Climate Dynamics (2020) 54:4891–4918
https://doi.org/10.1007/s00382-020-05267-6

Improving US extreme precipitation simulation: sensitivity to physics parameterizations

Chao Sun[1,2] · **Xin-Zhong Liang**[1,2]

Received: 13 November 2019 / Accepted: 23 April 2020 / Published online: 28 April 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
Climate models tend to underestimate rainfall intensity while producing more frequent light events, leading to significant bias in extreme precipitation simulation. To reduce this bias and better understand its underlying causes, we tested an ensemble of 25 physics configurations in the regional Climate-Weather Research and Forecasting model (CWRF). All configurations were driven by the ECMWF-Interim reanalysis and continuously integrated during 1980–2015 over the contiguous United States with 30-km grid spacing. Together they represent CWRF's ability to simulate characteristics of US extreme precipitation, and their spread depicts the structural uncertainty from alternate physics parameterizations. The US extreme precipitation simulation was most sensitive to cumulus parameterization among all physics configurations. The ensemble cumulus parameterization (ECP) was overall the most skilled at reproducing seasonal mean spatial patterns of daily 95th percentile precipitation (P95). Other cumulus schemes severely underestimated P95, especially over the Gulf States and the Central-Midwest States in convective prevailing seasons. CWRF with ECP outperformed the driving reanalysis, which substantially underestimated P95 despite its daily atmospheric moisture data assimilation. The CWRF improvement over ERI is much larger in warm than cold seasons. Changing alone ECP closure assumptions produced two distinct clusters of convective heating/drying effects: one altered P95 mainly by changing total precipitation intensity and another by changing rainy-day frequency. Microphysics, radiation, boundary layer, and land surface processes also impacted the result, especially under mixed synoptic and convective forcings, and some of their parameterization schemes worked with ECP to further improve P95.

Keywords Extreme precipitation · Numerical modeling · Climate model

1 Introduction

Since 1980, 258 billion-dollar weather disasters have occurred in the United States, causing more than $1.75 trillion in total economic losses (NOAA 2019). Of the damage caused by these weather disasters, more than 70% was due to extreme precipitation (Smith and Matthews 2015). Greater risks are anticipated in the future, since the frequency and intensity of extreme precipitation events are increasing over the United States (Kunkel et al. 1999; Groisman et al. 2005; Kunkel et al. 2013; Wuebbles et al. 2014; Anderson et al. 2015; USGCRP 2017). Despite the profound impact on society, prediction of extreme precipitation remains highly uncertain (Kunkel et al. 2002; Trenberth et al. 2003; Xie et al. 2015; Stegall and Kunkel 2019). The uncertainty can be associated with observational data scarcity and discrepancy (Donat et al. 2016; Herold et al. 2016), data-to-model spatial scale mismatch (Chen and Knutson 2008; Herold et al. 2017), and unpredictable natural variability (Fischer et al. 2013). The prediction uncertain can also arise from model deficiencies in representing organized convection systems and other physical processes (Wilcox and Donner 2007; Stephens et al. 2010; Evans et al. 2012; Yuan et al. 2012; Qian et al. 2015; Kang et al. 2015).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00382-020-05267-6) contains supplementary material, which is available to authorized users.

✉ Xin-Zhong Liang
xliang@umd.edu

[1] Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA

[2] Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

A major problem is that most climate models tend to underestimate extreme precipitation (Dai 2006; Sun et al. 2006; Stephens et al. 2010; Zobel et al. 2018). The causes of the underestimation are still debated. Some studies attributed the underestimation to a 'drizzling problem', where models tended to overestimate light rain events (Sun et al. 2003; Brown et al. 2010). Others reported simultaneous overestimation of light rain and underestimation of extreme precipitation (Frei et al. 2003; May 2004; Allan and Soden 2008). However, more light rain could consume only slightly more energy, which would not noticeably weaken extreme precipitation (Pendergrass and Hartmann 2014). Xie et al. (2004) argued that precipitation was underestimated because the threshold in the trigger function was set too low so that convective instability was released too early. On the other hand, Choi et al. (2015) showed that a modified trigger function suppressed not only light rain but also heavy rain. Kang et al. (2015) attributed the underestimation to a lack of ice phase processes in cloud microphysics schemes. While ice-phase processes are necessary, given the complexity of numerical models and especially convection-cloud-radiation interactions (Liang and Zhang 2013), they are not sufficient to resolve the underestimation. Many others attempted to address the problem by increasing model resolution, but large extreme precipitation biases still existed in global and regional climate simulations at grid spacing of 10–50 km (Iorio et al. 2004; Boyle and Klein 2010; Wehner et al. 2010; Li et al. 2011; Tripathi and Dominguez 2013; Wang and Kotamarthi 2015). Prein et al. (2017) found that even cloud-permitting models with grid spacing of 3–4 km could still underestimate extreme precipitation as their low-resolution counterparts. In contrast, Herman and Schumacher (2016) showed that cloud-permitting simulations could produce more heavy rains than lower resolutions, but they over-forecasted (by an order of magnitude) occurrences of extreme rainfall events.

None of these studies have systematically investigated the model sensitivity of extreme precipitation simulation to varying physics representations and the underlying mechanisms. For climate prediction, the most significant uncertainty lies in representing physical processes (Allen and Ingram 2002). This is especially the case for extreme precipitation, which by definition is rare and is typically not well tested during model development and evaluation. The choice of physical parameterization schemes could have greater impact on model performance during more intensive rainfall events (Evans et al. 2012). The regional Climate-Weather Research and Forecasting model (CWRF) has built in many alternate schemes with consistent coupling for each major physical process, including cumulus, microphysics, cloud, aerosol, radiation, planetary boundary layer, and surface processes (Choi et al. 2007, 2013; Choi and Liang 2010; Yuan and Liang 2011; Liang et al. 2012, 2015; Liang and Zhang 2013; Xu et al. 2014; Gan et al. 2015; Qiao and Liang 2015, 2016, 2017). CWRF is superior to the driving reanalysis and a popular regional climate model in capturing extreme precipitation characteristics over China (Liang et al. 2018). More importantly, its built-in ensemble of physics configurations offers a unique opportunity to systemically investigate the responsive processes to which extreme precipitation simulation is sensitive.

As the first part of a pair, this paper examines CWRF's improvements in simulating US extreme precipitation, and investigates the responsive processes by comparing a large ensemble of long integrations using multiple physics configurations. Section 2 describes CWRF and its selected physics configurations, the observational data used for evaluation, and the experiment design for the sensitivity analysis. Section 3 analyzes CWRF's performance at predicting seasonal mean and extreme precipitation distributions over the entire contiguous United States, relative to that of the driving reanalysis. Section 4 focuses on two key regions in which the reanalysis underestimated extreme precipitation and CWRF offered a significant improvement, and thereby examines the sensitivity of simulation skill to physics parameterization schemes. Section 5 compares the effects of vertical heating/drying profiles and closure assumptions among different cumulus parameterizations. Section 6 concludes the results. The companion paper (Sun and Liang 2019) will demonstrate how cumulus parameterization dominates extreme precipitation simulation and explore the potential physical contributors to extreme event modeling ability.

2 Model description, experiment design, observations, and extreme indices

CWRF has been systematically advanced as a Climate extension to the Weather Research and Forecasting model (WRF, Skamarock et al. 2008) since 2002 (Liang et al. 2012), with several important updates including terrestrial hydrology (Choi et al. 2013), cloud-aerosol-radiation interaction (Liang and Zhang 2013; Zhang et al. 2013), land surface characteristics (Xu et al. 2014), and upper oceans (Ling et al. 2015). Of particular relevance to this study, CWRF incorporates an ensemble cumulus parameterization (ECP) based on Grell and Dvénéyi (2002), which has outstanding performance in precipitation simulation, including extreme events and flooding (Qiao and Liang 2015), over oceans (Qiao and Liang 2016) and land (Qiao and Liang 2017), and in different climate regimes (Liang et al. 2018). CWRF is good fit for this study because it incorporates alternative parameterization schemes for each of the surface (land, ocean), planetary boundary layer, cumulus (deep, shallow), microphysics, cloud, aerosol, and radiation processes. Moreover, these schemes were coupled systematically to maximize

consistency between interactive components critical to regional climate simulation (Liang et al. 2012).

This study selected multiple schemes from each of the nine key CWRF parameterization processes (Table 1, Liang et al. 2012) to form 25 CWRF physics configurations, as listed in Table 2. Given the limited computing resource, these configurations present only a small subset of the CWRF full ensemble, focusing on the control version (Liang et al. 2012, 2018) and all major alternate schemes across each of the seven physic processes. They include five cumulus, four microphysics, two aerosol, two cloud, four radiation, four boundary layer, and four surface schemes. These schemes were found sensitive to the overall model performance (Liang et al. 2012). The first 17 configurations (B–R) were centered around the control version (A), in which only one process' scheme was replaced by an alternative scheme while others were kept identical. Configuration (S) replaced both radiation and boundary layer schemes, while the rest also replaced a third process' scheme, (T–W) for cumulus and (X–Y) for microphysics.

All CWRF simulations were conducted on a well-tested North American domain including the contiguous United States (Liang et al. 2001, 2004a, b, 2006, 2007, 2012). Horizontally, the domain was centered at (37.5° N, 95.5° W), containing 138×195 points at grid spacing of 30 km using the Lambert conformal map projection. There were 36 vertical terrain-following sigma (σ) levels, denser near the surface, and the top of the model was at 50 hPa. All simulations were driven by the European Center for Medium-Range Weather Forecasts Interim Reanalysis (ERI), with 6-hourly data available at horizontal grid spacing of approximately 80 km and 60 vertical levels up to 0.1 hPa (Dee et al. 2011). The simulation began in October 1, 1979 and ran continuously until the end of 2015. Considering the first 2 months as the spin-up, our analysis below is based on 1980–2015, a total of 36 years. Since ERI assimilated pseudo-observations of rainfall and surface analyses of temperature and humidity measurements, its resulting precipitation and surface air temperature can be considered as a realistic proxy of observations. For comparison with CWRF, ERI's cumulus scheme

Table 1 CWRF's physical parameterization processes and their major references

Physical processes	Parameterizations	References
Cumulus (CU)	ECP	Grell and Dvénéyi (2002), Liang et al. (2012), Qiao and Liang (2015, 2016, 2017)
	NKF	Kain and Fritsch (1993), Kain (2004)
	TDK	Tiedtke (1989), Nordeng (1994), Gregory et al. (2000), Bechtold et al. (2004), Zhang et al. (2011a)
	NSAS	Han and Pan (2011)
	BMJ	Betts and Miller (1986), Janjic (1994, 2000)
Microphysics (MP)	TAO	Tao et al. (1989, 2003)
	THO	Thompson et al. (2004, 2008), Thompson and Eidhammer (2014)
	MOR	Morrison et al. (2009), Morrison and Milbrandt (2010, 2015)
	WD6	Lim and Hong (2010)
Aerosol (AE)	A3D	Liang et al. (2012), Liang and Zhang (2013), Zhang et al. (2013). A3D uses aerosol mass loadings and optical properties or observed (MISR, MODIS), A2D uses aerosol optical depth distributions (Liang and Zhang 2013)
	A2D	
Cloud (CL)	XRL	Diagnostic cloud cover based on Xu and Randall (1996) with modifications by Liang et al. (2004b)
	CPL	Prognostic cloud cover based on Wilson et al. (2008)
Radiation (RA)	GSFC	Chou and Suarez (1999) and Chou et al. (2001)
	CCCMA	Li and Barker (2005), Li et al. (2005), Li and Shibata (2006)
	FLG	Fu and Liou (1992, 1993)
	RRTMG	Iacono et al. (2008)
Boundary layer (BL)	CAM	Holtslag and Boville (1993) with updates to include gravity wave drag effect and orographic turbulence stress (Liang et al. 2012)
	MYNN	Nakanishi and Niino (2006, 2009)
	ACM	Pleim (2007) with updates on MOL calculation following WRF3.7.1
	UW	Park and Bretherton (2009)
Surface (SF)		
Land	CSSP	Dai et al. (2003, 2004), Liang et al. (2005a, b), Choi (2006), Choi et al. (2007, 2013), Oleson et al. (2008), Choi and Liang (2010), Yuan and Liang (2011), Choi et al. (2015), Xu et al. (2014)
	NOAH	Ek et al. (2003), Niu et al. (2011), Yang et al. (2011)
Ocean	SST	Reynolds et al. (2007), Dee et al. (2011)
	XOML	Ling et al. (2011, 2015)

Table 2 The ensemble of CWRF physics configurations combining multiple alternate schemes from one to three key parameterization processes listed in Table 1

Configuration	Process	CU					MP				AE		CL		RA				BL				Land		Ocean	
	Scheme	ECP	NKF	TDK	NSAS	BMJ	GSFC	THO	MOR	WD6	A2D	A3D	XRL	CLP	GSFC	CCCMA	FLG	RRTMG	CAM	MYNN	ACM	UW	CSSP	NOAH	SST	XOML
A	ECP	◦					◦				◦		◦		◦				◦				◦		◦	
B	NKF		√				◦				◦		◦		◦				◦				◦		◦	
C	TDK			√			◦				◦		◦		◦				◦				◦		◦	
D	NSAS				√		◦				◦		◦		◦				◦				◦		◦	
E	BMJ					√	◦				◦		◦		◦				◦				◦		◦	
F	THO	◦						√			◦		◦		◦				◦				◦		◦	
G	MOR	◦							√		◦		◦		◦				◦				◦		◦	
H	WD6	◦								√	◦		◦		◦				◦				◦		◦	
I	A3D	◦					◦					√	◦		◦				◦				◦		◦	
J	CLP	◦					◦				◦			√	◦				◦				◦		◦	
K	CCCMA	◦					◦				◦					√			◦				◦		◦	
L	FLG	◦					◦				◦						√		◦				◦		◦	
M	RRTMG	◦									◦							√	◦				◦		◦	
N	MYNN	◦					◦				◦		◦		◦					√			◦		◦	
O	ACM	◦					◦				◦		◦		◦						√		◦		◦	
P	UW	◦					◦				◦		◦		◦							√	◦		◦	
Q	NOAH	◦					◦				◦		◦		◦				◦					√	◦	
R	XOML	◦					◦				◦		◦		◦				◦				◦			√
S	FM	◦					◦				◦		◦		◦				◦				◦		◦	
T	FMNKF		√				◦				◦		◦				√			√			◦		◦	
U	FMTDK			√			◦				◦		◦				√			√			◦		◦	
V	FMNSAS				√		◦				◦		◦				√			√			◦		◦	
W	FMBMJ					√	◦				◦		◦				√			√			◦		◦	
X	FMTHO	◦						√			◦		◦				√			√			◦		◦	
Y	FMMOR	◦							√		◦		◦				√			√			◦		◦	

The control schemes are identified by a solid circle, while the alternate ones are depicted by a check mark

was originally described by Tiedtke (1989), and has since been updated, including modifications in entrainment formulation (Bechtold et al. 2008) and parameterization closure (Bechtold et al. 2014).

Given the non-Gaussian temporal and inhomogeneous spatial distributions of extreme precipitation (Groisman et al. 1999; Easterling et al. 2000), an extended observational period in a dense monitoring network is required to properly capture the statistical characteristics of extreme events (Groisman et al. 2005). The daily precipitation observations were based on quality-controlled records from 8516 stations in the National Weather Service Cooperative Observer network (COOP), which are updated continuously (Durre et al. 2010; Menne et al. 2012). These stations each contained at least 40% available daily data for 1951–2012 and were kept the same for all subsequent years (personal communication with Kenneth Kunkel 2019). Following Liang et al. (2004a), they were adjusted for topographic dependence using monthly mean data from the Parameter elevation Regression on Independent Slopes Model (PRISM, Daly et al. 1997). This adjustment was necessary because elevation and precipitation correlate strongly, and observations over mountain areas are usually at lower elevations and thus may underestimate precipitation. The station data were mapped onto the CWRF 30-km grid following the mass-conservative Cressman objective analysis method of Liang et al. (2004a). For consistency, ERI daily precipitation values were mapped onto the same CWRF grid by a conservative algorithm from the Earth System Modeling Framework regridding package. These remapping procedures were applied to alleviate the impact of data scale mismatch on extreme event comparison (Chen and Knutson 2008). We recognized that PRISM had recently provided 4-km grid data of precipitation analysis merging rain gauge and radar measurements (Daly et al. 2017). As discussed in Sect. 6, the PRISM-COOP difference in P95 was minor, and hence, we chose COOP as the observational reference due to its longer records. We will discuss the impact of data spatial resolution and mapping procedure on extreme precipitation comparison in Sect. 4.

