimation by 0.2 mm day^{-1} on average. BMPE further improves the CWRF skill over its CTL, increases the correlation by 0.11 and reduces RMSE by 4%, although it shrinks the area of the core with a larger underestimation of average 1.0 mm day^{-1}.

For R95P (Fig. 7b), ERI still produces insufficient amounts, causing a large dry bias of 7.6 mm as averaged over the YRB. Again, CWRF CTL better captures the spatial distribution, increasing the pattern correlation over ERI by 0.14, but substantially overestimates the amount by 14.1 mm on average, increasing RMSE by 53%. BMPE further improves the CWRF skill over its CTL, increases the correlation by 0.12 and largely reduces both RMSE (by 19%) and average wet bias down to 1.1 mm.

For R95T (Fig. 7c), CWRF CTL outperforms ERI, increasing the pattern correlation by 0.14 and reducing RMSE by 10%. As compared with observations, CTL simulates an expanded coverage of strengthened R95T, causing a positive bias of 0.8% as averaged over the YRB, whereas ERI substantially underestimates both the coverage and strength, causing a larger negative bias of 3.1%. BMPE significantly improves the CWRF skill over its CTL, increasing the correlation by 0.14 and reducing both RMSE by 18% and overestimation bias down to 0.5%.

For CDD (Fig. 7d), ERI systematically underestimates the magnitude by 2.7 days as averaged over the YRB, leading to the common drizzling problem. On the other hand, CWRF CTL overestimates the magnitude by 2.0 days on average, eliminating the drizzling problem albeit overdoing it somewhat. BMPE improves the CWRF skill over its CTL, increasing the pattern correlation by 0.03 and reducing RMSE by 11%, but still has a large overestimation bias by 2.2 days.

Fig. 6 Scatter diagrams of models' MR index on account of Taylor diagrams for spatial (x axis) and interannual (y axis) variation over the YRB. Labeled is the correlation coefficient (CC), which is statistically significant at the 5% significance level. The models in the upper-right quadrant perform well for both conditions
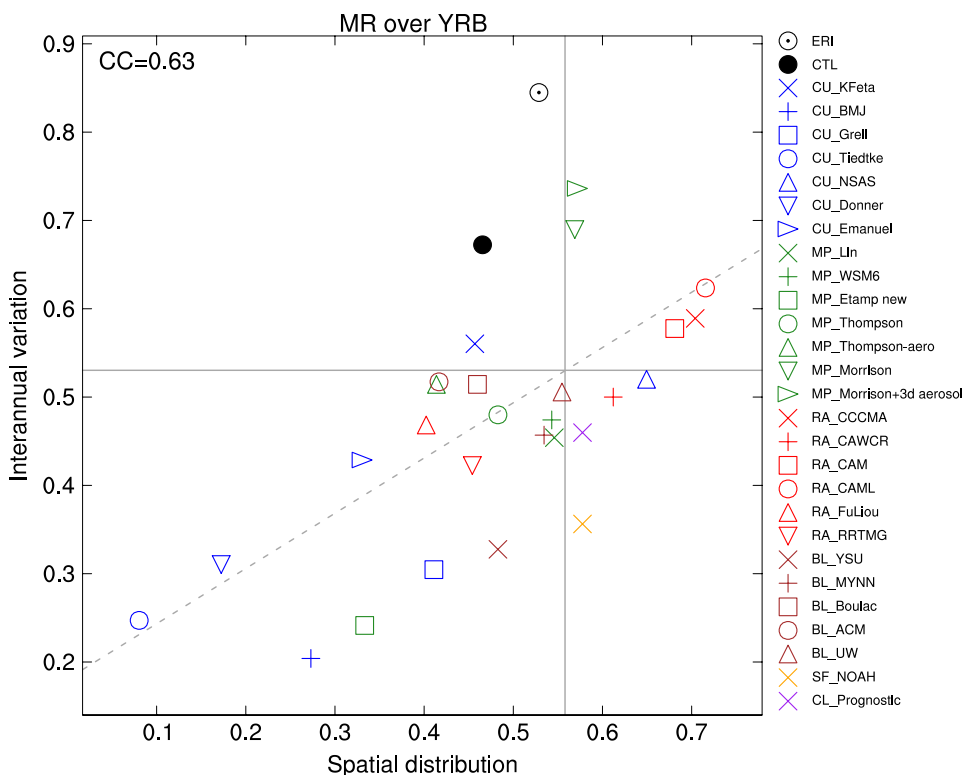
In summary, CWRF CTL enhances ERI skill for extreme precipitation, while BMPE further advances CTL ability. This consistent improvement from ERI to CTL to BMPE spans all eleven indices of extreme precipitation, including the additional P95, AEPI, R95N, R10, CWD, RX5day, and PMAX (Table S1). A detailed discussion of skills on these additional indices is provided in the Supplementary Information.

Figure 8 compares the geographic distributions of summer interannual correlations with observations between CWRF CTL and BMPE simulated four extreme precipitation indices during 1980–2015. Shown also are the density functions that depict the frequency distributions of the correlations at all CWRF grids within the YRB along with their respective percentages of areas that have significant correlations. We consider correlations greater than 0.28 to be indicative of skillful signals as they are statistically significant at the 5% significance level by the one-tail student's t-test. CWRF CTL captures observed interannual anomalies over 28.5%, 24.8%, 15.9%, and 26.9% areas of the YRB for SDII, R95P, R95T, and CDD, respectively. Most of these signal areas occur along the Yangtze River and to the south of its middle and lower reaches. Skills are lacking mainly in the regions between the Yellow and Yangtze Rivers. BMPE largely improves the skills over CTL for all the four indices as clearly shown by the systematic shift of the frequency density curve toward the

higher correlation end. BMPE captures observed interannual anomalies over 47.6%, 40.8%, 28.9%, and 33.5% areas of the YRB for SDII, R95P, R95T, and CDD, respectively. The added values of BMPE to CTL are the expanded coverages of significant correlations with observations by 19.1%, 16.0%, 13.0%, and 6.7% areas of the YRB for the four indices. There remain large areas (52.4–71.1%) where correlations are insignificant. This indicates the big challenge in capturing extreme precipitation interannual variations over the YRB, where prevailing convective systems during the summer monsoon are difficult to predict (Liang et al. 2019b; Li et al. 2020).