Numerous extreme indices have been recommended by the Expert Team on Climate Change Detection, Monitoring and Indices (Frich et al. 2002; Zhang et al. 2011b), and some are often used for observational analysis (e.g. Alexander et al. 2006) and model evaluation (e.g. Liang et al. 2018). Here we used daily 95th percentile precipitation (P95) to analyze climatic extreme simulation skills in different seasons and distinct regions. The P95 is defined as the wet-day percentile as in Schär et al. (2016), in which only rainy days with precipitation greater than 1 mm day^{-1} are counted. The geographic distribution of P95 in each season is preferred to other indices, since it was designed to represent the climatology characteristics and regime dependence of extreme precipitation (Karl and Knight 1998; Frich et al. 2002). P95 is also a more robust statistic than a maximum or other value, as it can show key features of a sample's distribution without being distorted by abnormal outliers (Fan et al. 1994; Cairo 2016; Jamili 2016). However, improved P95 performance could be due to an overestimate of drizzling events (Haylock and Nicholls 2000; Schär et al. 2016) or to an artificial shift in the precipitation intensity distribution (Sillmann et al. 2017), rather than improvement in the underlying physics processes. Therefore, to improve the reliability of the P95 analysis, we also evaluated total number of rainy days (NRD) and average daily rainfall intensity (DRI = total accumulated precipitation amount/NRD). Together they provide additional information on whether P95 biases are associated with deficiencies in clear-day frequencies or in rainfall magnitudes.

3 General performance of seasonal mean and extreme precipitation

We first analyze the performance of the control CWRF (CTL) in simulating precipitation-related fields over the contiguous United States, relative to ERI. Figure 1 compares observed and simulated 36-year (1980–2015) mean seasonal precipitation distributions. In all seasons, CWRF captured precipitation distribution details over mountain areas with a finer structure and more realistic intensity than ERI. This was especially obvious in the western United States. In winter, observed precipitation was more than 4.5 mm day^{-1} over the Gulf States. ERI did not produce enough intensity over that region, whereas CWRF captured the intensity, though its maximum center shifted inland. In spring, observations showed that the maximum center moved inward. While ERI missed this feature, CWRF produced sufficient intensity and the correct location. CWRF had a larger area of precipitation greater than 4.5 mm day^{-1}, which was more realistic than ERI. In summer, observations showed a strong, broad center over the Midwest. ERI significantly underestimated precipitation over that, while CWRF produced both better intensity and a more reasonable distribution. ERI overestimated precipitation over the Gulf States, averaging over 4.5 mm day^{-1} compared to the less than 4 mm day^{-1} observed. In between two rain belts over the Midwest and the Southeast, observations showed a narrow region with relatively weak precipitation. CWRF realistically simulated the intensity in this region, but ERI overestimated it. In autumn, observations showed peaks in Arkansas and along the Texas-Louisiana coast. CWRF captured these peaks with some overestimation near the Gulf Coast, while ERI produced insufficient intensity.

The above comparison shows that CWRF simulated climatological mean precipitation distributions better than ERI, even though ERI assimilated pseudo-rainfall (i.e. total

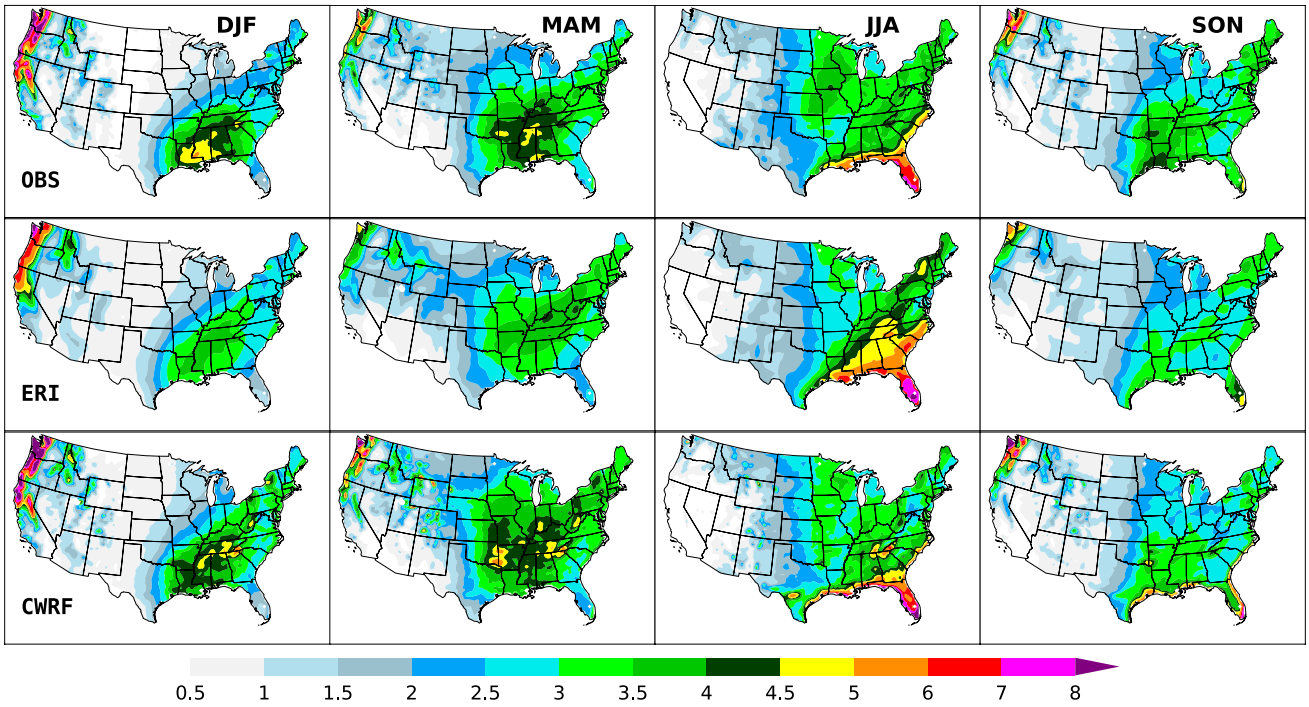


Fig. 1 Geographic distributions of 1980–2015 mean seasonal precipitation amount (mm day^{-1}) observed (OBS), assimilated (ERI), and simulated by CWRF control ECP for winter (DJF), spring (MAM), summer (JJA), and autumn (SON)

column water vapor) observations and surface measurements (Dee et al. 2011), which should have incorporated most realistic features. Due to its more comprehensive physics representation and finer resolution, CWRF showed useful added value in precipitation simulation, providing greater detail and higher accuracy than ERI across the majority of the United States.

Figure 2 compares 36-year mean seasonal NRD distributions. CWRF captured finer structural details than ERI over the western US mountain regions in all seasons, especially the rain shadow areas. In these regions, the gradient of rainy days tends to be large, and predicting detailed distribution is vital for management decision-making and planning processes (Daniels et al. 2012). In winter, both ERI and CWRF simulated NRD peaks well from the Midwest to Northeast, but underestimated them over the Gulf States. Given the dominance of stratiform precipitation, addressing this regional underestimation may require improving microphysics representation. In spring, both ERI and CWRF realistically captured the pattern and magnitude of NRD distribution over the entire Central to Eastern States. In summer, ERI overestimated NRD in the Central to Midwest States, exhibiting its drizzling problem, whereas CWRF reduced this overestimation. However, CWRF underestimated NRD near the Great Lakes, suggesting that its interactive lake model (Subin et al. 2012) needs refinement. In autumn, ERI overestimated NRD in the Southwest again due to its significant drizzling problem, while CWRF continually underestimated near the Great Lakes. Overall, CWRF captured the essential spatial features of NRD, demonstrating the potential to resolve the ERI's drizzling problem.

Figure 3 compares 36-year mean seasonal P95 distributions. In all seasons, CWRF outperformed ERI over the western US mountainous regions, including the Coastal Ranges, the Cascade Range-Sierra Nevada region, and the Rocky Mountains. The windward slopes of these mountains are prone to cyclogenesis-induced heavy rainfall (Catto and Pfahl 2013), while the eastern sides are typically dry transitional zones partly controlled by the precipitation-shadowing effect (Liang et al. 2004a). ERI underestimated P95 peaks with no clear dry zones in every season, especially in winter. On the other hand, CWRF well captured both the intensity and the wet-dry pattern distribution. This is an important improvement, as precipitation prediction in these mountainous regions is notoriously difficult (Leung et al. 2004; Siler and Roe 2014). Since general circulation models, including ERI, do not resolve topographical details, they are at a disadvantage when it comes to such regions. In contrast, CWRF incorporates finer details and so realistically captured precipitation variations (Liang et al. 2006).

CWRF also substantially outperformed ERI for P95 over the Central to Eastern States, demonstrating its ability to improve even in regions not dominated by topographic forcing. In winter, observations showed a broad region of

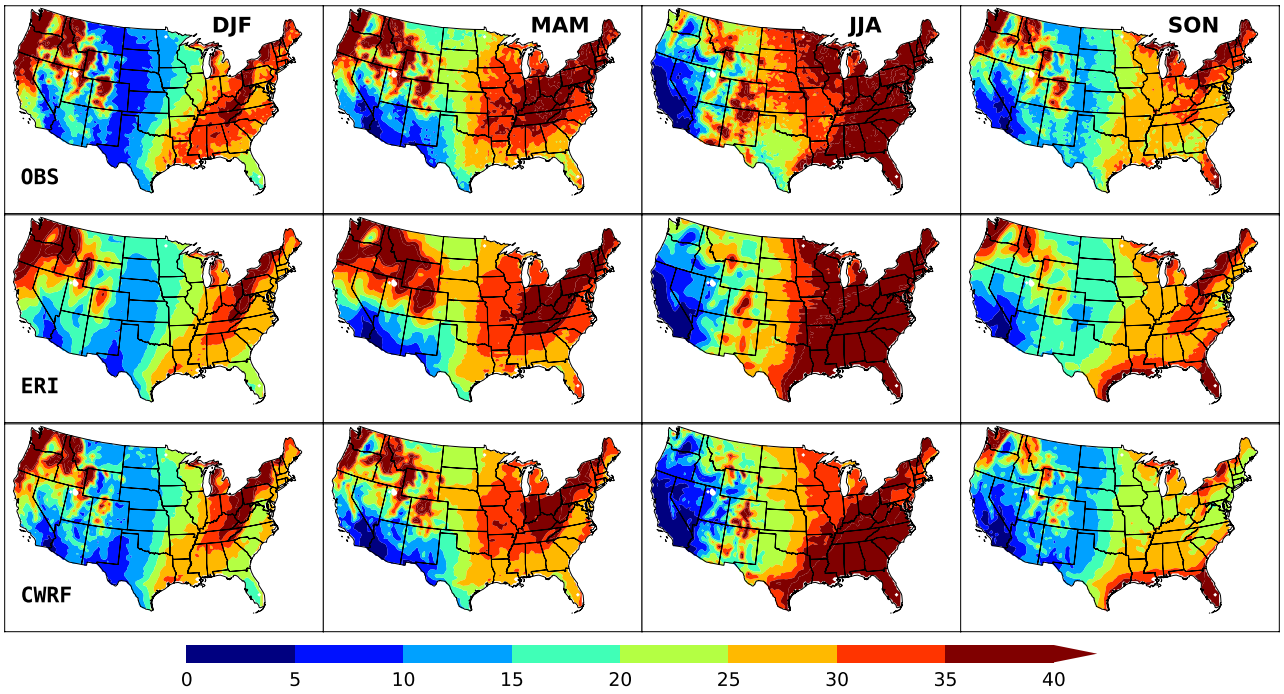


Fig. 2 Same as Fig. 1 except for the number of rainy days (NRD)

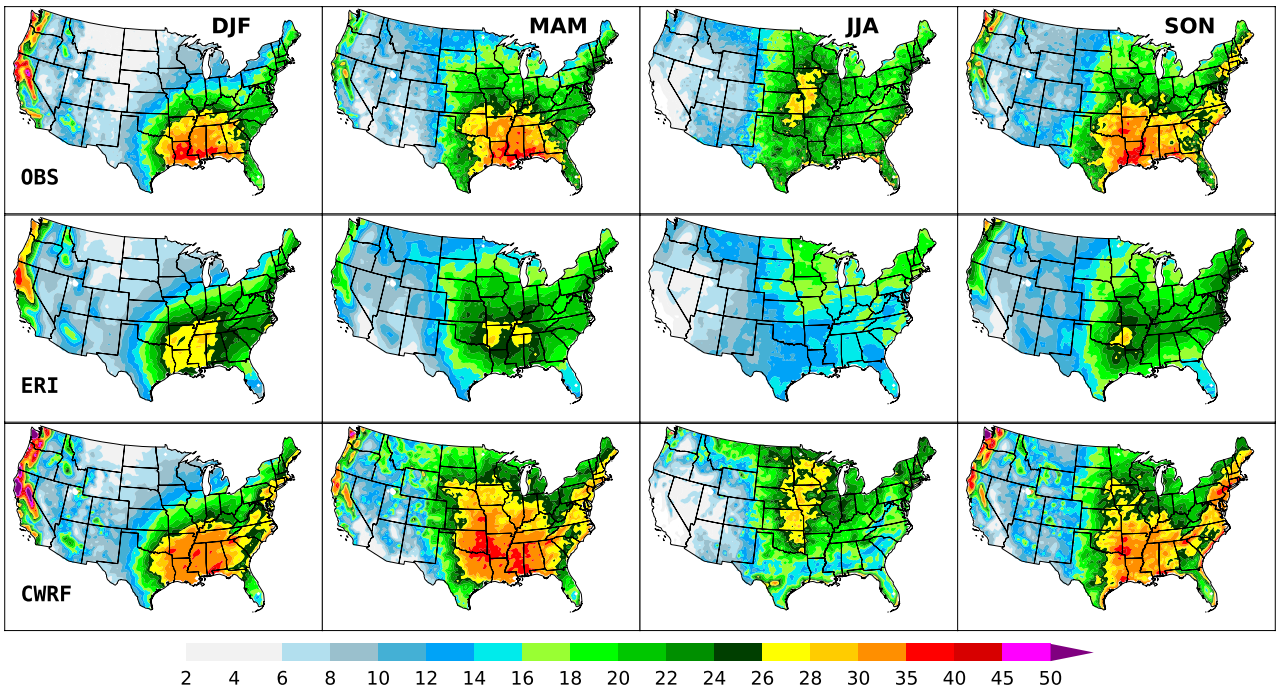


Fig. 3 Same as Fig. 1 except for the daily 95th percentile precipitation (P95) (mm day^{-1})

high values over the southern Central States, with extremes greater than 35 mm day^{-1}. ERI systematically underestimated the magnitude and reduced the coverage of these regional extremes. CWRF accurately captured both the magnitude and coverage, though it overestimated rainfall along the eastern coastal States. In spring, the observed P95

peak region expanded north and west. Again, ERI systematically underestimated that, with the area of P95 greater than 28 mm day^{-1} reduced substantially. In contrast, CWRF further expanded and strengthened the region, resulting in overestimation in the Midwest and along the eastern coastal States. In summer, observations showed high P95 values in the northern Central States, exceeding 26 mm day^{-1} across Iowa, Missouri, Kansas, and Oklahoma. ERI failed to produce any extreme precipitation greater than 20 mm day^{-1}, which is consistent with the drizzling problem identified from its NRD overestimation (Fig. 3). On the other hand, CWRF realistically captured the magnitude and pattern of observations. CWRF underestimated P95 in the Gulf States by up to 5 mm day^{-1}, while ERI underestimated by up to 10 mm day^{-1}. In autumn, the observed pattern resembled that of spring, except high values also occurred along the eastern coastal States. As such, CWRF performed even better in autumn than in spring, whereas ERI further deteriorated.

4 Physics sensitivity of regional extreme precipitation simulation

The above comparison identified two distinct regions in which the driving ERI substantially underestimated precipitation extremes that were realistically captured by the downscaling CWRF: the Gulf States (GS) and the Central to Midwest States (CM). Since the driving large-scale synoptic conditions were the same and the regional topographic forcing effects are expected to be small, the CWRF's superior performance to ERI in these regions likely resulted from improved physics parameterizations under a refined resolution. This downscaling ability presents a unique opportunity to explore the sensitivity of P95 simulation to CWRF's configuration of physics parameterizations, and therefore the key model improvement needed to alleviate the extreme precipitation underestimation and related drizzling problem. The following analyses focus on these two regions, comparing CWRF's performance among its ensemble of 25 physics configurations to simulate extreme precipitation features. The comparison identifies which physical processes CWRF extreme precipitation simulation is most sensitive to, and which schemes or combinations best capture those processes.

CWRF's improved skill over ERI is particularly evident in GS spring and CM summer. Therefore, we used ERI spring and summer P95 bias distributions to define the boundaries of the two regions, as illustrated in Fig. 4. The GS region encompasses all grids where ERI significantly underestimated spring P95. Similarly, the CM region encompasses all grids where ERI significantly underestimated summer P95, excluding those overlapping with GS. Scattered areas

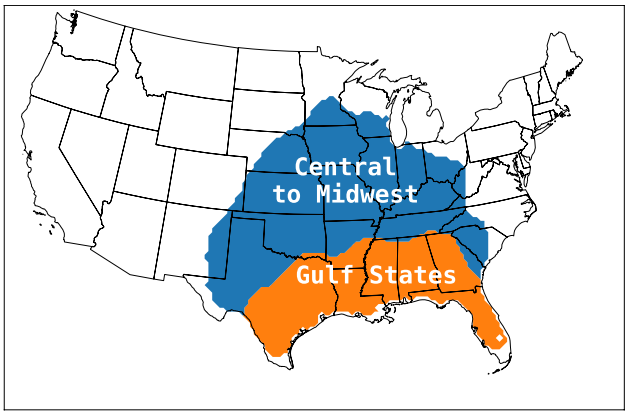


Fig. 4 Boundary specification of the two key regions where ERI severely underestimated extreme precipitation: the Central to Midwest States (CM) in summer and the Gulf States (GS) in spring

with radii smaller than three grids were discarded. The two regions differ in prevailing precipitation systems and dominant physical mechanisms.

A question is whether these P95 contrasts between ERI and CWRF were caused by the spatial resolution mismatch in data analysis (Chen and Knutson 2008; Herold et al. 2017). The analysis result may differ by merely data manipulation method (all other things being equal), that is, based on the bilinear versus conservative mapping between different resolution grids. Figure 5 compares 36-year mean seasonal P95 biases in the GS and CM regions between the two mapping methods and between the CWRF 30-km and ERI 80-km grids. Observed and simulated daily precipitation values were first interpolated to the common CWRF or ERI grid bilinearly or conservatively (i.e. area-averaging). P95 was then calculated from these gridded data for each season of a year, and finally its biases (simulated minus observed) were averaged over all years and all grids within the GS and CM regions. Differences in calculated biases do exist between using the four data manipulation procedures (two interpolation methods by two analysis grids), but they are all within ± 10% of the observed P95. These differences are significantly smaller than the contrasts between CWRF and ERI, especially in spring and summer for both regions which range in 20–30%. Therefore, the model physics representation at a finer resolution plays the major role for CWRF improvement over the ERI, which is examined in more detail below.

Figure 6 compares 36-year mean seasonal P95 biases (from observations) between ERI and the 25 CWRF physics configurations averaged over the two regions. The control CWRF corresponds to the ECP run, called in short the control ECP. There were five runs, in which only the ECP was replaced by other cumulus schemes. For brevity, we refer to these five CWRF runs directly by the name of the respective

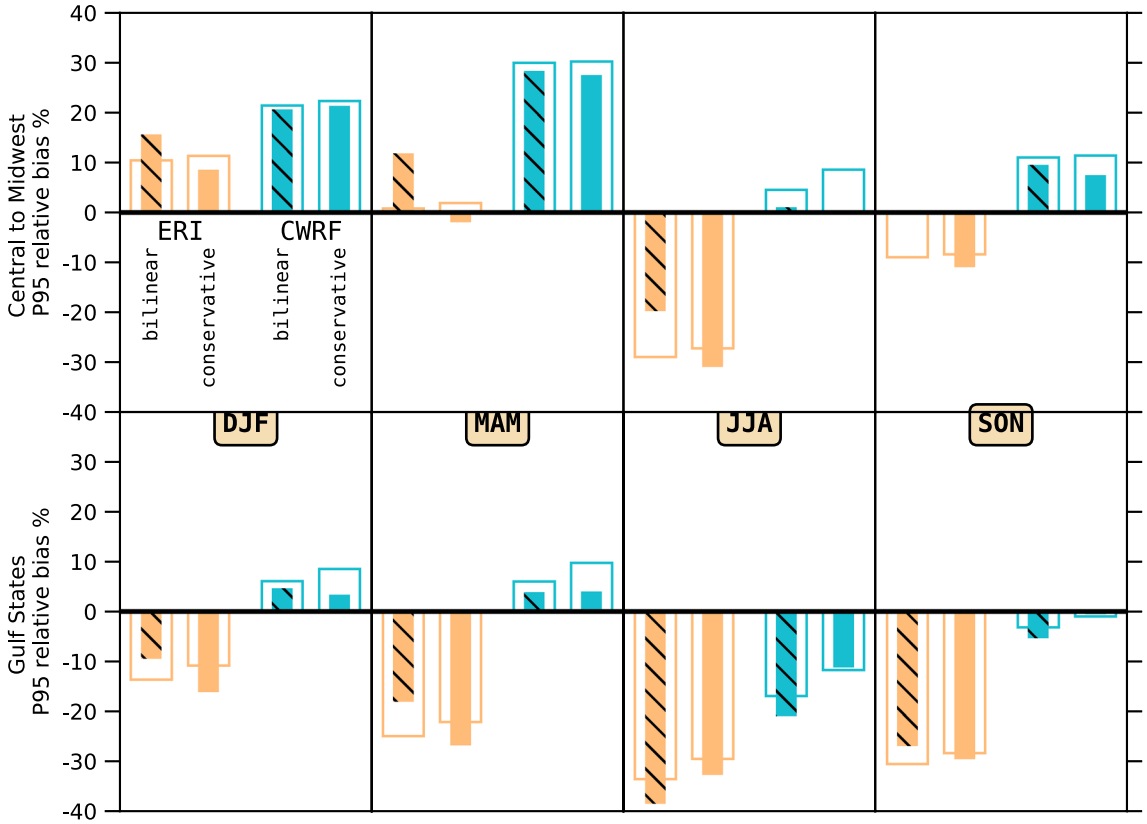


Fig. 5 Comparison between calculations using the bilinear and conservative mapping methods onto the CWRF 30-km (inner shaded bars) and ERI 80-km (outer hollow bars) grids for ERI and CWRF simulated 36-year mean seasonal P95 biases (in the percentages of observations) over the CM (upper) and GS (lower) in winter (DJF), spring (MAM), summer (JJA), and autumn (SON)

cumulus scheme they used. Similarly, any run differed from the control by a single process' parameterization scheme is referred by that scheme's name. For example, MOR denotes for the CWRF run replacing the microphysics scheme of Tao et al. (1989, 2003) with that of Morrison et al. (2009), Morrison and Milbrandt (2010, 2015) while the rest is identical to the control configuration. Other runs replacing two or more processes' parameterization schemes differed from the control CWRF are referred by the experiment name listed in Table 2. We often include both the process and scheme names to avoid confusion. A more general term like "ECP members" denotes for all runs using the CWRF configurations that include ECP cumulus scheme.

As apparent in Fig. 6, CWRF members using different radiation or microphysics but the same cumulus schemes had similar P95 biases. Likewise, the P95 bias spread among different boundary layer and land surface schemes was generally not large. On the other hand, P95 bias differences between members using different cumulus schemes were substantial. This suggests that cumulus parameterization plays a crucial role in extreme precipitation simulation.

According to the simulated P95 biases, CWRF physics configurations consisted of two broad types. Type I did not significantly underestimate P95 in either GS or CM region, and included members using the ECP and NKF cumulus schemes. Type II produced significant underestimates in either GS spring or CM summer, and included members using the TDK, NSAS, and BMJ cumulus schemes. In the GS region, where ERI substantially underestimated P95 in all seasons (by 15–32%), type I members produced reasonable extreme precipitation and relatively small biases, especially in autumn (mostly less than 5%). In particular, the control ECP had the least bias and a relatively stable performance, with no outliers in any season. It outperformed others most notably in spring, when ERI underestimation was the greatest. Other ECP members slightly (by less than 5–10%) overestimated in spring and underestimated in summer, while NKF members generally overestimated (by 18% in spring) except in winter. Type II members severely underestimated in all seasons, with the exception of TDK members, which produced small biases in winter, spring, and autumn. As discussed later, the exception for TDK in spring and autumn was identified with incorrect spatial patterns and excessive variability.

In the CM region, ERI substantially (by 30%) underestimated summer P95, whereas type I members of CWRF

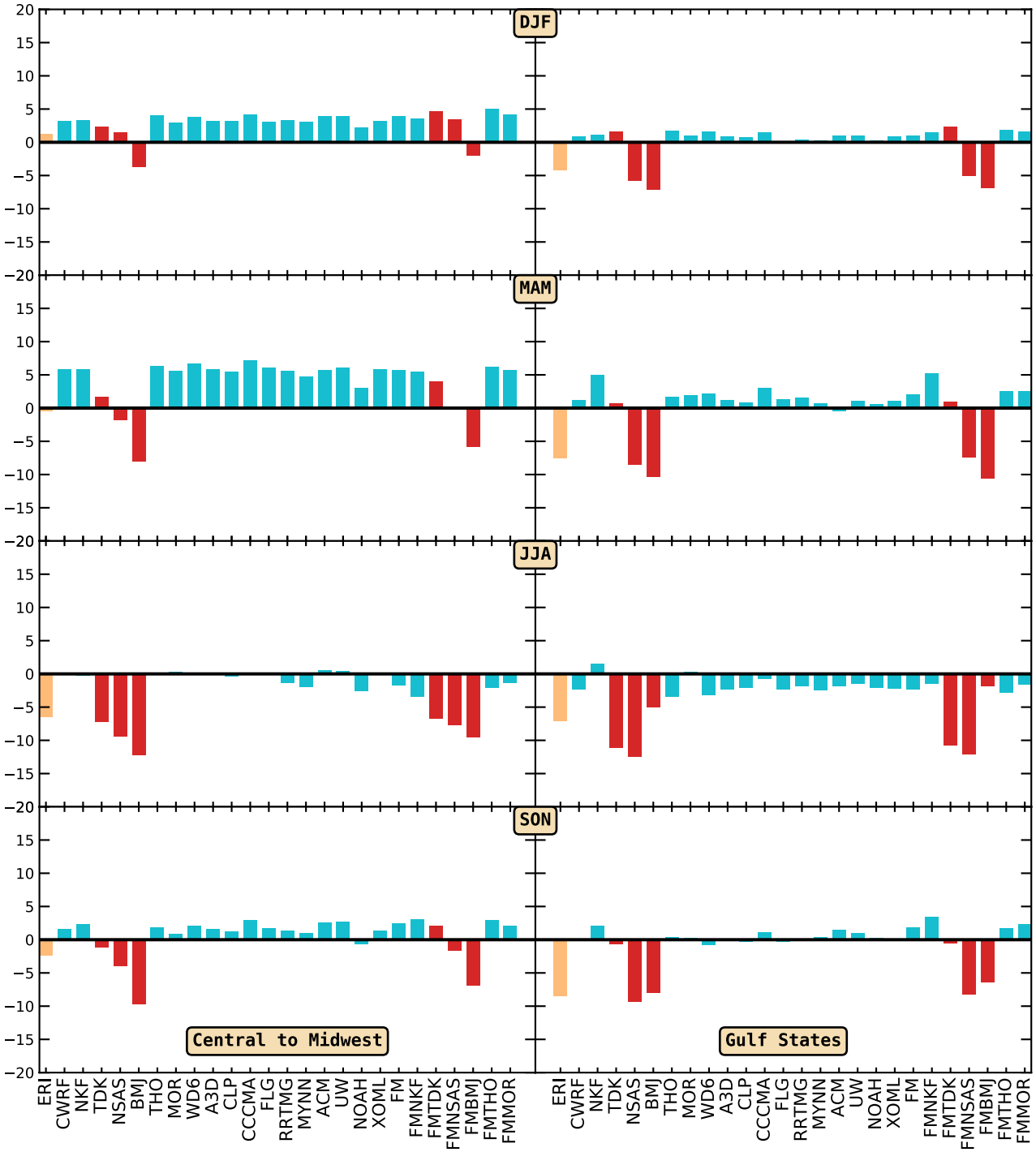


Fig. 6 Comparison among ERI and all CWRF physics configurations in simulating 1980–2015 mean seasonal P95 (mm day^{-1}) biases (from observations) averaged over the CM (left) and GS (right) for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). They are separated by color into type I (blue) and type II (red) members depending on their cumulus schemes

produced quite realistic simulations. ERI also underestimated autumn P95 by about 10%, which CWRF type I members overestimated by a similar magnitude. On the other hand, type II members significantly underestimated in both summer and autumn (when convective precipitation is dominant), except for TDK members, which had small biases in autumn. Once again, TDK members displayed incorrect spatial patterns (discussed below). ERI was most realistic in spring and overestimated slightly (by ~10%) in winter. Type I members significantly (by 15–30%) overestimated in both

winter and spring, when convective activities are relatively infrequent. In contrast, the biases of type II members were mixed in these two seasons: BMJ members still underestimated (especially in spring by 25–40%) and TDK members overestimated (by 10–30%), whereas NSAS members had small spring biases and moderate winter overestimates (by 10–20%).

Figure 7 compares geographic distributions of seasonal P95 biases from observations over the contiguous United States for ERI and the five CWRF members that differed

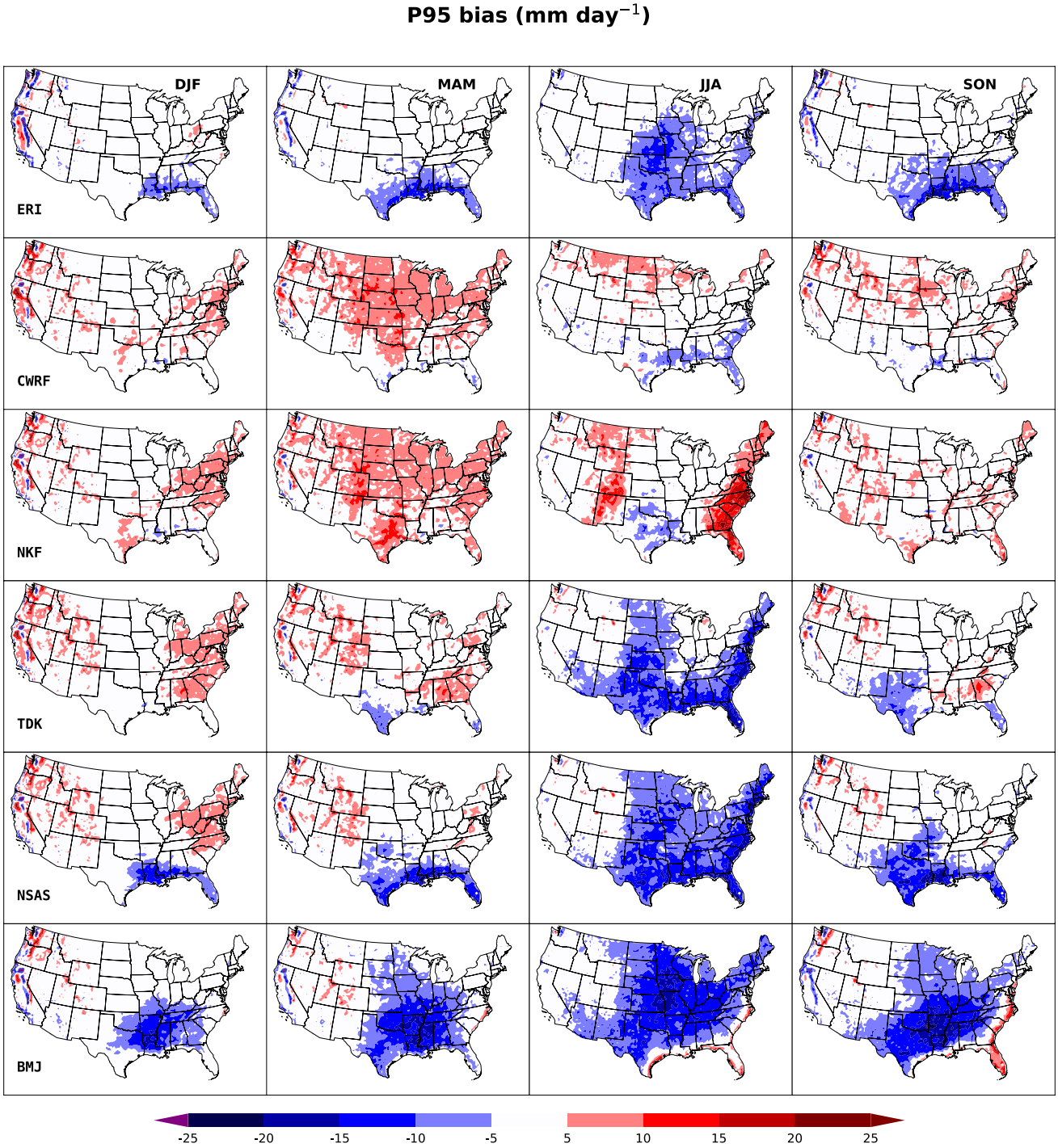


Fig. 7 Geographic distributions of 1980–2015 mean daily 95th percentile precipitation (P95) biases from observations (mm day^{-1}) for winter (DJF), spring (MAM), summer (JJA), and autumn (SON) as assimilated (ERI) and simulated by five CWRF members varying only the cumulus scheme (ECP, NKF, TDK, NSAS, BMJ)

only in cumulus schemes (ECP, NKF, TDK, NSAS, BMJ). In the GS region, the observed P95 maxima were higher than 25 mm day^{-1} in summer and even greater than 35 mm day^{-1} in other seasons (Fig. 1). These extreme precipitation events usually happened near the coastline. ERI failed to capture this intensity in all seasons, never exceeding 28 mm day^{-1}, and its maximum center shifted much to the inner mainland (see also Kunkel and Liang 2005). In contrast, the control CWRF produced sufficiently strong intensity as well as the correct location of the center. Both center dislocation and intensity underestimation caused ERI's substantial dry P95 biases (Fig. 7). On the other hand, in summer, NKF shifted the center eastward, causing substantial overestimations in the eastern coastal States but large underestimations in Texas, Oklahoma and Louisiana. These large opposite biases canceled each other to produce a smaller overestimation when averaged other the GS region. Similarly, in spring and autumn, TDK produced large overestimations in Georgia and Alabama but underestimations in Texas, which canceled each other to yield smaller GS average biases than ECP. In all seasons, NSAS systematically underestimated P95 over the GS region, while BMJ had more substantial underestimations over more extensive areas except for overestimations along the southern and eastern coastlines in summer and autumn.