Figure 9 compares interannual variations of CWRF CTL and BMPE simulated with observed anomalies of the four extreme precipitation indices averaged over the YRB. Also shown are the interannual correlation coefficient and RMSE between the simulated and observed anomalies during 1980–2015 for each index. The CTL captures well observed anomalies with correlations of 0.59, 0.64, 0.51, and 0.41 for SDII, R95P, R95T, and CDD, respectively. BMPE improves the SDII, R95P, and R95T skills over CTL, increasing the correlations with observations by 0.07, 0.04, and 0.07 and reducing RMSE by 17, 12, and 17%, respectively. In contrast, for CDD, BMPE reduces from CTL RMSE by 30% but also the correlation by 0.05. As discussed earlier, it is more difficult to capture CDD variations.
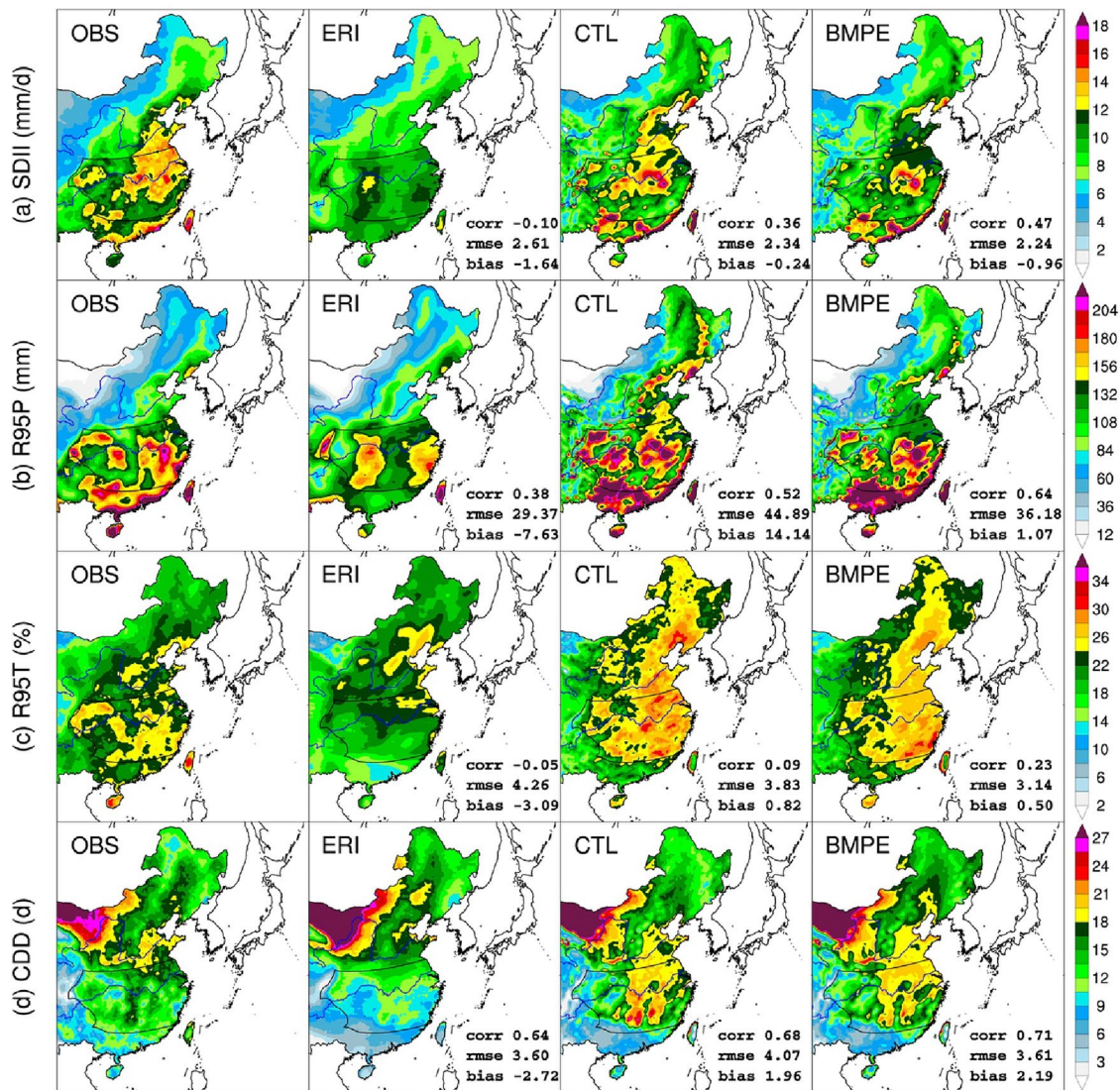
Fig. 7 Spatial distributions of summer four indices as observed (OBS), ERI, CWRF control (CTL), and the best multi-physics ensemble mean (BMPE): **a** SDII, **b** R95P, **c** R95T, and **d** CDD. Listed are the corresponding spatial pattern correlation (corr), root-mean-square error (rmse), and bias over the YRB between each simulation and observations

3.4 Regional circulations associated with extreme precipitation biases

To explore possible causes for extreme precipitation biases, CWRF modeled atmospheric circulations in CTL and the five top ranked configurations that differ only in microphysics (Morrison, Morrison plus 3d aerosol) and radiation (CCCMA, CAML, CAM) schemes as well as BMPE are compared with ERA5. See the Supplementary Information for the reason selecting ERA5 as the reference. Note that the changes in CWRF experiments are only by switching one scheme from CTL, so the circulation differences are induced by that specific physics representation. The EAJ and the Hadley cell are two distinct circulation systems dominating

east China monsoon rainfall (Liang and Wang 1998), and hence are elaborated in the comparison below. Figures 10 and 11 compare the summer mean circulation characteristics, including 200/850 hPa wind, 500 hPa geopotential height, and column moisture flux geographic distributions, as well as latitude-altitude cross-sections of wind and latitudinal variations of R95P biases from observations averaged across eastern China (105°-122°E), where the YRB covers latitudes around 24°–34°N.

As revealed in ERA5, the strong westerly jet stream prevails at 200 hPa, with the maximum speed exceeding 30 m s^{-1} over Xinjiang and the jet axis located at approximately 40°N (Fig. 10). The EAJ exit stretches across North China to Japan, having the YRB persistently beneath its core
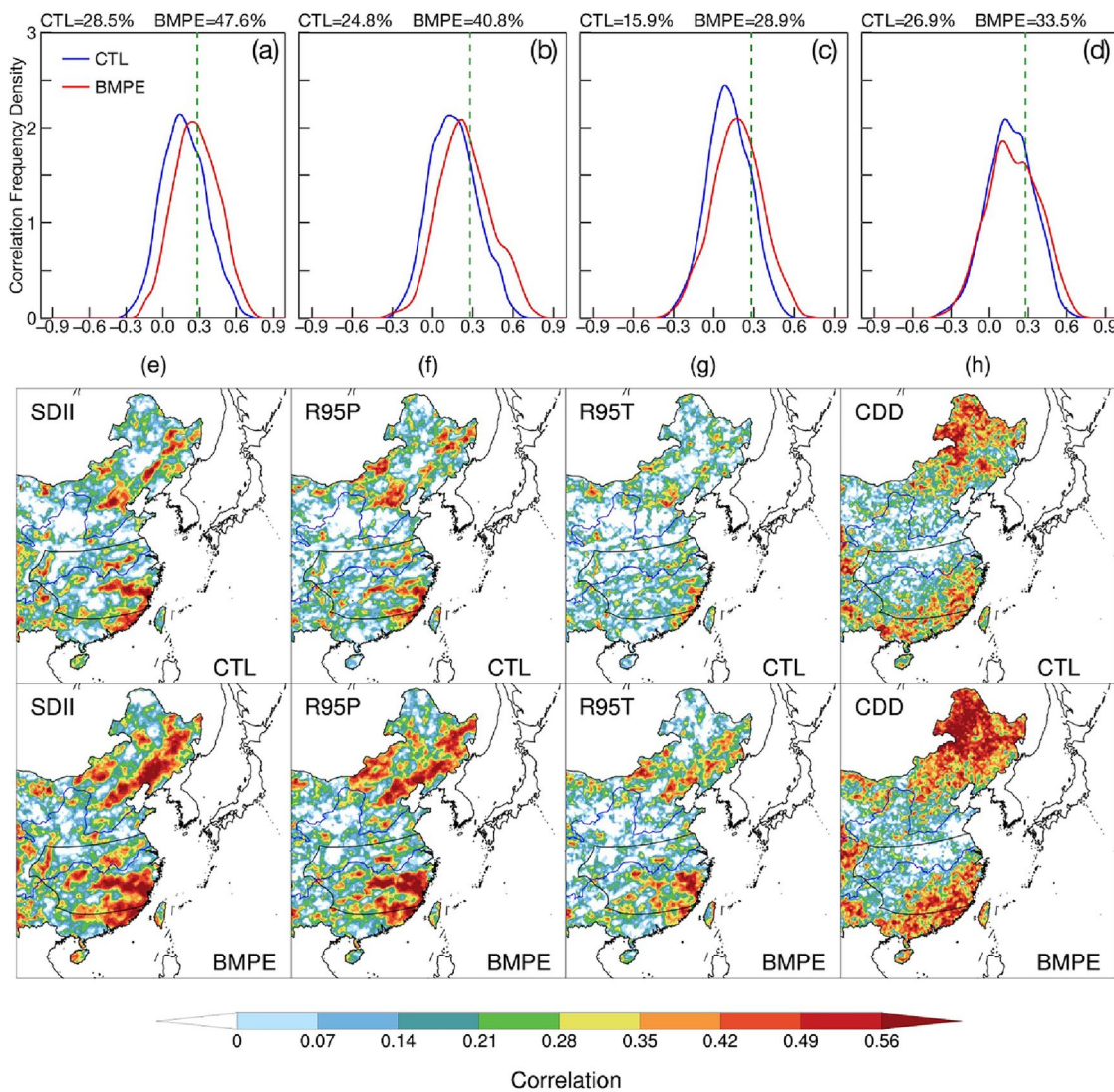
Fig. 8 Spatial distributions of CWRF CTL and BMPE simulated 1980–2015 summer four indices interannual correlations with observations: **e** SDII, **f** R95P, **g** R95T, and **h** CDD. Also shown (**a–d**) are the corresponding frequency density functions at all grids over the YRB. The correlations greater than 0.28 as marked by the vertical lines denote the 5% significance level with the one-tail student's t-test. Labeled at the top of each panel are the respective percentage of areas over the YRB that have significant interannual correlations