One potential factor contributing to CWRF's improvement in representing P95 is that its ECP used different sets of cumulus parameterization closure assumptions to distinguish land versus oceans (Qiao and Liang 2015, 2016, 2017), which more realistically represented the regional-specific processes governing extreme precipitation in both GS and CM. The ECP better captured coastal baroclinicity-generating fronts in GS and CM, both of which were linked to most heavy precipitation events in the respective region (Kunkel et al. 2012). Therefore, the ECP scheme, with a more comprehensive treatment of the land–ocean contrast helped produce sufficient convective activity and better extreme precipitation simulation. See Sect. 6 for more discussion on the effect of the ECP closures.

Figure 8 compares geographic distributions of NRD biases among ERI and the five CWRF members. ECP biases were generally between ± 10 days, and the lowest among all simulations. TDK also did reasonably well, except for large underestimations in both the CM and GS regions. This exception was coincident with small P95 underestimations. On the other hand, both TDK and NSAS substantially underestimated NRD in both the CM and GS regions throughout the year. The underestimations were especially large and systematic in summer, by more than 25 (days) over most regions of the central to eastern Unites States. Interestingly, BMJ did very well and was comparable to ECP in winter, spring, and autumn. In summer, BMJ resembled other type II members in great underestimations, except for a realistic simulation in the Great Plains.

Figure 9 compares geographic distributions of DRI biases among ERI and the five CWRF members. These DRI biases were highly correlated with P95 biases in all seasons. Their spatial pattern correlations over the entire CONUS (Table 3) varied weakly in seasons, with the annual average of 0.92 (ERI), 0.90 (ECP), 0.89 (NKF), 0.83 (TDK), 0.94 (NSAS), and 0.97 (BMJ). Strong correlations indicated that underestimations (overestimations) of extreme precipitation occurred mostly because rainfall intensities were systematically reduced (increased). This was especially the case for BMJ. Among all simulations, TDK had the lowest correlations, especially in summer (0.74) when substantial P95 underestimations corresponded to large DRI overestimations over most regions except the southern coastlines. Such summer P95 and DRI opposite biases were coincident with substantial NRD underestimations (Fig. 8), indicating that TDK simulated not only much lighter rains but also drastically more clear days. A similar situation occurred in autumn, though it was limited to Texas, Oklahoma and Florida.

Figure 10, using Taylor (2001) diagram, compares seasonal P95 spatial pattern correlations and standard deviations of ERI and the 25 CWRF physics configurations relative to observations. All statistics are based on 36-year mean distributions separately over the GS and CM regions. The simulation's distance from the observation represents its root mean square error (rmse).

In the GS region, ERI had almost no pattern correlation and the most severely underestimated standard deviation (0.3) in summer. ERI produced higher correlations in autumn (0.6), spring (0.7), and winter (0.9), but still significantly underestimated deviations (0.7–0.8). CWRF type I members showed improved skill over ERI, with generally higher correlations in summer (0.3–0.4), spring (0.6–0.7), autumn (0.5–0.7), and in winter (0.85–0.9), as well as larger deviations (0.8–1.2). In particular, the control ECP correlated most strongly to observations in all seasons. Other ECP members also performed consistently to each other, implying that combining ECP with other physical process schemes had little impact on simulation ability. One exception was with the boundary layer schemes ACM, UW, and MYNN, which had less skill than CAM in the control CWRF. Meanwhile, NKF members had lower scores (less correlation or larger variability) than ECP, especially in summer. On the other hand, type II members generally produced lower correlations and substantially underestimated (BMJ) or overestimated (TDK) deviations; these errors were especially excessive in summer, falling off the chart as outliers. TDK members simulated large positive and negative local errors (Fig. 7), which canceled each other to yield small regional mean P95 biases (Fig. 6) with significantly overestimated spatial deviations (Fig. 10).

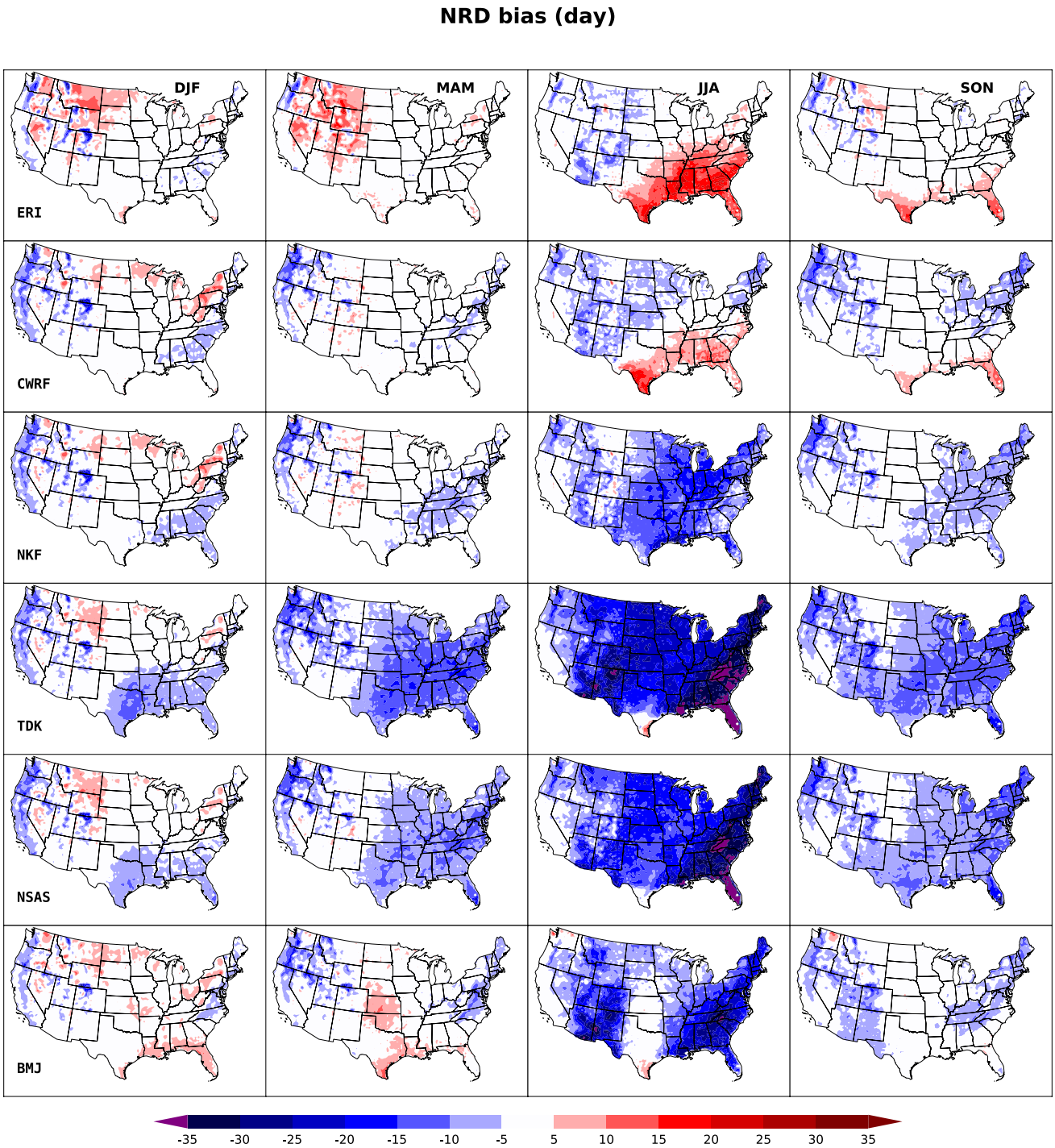


Fig. 8 Same as Fig. 7 except for the number of rainy days (NRD) biases

ERI performed better in the CM than GS region, with increased pattern correlations in summer (0.6), autumn (0.8), spring (0.9), and winter (0.9), but still underestimated deviations (0.5–0.9). The control ECP continually outperformed ERI, with comparable correlations but correct deviations (1.0–1.1) in all seasons. Other ECP members performed similarly well, especially in winter and spring, though deviations varied widely (0.95–1.4) in summer and autumn. One outlier was the member combining ECP with boundary layer scheme ACM (replacing CAM in the control CWRF) in summer, whose spatial variability (standard deviation) was about 1.6 times that of the observation. NKF members performed poorly in summer, with lower correlations (0.5) and excessive deviations (1.5);

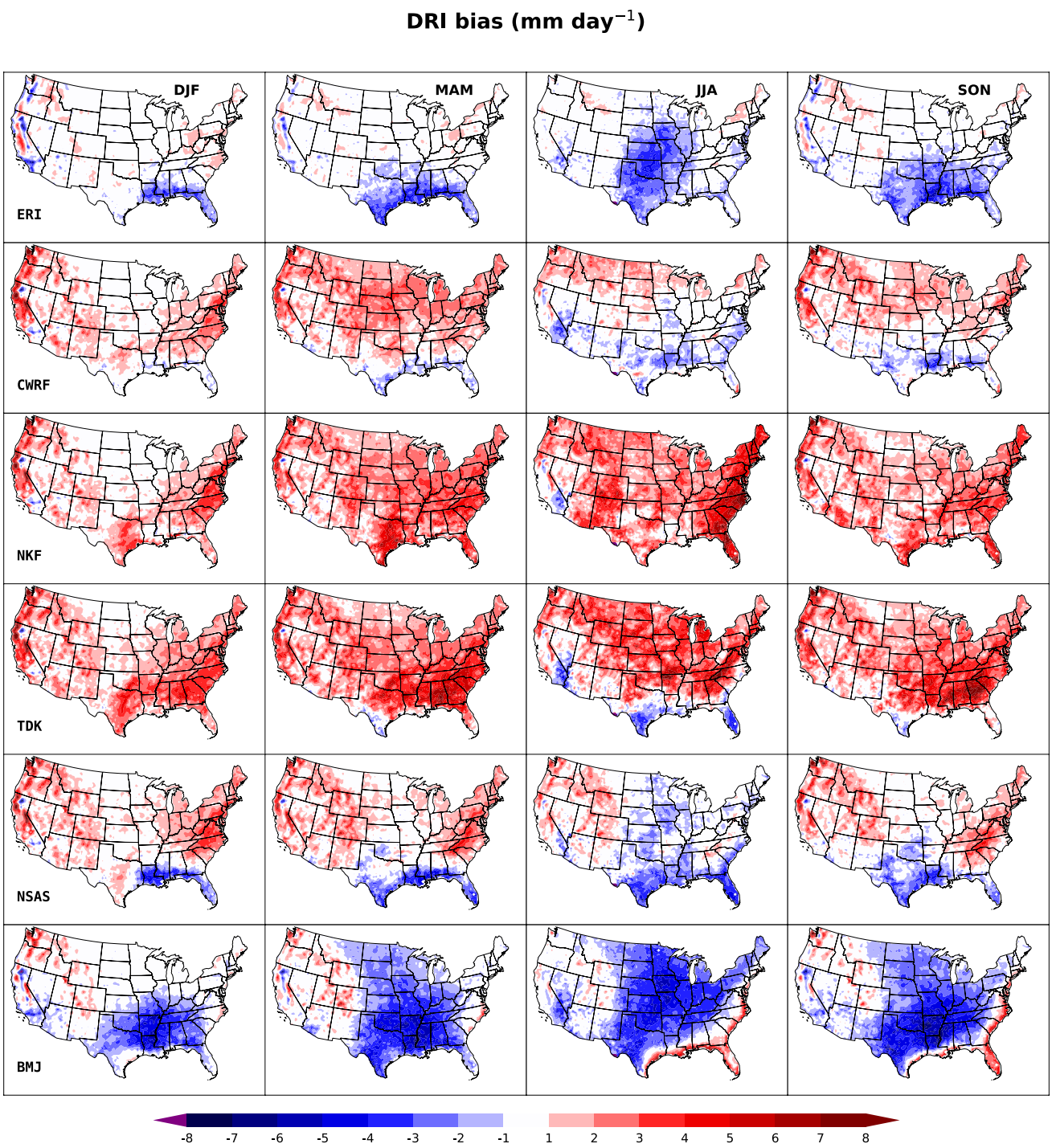


Fig. 9 Same as Fig. 7 except for the daily rainfall intensity (DRI) biases (mm day^{-1})

they were comparable to ECP members in other seasons. On the other hand, in all seasons type II members generally had lower correlations and were more scattered, with abnormally high or low deviations. In particular, TDK members substantially overestimated spatial variability in all seasons (Fig. 10), with large positive and negative local errors (Fig. 7) canceled each other to yield small regional mean biases in spring and autumn (Fig. 6).

To examine the model skill dependence on P95 magnitude, we adopt the widely used categorical equitable threat score (ETS), defined as the ratio of *hits minus hits expected by chance* divided by *hits plus false alarms plus misses*

Table 3 Spatial pattern correlations over CONUS between model DRI and P95 seasonal mean biases during 1980–2015

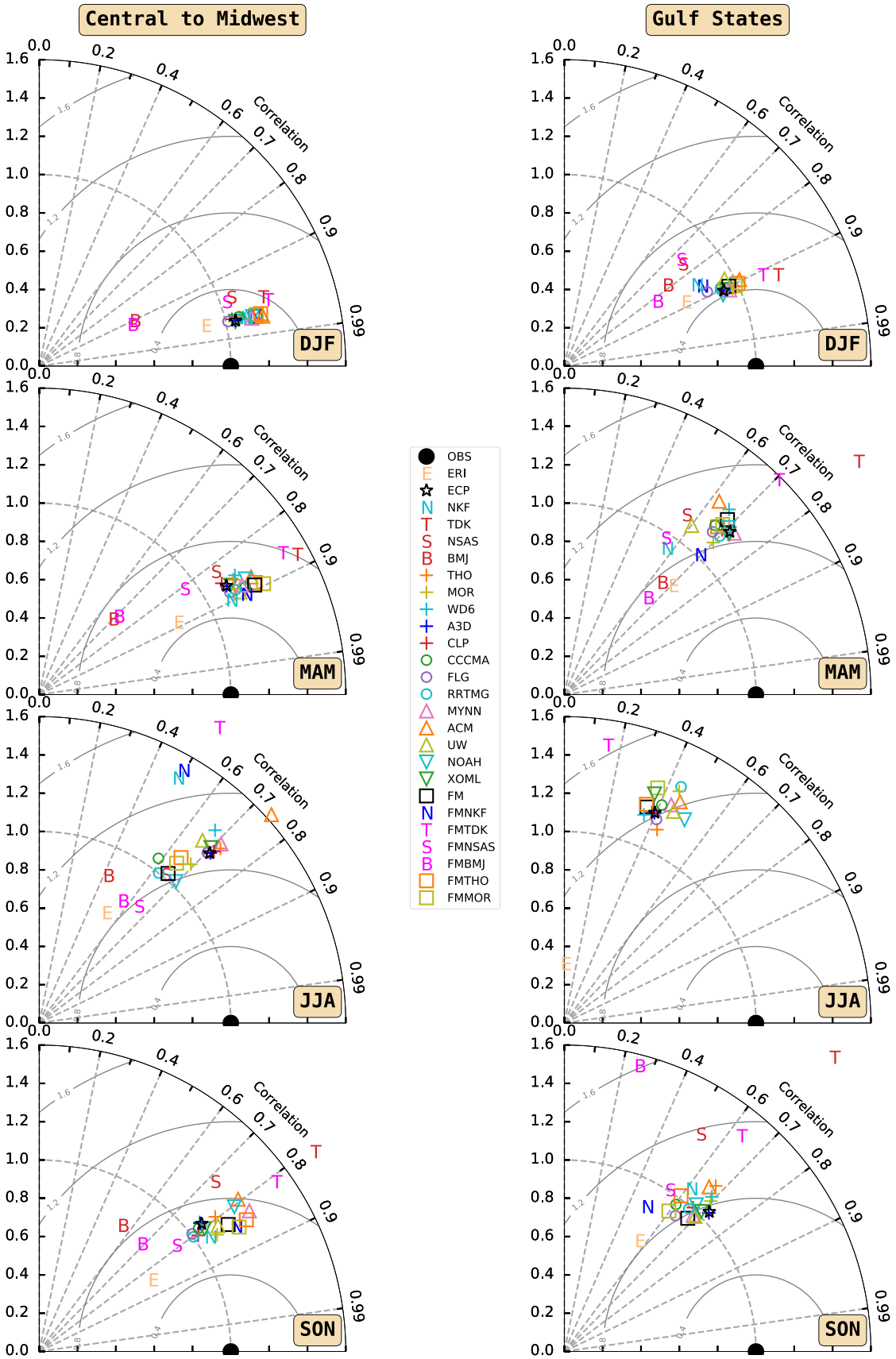
Season	Scheme					
	ERI	ECP	NKF	TDK	NSAS	BMJ
DJF	0.93	0.91	0.88	0.85	0.94	0.96
JJA	0.88	0.89	0.90	0.74	0.92	0.96
MAM	0.94	0.91	0.88	0.90	0.96	0.98
SON	0.93	0.89	0.88	0.82	0.94	0.97

minus hits expected by chance (Gandin and Murphy 1992). Given a specific P95 range where an event is observed, it is a hit if the modeled value also falls in that range or otherwise it is a miss. The numbers of these hits and misses are counted to calculate the ETS in each bin. Figure 11 compares ETS at a 1.0 (mm day^{-1}) bin interval of 36-year mean seasonal P95 spatial distribution over the GS and CM regions among ERI and the 25 CWRF physics configurations. In general, ETS skills were highest in winter and lowest in summer. This highlights the difficulty of simulating extreme events in summer. On the other hand, CWRF exhibited overall greater improvements over ERI in warm seasons (e.g. JJA and SON in both two regions and MAM in GS) than in cold seasons (e.g. winter and spring over CM and winter over GS). This seasonal contrast in CWRF's added values may partly result from the predominance of convective rainfall events in warm seasons, in which the primary model sensitivity comes from the cumulus parameterization (Di Luca et al. 2012; Liang et al. 2018). Another factor is that the microphysics scheme adopted in CWRF tends to overestimate stratiform precipitation (see discussion below), which may reduce the added values in cold seasons.

In the GS region, the control ECP outperformed ERI in all seasons for the entire range of observed P95 magnitudes, except in winter for light precipitation (~ 10 mm day^{-1}). This ETS enhancement was significant, especially in spring and autumn for the entire P95 range, when ERI showed low skill. The improvement was moderate in summer, when ERI scored zero, and also substantial in winter for rainfall above 25 mm day^{-1}. While most ECP members combining with other processes' parameterization schemes were fairly similar to the control, some had notable skill improvements. In particular, the members combined with land surface scheme NOAH replacing CSSP increased ETS and microphysics scheme MOR replacing TAO both increased ETS in winter for P95 of 12–28 mm day^{-1}, in summer for heavy rain above 25 mm day^{-1}. The radiation scheme CCCMA or RRTMG replacing GSFC, and the combined radiation-boundary layer-microphysics scheme FLG-MYNN-MOR replacing control GSFC-CAM-TAO also showed improved skill in summer. Thus, there is still room for further skill enhancement in summer through physics refinement.

CWRF members using cumulus schemes other than ECP generally had lower ETS. NKF members scored systematically lower in winter for the entire P95 range, and also in other seasons except for some improvement to rainfall above 35 mm day^{-1}, especially in autumn. The improvement was limited to a small area along the Texas-Louisiana coast, where ECP underestimated P95 (Fig. 7). Replacing ECP with TDK reduced ETS systematically, except for improvements in the middle range between 12–20 mm day^{-1} in winter, 20–22 mm day^{-1} in spring, and 22–25 mm day^{-1} in autumn. The improvements occurred in part of Texas, where ECP overestimated P95. Both NSAS and BMJ members substantially underestimated P95 (Fig. 7) and scored persistently lower than ECP in all seasons.

In the CM region, ERI generally scored higher than all CWRF members in winter and spring, mainly because the latter systematically overestimated P95 (Fig. 6). Most CWRF members had higher ETS for rainfall events above 27 mm day^{-1}, which were totally missed by ERI. Notably, NSAS had very high ETS in spring, substantially larger than other CWRF members. It produced systematically higher ETS than ECP, with especially large score increases between 20–30 mm day^{-1}. Figures 7, 8, 9 indicates that NSAS significantly underestimated NRD, while ECP was more realistic. Consequently, NSAS showed high skill for P95 and DRI, but at the cost of overestimating clear days. In spring, TDK also produced slightly higher ETS than ECP. In summer, for the entire P95 range, CWRF significantly outperformed ERI, which had almost zero ETS. The control ECP yielded the highest ETS, except for a slight improvement in rainfall above 23 mm day^{-1} achieved by its combination with boundary layer scheme ACM. All TDK and NKF members scored lower across the entire P95 range, with NSAS members performing poorly, and BMJ members failing completely. In autumn, ERI had higher ETS than most CWRF members for P95 between 15–22 mm day^{-1}. ERI skill dropped abruptly above this range, and thus was increasingly outperformed by CWRF members as P95 rose. The control ECP generally scored highest, but was exceeded by several other members between 18–27 mm day^{-1}, including those using cumulus NSAS and microphysics MOR. This again indicated the potential for further physics improvement. The BMJ members were persistently outliers, with little skill.


Central to Midwest
Gulf States
DJF
MAM
JJA
SON
Correlation
OBS
ERI
ECP
NKF
TDK
NSAS
BMJ
THO
MOR
WD6
A3D
CLP
CCCMA
FLG
RRTMG
MYNN
ACM
UW
NOAH
XOML
FM
FMNKF
FMTDK
FMNSAS
FMBMJ
FMTHO
FMMOR

◂**Fig. 10** Taylor diagram of pattern statistics comparing the overall performance among ERI and all CWRF physics configurations in simulating 1980–2015 mean seasonal P95 geographic distributions over the CM (left) and GS (right) regions for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). Shown are the pattern correlation (azimuthal) and normalized standard deviation (radius) compared with observations. The black dot (OBS) marks the perfect score with a unit correlation and deviation. Off the chart are outliers performing poorly, with correlations and deviations indicated in the parentheses

5 Effects of cumulus parameterization

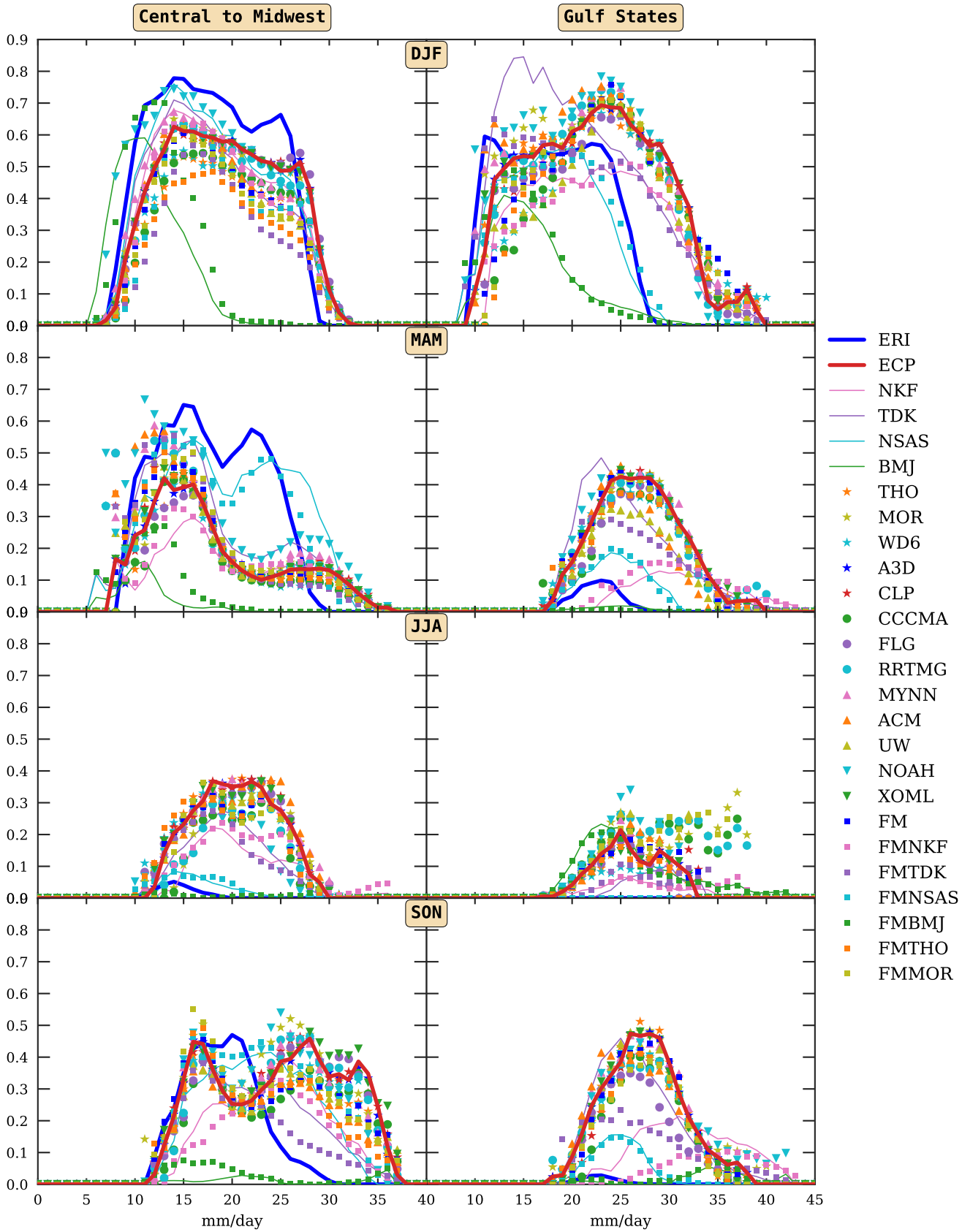
Figure 12 compares vertical temperature and humidity tendency profiles among the five cumulus schemes averaged in the summers of 2005–2009. These profiles were calculated (from re-runs initialized on March 1) using 3-hourly samples and averaged over all the grids having rainfall greater than 50 mm day^{-1} within the CM and GS regions separately. Also shown is the vertical integral of each profile, representing the overall strength of the convection. For all schemes, the tendency magnitudes were generally greater in the GS than CM region, indicating stronger convection in the former. While BMJ is based on the convective equilibrium that adjusts unstable model column to a moist adiabat, all other schemes are based on the mass flux concept but differ in their formulations of subgrid plume entrainment and detrainment as well as trigger function and closure assumption (Qian et al. 2015). Thus, for both regions, BMJ produced a greater warming peak at a higher altitude than other schemes in order to partly cancel a strikingly large cooling layer below ~600 hPa. A much weaker and shallower cooling layer occurred in ECP, which simulated a unique warming profile of all moderate attributes including the overall magnitude, layer thickness, and peak altitude as compared to other schemes. In contrast, NKF produced a much greater and deeper warming layer than ECP, and a tiny cooling at cloud base. Hence, NKF expanded the ECP warming profile toward the surface. NKF also had a significant cooling peak in the cumulus top, which could represent an outlier or formulation deficiency likely due to excessive heat loss from large detrainment and associated evaporation. On the other hand, NSAS generated a parabolic warming profile (without cooling near the cloud base) that is weaker above and stronger below ~750 hPa than ECP, especially in the GS region. In the entire cumulus tower, TDK yielded an extremely weak parabolic warming.

All cumulus schemes resulted in drying throughout the cloud column as water vapor was depleted by precipitation. However, the vertical distributions of the drying differed substantially. In the CM region, BMJ had distinct double peaks near 925 and 550 hPa, corresponding to respectively the shallow and midlevel convection, with the latter relatively stronger. ECP also simulated double peaks near 875 and 600 hPa, but the midlevel drying was predominant. In contrast, NKF produced a parabolic drying layer that is much stronger and deeper than ECP, with the predominant peak near 700 hPa. These features were similarly presented in the GS region, except that the drying peaks were closer to the surface by 50 hPa, indicating deeper convection, and that the overall strength were increased by 51% (ECP) and 62% (NKF) but decreased by 11% (BMJ). On the other hand, in the CM region, NSAS generated a deep drying layer that was overall stronger than ECP but weaker than NKF, with the predominant peak near 825 hPa, whereas in the GS region, it produced a much weaker drying than both ECP and NKF, with the peak further down near 950 hPa. In both regions, TDK yielded a tiny drying and very light rainfall.

The above comparison exemplified the complexity of convective effects as parameterized by different cumulus schemes. Among all schemes, NKF produced the strongest warming and drying rates and correspondingly largest P95, whereas NSAS and especially TDK generated drastically weaker rates and so substantially underestimated P95. On the other hand, ECP and BMJ both simulated moderate rates, but the former realistically captured P95 whereas the latter substantially underestimated it. Therefore, it is difficult to identify a general correspondence in the strength between convection and P95.

To simplify the complexity, a sensitivity experiment was conducted using CWRF control ECP with 16 closure assumptions individually turned on at a time. All 16 runs were similarly integrated for the five summers in 2005–2009. Figure 13 compares 5-year mean summer P95 biases (from observations) between these ECP sensitivity runs along with the ERI, CWRF control and four other cumulus schemes (configurations A–E) averaged over the two regions. Table 4 lists the values of these corresponding mean and relative biases. The performance for ERI and ECP was persistent from the 36-year to 5-year average in both (GS, CM) regions, with the former substantially underestimated and the latter fairly realistic. However, the performance for other four cumulus schemes was more variable: NKF strongly overestimated, TDK mixed, NSAS strongly underestimated, and BMJ similarly underestimated. Nonetheless, the 5-year average captured the major differences in the 36-year average between ERI and CWRF cumulus schemes. Notably, changing alone the ECP closure assumption could produce a similar range of sensitivity among different cumulus schemes, with biases varied from −6.5 to 10.6 mm day^{-1} or −28 to 46% in GS and from −2.5 to 7.6 mm day^{-1} or −11 to 33% in CM.

Figure S1 (*Supplementary Information*) compares vertical temperature and humidity tendency profiles among the ECP closure assumptions averaged in the summers of 2005-2009. These closures were separated into five main groups based on: moisture convergence (MC), vertical


Central to Midwest
Gulf States
DJF
MAM
JJA
SON
mm/day
mm/day
ERI
ECP
NKF
TDK
NSAS
BMJ
THO
MOR
WD6
A3D
CLP
CCCMA
FLG
RRTMG
MYNN
ACM
UW
NOAH
XOML
FM
FMNKF
FMTDK
FMNSAS
FMBMJ
FMTHO
FMMOR

◂**Fig. 11** The equitable threat score (ETS) that measures the overall skill dependence on rainfall intensity for ERI and all CWRF physics configurations in simulating 1980–2015 mean seasonal P95 geographic distributions over the CM (left) and GS (right) regions for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). The *x*-axis depicts the P95 thresholds at a 1.0 mm day^{-1} bin interval, while the y-axis scores the ETS values

velocity (W), quasi-equilibrium (AS), total instability (KF), and instability tendency (TD) (see details in Qiao and Liang 2015). They produced similar vertical tendency distributions as ECP, NKF, NSAS, and TDK discussed earlier (Fig. 12), with the column integrated total heating of even larger spreads: 1716 versus 1329 K day^{-1} in CM and 2539 versus 1816 K day^{-1} in GS. Although the strong lower-cooling/upper-warming dipole pattern of BMJ was absent, the systematic weak warming in TDK was simulated by two ECP closures (minimum MC and W of 9 surrounding grids). Thus, the sensitivity among the ECP closures may be considered as a reasonable illustration of that among the four mass-flux based cumulus schemes.

Figure 14 shows scattering relationships of the total convective heating and drying rates (from strong cumuli of rainfall greater than 50 mm day^{-1}) with P95 and mean precipitation (PRA) biases in individual summers of 2005–2009 averaged over the two regions, comparing all five cumulus schemes and 16 ECP closures. The corresponding relationships with NRD and CDD (consecutive dry days) biases are present in Fig. S2. Two clusters were identified with opposite relationships, which were more obvious in GS than CM. One cluster consists of ECP, NKF, NSAS schemes and ECP's MC (maximum, local, average), W maximum, AS, and KF closures. This cluster depicts that stronger convective heating/drying rates were associated with larger P95 and PRA, while the correspondences with NRD and CDD were much weaker. Thus, these schemes or closures altered P95 mainly through their impacts on total precipitation intensity rather than rainy-day frequency. Another cluster consists of the TDK scheme and ECP's MC minimum, W (average at cloud base or updraft, minimum) and TD closures. This cluster depicts that stronger convective heating/drying rates were associated with smaller P95 but more NRD and fewer CDD, while the correspondences with PRA were much weaker. Hence, these schemes or closures altered P95 mainly through their impacts on rainy-day frequency rather than total precipitation intensity. Note that the ECP scheme currently used an ensemble of the MC average and the mean of all four W closures as weighted by 0.5 and 1.5. Its result was near the margin between the two clusters. In addition, BMJ was an outlier for both P95 and PRA, most obviously in CM, suggesting its inability to capturing the regional cumulus characteristics.