to the right (south) side. The western Pacific subtropical high, depicted by the 5,860-gpm contour of 500 hPa geopotential height, occupies the southeast coasts. Meanwhile, two branches of low-level southerly monsoon flows sweep eastern China, one carrying water vapor from the Bay of Bengal and the other from the South China Sea and the western Pacific Ocean, which generate high moisture flux convergence over the south of the Yangtze River. The secondary meridional circulation crossing the EAJ exit in accord with the Hadley circulation produces prevailing ascent motions, resulting in major precipitation in the YRB (Figs. 11 and S4). ERA5 tends to overestimate the ascent strength, causing significant wet R95P biases along the Yangtze and Pearl Rivers.

Compared with ERA5, the control CWRF physics configuration weakens the jet stream, shrinking the EAJ exit westward to Hebei (Fig. 10). The subtropical high is slightly extended to northwest and accompanied with stronger low-level southwesterlies along its western ridge, causing stronger moisture fluxes to the upper reaches of the Yangtze and Peral Rivers but weaker fluxes in the lower reaches near the coast. The EAJ underestimate, up to 3 m/s over the southern North China and northern YRB areas, leads to stronger ascents directly beneath its exit, while the Hadley circulation intensification causes stronger ascents in South China (Fig. 11). Between the two ascending branches, descending motions occur in the southern YRB. As a result, the CWRF CTL reduces ERA5 wet R95P biases in the YRB
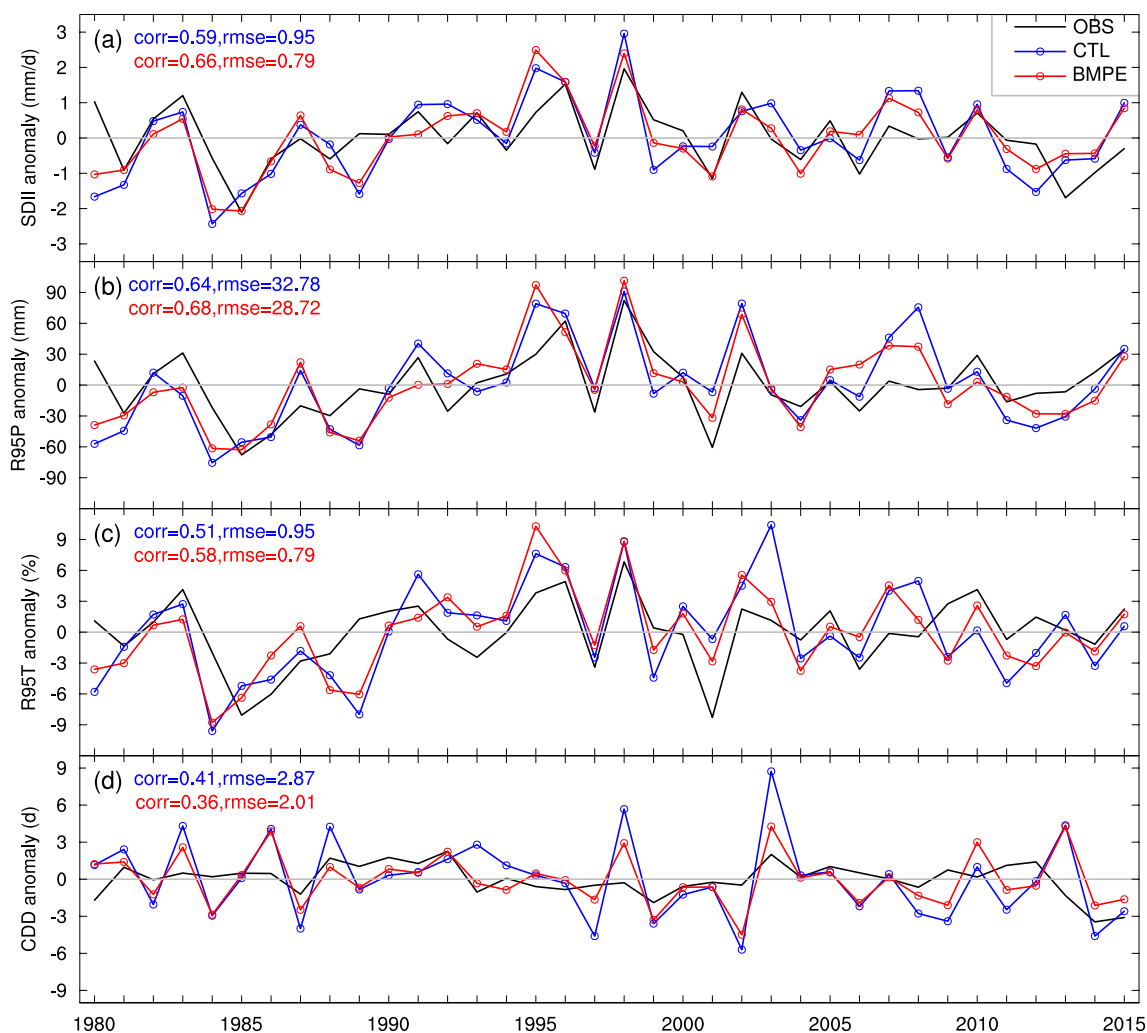
Fig. 9 Interannual anomalies of summer four indices during 1980–2015 averaged over the YRB as observed (OBS), CWRF CTL, and BMPE: **a** SDII, **b** R95P, **c** R95T, and **d** CDD. Listed are the corresponding interannual correlation (corr) and root-mean-square error (rmse) with observations

but increases them in North and South China (Figs. 11 and S4).

When CWRF couples the ECP cumulus and Morrison or Morrison plus 3d aerosol microphysics schemes, the jet stream is further weakened and its EAJ exit is shifted farther westward to Shanxi. Although the subtropical high and associated low-level southerlies over eastern China are better simulated, the moisture convergence is underestimated over larger YRB areas than CTL. The larger EAJ underestimate, up to 4 m/s over the southern North China and northern YRB areas, reduces upward motions and rainfall beneath its exit, while the more intensified Hadley circulation enhances upward motions and rainfall in South China, compared to CTL. These lead to small dry R95P biases in the YRB and decreased wet biases over North China but increased wet biases over South China. Using the Morrison versus Morrison plus 3d aerosol scheme results in marginally reduced

overall circulation biases. Consequently, the former exhibits comparatively improved performance in simulating extreme precipitation.