6 Conclusion and discussion

We analyzed CWRF extreme precipitation simulations in 1980–2015 over the contiguous United States and selected two key regions (GS and CM) of substantial ERI underestimation with weak orographic forcing to focus on the sensitivity to physical process parameterizations. By comparing an ensemble of 25 simulations downscaled from ERI during 1980–2015 with CWRF physics configurations of varying parameterization schemes, we investigated the responsive processes to which regional precipitation extremes are sensitive. We found that of all the physics configurations, CWRF's P95 simulation was most sensitive to cumulus parameterization. Accordingly, we classified the CWRF configurations into two broad types based on their cumulus schemes. Type I (using ECP and NKF) did not significantly underestimate P95 in either region, while type II (using TDK, NSAS, and BMJ) produced substantial underestimations in either GS spring or CM summer. They differed substantially in model biases and skill scores, depending on regions and seasons, as summarized below.

In the GS region, ERI substantially underestimated P95 in all seasons, while CWRF type I members produced general improvement. In particular, CWRF control ECP had the highest ETS for the entire observed P95 range, and outperformed others most significantly in spring when ERI's underestimation was the largest. NKF members generally overestimated P95, except in winter, and scored systematically lower than ECP, albeit for some improvement to heavy rainfall (especially in autumn). Type II members generally had lower ETS, severely underestimated P95 in all seasons, and produced generally lower pattern correlations and substantially smaller (BMJ) or larger (TDK) spatial variations (especially in summer). One exception was TDK's small biases in winter, spring, and autumn, which were the result of incorrect spatial patterns. The members of TDK replacing ECP reduced ETS systematically in summer, and with the exception of the middle P95 range, also in other seasons. NSAS members scored persistently lower than ECP, and BMJ members had even lower skill.

In the CM region, ERI substantially underestimated summer P95, while CWRF type I members produced the most realistic simulations. The control ECP had the highest ETS and significantly outperformed ERI for the entire P95 range. ERI underestimated autumn P95, while CWRF type I members slightly overestimated it and hence showed more skill for heavier rainfall. In winter and spring, when convective activities are relatively infrequent, ERI scored higher for light to moderate P95 than CWRF, but increasingly lower for heavier precipitation, mainly because of the systematic CWRF overestimation

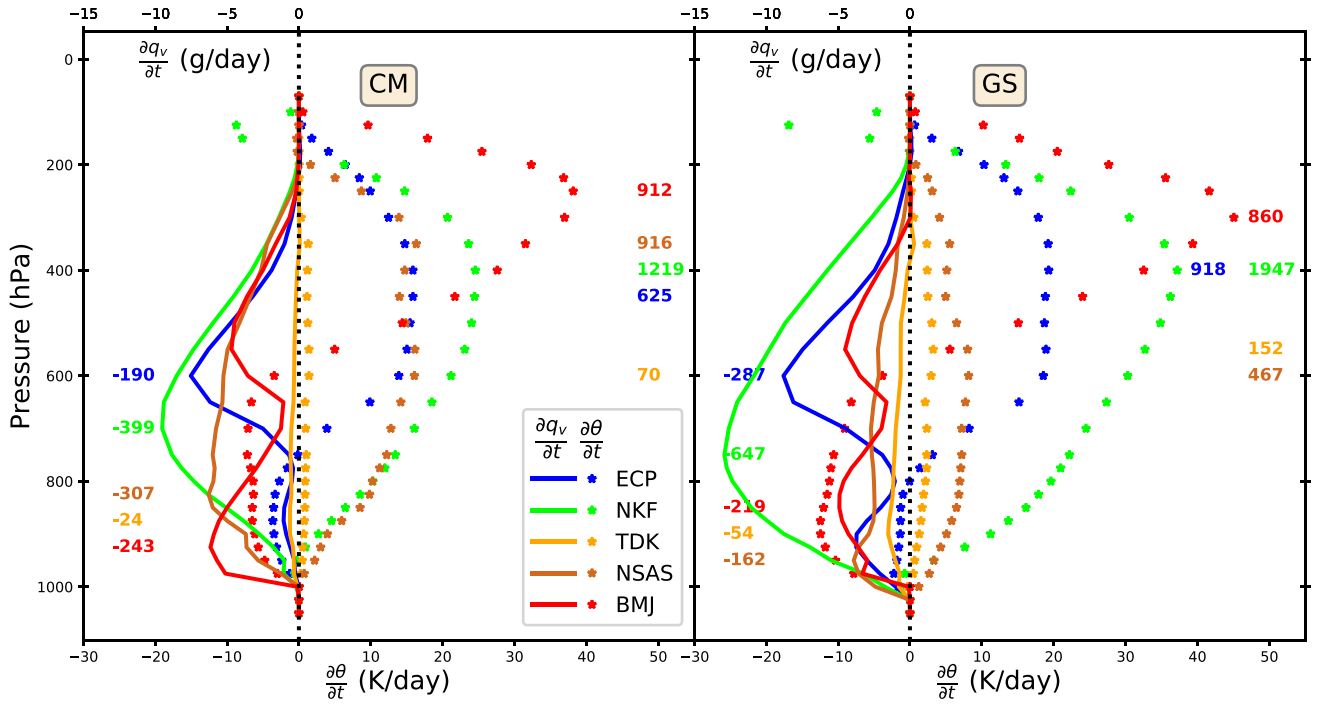


Fig. 12 Summer mean vertical potential temperature (θ, star) and water vapor (q_v, curve) tendency profiles among the five cumulus schemes (color) as averaged in 2005–2009 and over all the grids having rainfall greater than 50 mm day^{-1} within the CM (left) and GS (right) regions. Also labeled at the altitude of the profile peak is the number that depicts the tendency's vertical integral for the respective scheme coded with the same color

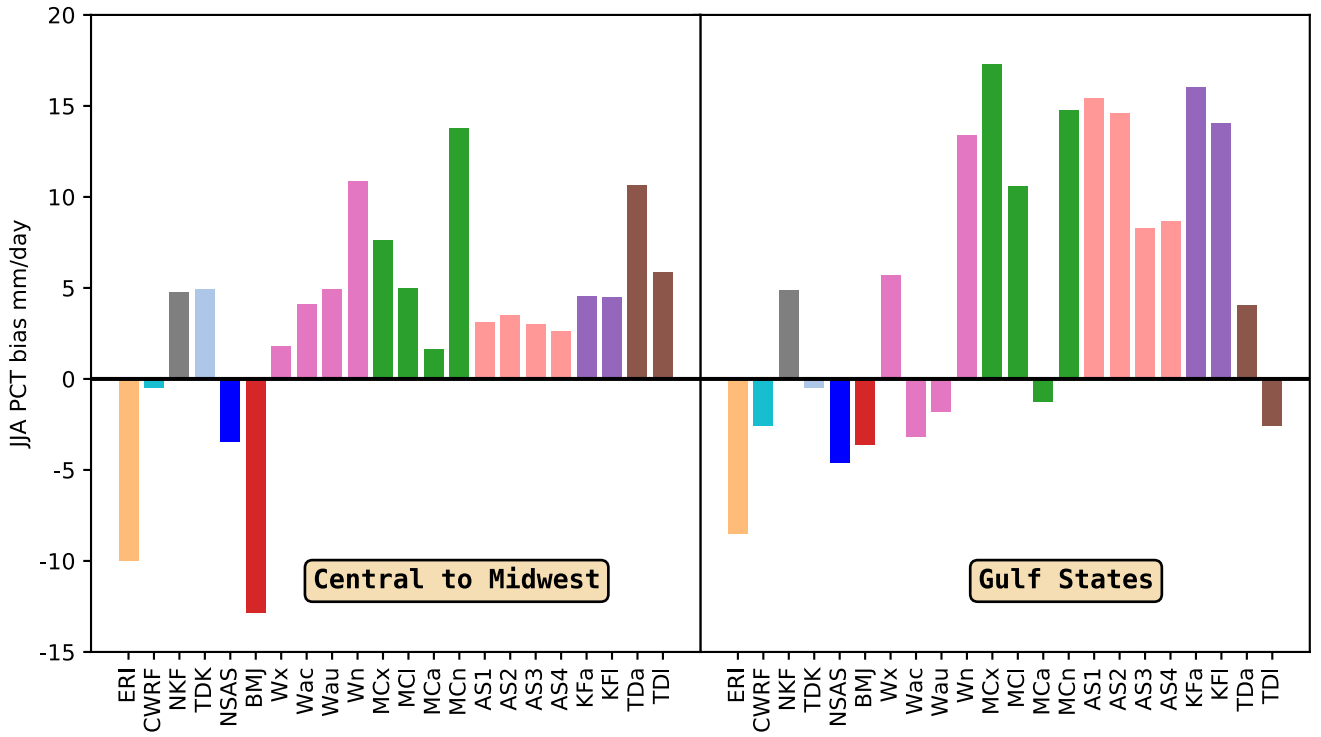


Fig. 13 Comparison among the 16 ECP closures' sensitivity runs along with the ERI and CWRF five cumulus schemes in simulating 2005–2009 mean summer P95 (mm day^{-1}) biases (from observations) averaged over the CM (left) and GS (right). They are separated by different colors into ERI, CWRF control and four other cumulus schemes, and five main cumulus closure groups

Table 4 Five-year averaged summer P95 mean (mm day^{-1}) and relative (%) biases (from observations) over the CM and GS regions am, comparing ERI, CWRF control run, and CWRF five cumulus schemes and five main cumulus closure groups

	ERI	CWRF	NKF	TDK	NSAS	BMJ	Wx	Wac	Wau	Wn	MCx	MCl	MCa	MCn	AS1	AS2	AS3	AS4	KFa	KFl	TDa	TDl
Bias																						
CM	−10.0	−0.5	3.9	−0.1	−4.2	−12.7	−2.3	−0.4	0.2	5.2	2.5	0.3	−2.5	7.6	−1.3	−0.9	−1.4	−1.6	−0.1	−0.1	5.0	1.1
GS	−8.5	−2.6	8.3	−7.3	−7.4	−2.6	0.9	−6.5	−5.3	7.3	10.6	5.0	−4.9	8.4	9.0	8.3	3.1	3.4	9.5	7.8	−0.5	−6.0
Rel. bias (%)																						
CM	−43.4	−2.2	17.1	−0.5	−18.4	−55.4	−10.1	−1.8	1.1	22.7	11.0	1.3	−10.8	33.2	−5.5	−3.9	−5.9	−7.2	−0.3	−0.5	22.0	4.6
GS	−36.8	−11.1	35.8	−31.7	−32.2	−11.2	3.9	−28.0	−23.2	31.6	45.8	21.4	−21.2	36.6	39.0	35.9	13.3	14.6	41.3	34.1	−2.1	−26.0

and ERI underestimation. In all seasons, type II members significantly underestimated P95, and produced generally lower pattern correlations and substantially smaller (BMJ) or larger (TDK) spatial variations. TDK scored lower than ECP and NKF, while NSAS had even less skill and BMJ failed totally. Two exceptions were that NSAS had outstanding ETS for spring P95, but due to its overestimation of clear days, and that TDK had small regional mean biases in spring and autumn, but due to the cancelation of large positive and negative local errors.

Some cumulus schemes may have the potential to capture precipitation extremes under mixed synoptic and convective forcings. For example, CWRF using TDK simulated spring and autumn P95 reasonably well in both regions, though it did not correctly capture spatial patterns, while NSAS even outperformed ECP for most P95 ranges in spring CM, though it overestimated clear days. Other parameterization schemes may be able to work with ECP to further improve CWRF skills. In particular, combining the ECP cumulus with MOR microphysics schemes significantly enhanced CWRF's ability to capture summer P95 in the GS region. For heavy summer rainfall events, the ECP cumulus combining with CCCMA radiation, MYNN boundary layer, and NOAH land surface schemes also produced scores higher than the control. Thus, there is still room for further skill enhancement through physics refinement of the whole model system.

The P95 improvement of the control CWRF over ERI, especially over the GS and CM regions in seasons dominated by convective precipitation, may reflect its use of ECP with an optimized closure ensemble based on the framework of Grell and Dvénéyi (2002) to represent convection variations between land and oceans (Qiao and Liang 2015, 2016, 2017). In contrast, CWRF members using TDK, NSAS, and BMJ underestimated P95 substantially in summer, and also largely in autumn and spring. ERI, which used a variant of the TDK cumulus scheme (Bechtold et al. 2004), similarly underestimated P95. Additional experiments of five summers in 2005–2009 indicated the complex P95 effects of convective heating/drying profiles and closure assumptions in various parameterization schemes. Changing alone the ECP closure assumption could produce the same range of sensitivity among different cumulus schemes. They formed two clusters of opposite relationships between convective heating/drying rates and P95 biases. One cluster cumulus schemes or closures altered P95 mainly by changing total precipitation intensity, while another by changing rainy-day frequency. The result suggests an opportunity to improve extreme precipitation simulation by refining cumulus closure and other assumptions.

One important question raised by a reviewer is whether a higher resolution can be beneficial to extreme precipitation simulation. This is a challenging question to answer

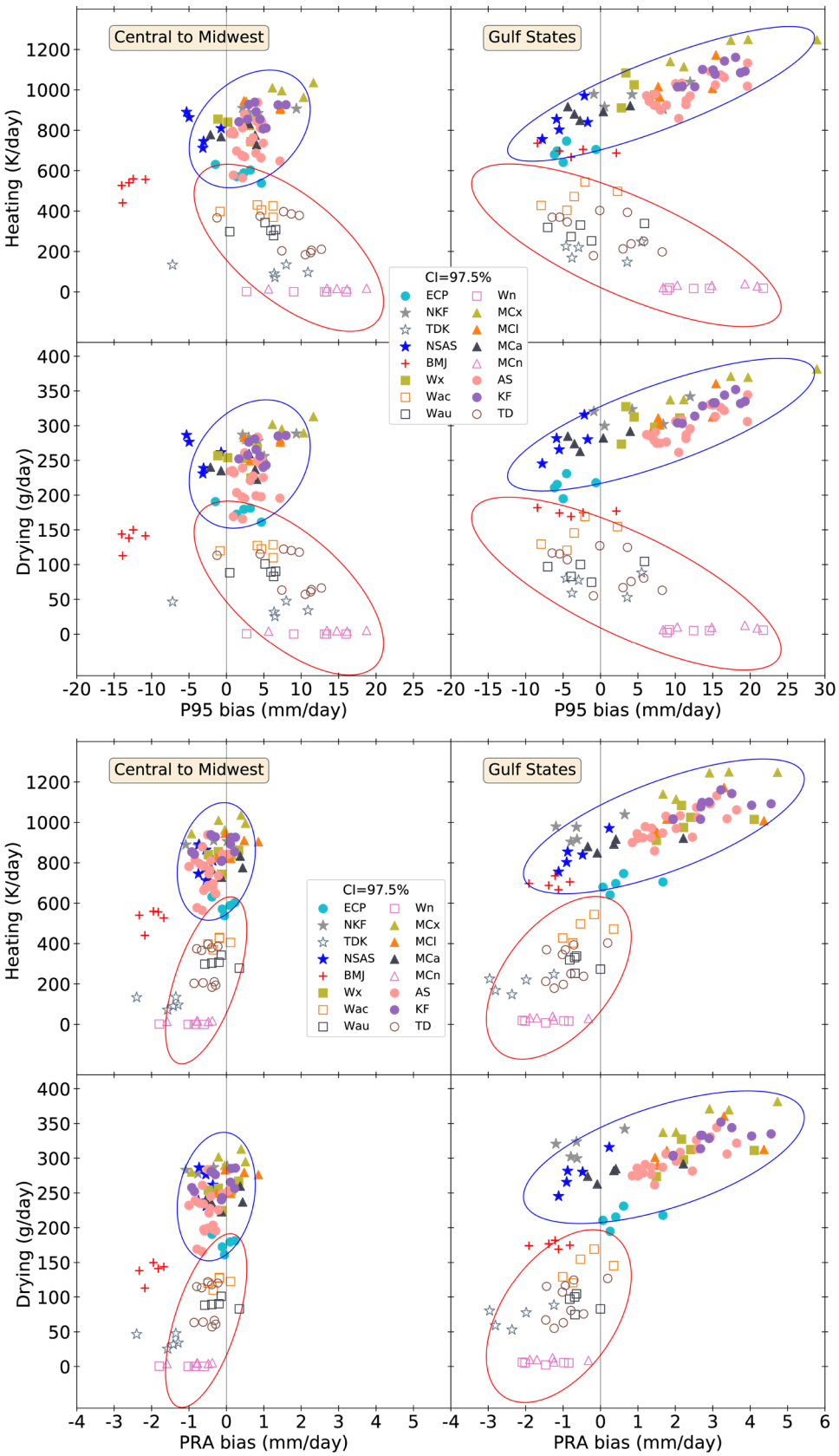


Fig. 14 The scattering diagram of total convective heating and drying rates with P95 and PRA biases among 16 ECP closures along with the ERI, CWRF control and five cumulus schemes in individual summers of 2005–2009 averaged over the CM (left) and GS (right) regions. Two ellipses in each plot enclose two clusters of closures within a confidence interval (CI) of 97.5%, using a solid and empty center mark separately

given our still very limited understanding of how precipitation extremes occur and the huge demand for computational resources running CWRF climate simulations at finer grids. The present CWRF ensemble experiment at 30-km with 25 members for a 36-year integration consumed over four million core hours of CPU time and occupied over 500 terabytes of storage, both of which were at the limit our group could afford. As a sensitivity test, we conducted an integration using the control CWRF for year 1993, when the record flood occurred in the Midwest in summer. (This single 1-year run took 180 K core hours and 2.3 terabyte storage.) Figure 15 compares seasonal P95 geographic distributions in 1993 among observations, ERI, and the control CWRF runs at 30-km and 10-km. In agreement with our earlier finding from the 36-year average, ERI systematically underestimated the summer P95 intensity over much of the CONUS, failed to capture the flood records over most of the Midwest. In contrast, both CWRF simulations well represented the Midwest summer flood, although the 10-km run produced generally weaker P95 than the 30-km run. Such P95 reduction from 30- to 10-km grid was also seen in other seasons. In particular, the large reduction in summer, especially over Texas and Louisiana (within GS), was a degradation as compared with observations, whereas that in winter and spring was an improvement in the GS region because of the overestimation at 30-km. The sensitivity analysis suggests that the ECP cumulus parameterization captures convective P95 (prevailing in summer) more realistically at 30-km than 10-km, while the microphysics scheme of Tao et al. (1989) reduces but improves stratiform P95 (dominant in winter and spring) as the grid resolution increases. The resolution dependence of cumulus parameterization versus explicit convection (using a microphysics scheme alone) has been evaluated under the multigrid nesting framework of whether forecasts by Liang et al. (2019). They concluded that nesting an outer domain (like our CWRF domain) parameterizing cumuli at 15-km with an inner domain explicitly resolving convections at 1-km can achieve the best overall rainfall forecast at local to regional scales. Therefore, it would require further running CWRF at 10-km finer to 1-km in order to fully address the reviewer's question, which is beyond the scope of this study.

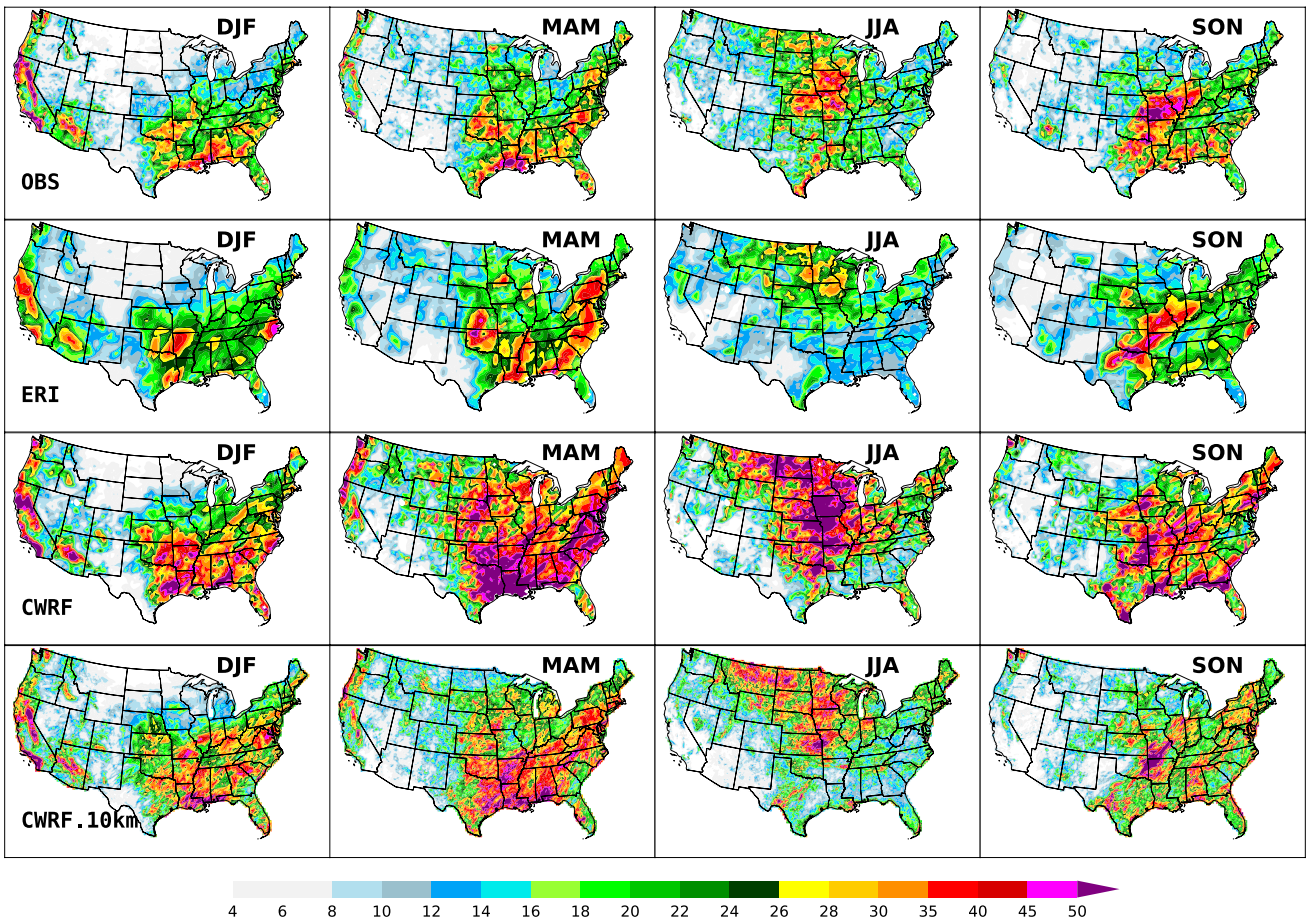


Fig. 15 Same as Fig. 1 except for P95 (mm day^{-1}) in 1993 alone as observed (OBS), assimilated (ERI), and simulated by CWRF control at 30-km and 10-km

Along the same line, the finer-resolution PRISM difference from the COOP data was questioned for potentially affecting the model bias analysis. In addition, the newly released ECMWF reanalysis ERA5 (Hersbach and Dee 2016) replaced ERI, reducing the grid spacing from 78- to 31-km. Both raised a concern on the robustness of our finding about the physics parameterization sensitivity. Figure S3 shows geographic distributions of seasonal P95 biases from observations (COOP) for PRISM and ERA5. Per the discussion in Sect. 4, these two datasets were interpolated onto the CWRF 30-km grid using the same conservative method. Given the incorporation of high-resolution radar measurements, PRISM overestimated P95 by smaller than 10 mm day^{-1} over scattered areas. As averaged over the CM and GS regions, the overestimation was only (2.5, 3.0, 2.7, 1.6) and (3.5, 2.4, 3.3, 2.8) from spring to winter, much smaller than the range of differences among ERI and CWRF five cumulus members in each season (Fig. 6). Would PRISM be chosen as the reference instead of COOP, all model underestimations, which are our main concern, would be somewhat larger, but the sensitivity among physics configurations would remain. In contrast, ERA5 substantially reduced P95 biases over ERI (Fig. 7) in all seasons, which were only (1.4, −0.9, 0.4, 1.8) in CM and (−2.3, −4.4, −3.6, −1.5) in GS. As such, ERA5 still had the underestimation problem only in GS, albeit reduced by about 2–3 times from ERI. While a detailed documentation of ERA5 is yet to come, its improvement in P95 over ERI may likely be attributed to its advanced physics representation and enhanced data assimilation in combination with refined spatial resolution (Hersbach and Dee 2016; Hersbach et al. 2019). In particular, ERA5 upgraded the cumulus scheme from Tiedtke (1989) and Bechtold et al. (2004) to incorporate new entrainment formulation (Bechtold et al. 2008) and parameterization closure (Bechtold et al. 2014), which caused breakthrough in forecasting convective precipitation (Bechtold et al. 2013). This is consistent with our finding that P95 simulation is sensitive to cumulus parameterization. On the other hand, ERA5 assimilated much more observational data than ERI, especially it newly incorporated radar and total-column water vapor data that directly restrict extreme precipitation amounts over land. Thus, ERA5 precipitation was constrained much more strongly than ERI, significantly reducing its P95 sensitivity to physics representation, and so inappropriate for this comparative study.

Note that all CWRF 36-year simulations, varying only in physics configurations, were driven by the same ERI forcing and run at 30-km grid spacing over the same North American domain. Thus, the result differences among these CWRF runs reflect solely the impacts of varying physics schemes. We have demonstrated that the model physics representation, especially the cumulus parameterization, plays a critical role in extreme precipitation simulation. This conclusion applies not only to the CONUS but also to the China domain (Liang et al. 2018). Along with our limited resolution sensitivity study just discussed and the finding in Liang et al. (2019), we would recommend that future regional climate modeling efforts should focus more on developing refined physics representation rather than simply adopting higher resolutions.

Given our sensitivity analysis above, it is imperative to understand how these cumulus parameterizations interact with other physics representations to alter extreme precipitation simulation. More comprehensive analyses and sensitivity experiments are needed to identify the physical mechanisms and feedback processes that are responsible for model failure or success. This will be the goal of our companion paper (Sun and Liang 2019).

Acknowledgements The CWRF simulations and analyses were conducted on the supercomputers, including the University of Illinois' Blue Water, the Maryland Advanced Research Computing Center, the Computational and Information Systems Lab of the National Center for Atmospheric Research, and the National Energy Research Scientific Computing Center of U.S. Department of Energy. We thank Kenneth Kunkel for providing the Cooperative Observer network station data. The research was supported by U.S. National Science Foundation Innovations at the Nexus of Food, Energy and Water Systems under Grants EAR-1639327 and EAR1903249, U.S. Department of Agriculture UV-B Monitoring and Research Program at Colorado State University under the National Institute of Food and Agriculture Grant 2015-34263-24070, and U.S. Environmental Protection Agency Science to Achieve Results under Assistance Agreement No. RD83587601. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the funding Agency. We sincerely thank Jennifer Kennedy for thorough editing.

References

Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Rupa Kumar K, Revadekar J, Griffiths G, Vincent L, Stephenson DB, Burn J, Aguilar E, Brunet M, Taylor M, New M, Zhai P, Rusticucci M, Vazquez-Aguirre JL (2006) Global observed changes in daily climate extremes of temperature and precipitation. J Geophys Res. https://doi.org/10.1029/2005JD006290

Allan RP, Soden BJ (2008) Atmospheric warming and the amplification of precipitation extremes. Science 321:1481–1484

Allen MR, Ingram WJ (2002) Constraints on future changes in climate and the hydrologic cycle. Nature 419:224–232

Anderson BT, Gianotti DJ, Salvucci GD (2015) Detectability of historical trends in station-based precipitation characteristics over the continental United States. J Geophys Res Atmos 120:4842–4859

Bechtold P, Chaboureau J-P, Beljaars A, Betts A, Kohler M, Miller M, Redelsperger J-L (2004) The simulation of the diurnal cycle of convective precipitation over land in a global model. Q J R Meteorol Soc 130:3119–3137

Bechtold P, Kohler M, Jung T, Doblas-Reyes F, Leutbecher M, Rodwell MJ, Vitart F, Balsamo G (2008) Advances in simulating atmospheric variability with the ECMWF model: from synoptic to decadal time-scales. Q J R Meteorol Soc 134:1337–1351

Bechtold P, Semane N, Lopez P, Chaboureau JP, Beljaars A, Bormann N (2013) Breakthrough in forecasting equilibrium and non-equilibrium convection. ECMWF Newsl 136:15–22

Bechtold P, Semane N, Lopez P, Chaboureau J-P, Beljaars A, Bormann N (2014) Representing equilibrium and nonequilibrium convection in large-scale models. J Atmos Sci 71:734–753

Betts A, Miller M (1986) A new convective adjustment scheme. Part II: single column tests using gate wave, BOMEX, ATEX and arctic air-mass data sets. Q J R Meteorol Soc 112:693–709

Boyle J, Klein SA (2010) Impact of horizontal resolution on climate model forecasts of tropical precipitation and diabatic heating for the TWP-ice period. J Geophys Res Atmos 115:D23113

Brown JR, Jakob C, Haynes JM (2010) An evaluation of rainfall frequency and intensity over the Australian region in a global climate model. J Clim 23:6504–6525

Cairo A (2016) The truthful art: data, charts, and maps for communication. New Riders, Indianapolis

Catto JL, Pfahl S (2013) The importance of fronts for extreme precipitation. J Geophys Res Atmos 118:1079–10801

Chen C-T, Knutson T (2008) On the verification and comparison of extreme rainfall indices from climate models. J Clim 21:1605–1621

Choi HI (2006) 3-D volume averaged soil-moisture transport model: a scalable scheme for representing subgrid topographic control in land-atmosphere interactions. Ph.D. dissertation, University of Illinois at Urbana, Champaign

Choi HI, Liang X-Z (2010) Improved terrestrial hydrologic representation in mesoscale land surface models. J Hydrometeorol 11:797–809

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

Choi HI, Liang X-Z, Kumar P (2013) A conjunctive surface–subsurface flow representation for mesoscale land surface models. J Hydrometeorol 14:1421–1442

Choi I-J, Jin EK, Han J-Y, Kim S-Y, Kwon Y (2015) Sensitivity of diurnal variation in simulated precipitation during east Asian summer monsoon to cumulus parameterization schemes. J Geophys Res Atmos 120:11–971

Chou M-D, Suarez MJ (1999) A solar radiation parameterization for atmospheric studies, vol 15

Chou M, Suarez MJ, Liang X-Z, Yan MM-H, Cote C (2001) A thermal infrared radiation parameterization for atmospheric studies

Dai A (2006) Precipitation characteristics in eighteen coupled climate models. J Clim 19:4605–4630

Dai Y, Zeng X, Dickinson RE, Baker I, Bonan GB, Bosilovich MG, Denning AS, Dirmeyer PA, Houser PR, Niu G, Oleson KW (2003) The common land model. Bull Am Meteorol Soc 84:1013–1024

Daly C, Taylor G, Gibson W (1997) The PRISM approach to mapping precipitation and temperature. In: Proceedings of 10th AMS conference on applied climatology, pp 20–23

Daly C, Slater M, Roberti JA, Laseter S, Swift L (2017) High-resolution precipitation mapping in a mountainous watershed: ground truth for evaluating uncertainty in a national precipitation dataset. Int J Climatol. https://doi.org/10.1002/joc.4986

Daniels AE, Morrison JF, Joyce LA, Crookston NL, Chen S-C, McNulty SG (2012) Climate projections FAQ. Gen. Tech. Rep. RMRS-GTR-277WWW. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer DP, Bechtold P (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597

Di Luca AR, de Elía R, Laprise R (2012) Potential for added value in precipitation simulated by high-resolution nested regional climate models and observations. Clim Dyn 38(5–6):1229–1247

Donat MG, Alexander LV, Herold N, Dittus AJ (2016) Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations. J Geophys Res Atmos 121:11–174

Durre I, Menne MJ, Gleason BE, Houston TG, Vose RS (2010) Comprehensive automated quality assurance of daily surface observations. J Appl Meteorol Climatol 49:1615–1633

Easterling DR, Evans JL, Groisman PY, Karl TR, Kunkel KE, Ambenje P (2000) Observed variability and trends in extreme climate events: a brief review. Bull Am Meteorol Soc 81:417

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

Evans JP, Ekstrom M, Ji F (2012) Evaluating the performance of a WRF physics ensemble over south-east Australia. Clim Dyn 39:1241–1258

Fan J, Hu T-C, Truong YK (1994) Robust non-parametric function estimation. Scand J Stat 21:433–446

Fischer EM, Beyerle U, Knutti R (2013) Robust spatially aggregated projections of climate extremes. Nat Clim Change 3:1033

Frei C, Christensen JH, Déqué M, Jacob D, Jones RG, Vidale PL (2003) Daily precipitation statistics in regional climate models: evaluation and intercomparison for the European Alps. J Geophys Res Atmos 108:4124

Frich P, Alexander LV, Della-Marta P, Gleason B, Haylock M, Tank AK, Peterson T (2002) Observed coherent changes in climatic extremes during the second half of the twentieth century. Clim Res 19:193–212

Fu Q, Liou K (1992) On the correlated k-distribution method for radiative transfer in non-homogeneous atmospheres. J Atmos Sci 49:2139–2156

Fu Q, Liou KN (1993) Parameterization of the radiative properties of cirrus clouds. J Atmos Sci 50:2008–2025