When CWRF adopts the CCCMA radiation scheme, the EAJ exit is shifted westward to Shanxi. Relative to CTL, the western ridge of the subtropical high exhibits a more accurate inland extension. This, coupled with strengthened low-level southwesterlies over the southern YRB and South China, contributes to intensified moisture fluxes south of the Yangtze River but weakened the fluxes to the north. Compared to CTL, the larger EAJ underestimate, up to 4 m/s over the southern areas of North China, produces weaker ascending motions in North China, while the widened Hadley circulation yields stronger ascending motions in the southern YRB and South China. As a result, the CCCMA reduces CTL's wet R95P biases in North China but increases them in the YRB and South China.
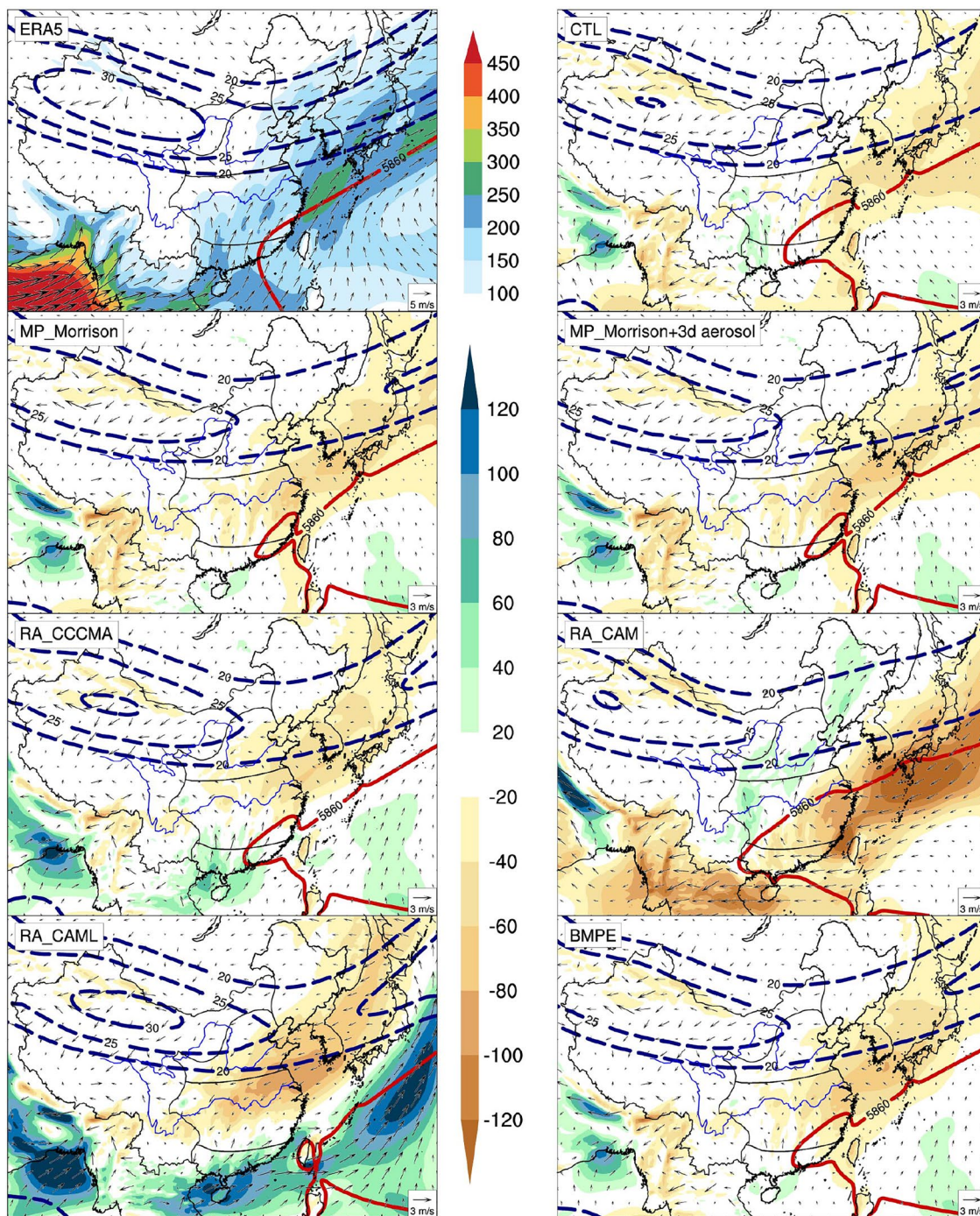
Fig. 10 Climatology of summer mean wind at 850 hPa (m s^{-1}, vectors) and vertically integrated (1000–300 hPa) moisture flux (kg m^{-1} s^{-1}, color shadings) based on ERA5 and their respective departures (from ERA5) simulated by seven CWRF physics configurations.

Overlaid are the corresponding wind speed at 200 hPa (m s^{-1}, dashed contours starting from 20 at an interval of 5) and 500-hPa geopotential height's 5860 gpm (red solid contour)

When CWRF adopts the CAML radiation scheme, the EAJ exit is shifted eastward to Hebei. This shift is accompanied with an eastward displacement of the subtropical high and its inland ridge extension as well as the low-level southwesterlies, causing northeasterly flow perturbations in North

China and north of the Yangtze River. Consequently, the moisture flux convergence is significantly underestimated in North China and the northern YRB but enhanced in South China. The EAJ underestimate, by 2.5 m/s in the southern areas of North China and 3 m/s in the YRB, produces
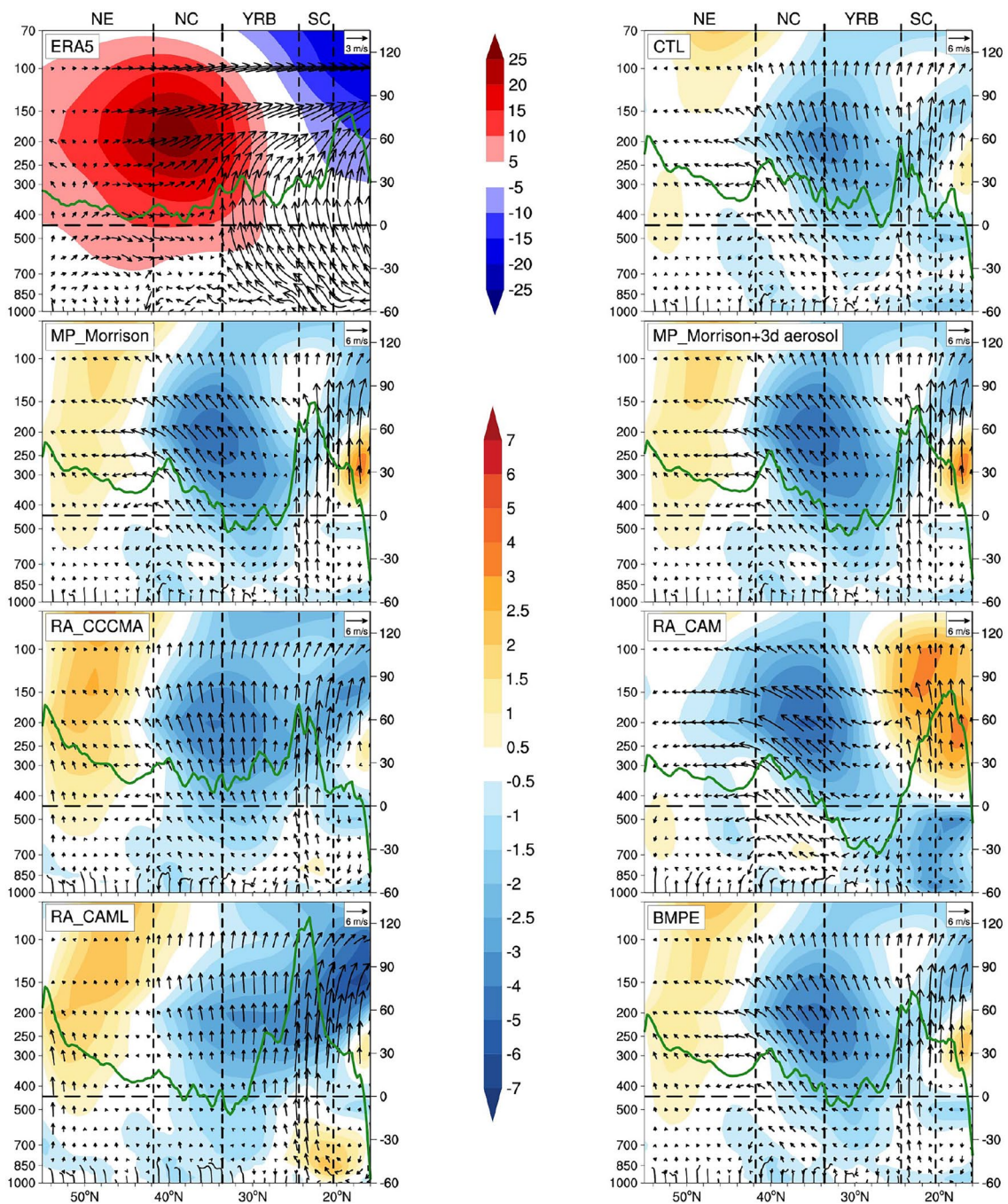
Fig. 11 Climatology of summer mean latitude-altitude wind circulation distributions averaged across 105°–122°E based on ERA5 and their respective departures (from ERA5) simulated by seven CWRF physics configurations. Color shadings and arrows denote the zonal (m s^{-1}) and meridional (m s^{-1})/vertical (10^{-3} m s^{-1}) wind components, respectively. Overlaid are the corresponding biases (from observations) in latitudinal variations of R95P (mm, solid curves with the scale on the right). The YRB spans latitudes around 24°–34°N, while the approximate bands for Northeast, North, and South China are marked by NE, NC, and SC, respectively