Gan Y, Liang X-Z, Duan Q, Choi HI, Dai Y, Wu H (2015) Stepwise sensitivity analysis from qualitative to quantitative: application to the terrestrial hydrological modeling of a conjunctive surface-subsurface process (CSSP) land surface model. J Adv Model Earth Syst 7:648–669

Gandin LS, Murphy AH (1992) Equitable skill scores for categorical forecasts. Mon Weather Rev 120:361–370

Gregory D, Morcrette J-J, Jakob C, Beljaars A, Stockdale T (2000) Revision of convection, radiation and cloud schemes in the ECMWF integrated forecasting system. Q J R Meteorol Soc 126:1685–1710

Grell GA, Dvényi D (2002) A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophys Res Lett 29:38–41

Groisman PY, Karl TR, Easterling DR, Knight RW, Jamason PF, Hennessy KJ, Suppiah R, Page CM, Wibig J, Fortuniak K, Razuvaev VN (1999) Changes in the probability of heavy precipitation: important indicators of climatic change. Clim Change 42:243–283

Groisman PY, Knight RW, Easterling DR, Karl TR, Hegerl GC, Razuvaev VN (2005) Trends in intense precipitation in the climate record. J Clim 18:1326–1350

Han J, Pan H-L (2011) Revision of convection and vertical diffusion schemes in the NCEP global forecast system. Weather Forecast 26:520–533

Haylock M, Nicholls N (2000) Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998. Int J Climatol 20:1533–1541

Herman GR, Schumacher RS (2016) Extreme precipitation in models: an evaluation. Weather Forecast 31:1853–1879

Herold N, Alexander L, Donat M, Contractor S, Becker A (2016) How much does it rain over land? Geophys Res Lett 43:341–348

Herold N, Behrangi A, Alexander LV (2017) Large uncertainties in observed daily precipitation extremes over land. J Geophys Res Atmos 122:668–681

Hersbach H, Dee DJEN (2016) ERA5 reanalysis is in production. ECMWF Newslett 147(7):5–6. Also ERA5 datasets were described at: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5

Hersbach H, Bell W, Berrisford P, Horányi A, Nicolas J, Radu R, Schepers D (2019) Global reanalysis: goodbye ERA-Interim, hello ERA5. ECMWF. https://doi.org/10.21957/vf291hehd7. https://www.ecmwf.int/node/19027

Holtslag A, Boville B (1993) Local versus nonlocal boundary-layer diffusion in a global climate model. J Clim 6:1825–1842

Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD (2008) Radiative forcing by long-lived greenhouse gases: calculations with the aero radiative transfer models. J Geophys Res Atmos 113:13103

Iorio J, Duffy P, Govindasamy B, Thompson S, Khairoutdinov M, Randall D (2004) Effects of model resolution and subgrid-scale physics on the simulation of precipitation in the continental United States. Clim Dyn 23:243–258

Jamili A (2016) Robust job shop scheduling problem: mathematical models, exact and heuristic algorithms. Expert Syst Appl 55:341–350

Janjic ZI (1994) The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon Weather Rev 122:927–945

Janjic ZI (2000) Comments on "Development and evaluation of a convection scheme for use in climate models". J Atmos Sci 57(21):3686. https://doi.org/10.1175/1520-0469(2000)057%3c3686:CODAEO%3e2.0.CO;2

Kain JS (2004) The Kain–Fritsch convective parameterization: an update. J Appl Meteorol 43:170–181

Kain JS, Fritsch JM (1993) Convective parameterization for mesoscale models: the Kain–Fritsch scheme. The representation of cumulus convection in numerical models. Springer, pp 165–170

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

Karl TR, Knight RW (1998) Secular trends of precipitation amount, frequency, and intensity in the United States. Bull Am Meteorol Soc 79:231–241

Kunkel KE, Liang X-Z (2005) GCM simulations of the climate in the central United States. J Clim 18(7):1016–1031

Kunkel KE, Andsager K, Easterling DR (1999) Long-term trends in extreme precipitation events over the conterminous United States and Canada. J Clim 12:2515–2527

Kunkel KE, Andsager K, Liang X-Z, Arritt RW, Takle ES, Gutowski WJ Jr, Pan Z (2002) Observations and regional climate model simulations of heavy precipitation events and seasonal anomalies: a comparison. J Hydrometeorol 3:322–334

Kunkel KE, Easterling DR, Kristovich DA, Gleason B, Stoecker L, Smith R (2012) Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. J Hydrometeorol 13:1131–1141

Kunkel KE, Karl TR, Brooks H, Kossin J, Lawrimore JH, Arndt D, Bosart L, Changnon D, Cutter SL, Doesken N, Emanuel K (2013) Monitoring and understanding trends in extreme storms: state of knowledge. Bull Am Meteorol Soc 94:499–514

Leung LR, Qian Y, Bian X, Washington WM, Han J, Roads JO (2004) Mid-century ensemble regional climate change scenarios for the western United States. Clim Change 62:75–113

Li J, Barker H (2005) A radiation algorithm with correlated-k distribution. Part II: local thermal equilibrium. J Atmos Sci 62:286–309

Li J, Shibata K (2006) On the effective solar pathlength. J Atmos Sci 63:1365–1373

Li J, Dobbie S, Raisanen P, Min Q (2005) Accounting for unresolved clouds in a 1-d solar radiative-transfer model. Q J R Meteorol Soc 131:1607–1629

Li F, Collins WD, Wehner MF, Williamson DL, Olson JG, Algieri C (2011) Impact of horizontal resolution on simulation of precipitation extremes in an aqua-planet version of community atmospheric model (CAM3). Tellus A Dyn Meteorol Oceanogr 63:884–892

Liang X-Z, Zhang F (2013) The Cloud–Aerosol–Radiation (CAR) ensemble modeling system. Atmos Chem Phys 13:8335–8364

Liang X-Z, Kunkel KE, Samel AN (2001) Development of a regional climate model for us Midwest applications. Part I: sensitivity to buffer zone treatment. J Clim 14:4363–4378

Liang X-Z, Li L, Dai A, Kunkel KE (2004a) Regional climate model simulation of summer precipitation diurnal cycle over the United States. Geophys Res Lett 31:L24208

Liang X-Z, Li L, Kunkel KE, Ting M, Wang JX (2004b) Regional climate model simulation of us precipitation during 1982–2002. Part I: annual cycle. J Clim 17:3510–3529

Liang X-Z, Xu M, Gao W, Kunkel K, Slusser J, Dai Y, Min Q, Houser PR, Rodell M, Schaaf CB, Gao F (2005a) Development of land surface albedo parameterization based on moderate resolution imaging spectroradiometer (MODIS) data. J Geophys Res Atmos 110

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

Liang X-Z, Xu M, Yuan X, Ling T, Choi HI, Zhang F, Chen L, Liu S, Su S, Qiao F, He Y (2006) Development of the regional climate-weather research and forecasting model (CWRF): treatment of subgrid topography effects. In: Proceedings of the 7th annual WRF users workshop, Boulder, pp 19–22

Liang X-Z, Xu M, Kunkel KE, Grell GA, Kain JS (2007) Regional climate model simulation of US–Mexico summer precipitation using the optimal ensemble of two cumulus parameterizations. J Clim 20:5201–5207

Liang X-Z, Xu M, Yuan X, Ling T, Choi HI, Zhang F, Chen L, Liu S, Su S, Qiao F, He Y (2012) Regional climate-weather research and forecasting model. Bull Am Meteorol Soc 93:1363–1387

Liang X-Z, Sun C, Zheng X, Dai Y, Xu M, Choi HI, Ling T, Qiao F, Kong X, Bi X, Song L (2018) CWRF performance at downscaling China climate characteristics. Clim Dyn 3–4:2159–2184

Liang XZ, Li Q, Mei H, Zeng M (2019) Multi-grid nesting ability to represent convections across the gray zone. J Adv Model Earth Syst

Lim K-SS, Hong S-Y (2010) Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. Mon Weather Rev 138:1587–1612

Ling T, Liang X-Z, Xu M, Wang Z, Wang B (2011) A multilevel ocean mixed-layer model for 2-dimension applications. Acta Oceanol Sin 33:1–10

Ling T, Xu M, Liang X-Z, Wang JX, Noh Y (2015) A multilevel ocean mixed layer model resolving the diurnal cycle: development and validation. J Adv Model Earth Syst 7:1680–1692

May W (2004) Simulation of the variability and extremes of daily rainfall during the Indian summer monsoon for present and future times in a global time-slice experiment. Clim Dyn 22:183–204

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG (2012) An overview of the global historical climatology network-daily database. J Atmos Ocean Technol 29:897–910

Morrison H, Milbrandt J (2010) Comparison of two-moment bulk microphysics schemes in idealized supercell thunderstorm simulations. Mon Weather Rev 139:1103–1130

Morrison H, Milbrandt JA (2015) Parameterization of cloud microphysics based on the prediction of bulk ice particle properties.

Part i: scheme description and idealized tests. J Atmos Sci 72:287–311
Morrison H, Thompson G, Tatarskii V (2009) Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one-and two-moment schemes. Mon Weather Rev 137:991–1007
Nakanishi M, Niino H (2006) An improved Mellor-Yamada level-3 model: its numerical stability and application to a regional prediction of advection fog. Bound Layer Meteorol 119:397–407
Nakanishi M, Niino H (2009) Development of an improved turbulence closure model for the atmospheric boundary layer. J Meteorol Soc Jpn Ser II 87:895–912
Niu G-Y, Yang ZL, Mitchell KE, Chen F, Ek MB, Barlage M, Kumar A, Manning K, Niyogi D, Rosero E, Tewari M (2011) The community NOAH land surface model with multiparameterization options (NOAH-MP): 1. model description and evaluation with local-scale measurements. J Geophys Res Atmos 116:D12109
NOAA (2019) NOAA national centers for environmental information (NCEI) U.S. billion-dollar weather and climate disasters. https://www.ncdc.noaa.gov/billions/
Nordeng TE (1994) Extended versions of the convective parametrization scheme at ECWMF and their impact on the mean and transient activity of the model in the tropics. Research Department Technical Memorandum 206:1–41
Oleson KW, Niu GY, Yang ZL, Lawrence DM, Thornton PE, Lawrence PJ, Stöckli R, Dickinson RE, Bonan GB, Levis S, Dai A (2008) Improvements to the community land model and their impact on the hydrological cycle. J Geophys Res Biogeosci 113
Park S, Bretherton CS (2009) The university of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the community atmosphere model. J Clim 22:3449–3469
Pendergrass AG, Hartmann DL (2014) The atmospheric energy constraint on global-mean precipitation change. J Clim 27:757–768
Pleim JE (2007) A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: model description and testing. J Appl Meteorol Climatol 46:1383–1395
Prein AF, Rasmussen RM, Ikeda K, Liu C, Clark MP, Holland GJ (2017) The future intensification of hourly precipitation extremes. Nat Clim Change 7:48
Qian Y, Yan H, Hou Z, Johannesson G, Klein S, Lucas D, Neale R, Rasch P, Swiler L, Tannahill J, Wang H (2015) Parametric sensitivity analysis of precipitation at global and local scales in the community atmosphere model CAM5. J Adv Model Earth Syst 7:382–411
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
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
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
Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG (2007) Daily high-resolution-blended analyses for sea surface temperature. J Clim 20(22):5473–5496
Schär C, Ban N, Fischer EM, Rajczak J, Schmidli J, Frei C, Giorgi F, Karl TR, Kendon EJ, Tank AMK, O'Gorman PA (2016) Percentile indices for assessing changes in heavy precipitation events. Clim Change, 1–16
Siler N, Roe G (2014) How will orographic precipitation respond to surface warming? An idealized thermodynamic perspective. Geophys Res Lett 41:2606–2613
Sillmann J, Thorarinsdottir T, Keenlyside N, Schaller N, Alexander LV, Hegerl G, Seneviratne SI, Vautard R, Zhang X, Zwiers FW (2017) Understanding, modeling and predicting weather and climate extremes: challenges and opportunities. Weather Clim Extremes 18:65–74
Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2008) A description of the advanced research WRF version 2. Tech. rep., National Center For Atmospheric Research Boulder Co Mesoscale and Microscale
Smith AB, Matthews JL (2015) Quantifying uncertainty and variable sensitivity within the us billion-dollar weather and climate disaster cost estimates. Nat Hazards 77:1829–1851
Stegall ST, Kunkel KE (2019) Simulation of daily extreme precipitation over the United States in the CMIP5 30-yr decadal prediction experiment. J Appl Meteorol Climatol 58(4):875–886
Stephens GL, L'Ecuyer T, Forbes R, Gettelmen A, Golaz JC, Bodas-Salcedo A, Suzuki K, Gabriel P, Haynes J (2010) Dreary state of precipitation in global models. J Geophys Res Atmos 115
Subin ZM, Riley WJ, Mironov D (2012) An improved lake model for climate simulations: Model structure, evaluation, and sensitivity analyses in CESM1. J Adv Model Earth Syst 4
Sun C, Liang X-Z (2019) Improving U.S. extreme precipitation simulation: Dependence on cumulus parameterization and underlying mechanism. Clim Dyn (a companion paper submitted)
Sun D-Z, Fasullo J, Zhang T, Roubicek A (2003) On the radiative and dynamical feedbacks over the equatorial pacific cold tongue. J Clim 16:2425–2432
Sun Y, Solomon S, Dai A, Portmann RW (2006) How often does it rain? J Clim 19:916–934
Tao W-K, Simpson J, McCumber M (1989) An ice-water saturation adjustment. Mon Weather Rev 117:231–235
Tao W-K, Starr D, Hou A, Newman P, Sud Y (2003) A cumulus parameterization work- shop. Bull Am Meteorol Soc 84:1055–1062
Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106:7183–7192
Thompson G, Eidhammer T (2014) A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. J Atmos Sci 71:3636–3658
Thompson G, Rasmussen RM, Manning K (2004) Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: description and sensitivity analysis. Mon Weather Rev 132:519–542
Thompson G, Field PR, Rasmussen RM, Hall WD (2008) Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part ii: implementation of a new snow parameterization. Mon Weather Rev 136:5095–5115
Tiedtke M (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon Weather Rev 117:1779–1800
Trenberth KE, Dai A, Rasmussen RM, Parsons DB (2003) The changing character of precipitation. Bull Am Meteorol Soc 84:1205–1217
Tripathi OP, Dominguez F (2013) Effects of spatial resolution in the simulation of daily and subdaily precipitation in the southwestern US. J Geophys Res Atmos 118:7591–7605
USGCRP (2017) Climate science special report: fourth national climate assessment, vol I. Page, US Global Change Research Program, Washington, DC
Wang J, Kotamarthi VR (2015) High-resolution dynamically downscaled projections of precipitation in the mid and late 21st century over north America. Earth's Future 3:268–288
Wehner MF, Smith RL, Bala G, Duffy P (2010) The effect of horizontal resolution on simulation of very extreme us precipitation events in a global atmosphere model. Clim Dyn 34:241–247
Wilcox EM, Donner LJ (2007) The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model. J Clim 20:53–69

Wilson DR, Bushell A, 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

Wuebbles DJ, Kunkel K, Wehner M, Zobel Z (2014) Severe weather in united states under a changing climate. Eos Trans Am Geophys Union 95:149–150

Xie S, Zhang M, Boyle JS, Cederwall RT, Potter GL, Lin W (2004) Impact of a revised convective triggering mechanism on community atmosphere model, version 2, simulations: results from short-range weather forecasts. J Geophys Res Atmos 109

Xie SP, Deser C, Vecchi GA, Collins M, Delworth TL, Hall A, Hawkins E, Johnson NC, Cassou C, Giannini A, Watanabe M (2015) Towards predictive understanding of regional climate change. Nat Clim Change 5:912

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

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

Yang ZL, Niu GY, Mitchell KE, Chen F, Ek MB, Barlage M, Longuevergne L, Manning K, Niyogi D, Tewari M, Xia Y (2011) The community NOAH land surface model with multi-parameterization options (NOAH-MP): 2. Evaluation over global river basins. J Geophys Res Atmos 116

Yuan X, Liang X-Z (2011) Evaluation of a conjunctive surface-subsurface process model (CSSP) over the contiguous United States at regional-local scales. J Hydrometeorol 12:579–599

Yuan X, Liang X-Z, Wood EF (2012) WWR ensemble downscaling seasonal forecasts of china winter precipitation during 1982–2008. Clim Dyn 39:2041–2058

Zhang C, Wang Y, Hamilton K (2011a) Improved representation of boundary layer clouds over the southeast pacific in ARW–WRF using a modified Tiedtke cumulus parameterization scheme. Mon Weather Rev 139:3489–3513. https://doi.org/10.1175/MWR-D-10-05091.1PDF

Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW (2011b) Indices for monitoring changes in extremes based on daily temperature and precipitation data. Wiley Interdiscip Rev Clim Change 2:851–870

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

Zobel Z, Wang J, Wuebbles DJ, Kotamarthi VR (2018) Evaluations of high-resolution dynamically downscaled ensembles over the contiguous united states. Clim Dyn 50:863–884

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.