weaker ascending motions north of the Yangtze River, while the significantly intensified Hadley circulation yields much stronger ascending motions south of the Yangtze River. Compared to CTL, these changes cause the monsoon rainband shifted southwards. As a result, the CAML scheme

notably reduces CTL's wet R95P biases in North China but increases them in the YRB and more substantially in South China.

Contrarily, the adoption of the CAM radiation scheme results in a westward shift of the EAJ exit to Ningxia and

an expansion of the subtropical high northwestward, with its ridge covering larger areas of southeastern China. Compared to CTL, these changes cause enhanced moisture fluxes across an extensive region stretching from the western YRB to Northeast China. Simultaneously, moisture fluxes are diminished in the eastern YRB and South China due to low-level perturbations of southerly and easterly flows, respectively. The larger underestimate of the EAJ exit by 5 m/s in southern North China, coupled with the low-level easterly flow overestimate by 3 m/s in South China, leads to weaker ascending motions north of the Yangtze River than CTL. Thus, the CAM scheme reduces CTL's overall wet R95P biases in North and South China but produces significant dry biases in the YRB.

On average of the five top-ranked CWRF configurations (Morrison, Morrison plus 3d aerosol, CCCMA, CAML, CAM), BMPE more accurately captures the subtropical high and low-level southerlies over eastern China, although still shifting the EAJ exit westward to Shanxi and underestimating moisture fluxes in the YRB. The EAJ exit underestimate by 4 m/s in southern North China leads to weaker ascending motions in North China and northern YRB than CTL, while the slightly intensified Hadley circulation causes stronger ascending motions in southern YRB and South China. Due to the error cancelation within the chosen configurations, BMPE yields minor wet R95P biases in the YRB and North China. However, large wet biases persist in South China.

Compared to ERA5, ERI produces systematically weaker westerly jet stream at 200 hPa over the entire CWRF domain, shrinking the EAJ exit westward to Hebei and simulating easterly departures in North China and northern YRB over 1.2 m/s as well as in South China over 2.5 m/s (Fig. S5). These alterations suppress ascending motions and decrease rainfall in the YRB and more strongly in South China. Therefore, the general underestimation of the EAJ exit in CWRF may be largely driven by the ERI forcing errors. Our results highlight the importance of physics representation for realistic regional extreme precipitation simulation.

4 Conclusions and discussion

This study employs the CWRF downscaling and its skillful multi-physics ensemble approach to enhance summer extreme precipitation prediction over the YRB. It quantifies the CWRF ability in downscaling spatial patterns and capturing interannual variations in four key extreme precipitation indices during 1980–2015, while comparing the results against the driving ERI reanalysis and ranking performance across 28 different combinations of physics parameterizations. We embrace a comprehensive evaluation strategy, incorporating multiple metrics across all indices alongside understanding of the linkages to regional

circulation patterns. The skill assessment comprises spatial and temporal correlation, root-mean-square-error (RMSE), standard deviation, and bias of each index, as well as mean absolute relative bias (MARB) of all indices and a comprehensive ranking metric (MR) based on relative scores of correlation, deviation, and RMSE among these indices. The finest CWRF physics configurations are identified through MR ranking of all four indices to construct the best multi-physics ensemble (BMPE). This ensemble is compared with the control CWRF to explore skill enhancement from varying physics representation and gain insights into model biases of YRB extreme precipitation and their connections to regional circulation patterns. The main findings are summarized below:

First, ERI notably underestimates all four indices, despite its extensive utilization of comprehensive data assimilation. Conversely, the control CWRF downscaling substantially enhances the ability to accurately capture observed spatial patterns of extreme precipitation. The skill enhancements are particularly remarkable for SDII and R95T, as these indices display spatial structures and magnitudes that ERI struggles to replicate. The CWRF downscaling demonstrates considerable added value in capturing distinctive regional characteristics of extreme precipitation, achieved through enhanced physics representations. Of particular note, CWRF integrates the ECP scheme, which employs dynamic selections and optimal cumulus parameterization closure assumptions, differentiating between land and oceans (Qiao and Liang 2015, 2016, 2017). This integration leads to a significant enhancement in total precipitation intensity and rainy-day frequency (Sun and Liang 2020b).

Second, the CWRF downscaling capability varies across different extreme precipitation indices. Broadly, CWRF tends to underestimate SDII and R95P, while overestimating R95T and CDD. Among most physics configurations, spatial distributions of R95P and CDD are better captured than SDII and R95T. These configurations also reasonably replicate interannual variations across all indices except CDD. The MR analysis underscores a noteworthy pattern: configurations that aptly capture spatial distributions tend to reproduce interannual variations more accurately in regional extreme precipitation, and vice versa.

Third, CWRF downscaling skills exhibit substantial variability among various physics configurations, with cumulus parameterization being the most influence, evident by its wide-ranging impact across six primary physical processes. Of all eight cumulus schemes, the control ECP overall demonstrates remarkable proficiency in capturing both spatial patterns and interannual variations of extreme precipitation. While NSAS excels in simulating mean spatial patterns, its performance for interannual variations is much weaker, resulting in insufficient intense precipitation amounts and an elevated number of consecutive dry days. The remaining

cumulus schemes generally show lower skills, with BMJ, Tiedtke, and Donner notably underperforming and displaying pronounced biases.

Fourth, the five highest MR-ranked CWRF configurations incorporate the Morrison and Morrison plus 3d aerosol microphysics schemes, alongside the CCCMA, CAML and CAM radiation schemes, which replace the control GSFCGCE and GSFCLXZ schemes respectively. Coupling ECP cumulus scheme with these microphysics and radiation schemes significantly enhances the CWRF capability to accurately capture observed spatiotemporal variations in extreme precipitation across the YRB. While certain underestimations of SDII and overestimations of CDD still persist in localized areas, the ensemble average of these skill-enhanced physics configurations (BMPE) more faithfully reproduces observed geographic distributions and interannual anomalies for all four indices, surpassing the performance of the control CWRF.

Fifth, differences in EAJ and Hadley cell circulations, as well as their associated vertical motions and moisture fluxes, exhibit strong correlations with YRB extreme precipitation biases. The control ECP and its members combined with the CCCMA and CAML radiation schemes simulate slightly weaker EAJ and expanded Hadley circulations, fostering stronger ascending motions in the YRB. These changes coincide with stronger low-level southerly flows over southeastern China, accompanied by enhanced moisture transport from the South China Sea and the western Pacific warm pool–resulting in notable wet R95P biases in the YRB. In contrast, the Morrison or Morrison plus 3d aerosol microphysics and CAM radiation schemes simulate relatively weakened EAJ and Hadley circulations that suppress strong ascending motions in the YRB. They also produce less intense moisture fluxes, contributing to significant dry biases in the YRB. On the other hand, BMPE adeptly captures the overall summer mean circulation features and displays minimal wet biases in the YRB. The favorable outcome of the ensemble mean primarily stems from the effective mitigation of errors among the chosen configurations.

To further explore the impact of model resolution on the CWRF's ability to downscale extreme precipitation, we considered the interlinkage between physics representation and spatial resolution, particularly the scale dependence of convection parameterization (Weisman et al. 1997; Jung and Arakawa 2004; Yu and Lee 2011; Field et al. 2017). Additional CWRF experiments are conducted using the control physics configuration at 30, 15, and 10 km grid spacings to determine the resolution sensitivity. In these experiments, all surface boundary conditions are constructed accordingly to match the increased resolution (Liang et al. 2005; Xu et al. 2014). Figure 12 contrasts summer geographic distributions of the four extreme precipitation indices for the year 2003 across these different grid spacings. Finer resolution

simulations are mapped onto the 30-km grid for uniform analysis, assessing mean biases, spatial correlations, and RMSEs with respect to observations over the YRB. The simulations at the 15-km grid spacing generally outperform those at 30-km across the YRB, increasing spatial pattern correlations with observations and generally reducing RMSEs, although some biases shift, notably for R95P. Conversely, the 10-km simulation shows a consistent decline in performance compared to the 15-km run, with reduced pattern correlations for most indices except SDII as well as increased RMSEs and biases except for CDD.

Convection-permitting model (CPM) simulations may improve extreme precipitation prediction. Liang et al. (2019a) systematically explored the efficacy of various WRF model configurations of grid nesting (from 30, 15, 9, 5, 3 to 1 km, single or double or triple nested grids) and convection treatment (the traditional or scale-aware cumulus parameterization or the explicit convection) for Jiangsu's Meiyu rainfall forecasts. They concluded that the double nested approach combining cumulus parameterization at a 15-km grid with explicit convection at a 1-km grid offers an effective solution to more accurate rainfall forecasting, particularly for clear and heavy to extreme rain events. This approach avoids the challenge in representing convections across scales. Such resolution sensitivity in the outer domain resulted from solely cumulus parameterization. Our result may indicate that the ECP cumulus scheme performs the best at 15 km. Dong et al. (2022) showed advantages of the WRF at the 1.5-km inner grid (resolving convection) over the 9-km outer grid (parametrizing cumulus) in simulating extreme precipitation on sub-daily timescales in the Yangtze River Delta. However, they also identified limitations for the CPM to capture the duration and coverage of heavy precipitation and the occurrence of longer-duration events. These studies underscore the complexity of high-resolution model abilities in capturing extreme precipitation characteristics. Improvements cannot be made by CPMs without enhancing physics representations to fit the refined resolution. While important to a complete understanding of extreme precipitation predictability, CPM simulations are impossible due to the lack of computing resources and beyond the scope of this study, whose main objective is to determine skill dependence on model physics representation at 30-km and identify those configurations that can enhance their ensemble performance. Nonetheless, our results provide solid evidence for the China's National Climate Center to improve its operational seasonal forecasts by optimizing CWRF multi-physics ensemble with a reduced number of but skillful configurations than the current suite, preferably at 15-km as computing resources permit.

Further endeavors are essential to enhance the ensemble's performance by fine-tuning the weights for a broader suite of superior and diverse configurations, guided by
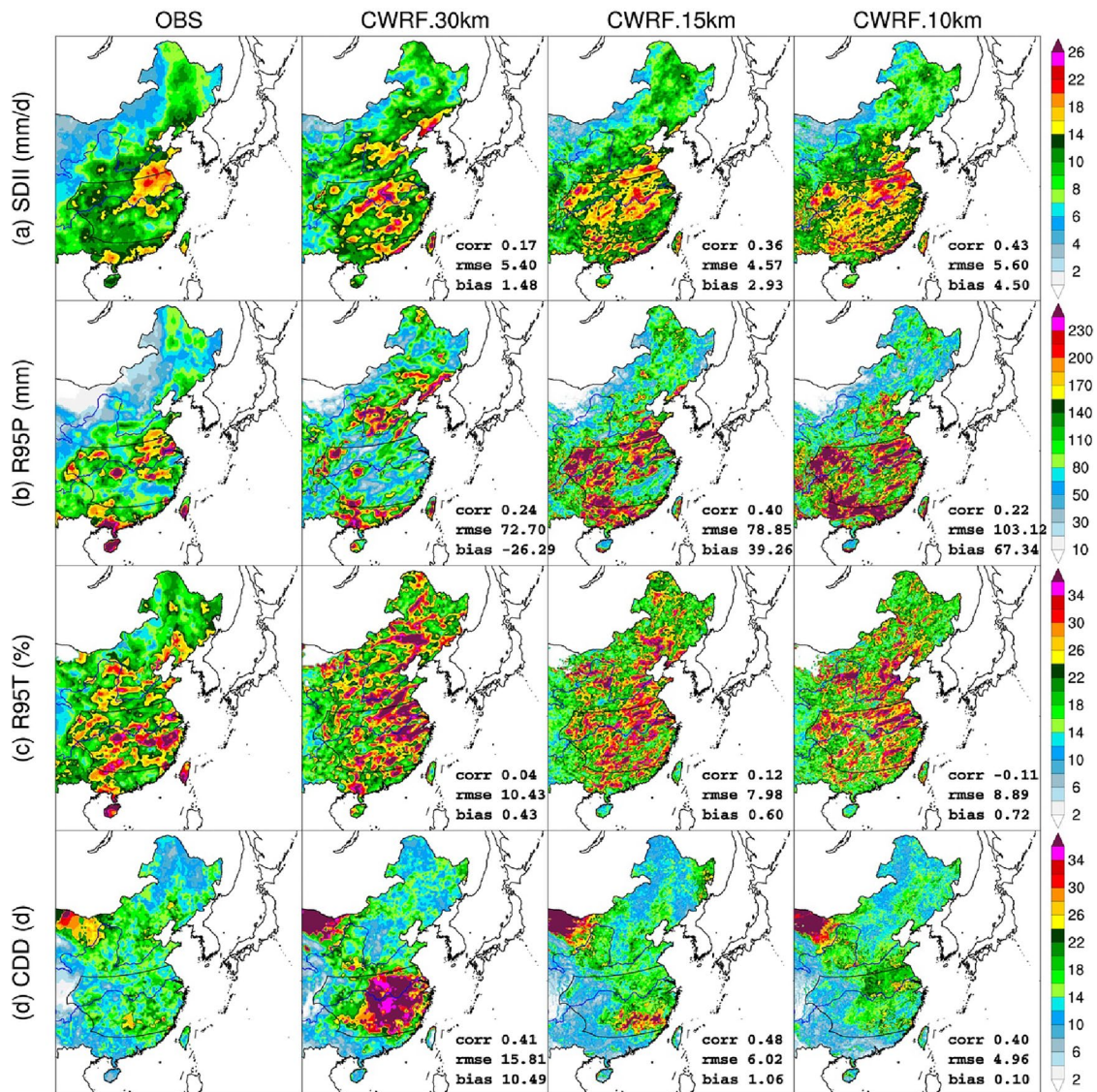
Fig. 12 Same as Fig. 7 except for the four indices in the summer of the year 2003 as observed (OBS) and simulated by CWRF control (CTL) at the respective grid spacings of 30, 15, and 10 km: **a** SDII, **b** R95P, **c** R95T, and **d** CDD

comprehensive model rankings across both spatial and temporal dimensions (Liang et al. 2007, 2012; Tang et al. 2021). The current study has confined its selection to only five top-ranked configurations, which, due to their limited diversity, may not offer sufficient spread to adequately address compensating errors for an optimal ensemble outcome. Moreover, while also discussing seven other indices, the focus of our skill assessment has centered around the four indices that capture the fundamental characteristics of extreme precipitation. Expanding the range to encompass more representative metrics, not solely for extreme precipitation but also for other statistical moments and even the entire daily frequency distribution, can lead to a more robust optimization. Despite these considerations, our findings stand as an encouraging

testament, underscoring the substantial potential of CWRF downscaling to elevate extreme precipitation predictions through the enhancement of its physics representations and the strategic optimization of its multi-physics ensemble. These physics enhancement and ensemble optimization can be integrated with resolution refinement to further increase prediction skill.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval We declare that the work presented in this manuscript is original and has not been currently under consideration or published in any other journals. We adhere to all other Ethical Responsibilities of Authors specified in COPE.

References

Chan SC, Kendon EJ, Fowler HJ, Blenkinsop S, Ferro CA, Stephenson DB (2013) Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation? Clim Dyn 41:1475–1495. https://doi.org/10.1007/s00382-012-1568-9

Choi HI, Liang X-Z (2010) Improved terrestrial hydrologic representation in mesoscale land surface models. J Hydrometeorol 11:797–809. https://doi.org/10.1175/2010JHM1221.1

Choi HI, Kumar P, Liang X-Z (2007) Three-dimensional volume-averaged soil moisture transport model with a scalable parameterization of subgrid topographic variability. Water Resour Res 43:W04414. https://doi.org/10.1029/2006WR005134

Choi HI, Liang X-Z, Kumar P (2013) A conjunctive surface–subsurface flow representation for mesoscale land surface models. J Hydrometeorol 14:1421–1442. https://doi.org/10.1175/JHM-D-12-0168.1

Cui D, Wang C, Santisirisomboon J (2019) Characteristics of extreme precipitation over eastern Asia and its possible connections with Asian summer monsoon activity. Int J Climatol 39:711–723. https://doi.org/10.1002/joc.5837

Dee DP, Uppala SM, Simmons A, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda M, Balsamo G, Bauer P (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

Dong T, Dong W (2021) Evaluation of extreme precipitation over Asia in CMIP6 models. Clim Dyn 57:1751–1769. https://doi.org/10.1007/s00382-021-05773-1

Dong G, Jiang Z, Wang Y, Tian Z, Liu J (2022) Evaluation of extreme precipitation in the Yangtze River Delta Region of

China using a 1.5 km mesh convection-permitting regional climate model. Clim Dyn 59:2257–2273. https://doi.org/10.1007/s00382-022-06208-1

Ek M, Mitchell K, Lin Y, Rogers E, Grunmann P, Koren V, Gayno G, Tarpley J (2003) Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. J Geophys Res Atmos 108:D22

Field PR, Brozková R, Chen M, Dudhia J, Lac C, Hara T, Honnert R, Olson J, Siebesma P, de Roode S (2017) Exploring the convective grey zone with regional simulations of a cold air outbreak. Q J R Meteorol Soc 143:2537–2555

Gao Z, Zhu J, Guo Y, Luo N, Fu Y, Wang T (2021) Impact of land surface processes on a record-breaking rainfall event on May 06–07, 2017, in Guangzhou. China. J Geophys Res Atmos 126:e2020JD032997. https://doi.org/10.1029/2020JD032997

He S, Yang J, Bao Q, Wang L, Wang B (2019) Fidelity of the observational/reanalysis datasets and global climate models in representation of extreme precipitation in East China. J Clim 32:195–212. https://doi.org/10.1175/JCLI-D-18-0104.1

Hersbach H, Bell B, Berrisford P, Horányi A, Sabater JM, Nicolas J, Radu R, Schepers D, Simmons A, Soci C (2019) Global reanalysis: goodbye ERA-Interim, hello ERA5. ECMWF Newsl 159:17–24

Huang D, Gao S (2017) Impact of different cumulus convective parameterization schemes on the simulation of precipitation over China. Tellus A Dyn Meteorol Oceanogr 69:1406264. https://doi.org/10.1080/16000870.2017.1406264

Jiang Z, Li W, Xu J, Li L (2015) Extreme precipitation indices over China in CMIP5 models. Part I: model evaluation. J Clim 28:8603–8619. https://doi.org/10.1175/JCLI-D-15-0099.1

Jiang R, Sun L, Sun C, Liang X-Z (2021) CWRF downscaling and understanding of China precipitation projections. Clim Dyn. https://doi.org/10.1007/s00382-021-05759-z

Jung J-H, Arakawa A (2004) The resolution dependence of model physics: illustrations from nonhydrostatic model experiments. J Atmos Sci 61:88–102

Kang I-S, Yang Y-M, Tao W-K (2015) GCMs with implicit and explicit representation of cloud microphysics for simulation of extreme precipitation frequency. Clim Dyn 45:325–335. https://doi.org/10.1007/s00382-014-2376-1

Kong X, Wang A, Bi X, Sun B, Wei J (2022) The hourly precipitation frequencies in the tropical-belt version of WRF: sensitivity to cumulus parameterization and radiation schemes. J Clim 35:285–304. https://doi.org/10.1175/JCLI-D-20-0854.1

Kopparla P, Fischer EM, Hannay C, Knutti R (2013) Improved simulation of extreme precipitation in a high-resolution atmosphere model. Geophys Res Lett 40:5803–5808. https://doi.org/10.1002/2013GL057866

Li F, Lin Z (2015) Improving multi-model ensemble probabilistic prediction of Yangtze River valley summer rainfall. Adv Atmos Sci 32:497–504. https://doi.org/10.1007/s00376-014-4073-8

Li X, Lu R (2017) Extratropical factors affecting the variability in summer precipitation over the Yangtze River basin, China. J Clim 30:8357–8374. https://doi.org/10.1175/JCLI-D-16-0282.1

Li P, Guo Z, Furtado K, Chen H, Li J, Milton S, Field PR, Zhou T (2019) Prediction of heavy precipitation in the eastern China flooding events of 2016: added value of convection-permitting simulations. Q J R Meteorol Soc 145:3300–3319. https://doi.org/10.1002/qj.3621

Li Q, Wang T, Wang F, Liang XZ, Zhao C, Dong L, Zhao C, Xie B (2020) Dynamical downscaling simulation of the East Asian summer monsoon in a regional Climate-Weather Research and Forecasting model. Int J Climatol 41:E1700–E1716. https://doi.org/10.1002/joc.6800

Liang X-Z (2022) Extreme rainfall slows the global economy. Nature 467:193–194. https://doi.org/10.1038/d41586-021-03783-x

Liang X-Z, Wang WC (1998) Associations between China monsoon rainfall and tropospheric jets. Q J R Meteorol Soc 124:2597–2623. https://doi.org/10.1002/qj.49712455204

Liang X-Z, Zhang F (2013) The cloud–aerosol–radiation (CAR) ensemble modeling system. Atmos Chem Phys 13:8335–8364. https://doi.org/10.5194/acp-13-8335-2013

Liang X-Z, Choi HI, Kunkel KE, Dai Y, Joseph E, Wang JX, Kumar P (2005) Surface boundary conditions for mesoscale regional climate models. Earth Interact 9:1–28

Liang X-Z, Xu M, Kunkel KE, Grell GA, Kain JS (2007) Regional climate model simulation of U.S.–Mexico summer precipitation using the optimal ensemble of two cumulus parameterizations. J Clim 20:5201–5207. https://doi.org/10.1175/JCLI4306.1

Liang X-Z, Xu M, Yuan X, Ling T, Choi H, Zhang F, Chen L, Liu S, Su S, Qiao F, He Y, Wang J, Kunkel K, Gao W, Joseph E, Morris V, Yu T-W, Dudhia J, Michalakes J (2012) Regional climate-weather research and forecasting model (CWRF). Bull Am Meteorol Soc 93:1363–1387. https://doi.org/10.1175/BAMS-D-11-00180.1

Liang X-Z, Li Q, Mei H, Zeng M (2019a) Multi-grid nesting ability to represent convections across the gray zone. J Adv Model Earth Syst 11:4352–4376. https://doi.org/10.1029/2019MS001741

Liang X-Z, Sun C, Zheng X, Dai Y, Xu M, Choi HI, Ling T, Qiao F, Kong X, Bi X (2019b) CWRF performance at downscaling China climate characteristics. Clim Dyn 52(3):2159–2184. https://doi.org/10.1007/s00382-018-4257-5

Liu S, Liang X-Z, Gao W, Zhang H (2008) Application of climate-weather research and forecasting model (CWRF) in China: domain optimization. Chin J Atmos Sci (in Chinese) 32(3):457–468

Ma S, Zhou T, Dai A, Han Z (2015) Observed changes in the distributions of daily precipitation frequency and amount over China from 1960 to 2013. J Clim 28:6960–6978. https://doi.org/10.1175/JCLI-D-15-0011.1

Merino A, García-Ortega E, Navarro A, Sánchez JL, Tapiador FJ (2022) WRF hourly evaluation for extreme precipitation events. Atmos Res 274:106215. https://doi.org/10.1016/j.atmosres.2022.106215

Murray V, Ebi KL (2012) IPCC special report on managing the risks of extreme events and disasters to advance climate change adaptation (SREX). J Epidemiol Commun Health 66:759. https://doi.org/10.1136/jech-2012-201045

O'Brien TA, Collins WD, Kashinath K, Rübel O, Byna S, Gu J, Krishnan H, Ullrich PA (2016) Resolution dependence of precipitation statistical fidelity in hindcast simulations. J Adv Model Earth Syst 8:976–990. https://doi.org/10.1002/2016MS000671

Peterson T (2005) Climate change indices. WMO Bull 54:83–86

Piao S, Ciais P, Huang Y, Shen Z, Peng S, Li J, Zhou L, Liu H, Ma Y, Ding Y (2010) The impacts of climate change on water resources and agriculture in China. Nature 467:43–51. https://doi.org/10.1038/nature09364

Qiao F, Liang X-Z (2015) Effects of cumulus parameterizations on predictions of summer flood in the Central United States. Clim Dyn 45:727–744. https://doi.org/10.1007/s00382-014-2301-7

Qiao F, Liang X-Z (2016) Effects of cumulus parameterization closures on simulations of summer precipitation over the United States coastal oceans. J Adv Model Earth Syst 8:764–785. https://doi.org/10.1002/2015MS000621

Qiao F, Liang X-Z (2017) Effects of cumulus parameterization closures on simulations of summer precipitation over the continental United States. Clim Dyn 49:225–247. https://doi.org/10.1007/s00382-016-3338-6

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2008) A description of the Advanced Research WRF version 3. Tech Rep. https://doi.org/10.5065/D6DZ069T

Sun C, Liang X-Z (2020a) Improving US extreme precipitation simulation: dependence on cumulus parameterization and underlying mechanism. Clim Dyn 55:1325–1352. https://doi.org/10.1007/s00382-020-05328-w

Sun C, Liang X-Z (2020b) Improving US extreme precipitation simulation: sensitivity to physics parameterizations. Clim Dyn 54:4891–4918. https://doi.org/10.1007/s00382-020-05267-6

Sun C, Liang X-Z (2023a) Understanding and reducing warm and dry summer biases in the central United States: analytical modeling to identify the mechanisms for CMIP ensemble error spread. J Clim. https://doi.org/10.1175/jcli-d-22-0255.1

Sun C, Liang X-Z (2023b) Understanding and reducing warm and dry summer biases in the central United States: improving cumulus parameterization. J Clim. https://doi.org/10.1175/JCLI-D-22-0254.1

Sun H, Wang A, Zhai J, Huang J, Wang Y, Wen S, Zeng X, Su B (2018) Impacts of global warming of 1.5 C and 2.0 C on precipitation patterns in China by regional climate model (COSMO-CLM). Atmos Res 203:83–94. https://doi.org/10.1016/j.atmosres.2017.10.024

Tang B, Hu W, Duan A (2021) Future projection of extreme precipitation indices over the Indochina Peninsula and South China in CMIP6 models. J Clim 34:8793–8811. https://doi.org/10.1175/JCLI-D-20-0946.1

Wang Y, Zhou L (2005) Observed trends in extreme precipitation events in China during 1961–2001 and the associated changes in large-scale circulation. Geophys Res Lett. https://doi.org/10.1029/2005GL022574

Weisman ML, Skamarock WC, Klemp JB (1997) The resolution dependence of explicitly modeled convective systems. Mon Weather Rev 125:527–548. https://doi.org/10.1175/1520-0493(1997)125%3c0527:TRDOEM%3e2.0.CO;2

Wilson DR, Bushell AC, Kerr-Munslow AM, Price JD, Morcrette CJ, Bodas-Salcedo A (2008) PC2: A prognostic cloud fraction and condensation scheme. II: climate model simulations. Q J R Meteorol Soc 134:2109–2125

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

Xu K-M, Randall DA (1996) A semiempirical cloudiness parameterization for use in climate models. J Atmos Sci 53:3084–3102

Xu C, Luo Y, Xu Y (2011) Projected changes of precipitation extremes in river basins over China. Quat Int 244:149–158. https://doi.org/10.1016/j.quaint.2011.01.002

Xu M, Liang X-Z, Samel A, Gao W (2014) MODIS consistent vegetation parameter specifications and their impacts on regional climate simulations. J Clim 27:8578–8596. https://doi.org/10.1175/JCLI-D-14-00082.1

Yang H, Jiang Z, Li L (2016) Biases and improvements in three dynamical downscaling climate simulations over China. Clim Dyn 47:3235–3251. https://doi.org/10.1007/s00382-016-3023-9

Yu X, Lee T-Y (2011) Role of convective parameterization in simulations of heavy precipitation systems at grey-zone resolutions—case studies. Asia-Pac J Atmos Sci 47:99–112

Yuan X, Liang X-Z (2011) Improving cold season precipitation prediction by the nested CWRF-CFS system. Geophys Res Lett 38:L02706. https://doi.org/10.1029/2010GL046104

Yuan X, Liang X-Z, Wood EF (2012) WRF ensemble downscaling seasonal forecasts of China winter precipitation during 1982–2008. Clim Dyn 39:2041–2058. https://doi.org/10.1007/s00382-011-1241-8

Zeng M-J, Lu W-S, Liang X-Z, Wang X-L (2008) Ensemble forecast experiment on precipitation in summer by CWRF numeric model. Plateau Meteorol (in Chinese) 27(6):1218–1228

Zhai P, Zhang X, Wan H, Pan X (2005) Trends in total precipitation and frequency of daily precipitation extremes over China. J Clim 18:1096–1108. https://doi.org/10.1175/JCLI-3318.1

Zhang W, Zhou T (2020) Increasing impacts from extreme precipitation on population over China with global warming. Sci Bull 65:243–252. https://doi.org/10.1016/j.scib.2019.12.002

Zhang F, Liang X-Z, Li J, Zeng Q (2013) Dominant roles of subgrid-scale cloud structures in model diversity of cloud radiative effects. J Geophys Res Atmos 118:7733–7749. https://doi.org/10.1002/jgrd.50604

Zhang S, Wang M, Wang L, Liang X-Z, Sun C, Li Q (2023) Sensitivity of the simulation of extreme precipitation events in China to different cumulus parameterization schemes and the underlying mechanisms. Atmos Res. https://doi.org/10.1016/j.atmosres.2023.106636

Zhao G, Mu X, Hörmann G, Fohrer N, Xiong M, Su B, Li X (2012) Spatial patterns and temporal variability of dryness/wetness in the Yangtze River Basin, China. Quat Int 282:5–13. https://doi.org/10.1016/j.quaint.2011.10.020

Zhaoye P, Yang K, Wang C (2022) Impacts of cumulus parameterizations on extreme precipitation simulation in semi-arid region: a case study in Northwest China. Atmosphere 13:1464. https://doi.org/10.3390/atmos13091464

Zong Y, Chen X (2000) The 1998 Flood on the Yangtze, China. Nat Hazard 22:165–184. https://doi.org/10.1023/A:1008119805106